Blur perception:
An evaluation of focus measures

Robert Thomas Shilston

Department of Computer Science, UCL

September 2011

I, Robert Thomas Shilston confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Since the middle of the 20th century the technological development of conventional photographic cameras has taken advantage of the advances in electronics and signal processing. One specific area that has benefited from these developments is that of auto-focus, the ability for a cameras optical arrangement to be altered so as to ensure the subject of the scene is in focus. However, whilst the precise focus point can be known for a single point in a scene, the method for selecting a best focus for the entire scene is an unsolved problem. Many focus algorithms have been proposed and compared, though no overall comparison between all algorithms has been made, nor have the results been compared with human observers.

This work describes a methodology that was developed to benchmark focus algorithms against human results. Experiments that capture quantitative metrics about human observers were developed and conducted with a large set of observers on a diverse range of equipment. From these experiments, it was found that humans were highly consensual in their experimental responses. The human results were then used as a benchmark, against which equivalent experiments were performed by each of the candidate focus algorithms.

A second set of experiments, conducted in a controlled environment, captured the underlying human psychophysical blur discrimination thresholds in natural scenes. The resultant thresholds were then characterised and compared against equivalent discrimination thresholds obtained by using the candidate focus algorithms as automated observers. The results of this comparison and how this should guide the selection of an auto-focus algorithm are discussed, with comment being passed on how focus algorithms may need to change to cope with future imaging techniques.

Forward

As part of my undergraduate studies, I invented a new focus measure. This was refined and implemented in hardware such that it was used to focus a video camera, and appeared to outperform the autofocus function built into the camera. My novel auto focusing algorithm was duly patented [1] and licensed to British Telecom PLC. However, my curiosity was not sated, and the question remained in my mind as to whether this new method was 'better' than existing methods, or even what constituted 'better' when considering focus. This thesis is the result of my exploration into focus.

I'd like to extend sincere thanks to Fred Stentiford and Steven Dakin for their advice, and also to my friends, family, and especially my wife for their support and patience.

Contents

List of Figures

List of Acronyms and Abbreviations

AE auto-exposure

AF auto-focus

2AFC two-alternative forced choice

ASIC application-specific integrated circuit

AWGN additive white gaussian noise

CCD charge-coupled device – An image sensor

CIE International Commission on Illumination (Commission internationale de l'clairage)

CIEDE CIE difference equation

CSV comma separated values

D Dioptres

DAC digital to analogue converter

DTD document type definition

EW Edinger-Westphal

FFT fast Fourier transform

GPU graphical processing unit

GUI graphical user interface

HDR high dynamic range

HSV hue, saturation, value

HVS human visual system

MPEG Motion Pictures Experts Group

MSE Mean square error

NTSC National Television System Committee, and name of the analogue television system used predominantly in North America

PAL phase alternate line, an analogue television system used in the UK and many other countries

PDF probability distribution function

PSF point spread function

PTP picture transfer protocol

RGB red, green, blue

RMS root mean square

SLR single-lens reflex

VA Visual Attention

VEP visual evoked potential

XML extensible markup language

Glossary

Amblyopia is a developmental anomaly of spatial vision, which is characterised by reduced visual acuity and reduced contrast sensitivity, colloquially known as a "lazy eye".

Emmetrope is a perfectly sighted individual

Hypermetropia long-sighted

Myopia short-sighted

Pedestal condition is the magnitude of the stimulus for which a psychophysical threshold is being determined. Typically the threshold is determined for multiple pedestal conditions.

Plenoptic camera is one which uses a microlens array to capture 3D light field information about a scene. The microlens array sits between the lens of the camera and the image sensor. It refocuses light onto the image sensor to create many small images taken from slightly different viewpoints, which are manipulated by software to extract depth information.

Presbyopia is the condition where the eye exhibits a progressively diminished ability to focus with age

Chapter 1

Introduction

Every day, countless millions of photographs are taken around the world. The process of taking photos has changed over the past decade from using the relatively expensive silver-halide based approach to digital capture, with cameras being embedded into ever more devices. This change to digital capturing, where there is no per-image cost and the ability to instantly review images, is enabling and encouraging people to take more and more photos. Increased processing power means advanced features can be incorporated into cameras, such as cameras that wait for people to smile before automatically taking a photo. Beyond domestic photography, improvements in storage, communication and processing technologies have enabled growth in automated imaging applications, ranging from security to quality control in manufacturing processes.

Despite this rapid growth and development of additional features, there remains a fundamental aspect of imaging that merits investigation. To capture a photo, light emanating from the scene goes through one or more lenses to form an image. The lenses need to be arranged such that the image formed on the camera's sensor is in focus, a requirement that is just as important in human vision as it is in a camera. Numerous focus measures have been proposed, each of which assigns a score to an image such that the image with the highest score is the image that is best focussed. This information is used to help cameras adjust their optics to ensure that their image is in focus. However, a comprehensive review of these measures comparing their performance with the judgements made by humans has not been conducted.

It is possible to establish the precise configuration of lenses required to ensure light rays at a particular point in a scene are focussed. However, the method for selecting a best overall focus for an entire scene is an unsolved problem. This work starts from the assumption that a camera (or other imaging device) can produce images of a given scene at a range of focal distances, and explores how these images

can be compared to determine which is the 'best' for the entire scene. Such an approach is independent of the optics of the camera, and other device-specific artifacts arising from image acquisition.

1.1 Objectives

This thesis proposes that there is an algorithm that can reproduce the results obtained from human perceptual and subjective experiments. To explore this statement, several areas are examined. Firstly, existing comparisons of focus measures do not have a robust approach to establishing the ground truth as to which candidate image is best focussed. Secondly, previous comparisons have not been comprehensive – no overall comparison between all algorithms has been made, nor have the results been compared with human observers. Given the complexities of real-life scenes, an approach for identifying the truth needs to be established, and used to explore whether human observers are consensual in their opinion. Beyond identifying the 'best' image, other ways of characterising human blur opinions and perception need to be explored and performed.

The various focus measures found in the literature and developed during the course of this work then need to be compared with the results obtained from human experiments, to establish which measure is best able to reproduce the human results. To permit such a comparison, it might be necessary to consider how the methodologies need adapting to suit the differences in behaviour between human and model observers.

To narrow the scope of this work, sample images used will be normal scenes – scenes that the man-in-the-street might want to photograph. Skilled photographers who make clever use of depth-of-field or bokeh[1] are well able to configure their camera to achieve the artistic effect they desire, and so are unlikely to need help with focus. However, most camera owners do not have such a level of skill, and for them it is important that their camera tries to help them take the photo they want. Given this attention on normal photographs, the observers used for establishing human results should be normal people who are neither highly skilled photographers nor with experience of psychophysics.

[1]The result of the camera's optics on out-of-focus points of light

1.2 Contribution

A comprehensive account of the state of existing focus measures is provided in the literature review, and software implementations for all focus measures encountered during this review are included in the appendix for future research to build upon. For the first time, a full comparison of all focus measures in the literature is performed both with a single scene and assessing the overall performance across a number of scenes. To assist with future focus measure assessment (and related studies), the image library developed during this work has also been made available.

The results from the experiments described in this thesis increase our understanding of human opinions of image blur. Two key results from human experiments are reported: Firstly, it is shown that when humans are asked to select the best focused image of a scene, they are highly consensual, and there appears to be no subset of the population with a significant difference of opinion.

Secondly, blur discrimination thresholds have been measured in natural scenes, and these are found to be of a 'dipper' shape of similar magnitude to the discrimination thresholds measured in synthetic images.

Finally, a methodology for assessing focus measures is proposed. Using this methodology shows that few focus measures are able to reproduce human blur perception characteristics. This suggests that none of the focus measures reviewed in this work are the sole method used by the human vision system.

1.3 Structure

The structure of the thesis is as follows: Chapter 2 presents a review of literature relating to the human visual system, models of focus and perception, and recent developments in image acquisition technology. Chapter 3 provides more details about the selection and preparation of sample images, and the experimental methodologies deployed.

Chapter 4 describes and discusses experiments that capture quantitative metrics about human observers that were conducted with a large population of observers on a diverse range of equipment. These human results were then used as a benchmark, against which equivalent experiments were performed by each of the candidate focus algorithms.

A second set of experiments are described in Chapter 6, which were conducted in a controlled environment to capture the underlying human psychophysical blur discrimination thresholds in natural scenes. The resultant thresholds were then characterised and compared in Chapter 7 against equivalent discrimination thresh-

olds obtained by using the candidate focus algorithms as automated observers. The results of this comparison and how this should guide the selection of an auto-focus algorithm are discussed, with comment being passed on how focus algorithms may need to change to cope with future imaging techniques.

In Chapter 8, a summary of all the results is presented and compared with the thesis objectives. Significant findings are presented, and an indication of the direction of future work is provided.

Finally, the appendix includes software implementations of the focus measures used throughout this work, providing a solid base from which this work can be continued.

Chapter 2

Review of literature

This section introduces material that sets the context to, and underpins the theory and history of the work that has been carried out in a variety of areas that are used through this thesis, as well as providing an overview of the state-of-the-art of related fields. It leads on to a discussion, conclusions and direction for the research that follows.

2.1 Human vision mechanisms

Many living species have the ability to see, and their various vision structures first evolved during the Cambrian period. The human vision system is very good for our needs, but is not a perfect seeing-device. Other species can do far better – owls, for example, have greater visual acuity, and bees can see ultra-violet light which is invisible to ourselves.

Debate about how the eye worked started with Plato, in the fourth century BC who wrote that light emanated from the eye, though Aristotle advocated the opposite, that the eye receives rays of light. In the second century AD, Galen identified many parts of the eye, including the retina, cornea and iris, though concluded that the crystalline lens was the principal instrument of vision, based on the evidence of cataracts – where clouding of the lens reduces image clarity. However, real progress into the understanding of the eye only started during the 16th century, leading to Kepler proposing the theory of a retinal image in 1604. More recently, most notable were the works by Young, Helmholtz and Maxwell exploring how the retina functions, and which culminated in the trichromatic theory of colour vision [2, 3].

2.1.1 Physical mechanism for focusing the eye

The biological term for the way the eye focuses is *accommodation*, but how the eye actually does this was only resolved in the mid 19th century. It was the suggestion by Helmholtz in 1856 that the lens is elastic and held under tension by various muscles (the ciliary muscles) which became the accepted theory of accommodation [4]. This is shown in Figure 2.1 which demonstrates how the lens changes shape to focus the image on the retina. Table 2.1 lists key individuals and the theories of accommodation they proposed.

Proponent	**Focusing mechanism**
Home (1795)	Cornea curvature changes
Magendie (1816)	Eye has universal focus
Burow (1841)	Lens moves back and forth
Donders (1846)	Pupil changes in size
Listing (1853)	Eye ball elongates
Helmholtz (1856)	Lens changes shape

Table 2.1: Proposed accommodation methods (adapted from [4])

Accommodation is present in a wide range of animals, each of which has a different accommodative range, as required by their various habitats and behaviour. This is summarised in Table 2.2. Interestingly, the cat has proven most challenging to assess, "with studies reporting amplitudes between less than 2 Dioptres (D), and as much as 12D ... some have doubted that cats accommodate at all. ... One obvious possibility is that there is a genetic variation, with in-bred house cats having poorer accommodation than animals that have to catch mice to stay alive!" [6].

Figure 2.1: This diagram shows how the lens changes shape to ensure that the image is focussed on the retina (adapted from [5])

Animal	Accommodative power
Pigeon	9D
Chicken	17D
Macaques (young)	18D
Marmosets	20D
Humans (young)	14D
Cat	<2D to 12D

Table 2.2: Accommodative power varies between species (adapted from [6]), which also notes that the accommodative range is very large in animals that need to see in both air and water, such as the sea otter and the cormorant.

2.1.2 Source of the control signals

The mechanisms above solely concern the mechanics of how to focus; that is, how the eye's structure adapts to ensure that the image formed on the retina is in focus. The actual signals that control the ciliary muscles were not considered for nearly a century. Pertinent work exploring these signals is now introduced, considering in turn the fundamental sources of information that the eye has available:

Mechanism 1: Convergence

The eye uses several cues for driving the ciliary muscles, the simplest of which is to use the information available from the convergence mechanisms in binocular vision [7]. The closer the object is to the eye, the greater the eyes must be oriented towards the nose, and hence by determining the amount of convergence the brain can send an appropriate signal to the ciliary muscles. However, this is clearly not the sole means by which the eye focuses, as it is entirely possible for a person to focus on an object even if they cover one eye.

A widely cited study showed that the human eyes are consensual (ie the left eye does not focus differently to the right) when presented with unequal accommodation demands to each eye, by the use of targets at different distances [8]. Along these lines, Flitcroft (1992) presented subjects with a series of dynamic aniso-accommodative stimuli and observed that the response was equal in both eyes, and was a compromise between the inputs to the two eyes, with no evidence of a random alternation of eye dominance [9]. However, work in 1998 showed that in extreme circumstances, the eyes can focus differently – Marran measured "an average 0.75D aniso-accommodative response [for a] 3.0D aniso-accommodative stimulus" [10].

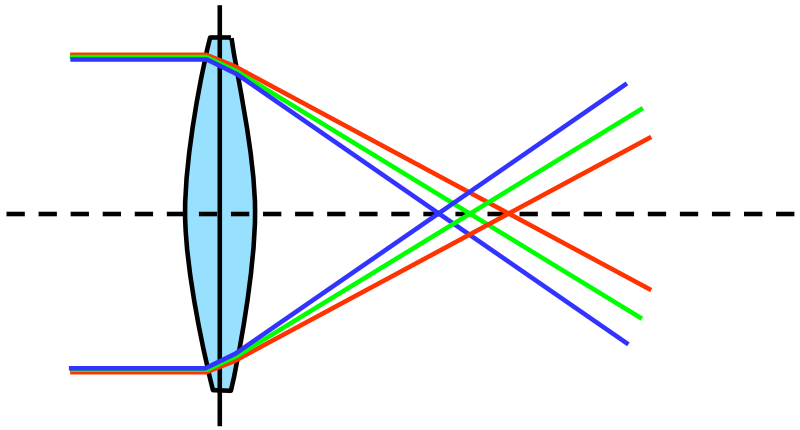


Figure 2.2: Chromatic aberration occurs when light of different wavelengths is refracted by a lens

Mechanism 2: Chromatic aberration

Another cue used by the visual system is chromatic aberration [11]. Chromatic aberration is the result of light of different wavelengths refracting by slightly different extents (see Figure 2.2). In photographs this exhibits itself as coloured fringes at the edges of objects that are not in perfect focus. For example Figure 2.3 exhibits noticeable chromatic aberration where blue and red fringes are evident. It is, of course, undesirable for photographic images to exhibit such aberration, and its impact can be greatly reduced by the use of an achromatic doublet lens invented by Hall and Dolland early in the 18th century [12].

The eye, however, could make use of the chromatic aberration. If the eye is focused beyond the object (hypermetropia), then the object will have a red edge. Whereas, if the object is behind the eye's focus position (myopia), then the object will have blue fringes. Recent work has indeed shown that this chromatic aberration can be used to provide an odd function signal[1] for ciliary control [13]. The use importance of an odd-functioned signal is discussed in Section 2.1.5.

As with convergence, it is possible to establish whether this is used as a focusing cue by investigating whether focusing can be achieved when the cue is absent. If a scene is illuminated with monochromatic light, then no chromatic aberration will occur, as there is only a single frequency of light. Experiments by Fincham with 55 human subjects under such conditions found that 19 of these were unable to focus, and were conscious of a blurred image, 22 focused as well as when the object was illuminated with white light, and the remaining group accommodated in the same direction regardless of whether a positive or negative lens was inserted [7, 15]. More

[1] An odd function is one where $f(-x) = -f(x)$, whilst an even function behaves such that $f(x) = f(-x)$

(a) Photograph of a block of flats (b) Detail shows chromatic aberration

Figure 2.3: A photo of a block of flats [14] which shows significant chromatic aberration around the window frames

recently, Chen showed that there is no significance between subjects who are short sighted, and those with correct vision (emmetropes) when focusing in cue-absent monochromatic conditions [16].

These results suggested that, at least for some subjects, broad spectrum illumination (and thus chromatic abberration) is critical in providing directional information regarding the level of defocus. This is reinforced by Campbell's work which explored the minimum quantity of light required for humans to focus, concluding "the mechanism of accommodation was found to take up a relatively fixed position approximately 0.6D greater than the minimum refractive power of the eye when the luminance of the test object was below the cone threshold for visibility. It is concluded that the receptors involved in the accommodation reflex are the foveal cones and that in the absence of a foveal stimulus the mechanism of accommodation takes up a relatively fixed focus greater than the minimum refractive power of the eye." [17].

Mechanism 3: Minimising blur

Whilst it is relatively easy for a person to manually adjust a camera, telescope or microscope etc to bring the object into focus, actually specifying what constitutes in-focus is difficult. The dictionary defines focus as "that point or position at which an object must be situated, in order that the image produced by the lens may be clear and well-defined" [18], but this then leads to the challenge of the defining 'clear' or 'well-defined'. One view is that in-focus means that the image has the least possible blur, but yet again, this poses the difficulty of defining the readily understood concept of blur in a mathematically meaningful way. Pentland stated that 'exact

focus' means the point spread function has minimum variance [19], but this simply anchors the definition to a mathematical function without any investigation to show this is correct.

Just as a human will tend to hunt backwards and forwards whilst attempting to focus a camera, gradually refining the focus position, the eye appears to perform a similar hunting motion. In some fascinating work by Fender in the 1960s to investigate the control system of the eye's tracking mechanism, he observed that a 'hunting' motion is superimposed on the accommodative mechanism, and that this continually lengthening and shortening of the focal length will improve (or worsen) the image and thereby provide information to control the accommodation system [20].

Fender's work was cited by Crane, who concluded that it was important to determine exactly what measure of the image is fed back to the accommodation system [15, p17]. Crane proceeds to propose that, as "the effect of blurring and defocus are to reduce the high-frequency spatial components in an image", an approach which involves the two-dimensional spatial derivative can reproduce many experimental results. He looks at models based on both the summation of, and the peak amplitude of the spatial derivative. Whilst the former does have good performance at focusing, it is only the latter that, he claims, can provide fine-focus control [15, p29].

Two decades later, and in marked contrast to the broadly gradient-based approaches that had been proposed to date, Morrone et al observed that 'lines' and 'edges' are the points in a waveform where the Fourier components are in phase with each other [21]. Such a phase coherence approach is explored further: Kovesi proposed a mechanism for determining when a maxima in phase coherence actually constitutes a feature [22], whilst Wang showed how blurring disrupts phase coherence [23][2].

Section 2.3 looks at mathematical focus measures in more detail, whilst the hunting motion is considered further in Section 2.1.4.

Troelstra and Stark both describe experiments wherein a target is randomly moved in the absence of cues (using monocular vision, with no change in perceived size or illumination, and no lateral movement). Both observed that the eye starts to refocus in the wrong direction 50% of the time. Figure 2.4 shows that, when the target is moved from its rest position to either the far or near position, the eye makes an erroneous initial accommodative response half of the time [24, 25]. Stark argues

[2] An attempt was made to contact Wang et al to obtain details of their algorithm. Unfortunately no reply was received.

Figure 2.4: Ten responses with five initial errors from a typical experiment. The arrows annotate where the initial accommodation was erroneous [24].

that this result shows there is not an odd-error component in the accommodative signal, and that as blur is an even signal, it is likely that blur is the metric that is used by the eye.

Philips and Stark compare the accommodative response when the target is blurred (that is, the eye is looking at an out-of-focus image projected onto a screen) with that when the eye is defocused, and observes the responses are equivalent. The paper concludes that: "without minimizing the role of vergence-accommodation, the rich optical and other clues to target distance, and the higher level 'volitional' and predictive control of accommodation in normal viewing, we can point to clear experimental evidence for blur as the sufficient neurological stimulus to accommodation" [26].

Finally, both Marshall and Mather show that blur plays a role in depth perception: Experiments with an ambiguous figure containing a blurred and sharp region divided by a wavy line showed that the relative distance of the two regions was perceived differently depending on whether the boundary was blurred or not [27, 28].

Other mechanisms

Other work has shown that a change in apparent distance (that is, by varying angular size of an object) can also cause trigger a change in accommodation [29], and that subjects can use audio biofeedback to exert voluntary control of accommodation to reduce myopia [30]. Both results demonstrate how versatile the accommodation system is at using any available information.

2.1.3 Neurological pathways

A thorough overview of the neurological pathways involved in accommodation is provided by Gamlin [31] who, in summary, describes how observations in the late 19th century led to the identification of the Edinger-Westphal (EW) nucleus, which has been shown to be responsible for the accommodation in both mammals and birds. This group of cells – containing 800-1200 neurons in primates – when subjected to electrical microstimulation evokes ocular accommodation in primates with a latency of around 75ms. Despite relatively few studies examining the inbound accommodation-related signals into the EW, there are confirmed links from the cerebellum (eg [32]). This could be the neurological means by which audio-biofeedback control of the accommodation system might work.

Hung et al assert that "the summated blur signals are transmitted through the magnocellular layer of the lateral geniculate nucleus to arrive at area 17 of the visual cortex. The summated cortical cell responses form a sensory blur signal", though provides no indication as to how that original blur signal is computed [33, p288]. Fylan presented preliminary results measuring visual evoked potentials when blur is applied to image stimuli, suggesting that it might be possible to measure this blur signal and understand image properties that cause perceptual blur, though no followup work has been published since these results from 1998 [34].

2.1.4 The 2Hz oscillations

Many papers discuss the fluctuation in lens position that was first observed by Fender and which has been measured to be around 0.1D at 2Hz. It has been proposed that this might be used either to perform hunting, to ensure that the subject remains in focus [15], or that it adds odd information to the otherwise even signal from blur [25]. Crane also shows, in an appendix, that these vibrations might play a role in increasing the perceived depth-of-field. However, none of these suggestions are conclusive.

Analysis of the stability and root-locii[3] of various proposed vision models predict an instability at 0.45Hz, which is observable in human subjects, but do not predict one at 2Hz. Gray showed that whilst the 0.5Hz fluctuation varied with pupil diameter (and presumably thus posits that it is associated with the vision mechanism), the high frequency did not. "Thus [Gray] concluded that the lower accommodative frequency peak which was predicted by the root locus analysis is most likely

[3]In addition to determining the stability of a system, the root locus can be used to design for the natural frequency of a feedback system (see, for example, [35])

associated with neurologically-controlled feedback instability oscillations, whereas the high frequency peak is an epiphenomenon due to the effect of arterial pulse on lens motion that is detected by the recording optometer" [33, p321]. It had been earlier shown that the 2Hz fluctuations are "significantly correlated with arterial pulse frequency" [36], though Charman suggests that thay may arise because of the mechanical and elastic characteristics of the lens, zonule and ciliary body [37].

However, these fluctuations remain an area of intrigue: Judge and Flitcroft, whilst considering this observed correlation with pulse note "If this is so, it is puzzling that macaques, which have considerable higher pulse rates than humans, have high frequency fluctuations of a similar frequency to humans. Indeed, fluctuations in humans at 2Hz would imply a heart rate within the definition of clinical tachycardia." [6].

That this source of the 2Hz signal is not certain does not, of course, mean that the visual system is not making use of such fluctuations in lens position for some purpose, whether its cause be deliberate or some artefact of other biomechanical activities.

2.1.5 Models of accommodation

Despite the wide range in apparent inputs to the accommodation mechanism, the identification of the neurons which control the ciliary muscles and the pyschophysical response to stimuli, the control system in the brain is not well understood.

Fender's exploration of the eye from the point of view of systems analysis, constrained itself to the systems associated with the control system that the eye uses when tracking a target, and whilst touching on accommodation did not itself suggest how it should be modelled [20]. The systems analysis approach is further pursued by Toates who begins "the nervous control of the ciliary muscle, a subject of secondary interest in previous papers, is central to [the proposed model]" [38]. He proposed a model in the form of a classic negative feedback proportional control system (see Fig 2.5). This model, as with other negative feedback control systems, relies upon an odd function to assess blur, and Toates uses Stark's 1965 function [39]. Stark's function is a linear response, but with limits which impose a constant output once the input exceeds some threshold.

Further work by Stark (1975) used a computer simulation of Toates' proportional control model, and demonstrated that it exhibited unstable rather than smooth responses to step stimuli, and proposed that a model based on a leaky integrator, rather than a proportional controller, would improve it [33, p300].

A useful review of the progress made in model development over the last quarter

Figure 2.5: Toates' negative feedback control system [38] uses Stark's measure of blur

of the 20th century, discussing their respective strengths and weaknesses is provided by Eadie and Carlin [40] and Khosroyani and Hung [41]. Khosroyani and Hung introduce a dual-mode model which connects the vergence and accommodation systems and can be used to explain the difference in behaviour for slow and fast ramping inputs. However, no reference is made to the source of the 'blur' signal, beyond acknowledging that it is important: "the vergence system acts to produce foveal registration and the accommodation system acts to reduce retinal blur" [41].

2.2 Visual psychophysics

Independent of the low-level neural and physical analysis of accommodation and its control, a separate body of work has looked at the psychophysics of blur; the study of the relationship between the perception of blur and actual blur present in the stimuli. A number of intertwining research themes have been pursued over the past decades, and are introduced approximately chronologically. These consider the nature of blur, sensitivity to contrast and frequency, as well as examining whether the human visual system is optimised for the statistical structure present in natural images.

Psychophysical experiments exploring the sensitivity to a wide range of physical stimuli, not just within vision science, tend to produce results that indicate a common perceptual response – that is, at low stimulus levels, the just-noticeable-difference decreases as the stimulus increases, then rises once beyond a certain stimulus magnitude.

2.2.1 Early psychometric experiments

The earliest literature relevant to the study of blur describes work done in the late 1960s. Campbell quantitatively investigated the psychometric responses of the human vision system to known stimulii, showing that in some experiments just-

(a) Original figure (b) Replotted

Figure 2.6: Results from Campbell showing the original reported just-noticeable frequency ratio as a function of the lower frequency [42], and the same data re-plotted using different axis to enable comparison with more recent work.

noticeable-difference of spatial frequency[4] is constant in terms of the ratio of the two frequencies being compared, unlike a dipper [42]. However, he commented that this might have been an artefact of the experimental design – giving subjects an unlimited period of time to adjust one frequency to match the stimulii might mean that they achieved the same degree of accuracy in the end, but that it was harder (and thus took longer) for certain frequencies.

To address this factor, a second experiment is described, which used a fixed time interval for the stimulii presentation. It showed that discrimination accuracy improved with the frequency ratio. Whilst not exhibiting a "dip", this result conforms with Weber's law:

$$\frac{\text{just noticeable difference}}{\text{stimulus intensity}} = \text{constant} \tag{2.1}$$

Whilst there is no evidence to suggest this is the case, it could be that a dipper was present, but that by virtue of the experimental design or data processing, it might be that the relevant pedestal points were not measured, or the results were considered to be anomalous, and discarded.

2.2.2 Preliminary investigations of blur

During this time, vision-centric theoretical work was also progressing, built upon quantitative results obtained in the late 1970s. In 1980, the first theory of edge detection was proposed by Marr [43]. This suggested that edges are defined as

[4]A spatial frequency is one which is expressed as the number of cycles per degree of visual angle. Experiments examining performance at different spatial frequencies typically use a stimuli based on a sinusoidal grating.

Figure 2.7: Blur extent difference thresholds for three blurring functions as a function of criterion blur extent. The error bars represent 2 standard errors. The space constants employed to express blur extent are: standard deviation for the Gaussian; ramp extent for the rectangular; and half-wavelength for the cosinusoidal blur. The data for two subjects are shown (R.J.W. *o*; C.C. *x*). In each case the rising portion of the functions corresponds to a power law with an exponent of 1.5. From [45]

zero-crossing points of the second derivative of intensity, and showed that the most sensible means of finding these is by searching for the zero values of the convolution $\nabla^2 G * I$, where ∇ is the Laplacian, and G the Gaussian operators, and I is intensity.

Hamerly measured the blur detection threshold, and showed that this was lower when both stimuli were blurred, than when one was unblurred [44]. However, the amount of blurring used was not sufficient to reach the other side of the dip, and encounter a Weber's law type relationship – the pedestal blur was not increased beyond 80 arc-seconds.

A few years later in 1983, a simple experiment using a step change in luminance was conducted by Watt [45]. A 1D band of light with a variety of blur functions applied was shown, and the observer indicated which of two stimulii had a broader blur extent. Three blur functions were used: gaussian, rectangular and half-wave cosinusoidal profiles were each applied to the the step change in luminance.

The primary finding from this work is that "blur comparison is most precise at some non-zero criterion blur for each blurring function. In each case the data shows a decrease in threshold as the criterion blur is increased from zero to an optimum level, beyond which threshold rises rapidly" – see Figure 2.7, which shows that what Hamerly perceived to be an increase in sensitivity, was simply a dip which ends after approximately 3 arc-minutes of pedestal blur. Thus, a decade after a dipper response was observed by Nachmias and Sansbury investigating contrast, the same shape of response was found in blur discrimination.

In exploring the possible cues that the subjects might use, Watt ruled out $\widehat{\delta I}$

Figure 2.8: (a) A hypothetical double step stimulus and the second derivative (b) of the resultant retinal image. The visual system is able to resolve the stationary point at I and thereby perceive two separate steps. (c) A narrow double step stimulus and the second derivative of the resultant retinal illumination (d). The visual system is not able to resolve the stationary point which would be at I, and the Chevreul illusion results. (e) A ramp edge stimulus and the resultant second derivative profile (f). The zero-valued stationary point at I causes a misrepresentation of the stimulus and Mach bands result. (Adapted from [45]).

(maximum change in intensity), simple Fourier transform model and zero-crossing of the second derivative. Instead, contradicting Marr's theory, he showed that more suitable primitives were stationary points (rather than zeros) in the second derivative corresponding to edges. As ever, being able to explain observed phenomena with a proposed theory is a useful validation technique. Watt showed that his primitives could explain the Mach band and Chevreul illusions. The Mach band illusion comprises a linear gradient between a light and dark uniform areas. There is the perception of a lighter stip on the light side of the gradient (and the converse on the dark side). The Chevreul illusion appears when a light to dark stepped sequence of bars are viewed – the bars tend not to look as if they are of a single colour, but instead appear graduated from light to dark in the opposite direction to the main steps. These illusions and explanation are shown in Figure 2.8.

2.2.3 Quantifying blur discrimination

After Watt's work, various papers were published which used similar experiments to extract more information about exactly how the eye (and human visual system) might work. For example, Levi investigated the effect of blur on line detection, spatial interval discrimination, the 2-line resolution, and developed models which represent the behaviour in terms of an equivalent intrinsic blur [46]. The experiments were then repeated in subjects with amblyopia, whereupon the altered behaviour was shown to be represented by a minor change to the models that had been developed [47].

The impact of exposure time upon performance was investigated [48], and shown that discrimination improves with duration, but plateaus after approximately 130ms.

Mather and Smith considered how image blur might be used as a depth cue in human vision [49, 50], and showed that blur discrimination should be more effective than convergence at larger distances.

Jacobs [51], whilst investigating sensitivity to defocus, considered the two ways in which it can be produced: "Either the source of the visual image, such as a photographic print or projected slide can be defocused or the observer can be defocused using positive lenses placed in the spectacle plane. These two methods have been called source and observer methods respectively...". In 1989, when this work was being done, blur thresholds had only been established with observer methods of defocusing. Jacobs showed that results of both source and observer methods correlated with a high coefficient – 0.994, and thus, giving "greater validity to the source method, which has a number of advantages over the observer method, especially because it is within the capability of current technology to present simulated defocused images generated by computer processing".

Whether the results achieved thus far could be reproduced in natural scenes (as all experiments described to date used relatively simple stimulii; typically a simple edge or pair of edges shown on an oscilloscope) was explored [52]. Walsh used 18 subjects whose accommodation was temporarily paralysed by the use of anaesthetic. The subjects were presented with a stimulus oscillating at 2Hz, and used a staircase procedure to find the just noticeable change in contrast with defocus. The results showed a dipper function that was symmetrical with both induced hyperopia and myopia (see Figure 2.9), and was similar for both synthetic images (sinusoidal gratings) and natural scenes (of a street). The same shaped responses were observed in monochromatic illumination of various frequencies, though the centre of the symmetrical dipper was moved "in a progressively more hypermetropic direction as the wavelength increased".

Figure 2.9: Typical plot of the minimum detectable amplitude of oscillation of defocus as a function of mean position of focus. Vertical bars represent the standard deviations in the thresholds. Subject GW (age 29); 2.55 c/deg grating; 3 mm diameter pupil; green light; 2Hz target oscillation frequency [52]. NB. In monochromatic illumination, the centre of the symmetrical dipper moved to the right with increased wavelength.

Without reference to Walsh's work, Flitcroft explored the effect of temporal modulations in luminance contrast on accommodation [53], and showed that fluctuations in the 1-4Hz region effected the greatest detriment on accommodation, a result which is "compatible with the ... hypothesis that flicker impairs the ability of the accommodation system to utilize temporal cues such as those derived from the higher frequency component (1-2Hz) of accommodative oscillations" (see Section 2.1.3).

2.2.4 Applications

The 1990s saw researchers trying to apply the well-established basic observations and explanations of blur discrimination to other areas and applications. In one experiment, subjects were required to judge the amount of blur in moving stimulii [54], which showed movement made blur discrimination harder. The actual impact of motion reduced as the pedestal blur increased – see Figure 2.10. Pääkkönen showed that motion produces equivalent spatial blur, and suggested a mechanism by which it might arise. However, these results are reviewed by Hammett who showed that "whilst discrimination performance for physically constant blur widths increases monotonically with speed, subjects' performance for constant perceived blur widths is virtually constant for speeds up to 6.3 deg/sec", and that this might be as a result of perceptual sharpening, rather than motion blur [55]. There was consensus that the blur width of a moving edge needed to be larger than for a static edge to achieve the same perceptual width, and that the extra width required was

Figure 2.10: Blur-discrimination data for an observer, plotted as a function of reference blur for six different velocities. The optimum blur is not at zero but at some higher reference-blur value for all velocities. [54]

proportional to speed.

Burr and Morgan investigated motion and blur, to explore why "moving objects look more blurred in brief than in long exposures". They showed that motion does not improve an observer's ability to discriminate, but that moving objects appear sharp because "the visual system is unable to perform the discrimination necessary to decide whether the moving object is really sharp or not" [56]. All three of these papers used gaussian blurred step functions as opposed to natural scenes or other blur methods.

Approaching this area from a different direction, Kayargadde and Martens [57, 58,59], proposed a strategy for determining the quantity of blur present in an image. They achieved a high degree of correlation between their algorithm's response and the mean-opinion-score of a number of subjects across a range of images, and thereby argue that their "blur index" can be considered a psychometric measure of sharpness. The blur index is determined by measuring the blur spread (an estimate of the kernel size that caused the blur) across an image, then producing a global estimate by combining an average of the blur spread with a weighting based on the edge strength and length.

However, Peli observes, when considering Kayargadde's edge detection performance, that "It is clear that the algorithm has high sensitivity, but also a high false alarm rate. The falsely detected edges are not only single noise pixels, but they also created false connections between real edges" [60] – a frequent problem in edge detection algorithms.

Similar blur estimators have been developed by others (especially Ferzli, Karam

and associates [61, 62, 63] and Marziliano [64, 65]), with the aim of being able to quantitatively assess blur scenes without a reference. That is, to to permit comparison of two images of completely different scenes and determine which of the two is most in focus.

2.2.5 Natural scenes and the human visual system

Early research by Campbell showed that the human visual system appears to contain multiple 'channels' for contrast detection that are tuned to different spatial frequencies, and that the sensitivity of these channels is not equal – see Figure 2.11 [66] (more can be found in a review by Klein [67]). Commenting on this, Billock suggests at low spatial frequencies, contrast discrimination performance is consistent with a Laplacian-like filter, whilst at higher frequencies there is evidence for multiple, overlapping, wavelet-like spatial mechanisms tuned to narrow bands of spatial frequencies [68]. The obvious question to draw from this is "Is this optimal for the environments in which the human visual system is used?".

Field observes (1987), from the opposite angle, that "there seems to be a belief that images from the natural environment vary so widely from scene to scene that a general description would be impossible" [69]. For any arbitrary non-random signal, there is a coding strategy which permits it to be transmitted more efficiently than simply transmitting the raw signal. An examination of a number of scenes that would be typical for the mammalian visual system to encounter, shows that there are generalities, so optimisation is possible. Looking at six different images, it is clear that the frequency spectra follows the relationship $g(f) = k/f^2$, as shown in Figure 2.12, a relationship which had been found earlier by Carlson in 1978 [70]. Such a relationship is expected if an image's energy were scale invariant – there is equal energy in each frequency octave.

Knill, Field and Kersten [71] explored whether the exponent of the frequency drop-off is always 2, or does is there a variation in power spectrum between images, and thus of the form:

$$P_I(f_r) \propto \frac{1}{f_r^{\beta}} \tag{2.2}$$

This can also be rewritten in terms of amplitude, where $\alpha = \beta/2$, and is termed the 'slope parameter':

$$amplitude(f) \propto \frac{1}{f^{\alpha}} \tag{2.3}$$

By rearranging, alpha can be defined as:

Figure 2.11: The thick curve represents the contrast sensitivity (defined as reciprocal threshold contrast) of the human visual system to a sinusoidal grating, plotted against spatial frequency. The shaded area must always remain invisible to us unless the spatial frequency content of the image is shifted into the visible domain by optical means, such as the microscope. The lighter curves represent channels sensitive to a narrow range of spatial frequencies (from [66]). It should be noted that "humans vary widely in the shape and overall sensitivity of the contrast sensitivity function (normal observers may vary by as large as a factor of 3) [68].

$$\alpha \propto \frac{-log\left(amplitude(f)\right)}{log(f)} \tag{2.4}$$

Knill et al asked subjects to discriminate between random noise textures based on their spectral drop-off, and compared this with an ideal observer. This *ideal observer* is one which takes the place of the subject in the same experimental setup, but can use all available information. So, rather than displaying a stimuli, then using a human observer to provide feedback as to their observation, the task can be put into a closed loop using an ideal observer. Using this approach, they showed that the ideal observer was uniformly able to discriminate changes in slope, but the human observers were best in the range $2.8 < \beta < 3.6$, within the range of $2 < \beta < 4$ exhibited by natural images. They conclude that their results are consistent with a visual system tuned to an ensemble of images with a β of approximately 3.

Alongside investigations into blur discrimination, insight was being gained into the nature of real-world images. However, there was obvious conflict between the observation of images showing $\alpha = 1$ and the tuning of the vision system. Tadmore and Tolhurst attempt to resolve this discrepancy, and also consider the similarity

Figure 2.12: Amplitude spectra for the six [natural scene] images A-F, averaged across all orientations. The spectra have been shifted up for clarity. On these log-log coordinates the spectra fall off by a factor of roughly $1/f$ (a slope of -1). Therefore the power spectra fall off as $1/f^2$. [69]

Figure 2.13: The image on the left has a flat spectrum, whilst the right has a $1/f$ spectrum. [72]

between slope discrimination and blur discrimination [73]. These are similar tasks, as a change in slope changes the amount of high frequency signal present, and thus impedes (or potentially assists) in discriminating blur. They suggest that "even if the visual system were 'tuned' for processing images with natural statistics, it is not clear why one would expect to be particularly good at discriminating small changes from natural statistics" and that instead, subjects would be good at discriminating large changes. That is, only when an image was outside the normal operating region would the vision system start to flag it as being especially unusual.

Further, Tolhurst and Tadmor [74] discuss that increasing the α value of an image causes it to look blurred. They show that their observed slope parameter discrimination threshold approximately matches those blur width discrimination thresholds, across a number of the papers described above. Continuing the tuning argument, and contradicting earlier thinking, they conclude:

> ...a higher threshold represents a greater *tolerance* of change in α. An optimised visual system *should* be most tolerant of image-distortions when the image is in focus... Perception would not then be disturbed by, for instance, any small accommodative errors or changes in pupil diameter that might change α. On the other hand, a high sensitivity to changes in α is required when the image is defocused, so that the appropriate accommodative responses can be evoked. In this sense, the present experiments have shown that the human visual system may be optimised for the processing of natural images.

(a) High edge density ($\alpha = -1.133$) [75] (b) Low edge density ($\alpha = -1.296$) [76]

Figure 2.14: As the number of edges increases, the slope parameter changes, despite the fact that both images are in focus.

Field proposed a model, "RCS", that was intended to overcome two main problems: Firstly, that α varies between images, and secondly that an image of white-noise with a uniform frequency distribution appears to have too much high frequency information (Figure 2.13). Field's experiment involved changing α for a series of images that were captured in-focus, then asking subjects to browse through values of α until they determined that the image was "just blurred". The results suggested that "the amplitude spectrum of natural scenes is not sufficient to predict when an image is in focus or when human observers will judge an image to be blurred. The slope of the amplitude spectrum is a result of both the amplitude of the structure at different frequencies (eg contrast of edges) as well as the density of the structure (number of edges). Blur, however, appears to depend entirely on the amplitude of the structure" [72]. For example, see Figure 2.14.

Parraga and Tolhurst added random contrast variation to images whilst asking the observer to discriminate changes in slope [77]. This was done in such a way to tease apart whether the observer was actually discriminating the slope change, or whether they were performing a single-frequency-band contrast discrimination. They showed that it was the latter that was happening, and that the change in slope could not be directly detected. Thomson and Foster, acknowledging this controversy, show that phase information clearly plays a big part in encoding the structural information of a natural scene [78]. They demonstrate that disrupting the phase information completely removes any the tuning that had previously been reported.

Billock reviews eleven sets of natural images that have been published, and shows that their average slope parameter is 1.08 across all 1176 images [68], observering that there is no significant difference between the slope, regardless of photographic technique, calibration or computation used in the underlying studies. He says that

adults are tuned to a slope parameter around 1.09-1.20, and that this is not present in infants suggests that either a genetic or adaptive influence affects the development of the visual system.

Later work by other researchers has shown that perception of natural images varies with crowding (ie the presence of distractors around the target) and whether the stimulus is viewed in the fovea or periphery of the retina (eg [79, 80, 81, 82].

2.2.6 Related work

Westheimer investigated the impact on blur discrimination on simple edges as contrast varied, and showed that thresholds rose significantly as constrast is reduced [83]. Wuerger examined blur discrimination when a monochromatic stimulus was presented in different colour channels [84]. She showed that her four subjects had approximately uniform blur discrimination threshold when the stimuli was yellow-blue, but that the usual dipper was seen when using the red-green stimuli. No comparison was made with Fincham's work (see Section 2.1.2).

More recently, a very interesting result whereby the vision system adapts to blur was demonstrated [85]. When asked to indicate, by means of a staircase procedure, whether each image was too sharp or too blurred, the results were significantly different if the adaptation field between stimuli was of a blurred or sharpened image. This result was demonstrated both for temporally and spatially separated adaptation fields (Figure 2.15), as well as temporal edges [86,87,88], but little effect was observed following adaptation to luminance and chromatic patterns [89].

The way different types of distortion affect the perception of images has also been measured [90]. The most interesting result is that if an image has experienced some structural distortion (eg manipulating a wavelet frequency), then the apparent distortion is **reduced** if additive white gaussian noise (AWGN) is added. Similar work used 37 subjects to explore the performance of 7 denoising filters, concluding: "If a trade-off needs to be made then the blurring is more a problem than the remaining noise or artefacts". That is, people are more bothered by blur than by noise [91].

2.3 Focus measures

In the mid 1970s, work started being published about computational autofocus, such as Erteza [92] and Muller and Buffingham [93]. The principle of computational autofocus is maximise a function which produces a 'figure-of-merit' based on its input (a two-dimensional array of pixel brightness values). Chern terms this approach

Figure 2.15: The two centre faces are identical and physically focused, but the right image appears blurry in the sharpened surround while the left image appears sharp in the blurry surround [86]

"pixel based focusing" [94]. Ultimately, a focus function should produce a small value for blurred images, and a big value for images which are in focus.

There is minimal discussion to explain how the focus meaures are constructed. Most appear to be the results of supposition or trial an error, though there are exceptions. Vollath performs analysis of some simple functions to extract the noise and scene-dependent terms, then synthesises new measures without these terms [95].

An early comparison of focus measures by Groen categorised those being evaluated by determining the underlying format of the equation, and proposed three equation types [96]:

$$F^1_{n,m,\theta} = \sum\sum E\left\{\left|\frac{\partial^n g(x,y)}{\partial x^n}\right| - \theta\right\}^m \tag{2.5}$$

where θ is an arbitrary threshold, $g(x,y)$ the grey level at coordinates (x,y) and $E(z) = z$ if $z > 0$; $E(z) = 0$ if $z < 0$. That is, measures of the form F^1 are derivatives of some sort.

$$F^2_{f,\theta} = \sum\sum f\left(g(x,y) - \theta\right) \tag{2.6}$$

In F^2, the function $f(z)$ is one which analyses some statistical property of the image, such as measuring the depths or sizes of peaks or valleys in the image.

$$F^3_{m,c} = \frac{1}{c}\sum\sum |g(x,y) - \overline{g}|^m \tag{2.7}$$

Functions of the class F^3 are the 2D summation of difference between pixel value and mean, raised to an arbitrary power, and subject to arbitrary scaling (typically to normalise the measure).

An alternative set of categories was used in Sun's review [97]. However, neither approach is sufficiently comprehensive to cover all the measures discovered in the literature.

In this work, measures have been grouped based on their underlying core function. Measures which operate on gradient (ie the simple difference between two pixels, perhaps subject to an arbitrary sampling interval) could be represented as a convolution with $[1, 0, -1]$ or similar. However, for clarity, they are listed as gradient measures, not convolutions. All the measures discovered in the literature are listed in Table 2.3 which also summarises a few of their key properties. Several new measures have been created during the course of this work, and have been listed in this table for completeness. They are denoted by a double-asterisk, and described in Appendix C.

Method	**Year**	**Category** [97]	**Eq** [96]	**Core function**	**Parameters**	**Summation**	**Thresholded**
cranepeak [15]	1966	Derivative		Gradient	2D	No	No
cranesum [15]	1966	Derivative		Gradient	2D	Yes	No
tenengrad [98]	1970	Derivative		Convolution	3x3 Sobel	Of square	Yes
brennergradient [99]	1971	Derivative		Gradient	1D, step=2	Of square	Yes
menmay [100, 101]	1972	Statistical*	F2	Histogram		Yes	No
thresholdedcontent [100, 96]	1973	Intuitive	F2	None		Yes	No
thresholdedpixelcount [96, 102]	1973	Intuitive	F2	None		Of number of pixels exceeding intensity	Yes
energylaplace [93, 96]	1974	Derivative	F1	Convolution	3x3 Laplace	Of square	No
squaredgradient [93, 96]	1974	Derivative	F1	Gradient	1D	Of square	Yes
masgrn [103, 101]	1975	Statistical*		Histogram		Of bins	Yes
absolutegradient [104, 96]	1976	Derivative*	F1	Gradient	1D	Of absolute value	No
thresholdedabsolute-gradient [96]	1985	Derivative*	F1	Gradient	1D	Of absolute value	Yes
imagepower [96, 99]	1985	Intuitive	F2	None		Of square	Yes
groenvariance [96]	1985	Statistical	F3	None		Of squared distance from mean	No
normalizedgroenvariance [96]	1985	Statistical	F3	None		Of squared distance from mean	No
absolutevariation [96]	1985	Statistical*	F3	None		Of distance from mean	No

Table 2.3: Comparison of focus measures (Part 1). (* denotes my categorisation along the same lines as [97])

Method	**Year**	**Category** [97]	**Eq** [96]	**Core function**	**Parameters**	**Summation**	**Thresholded**
autocorrelation [105, 97]	1987	Statistical		Autocorrelation		Yes	No
standarddeviationbased-autocorrelation [105, 97]	1987	Statistical		Autocorrelation		Of distance from mean	No
voll4 [105, 99]	1987	Statistical*		Autocorrelation	[0,1,-1]	Yes	No
voll5 [95, 99]	1988	Statistical*		Autocorrelation	[0,1]	Yes (less mean)	No
sml [106]	1989	Derivative		Convolution	3x3 Laplace	Of absolute value	Yes
rmscontrast [107]	1990	Statistical		None		Of squared distanct from mean	No
triakis7d [101]	1991			Voxel statistics		Of matching voxels	Yes
triakis11s [101]	1991			Voxel statistics		Of matching voxels	Yes
entropy [101]	1991	Histogram		Information		Yes	No
range [101]	1991	Histogram		Max-Min		None	No
chernfft [94]	2001			FFT		None	No
histogramentropy [94]	2001	Statistical*		Histogram		Of entropy	No
hlv [94]	2001	Statistical*		Histogram		None	No
laplace [94]	2001	Derivative*		Convolution	3x3 Laplace	Yes	Yes
smd [94]	2001	Derivative*		Gradient		None	No
nrbm [64]	2002	Derivative*		Convolution		Of edge widths	No
waveletw1 [108]	2003	Derivative*		Wavelet		Of absolute value of HL, LH and HH sections	No
waveletw2 [108]	2003	Derivative*		Wavelet		Of squared distance of absolute value from mean absolute, in each of HL, LH and HH sections	No

Table 2.4: Comparison of focus measures (Part 2). (* denotes my categorisation along the same lines as [97])

Method	**Year**	**Category** [97]	**Eq** [96]	**Core function**	**Parameters**	**Summation**	**Thresholded**
waveletw3 [108]	2003	Derivative*		Wavelet		Of squared distance to mean of each HL, LH and HH sections	No
kurtosis [109]	2004			Kurtosis		Of DCT	
JNBM [61]	2006			Statistics			
va [110]	2006			Attention		Yes	No
CPBD [63]	2010			Statistics			
energylaplace5a, based on [111]	**	Derivative*		Convolution	5x5 Laplace	Of square	No
energylaplace5b, based on [111]	**	Derivative*		Convolution	5x5 Laplace	Of square	No
energylaplace5c, based on [111]	**	Derivative*		Convolution	5x5 Laplace	Of square	No
rawlaplace	**			Convolution	3x3 Laplace	Yes	No
phasecongruence based on [22]	**			Phase congruence		Of features	Yes
phasecongruence2 based on [22]	**			Phase congruence		Of features	No
randomnumber	**			Random		None	No
alphaAdult	**	Statistical		Spectrum		None	No
alphaRedOnion	**	Statistical		Spectrum		None	No
alphaImageEnsemble	**	Statistical		Spectrum		None	No

Table 2.5: Comparison of focus measures (Part 3). (* denotes my categorisation along the same lines as [97], ** are new measures created during this work, details of which can be found in Appendix C.)

2.3.1 Timing and performance

Early work comparing focus measures paid particular attention to the computational requirements of the various algorithms – Santos reports that their focusing computations took 60% of the total time spent focusing [99]. But, this was done in 1997 when the computing resources they had available ran at 16MHz and had 12MB memory. Seven years later, Sun et al said: "Computation time does not exceed 30ms for almost all the focus algorithms tested. Therefore, the speed of the focus algorithms is not used as a criterion for comparing and ranking [the focus algorithms]" [97].

Even a low end computer is now 100 times faster than the 1997 computer, and modern computation strategies such as off-loading matrix calculations to the graphical processing unit (GPU) can boost performance by another order of magnitude [112]. Interestingly, Santos did conclude that if the execution time had been ignored from their evaluation, then the rankings are almost identical – just *tenengrad* moves places and rises up the ranking [99].

Santos et al excluded certain algorithms (those in the frequency domain) from their survey because "their complexity makes it difficult to produce fast algorithms" [99]. Again, with ever faster computations, such restrictions do not need to be imposed.

Sampling within captured images will also affect performance. As early work used images that were just 64x64 pixels (eg [101]), and modern cameras take significantly larger photos (sizes in excess of 4000x3000 pixels are common), then considering the full information content of the image will take longer. One possible solution is to use subsampling, though such a strategy may require tuning of thresholds and algorithm parameters.

2.3.2 Getting the images

In Groen's 1985 comparison of focus measures, he explains that the images acquired for a real-life image were taken by placing a photograph in front of the camera - "the in-focus distance between the lens and photograph was 520mm. Focussing took place by moving the camera with respect to the object in 20-mm steps. In this image sequence a relatively large change in image content is present related to the relatively large depth of field of the macro setup. This poses a separate problem from the variation in fixed-content images." [96]. Clearly such an approach, whilst it may have merits, is of limited value given both the enormous change in image content and that it is not a normal mechanism for focusing a camera – the photographer

will normally stay still, and change the optics rather than the other way around.

Most of the literature about mathematical focus measures surrounds the autofocusing of microscopes. Typically, a set of images are captured by changing the focus step of the optical system by a few μm, whilst maintaining constant illumination (eg [96, 97]). For example, Santos explains that their images were captured at steps of 0.025μm, had a depth of field of 0.16μm, and had an exposure time of 0.3s which permitted 185 grey levels to be captured [99]. In summary, by whatever method was appropriate to the object being imaged, a set of photos was captured.

2.3.3 The nature of blur

The exact nature of focal blur is of importance. Ultimately this is a result of the point spread function (PSF) of the optical system between the object and imaging sensor. A number of different functions are discussed in the literature, including cosine or gaussian operators, and manipulating the amplitude spectrum [92, 113, 19, 87, 114]. Murray and Bex suggested that the blur (and sharpening) caused by manipulating the amplitude spectrum does not simulate perceptual blur. However, both sinc (that is, $sin(x)/x$) and gaussian blurred images do – each produced dipper-shaped blur discrimination thresholds. Further, they showed that blur-equivalence between gaussian and sinc blurred images as determined by human observers was better reproduced by models based on luminance slope than those based on spatial frequency. Thus, they suggest that phase components of images are important when measuring perceived blur, indicating that a sinc operator might be more appropriate than a gaussian [115].

Indeed, in terms of physics, the optical blur at a given wavelength is a sinc shaped PSF[5], but the aggregate effect of summing this across multiple wavelengths is approximately gaussian [19, Appendix], [96, Fig 2]. Whilst Murray and Bex suggested that phase is important, they did not apply the sinc function at multiple wavelengths, thus their results are not directly applicable to simulating optical blur in non-monochromatic scenes. Accordingly, almost all work has used a gaussian kernel as the mathematical means of simulating blur, and equate the standard deviation of the gaussian kernel with the arc-minutes of blur extent. No description has been found in the literature regarding this relationship, nor how the blur extent could be compared to known physical changes in focus distance or lens strength when capturing real objects.

Whilst the behaviour of a man-made optical system can be calculated, the PSF

[5]Primarily due to diffraction effects which produce wave cancellation and reinforcement (see [19])

of the human eye has to be determined by measurement. It is not a trivial mathematical function, and a significant body of work has been done investigating it, such as that by Roorda et al and Navarro et al who scanned a laser point across the eye and used a camera to record the resultant image on the retina [116, 117].

Synthetic images have been used in several experiments; typically square waves, sinusoidal gratings and white random noise [84, 101] which have then had gaussian blur added. Bex went further and created a random noise image, then low-pass filtered and thresholded it to generate a binary monochrome image, which results in a random image with an amplitude spectrum equivalent to a natural scene [114].

2.3.4 Establishing the ground truth

An important property of a set of images captured for the purpose of analysing the performance of focus measures is to know which is the most in focus; that is, to establish the ground truth. The literature does not indicate the existence of an established means of doing this across a wide range of applications. Muller, when considering how to compensate for atmospheric perturbations of astronomical objects, mathematically proved that certain "sharpness functions reach their maxima only for a properly restored image" [93]. However, this application can use point light sources and zero depth of field to perform modelling.

Whilst it would technically be possible (if looking at a flat object perpendicular to the camera's axis) to compute the precise ground truth, there is no evidence in the literature that this has been done. Such an approach could only be used for three dimensional scenes, where there will be a multitude of distances, if the depth of field were also taken into account. Again, no discussion of such a strategy has been found. Instead, the ground truth appears to routinely be established by human decisions, and rarely discussed. For example, one experiment took 28 focus images, and simply stated that the middle of the sequence was the "visual in-focus image" [96]. The best image in other papers was "manually determined by proficient microscope technicians" [97] and "obtained by a trained operator" [99]. Rather than determine a ground truth, Chern et al reviewed the image of best focus for each measure, and noted that "visual inspection reveals virtually no difference between the frames" [94].

2.3.5 Quantitative comparisons

Once a set of source images has been captured and processed by a candidate focus measure, it is desirable to quantitatively describe the result. An early quantitative

comparison was made by Groen, who normalised the resulting focus scores to have a maximum of 1 (though no scaling to anchor the minima to a particular value was performed), then compared the width of the focus measure at 50% and at 80% of the maxima [96]. Building on this, a standard approach to characterise and rank candidate measures is used by a number of different researchers (eg [97, 101]). It uses the following parameters:

- **Accuracy (A):** Distance between maxima of the focus curve and the ground truth of 'best' image, measured in number of image frames of distance.
- **Range (R):** The distance (in number of images) between the first minima on either side of the global maxima. This should be large, as there should not be any local maxima on the focus curve.
- **Number of false maxima (F):** The number of maxima appearing in a focus cuve, excluding the global maximum.
- **Width (W):** The width of the curve (in number of images) at 50% of the maxima's height. Ideally this should be small.
- **Noise level (N):** This describes the speed of the direction of change between two false maxima of a focus curve. It is computed by taking the sum of squares of the second derivative obtained by convolving the curve (ommitting the peak value) with the kernel $(-1, 2, -1)$.

Figure 2.16 shows these parameters annotated on a typical focus curve.

An additional parameter was suggested by Ligthard and Groen in an early paper – that the algorithm should be able to use the same video signal as that used for the ultimate image capture so as to avoid any systemic errors that might arise from a hardware implementation using a different video source [118]. However, given the advances in digital image processing and acquisition, such a precaution is likely not to be necessary.

So as to measure the robustness of the algorithm under test, Sun et al also measure these parameters on three additional versions of the input – after subsampling, adding random noise and low-pass filtering. The overall score for each focus measure is then computed as the Euclidean distance from the perfect score, where the perfect score is where all parameters are zero, except for the range which is the number of images under test. Sun et al normalised each parameter before computing the score, giving all distances equal weights. Once normalised, the score is thus:

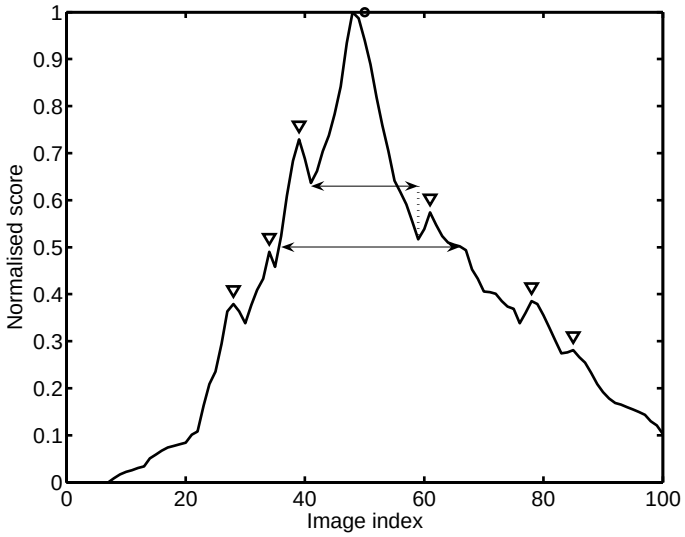


Figure 2.16: An example focus curve annotated with the attributes used to quantitatively compare this algorithm against others. Accuracy is 2 (the distance between peak of the curve and ground truth 'best' image denoted by a circle at $(50, 1)$. The range is indicated by the upper horizontal line, and is the distance between the first minima on either side of the peak (18). There are six false maxima, indicated by downwards arrows. The lower horizontal line shows the width of the curve at 50%, which is 30.

$$score = \sqrt{A^2 + (numimages - R)^2 + W^2 + N^2 + F^2} \tag{2.8}$$

None of the existing comparison studies in the literature have covered all focus measures, and indeed one study failed to include an overall table showing how the measures ranked across a number of different assessments [97]. Their results have been consolidated to produce an overall rank to show which measure performs best, as can be seen in Table 2.6. Sun concluded: Across a wide range of images, conditions and pre-processing steps, the *normalisedvariance* measure was determined to be the best. The only exception is when the images are subsampled [E7 in Table 2.6] where *tenengrad* is best.

Name	E1	E2	E3	E4	E5	E6	E7	E8	E9	Total	Rank
thresholdedabsolutegradient	7	8	7	7	6	10	3	8	4	60	6
squaredgradient	8	10	8	8	8	8	5	9	7	71	7
brennergradient	5	6	5	5	5	7	2	6	6	47	5
tenengrad	4	5	4	4	4	6	1	5	5	38	4
sml	14	14	14	13	14	13	7	14	12	115	14
energylaplace	11	11	10	9	10	14	16	12	9	102	11
waveletw1	16	15	15	14	15	12	12	17	14	130	16
waveletw2	15	12	16	12	13	16	11	16	16	127	15
waveletw3	12	13	12	10	12	15	10	15	15	114	13
groenvariance	3	3	3	3	3	3	13	2	2	35	3
normalisedgroenvariance	1	1	1	1	1	1	6	1	1	14	1
autocorrelation	6	9	6	6	7	9	17	7	8	75	9
stddevcorr	2	2	2	2	2	2	14	3	3	32	2
range	13	16	9	11	16	11	4	11	11	102	11
entropy	18	18	18	18	18	18	15	13	18	154	17
thresholdedcontent	10	7	13	16	11	5	9	10	13	94	10
thresholdedpixelcount	17	17	17	17	17	17	18	18	17	155	18
imagepower	9	4	11	15	9	4	8	4	10	74	8

Table 2.6: Score summary across the different tests performed by Sun [97]. E1 uses no magnification; E2 uses 100x magnification; E3 uses 400x magnification; E4 used bright field observations; E5 used phase contrast observations; E6 was observed with DIC; E7 used 5% subsampling; E8 added random noise; E9 applied low pass filtering. Each of these were performed with at least 18 image series.

When Sun et al aggregated their results for a particular experiment acrosss a number of different image series, they simply averaged the score. Santos proposed a different approach [99], though this appears not to have been used by any subseqent work. It is calculated as follows:

1. Compute the mean and standard deviation of each of the five metrics (introduced above) for each image series.
2. Normalise each metric by subtracting the mean and divide by the standard deviation.
3. From these normalised metrics, compute the score (the Euclidean distance from the ideal, as in Equation (2.8)
4. The global score for the function is now the mean of the scores of normalised metrics.

Table 2.7 summarises the ranks of all focus measures evaluated in the most widely cited review papers. It shows that a variance based measure typically performs best.

Focus measure	Sun's rank	Santos' rank	Firestone's rank
thresholdedabsolutegradient	6	7	
squaredgradient	7	11	
brennergradient	5	10	1
tenengrad	4	8	
sml	14		
energylaplace	11		
waveletw1	16		
waveletw2	15		
waveletw3	13		
groenvariance	3		2
normalizedgroenvariance	1	1	
autocorrelation	9		
stddevcorr	2		
range	11	12	7
entropy	17	9	9
thresholdedcontent	10	5	
thresholdedpixelcount	18	6	
imagepower	8	4	
voll4		1	
voll5		3	
triakis11s			3
spectral			4
triakis7d			5
menmay			6
masgrn			8

Table 2.7: Score summary across three review papers [97, 99, 101]

2.4 Cameras

Whilst the eye took many millions of years to evolve, the evolution of photography has been far quicker. The earliest camera is the camera obscura (a pin-hole camera), whose principle was known by Aristotle in 300BC, although the first recorded use much later was in a drawing by Leonardo da Vinci in 1519. The first actual photo was taken in the summer of 1827, by Nipce, and required an eight hour exposure. By the 1850s, exposure times had been reduced to just two or three seconds, but the process required that the plates were freshly made, and still wet, when the photographs were taken. However, this too was resolved in the coming decades, with first gelatin and then celluloid being used to replace the glass plates. Finally, in 1888, George Eastman released the box camera, and photography became available to the masses [119].

Since then, cameras have improved considerably. Some of the most important developments are listed below [120]:

Year	**Event**
1914	Leitz introduces the sprocketed 35mm film.
1932	Technicolor for movies is introduced. Three black and white films in the same camera capture the scene under different filters.
1936	The first multi-layered colour film (Kodachrome) is developed, as is the first 35mm single-lens reflex (SLR) camera.
1955	Minsky develops confocal microscope [121]
1963	The first instant colour film is developed by Polaroid.
1975	Kodak build the first CCD-based still camera
1985	Minolta markets the first autofocus SLR
1991	Kodak release first digital SLR, the DCS-100, which is a modified Nikon F3
1999	Nikon D1 SLR 2.74 mega pixel camera, the first ground-up digital SLR
2000	The first camera phone is introduced in Japan by Sharp and J-Phone
2004	Kodak stop producing film cameras
2005	Plenoptic (light field) camera realised
2009	FujiFilm consumer 3D camera launched

The following sections describe in more detail the evolution of certain features, such as auto-focus, as well as introducing some of the latest research into how cameras could be further improved.

2.4.1 Auto focus in conventional cameras

Despite the large number of focus measures described earlier, literature surrounding the actual techniques used in cameras is scarce, probably because of reasons of commercial confidentiality. However, there are fundamentally two approaches. Firstly, the camera can actively measure the distance to the subject, using a variety of mechanisms, and then use a look-up table to determine the lens position for a given subject distance. Alternatively, the camera can perform some image analysis and change the lens position until the image is determined to be in focus (eg [122]). As most scenes contain vertical lines, this is typically done by varying the lens position to maximise the contrast between two horizontally adjacent pixels. An improvement upon this, used by modern cameras, is to consider multiple pairs of pixels, arranged in a grid around the centre of the image, and even to allow the photographer to move the focus point to the subject of their photo [123].

More recently, manufacturers have released cameras with significantly more advanced focusing strategies, such as:

AiAF "Canon's 9-point AiAF (Artificial Intelligence Auto Focus) automatically scans and selects subjects from a set of nine focusing areas across the scene. This ensures accurately focused images even when subjects are not in the centre of the frame." [123]

FlexiZone AF/AE "FlexiZone AF/AE lets users manually select the focus point from almost any point in the frame by moving the auto focus window in the viewfinder. Exposure can be linked to the focus point to ensure that the chosen subject is both accurately focused and exposed." [123]

Face Detection AF/AE "Face Detection AF/AE ensures superb people shots by automatically detecting the subjects in the frame and setting the optimum focus and exposure" [124], though it is noted that faces may not be detected if they "appear small, large, dark or bright relative to the overall composition, [or] if the subjects are looking sideways, lying down, or their faces are partially obscured" [125, p45].

Face-priority Auto Focus "A special digital detection program ... [scans] for facial details and then controls autofocus operation based on the location of

the detected face in the scene." [126]

Auto depth-of-field Canon's digital SLR cameras incorporate A-DEP, where "aperture is determined to maximize depth of field so that all objects in the 9 focus points are sharply focused" [127].

Other autofocus strategies include using additional information sources, such as triangulating the location of the speaker in a video conference by analysing the time-delay-of-arrival of the speaker's voice to an array of microphones [128].

Future cameras may well abandon solid lenses for liquid ones, which are stimulated by electric fields to change shape (and hence focus), just as with the human eye. The advantages include the elimination of moving parts, lower power consumption and smaller size. Jung et al explain how a liquid lens can be improved such that it could be used in a portable device [129].

2.4.2 Other devices

Certain devices have been invented over the years which are immune to blur. The confocal camera, patented in 1955, uses a pinhole to eliminate out-of-focus light, in conjunction with point-wise illumination [121]. However, the trade-off is that image acquisition is accordingly slower, as the illumination has to be scanned over the sample.

A novel approach for image acquisition called Compressing Imaging was proposed by Wakin [130]. Its main benefit is that just a single-pixel sensor is required, allowing imaging to be performed in spectral regions where a CCD or other matrix-sensor is financially prohibitive or technically impossible to construct. The sensor is used in combination with a micromirror array showing pseudo-random binary patterns. The mirror is set to a particular pattern, and the sensor is instructed to acquire a data point. The pattern is then changed repeatedly and new data points acquired. This approach enables a rough image to be displayed after just a few samples, and for each subsequent sample to simply improve the image quality. This is shown in Figure 2.17.

2.4.3 Recent developments

In addition to changing the camera's parameters (such as shutter speed or aperture) to take an individual image, it is possible to perform post-processing on a range of photos to improve the output. Such techniques include high dynamic range (HDR)

(a) Ideal image (b) 819 measurements (c) 1600 measurements

Figure 2.17: Compressive imaging example: (a) shows an ideal image of 64x64 pixels (4096 pixels). This was reconstructed using compressive sensing using (b) 819 measurements and (c) 1600 measurements. (Adapted from [130]).

imagery, whereby the results of conventional exposure bracketing are then combined to produce a single image with high dynamic range [131].

Some of Fujifilm's digital cameras incorporate technology in their sensor to improve its performance in terms of dynamic range [132]. This is done by having two sensing elements per pixel, one with lower sensitivity than the other (see Figure 2.18), which attempts to recreate the variety of grain sizes found in conventional film photography (and is similar to the difference in sensitivity between the human eye's rod and code receptors).

Figure 2.18: Fujifilm's 4th generation Super CCD HR [133]

HP has added image processing, rather than specific sensor hardware, to attempt to improve the perceived dynamic range of the camera's images [134, 135]. Whilst they acknowledge that "a sophisticated user may be able achieve some of the benefits of HP Adaptive Lighting Technology in image editing packages", they go on to say that "while these techniques are possible, they are very difficult, time consuming and beyond the ability of most users.". HP's technology operates automatically, and is also able to make use of the greater dynamic range available in the camera. This means that their algorithm can process raw images before they are reduced to

a 24bit JPEG image, so can outperform computer based image manipulation.

Other work has looked at focus, and strategies for improving the depth-of-field beyond that of simply having a very small aperture. Most notably, the implementation by Ng et al's of a plenoptic camera [136] which yields superb results (see Figure 2.19). By placing an array of microlenses directly on top of the CCD sensor, plenoptic cameras can capture more of the light field inside the camera itself. The result of this is that, for a single exposure, it is possible to computationally refocus the image as well as being able to computationally move the observer both laterally with respect to the subject, and also in terms of distance. These manipulations are performed on the light field photograph **after** acquisition – that is, a scene can be refocussed years after the image was taken.

(a) Conventional photo, focused on the clasped fingers

(b) Extended depth of field computed from a stack of photographs focused at different depths.

Figure 2.19: Sample photograph from a plenoptic camera [136]

Arbitrarily large synthetic apertures have also been created, meaning that images can be captured with a very narrow depth of field. This can be used to "see through" objects, such as in Figure 2.20 [137].

Focus (or defocus) information can also be used to build a three-dimensional model of the scene [138,139], a technique known as "shape from focus". In essence, if a large number of photos are taken of the scene, they can be processed to determine the optimal focus position for each pixel, and hence produce a three-dimensional model of the scene, much as could be achieved with stereo cameras or laser range finders. Figure 2.21 shows an example of such processing.

In summary, camera manufacturers are keen to demonstrate differences between

(a) Conventional photograph

(b) Large aperture allows for narrow depth of field, eliminating the tree from the foreground.

Figure 2.20: Synthetic aperture photograph [137]

(a) Original image (near-focused)

(b) Final generated shape model

Figure 2.21: Depth from focus example [139]

their cameras and those of their competitors. They have done this by making changes both at the software level, and by improving and evolving the hardware. Some of the software techniques now being added to consumer cameras simply make it easier (or even make it practical at all) to achieve some desired end result, despite the fact that the end result has been technically achievable for some time, if the photographer has the time to spend working on his photos. Other changes will take a while to reach the consumer market, and when they do arrive could create a big change in the way photographs are taken and used.

2.5 Summary

Since Helmhotz's proposal that the eye accommodates by changing shape over 150 years ago, much investigative work has been done to understand how accommodation is controlled. The EW group of cells (see Section 2.1.3), mentioned above, has been found to be responsible for accommodation, but their exact behaviour has not yet been characterised. Many stimulii have been identified, yet accommodation is still possible in their absence. Around 1/3rd of the population can focus when stimulii are removed, such as viewing an unknown subject, monocularly, in monochromatic light. This leads to the conclusion that another stimulus is used; that the vision system tries to reduce blur.

Crane was the first to question what mathematical property of the image is used to calculate blur, and proposed the first focus measure. Since then, almost fifty years have passed, seeing the publication of a multitude of models of accommodation, each depending on quantifying 'blur' but no such quantitative measure being proven.

The psychophysical behaviour of a number of visual properties has been investigated and typically shows that discrimination exhibits a dipper response with increased base magnitude. The seminal work on blur, Watt and Morgan (1983), showed that the perceptual width of a blurred edge is related to stationary points in the second derivative. Watt also showed how such a model could explain various optical illusions, and was the first to publish a dipper shape when measuring blur discrimination.

Watt's experiments, and those of most other researchers since then, have been conducted with simple 1-dimensional stimulii. Just two experiments stand apart from this trend: Walsh found a dipper-function for blur discrimination using real photographs, though without using accommodation (the subjects were anaesthetised and moved the photographs which were blurred by the insertion of lenses) – this is not a natural viewing experience. Secondly, Kayargadde developed a means of

quantifying blur in real images, and showed this could produce similar mean opinion scores as humans, though did not try to use this to reproduce psychophysical results. Similarly, Ferzli and Karam developed metrics for quantifying blur such that images of different scenes could be compared to establish which was most in focus.

Statistical patterns in natural scenes were found by Carlson, which led to the observation that $amplitude \approx 1/f^{\alpha}$ – that is, across a wide range of natural scenes there was found to be a relationship between the amount of energy at each frequency, and the frequency itself, with the relationship characterised by the slope parameter, α. Many papers tried to understand α, and to show that human vision is optimised for the alpha values found in the natural world, though there is debate about what "optimised" would mean and what would be exhibited by such optimisation. Whilst changing α does change the degree of blurriness in the image, there is no α which universally corresponds to the sharpest image, nor does changing α produce the sort of defocus produced by optical systems.

Tadmore and Tolhurst suggest α drives accommodation: "a high sensitivity to changes in α is required when the image is defocused so that the appropriate accomodation response can be evoked". Though, Field subsequently shows "α is not sufficient to predict when an image in in focus", and thus it seems unlikely that it could be the measure used to focus the eye.

Approaching the subject of blur from the engineering perspective are the autofocus algorithms in cameras and microscopes. Since the 1970s a growing assortment of focus measures have been developed. Based on numerous different underlying approaches, these measures have been compared with a widely used evaluation methodology. Published results show that a normalised variance method performs well.

Modern cameras have moved beyond simple autofocus, adding exciting features that, one assumes, the manufacturers believe will drive sales. On the research front, image processing in cameras is showing impressive results. The plenoptic camera, by capturing the entire light field, means the image can be focussed after acquisition. Other fields depend on autofocus and have very specific requirements – automated screening can require time-sensitive reagents to be added to microscope slides, and for images to be captured automatically within a particular time frame. However, no general purpose algorithm has been proposed in the literature.

2.6 Discussion

Despite clear progress being made in the understanding of the human vision system, there remains an uncertainty about the method by which blur is quantified. The

exploration of the neurological pathways involved in accommodation has clearly progressed over the past century, though the difficulty of determining brain function means that some exploratory procedures appear very crude – for example, making lesions in regions of the brain and observing their impact (eg [31]). Preliminary work by Fylan measured the VEPs when looking at blurred stimulii and might shed more light on the neural interpretation in a less invasive manner. Regardless of the techniques employed, it is clear that whilst the pathways are known, there is no explanation from this line of research as to how the image formed on the retina is interpreted and appropriate signals formed for the ciliary muscles.

Various psychophysical behaviours, responses and thresholds have been established, but to-date a blur discrimination assessment has not been performed with real images - stimulii tend to be simple edges. Whether similar results are obtained with natural scenes when subjected to a natural point spread function has not been shown.

Many theories have been proposed to explain perceived blur. Bex summarises these theories saying: "For example, the perceived blur of the edges in an image could depend on the gradient at the zero crossings [43], on the separation between peaks in either the second derivative of luminance [45] or in the summed outputs of a bank of band-pass filters [113], the scale of the filter producing peak response to a blurred edge [140], the slope of the amplitude spectrum of the image [9] or the relative contrast at high spatial frequencies [74]". Each can explain different portions of the response, but no comprehensive or universal model has been derived or tested.

Another thread of model development has examined the human visual system's response times to quantified inputs - the same approach used when characterising physical systems. However, the latency, settling time, initial error, overshoot, damping etc whilst of value, are not necessarily applicable when looking at a stationary target. Philips showed that the same responses are elicited when showing a blurred stimulus as showing an image of a blurred stimulus [26], greatly increasing the practicality of conducting studies into blur. Thus, whilst there might be possible model improvements, the lack of a measure of perceptual blur (eg see Figure 2.5) is an area ripe for further investigation.

Measures of blur are widely used in cameras and microscopy though there appears to have been no attempt to use these measures to reproduce any psychophysical results. Instead, they have been compared by rating their accuracy and characteristics of the focus curve, with variance-based measures being found to be best. No one has proposed a methodology (or results) for establishing such a curve for humans

observing a scene, nor have real world scenes been examined, nor proper ground truth established.

Modern cameras have created a plethora of additional focus related features to help differentiate them in the consumer market place. However, there will always remain scenes whose optimal focus distance can only be found by knowing the interests of the observer; the top-down objective. Perhaps the point of interest in the scene (at a later date) differs from the original subject, or simply that different viewers of the same scene have different interests; a geologist and botanist looking at an alpine photo will perhaps be drawn to the rock or flora respectively. Advances in light field photography mean such a dichotomy of requirements can now be met by post-processing, but do mean that all need their focus distance establishing - optical focus is eliminated. Thus, despite the change to focusing after capture (rather than before) these advances still have a need to focus.

Establishing the consensual 'best' focus across a range of subjects is essential to better compare focus measures and understand population diversity, but appears not to have been done before.

The performance (in terms of computational time) is discussed in many papers, though with continuous improvements in general purpose computing and the ever present opportunity to design ASIC solutions mean these should not be considered a priority when comparing focus measures. The number of frames required to be captured, and any hill climbing strategies for finding a global maxima that reduces this figure are similarly areas for design and production optimisation, rather than of any immediate value when assessing the performance of focusing strategies.

2.7 Thesis statement

As Wang says: "Human observers are bothered by blur, and our visual systems are quite good at reporting whether an image appears blurred (or sharpened). However, the mechanism by which this is accomplished is not well understood" [23]. Mechanisms for autofocus are well developed, but have not been compared with human opinion or perception. The literature, discussions and conclusions lead to the thesis that will be addressed in this research, which can be stated as:

"It is hypothesised that it is possible to construct an algorithm that accurately replicates human perceptual and subjective experimental results"

2.8 Key research questions

The thesis statement, and the discussions from a review of the literature lead to key research questions which need to be addressed. Throughout this work, a range of real-world images should be used.

1. Can ground truth data be obtained that is suitable for testing focus measures?
2. How well do focus measures perform when compared against the ground truth?
3. How can human blur opinions and perception be measured?
4. Can human results be compared with those from focus measures?

The following chapter introduces the methodology, and more details of the various algorithms and processes that will be used in answering these questions. Chapters 4 onwards move on to provide the results of experiments that have been performed, and then Chapter 8 discusses the results and draws conclusions to answer the thesis, before discussing future work that could be done.

Chapter 3

Methodology

In order to answer the key research questions, different experiments need to be undertaken and analysis performed. Each step of the process requires careful preparation of both input data and experimental procedure as well as the selection of appropriate analytical techniques for understanding and interpreting the results.

This chapter describes experimental procedures that have been used in previous work, compares and contrasts their approaches, and describes practical considerations about how the activities are to be performed, then moves on to discuss the necessary data analysis.

3.1 Image selection and acquisition

To explore the estimation of defocus of humans and algorithms, a library of images is required. For each scene multiple images, each corresponding to different focus depths are required. Despite an extensive review of prior work, there appears to be no publicly available library of images captured at different focus positions. Such image sets have been captured for related work (eg [141,142] which attempt to merge multiple images of the same scene to achieve an infinite depth of field), but neither the source of these images nor a description of their acquisition is provided. In the future, it is likely that light field photography (eg [136]), will be able to produce the required set of images in a single exposure. Unfortunately, the present state of the art of light field photography is of relatively low resolution.

Accordingly, it was necessary to capture scenes for use in this work. As multiple images need to be captured for each scene, it was found to be impractical to acquire these manually. Despite great care, adjusting the camera's lens frequently resulted in slight movement of the tripod, and thus changed the view of the scene. Instead, it was necessary to capture images automatically.

Some vision research areas, typically those examining the visual pathway and processing of natural scenes, warrant careful camera calibration. Specifically, there is a danger that observed behaviour in humans could be a result of artifacts arising within the camera's image forming process, and not actually due to the intended natural scene stimulus. Brady and Legge provide an extensive camera calibration methodology [143]. They compare a number of lenses and camera bodies, and show (of the ten characterisation parameters they examine) some parameters vary as the camera's lens, zoom, exposure, aperture, ISO and object distance vary. However, they report that two variables, object distance and lens focus, cannot be characterised, thus rendering their procedure of doubtful use to this work. On this basis, camera calibration has not been performed, a decision which is especially mitigated by the fact that there is no requirement to vary camera parameters other than focal distance during image acquisition, thereby not needing the benefit of Brady and Legge's approach to handling different luminances, zoom positions, or other camera settings.

That is, whilst it is possible to establish the precise configuration of lenses required to ensure light rays at a particular point in a scene are focussed, characterising a given camera lens for its entire field of view is an arduous task, and not central to this work. Instead, it is assumed that a camera (or other imaging device) can produce images of a given scene at a range of focal distances. This work then explores how these images can be compared to determine which is the 'best' for the entire scene. Such an approach is independent of the optics of the camera, and other device-specific artifacts arising from image acquisition.

3.1.1 Capture

There are several approaches for automatically capturing images from a camera, as follows:

1. **On-camera script on the DC-290**: Certain cameras, such as Kodak's DC-290, allow for programs to be written on a computer and downloaded to the camera, using the Digita language [144]. Whilst the scripting language allows for a large number of possible focus positions to be specified, testing on the camera itself revealed that only seven different positions are achievable. See Appendix D.1 for further information.

2. **Programmable video cameras**: Sony have manufactured a range of video cameras, including the EVI-D30, designed for use in video conferencing applications. To facilitate their installation and use, they can be controlled via a

(a) Olympus C-2040 Zoom (b) Olympus 500-UZ

Figure 3.1: Cameras

serial communications protocol which (in addition to controlling pan, tilt and zoom) allows the in-built auto focus mechanism to be disabled and the camera focussed at an arbitrary position. See Appendix D.2 for further information.

3. **PTP camera control of still digital cameras**: Many cameras built since 2003 support the picture transfer protocol (PTP), which allows computers and printers to communicate with cameras to download and print photos. This protocol is not solely one-way, but does allow the host to control the camera to some extent. Whilst most cameras allow for the computer to request a photo be taken, and some allow for the shutter and aperture to be specified, only Olympus cameras allow their focus to be specified.

4. **Bespoke scientific cameras**: Some papers describe highly specialised equipment for automatically capturing images (eg [145]). However, because of the high cost associated with such equipment, and that calibration is not considered in this work (which is centred around image selection from a set of images with varying de-focus, rather than precise focus points), this route has not been pursued.

The Sony EVI-D30 responds rapidly to focus commands, which together with a PC-based video capture card permits images to be captured very rapidly. Unfortunately, image quality obtained from a video capture device connected to PAL cameras is of significantly lower quality than using digital still cameras, as a result of interlacing and ultimately of the sensor being a lower resolution. As such, this device was not used for any detailed studies described in the subsequent chapters.

Figure 3.2: Fresh flowers moved during the course of a photography session. Note how the top petal of the central blue flower moves in relation to the centre of the white flower

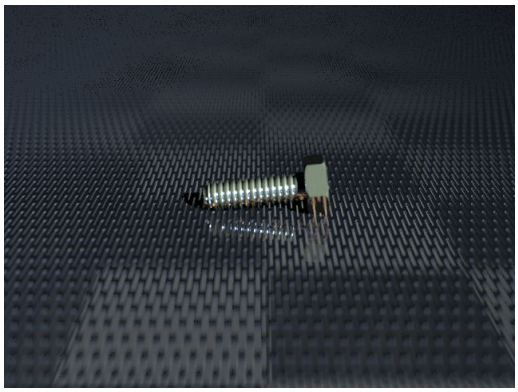
Two cameras were evaluated using the PTP protocol; Olympus's 2 mega-pixel C-2040Z and 6 mega-pixel 500UZ. Both of these were assessed using the open source ptpcam/libptp software libraries running under Linux and using Pine Tree Computing's Camera Controller on Windows. Both cameras allow their focus to be controlled to one of 240 positions, though only the 500UZ model has sufficient memory and battery life to actually capture that many images in a single session and was the camera used in this work. (A set of 60 images took approximately 12 minutes to acquire using Camera Controller).

3.1.2 Scene

Independent of the technique used for capturing images is the preparation of the scene that is being photographed. Ideal sample images should include a clear subject, which occupies and prominent and substantial portion of the scene, and feature relatively obvious edges.

As the acquisition takes a non-zero length of time, natural fluctuations in daylight (such as from the movement of clouds) cause brightness differences between images. To minimise the impact of this, it is necessary to ensure the scene is illuminated by artificial light and away from any air movement. Despite such precautions, certain objects did move. A bunch of fresh flowers was considered as a possible scene. Unfortunately, the flowers opened and changed shape during the course of the capturing, rendering them an unsuitable subject (see Figure 3.2).

Scenes were chosen from two categories. Firstly, scenes with minimal depth of field, such as a flat layer of coins or piece of material. Secondly, objects with a clear difference in depth, such that each image of the scene clearly had a portion of the object out of focus. To achieve this, small objects were used with the camera at



(a) Original [146] (b) Simplified

Figure 3.3: Rendering of a bolt with Povray

a distance typically between 25 and 50cm. All images were captured with a large aperture (thus small depth of field) using the camera configured to use its best quality.

As no technical information about the lens optics was available from the manufacturers, it is not possible to quantify the focal distance for each photograph. Instead, it is solely possible to sequence the photographs in order of software-requested focus position. Each scene was photographed from nearest to farthest focus to minimise any hysterisis, mechanical lag, or other confounding behaviour that the camera might exhibit if captured non-sequentially.

3.1.3 Other sources

Two further sources of images were used: First, the open source ray tracing software Povray was used to generate synthetic scenes. Ray tracing is a physically accurate means of modelling how light interacts with virtual objects within a scene. An image [146] was selected from the Povray Hall of Fame [147], and is shown in Figure 3.3(a). This was then simplied to produce an image for further experiments, as can be seen in Figure 3.3(b).

Povray's default behaviour is to render the scene with a pinhole aperture. This means that all points are in focus, and minimises the computation time required to render each scene, though this is clearly undesirable when trying to generate a stack of varying focused images. However, it is possible to build a stack of images focused at different distances, by specifying the `focal_point` parameter (this specifies the coordinates of the point at which the virtual camera should focus) and using a combination of `blur_samples`, `confidence` and `variance` to control the quality of the rendering.

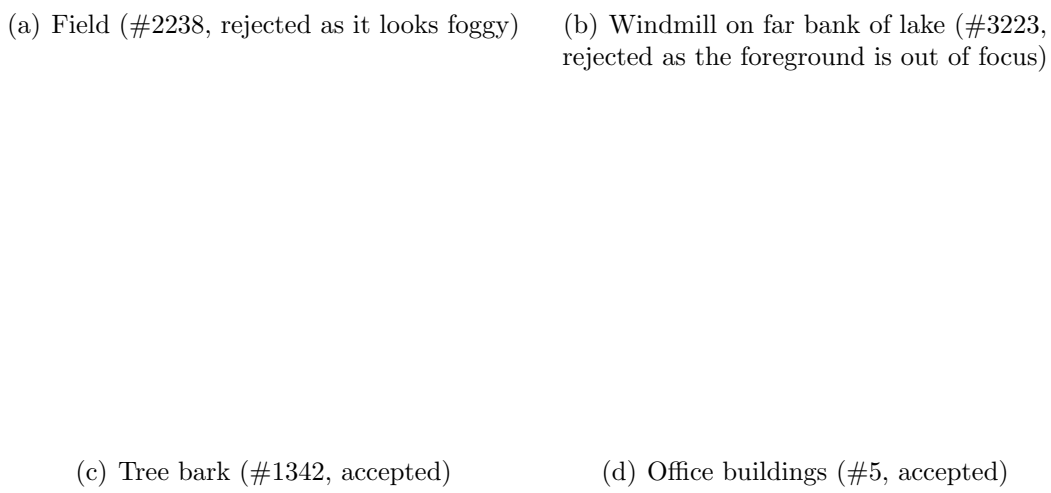
Secondly, to assist with comparing experimental results with published literature, existing calibrated images have also been considered. A calibrated monitor is used in experiments to view images so as to eliminate possible distortions that may confound the results. On such a device, a proportional increase in photon emission rate will result from a fixed increase in image brightness – ie the computer monitor will emit light directly proportional to the pixel value. Thus, the manner in which the mathematical measures will 'see' the image is the same as for humans, and so be a fairer method for comparing the human visual system with computational techniques.

Given this level of careful control, and the fact that work exploring the sensitivity of the human visual system to image properties within natural scenes (eg [148]) makes use of existing libraries of calibrated images, it is logical to consider them for comparison. Related work investigating human sensitivity to contrast use such libraries [149], whilst others explicitly choose to use non calibrated images from other sources (such as still frames from commercial DVDs [150]).

Two image libraries were considered: McGill [151] and Van Hateren [152]. Of these, Van Hateren was selected because its images are greyscale, and thus eliminate any confounding factors arising from the challenges of extracting perceptual brightness from colour images. The Van Hateren library is made available in two data formats, IML and IMC. The IML images are "slightly blurred by the point-spread function of the camera (in particular due to the optics of the lens)". This is reversed in the IMC images "by deconvolving the images with the point-spread function corresponding to the used lens aperture ... therefore this image set is best suited for projects where well-defined edges are of more importance than strict linearity" [152]. Accordingly, the IMC images were used in the experiments described in this thesis. Of the hundreds of photos within the Van Hateren set, images which appeared to be in focus were selected – those with a blurred foreground or background were rejected, as were those that appeared foggy. Figure 3.4 shows some accepted and rejected images.

3.2 Mathematical focus measures

One of the key research questions is to investigate whether autofocus measures can be used to reproduce psychophysical results. It is therefore essential that all measures are implemented in a consistent software interface to simplify their use as model observers in these experiments. The original literature was reviewed for each measure, using review papers from 1985 [96], 1991 [101], 1997 [99], 2001 [94] and 2004



(a) Field (#2238, rejected as it looks foggy)

(b) Windmill on far bank of lake (#3223, rejected as the foreground is out of focus)

(c) Tree bark (#1342, accepted)

(d) Office buildings (#5, accepted)

Figure 3.4: Selected images from the Van Hateren image library [152]

[97] as the primary starting points. Frequently this contained insufficient information to reproduce the algorithm, whereupon subsequent work citing the original approach was reviewed. This helped in the selection of parameter values, and in one instance to resolve ambiguities in the equations, most likely caused by poor typesetting in the equations (compare equation F8 in [97] with equation 5 in [108]). The net result is that over four dozen measures have been implemented in Matlab with a consistent type signature and output. (The full source code of each measure is included in Appendix H).

Several measures were not implemented due to unresolvable ambiguities in their description, and so are excluded from further analysis. Several new measures have been created during this work and are documented in Appendex C.

Certain measures (as described in literature) operate in the opposite direction, producing a lower score when in focus (eg the *entropy* measure). As part of the implementation, the polarity of measures was normalised such that a high score corresponded to most in focus for all measures.

The emphasis of these implementations has been on clarity and simplicity, so as to ensure they are an accurate representation of the intended mathematical functions. To maximise clarity, and reduce the chance of any potential errors in implementation, no effort has been made to optimise performance (see Section 2.3.1).

For example, Sun describes the Brenner Gradient algorithm as one which "computes the first difference between a pixel and its neighbor with a horizontal/vertical distance of 2". However, the mathematical equation only suggests a horizontal offset, as does the original paper [97, 153]. Similarly, Crane describes his algorithm as a "measure of derivative". However, care must be taken not to compute the 1D derivative in one direction, then the other, as this can render a simple step input invisible. Instead, the score must be the result of the ∇I. Such subtleties are not described in the original literature, but are critical for successful use and evaluation of the various methods.

Chern notes that the focusing window (ie "the region of the scene that is to be focused") must be considered [94]. In this work, the focus measures operate over the entire input image.

3.3 Colour

The ability to perceive different colour is something that humans rely upon in their everyday lives; from being able to appreciate scenery through to telling whether a traffic light is red or green. However, the perception of colours varies from person to

person. Computers, on the other hand, need to be able to store colours numerically, such that they can be reproduced at a later date. This is done by representing a colour as a set of co-ordinates in a colour space. There are a number of different colour spaces (such as specified in [154, 155, 156, 157]), each of which has its own particular benefits, applications and disadvantages. Computers typically use an RGB representation.

Once a colour has been represented numerically, it is then frequently desirable to establish the difference between two colours. This is trivially done by computing the Euclidean distance through the colour space. However, two equal colour distances at different points in the RGB colour space are not necessarily perceptually similar – the establishment of a perceptually uniform colour space is an area of active research (eg [158, 159]). When such a colour space is used, experimental results in a wide range of areas of research are better than when using simpler colour spaces such as RGB (see [160, 161, 162]).

Most of the focus measures described in the literature rely solely upon the grey level intensity of the constituent pixels. So, whilst Crane's proposed measure is to compute the derivative of the image, and this could be extended to be computed independently on each colour channel, there is then the challenge of determining how to combine the per-channel measure to give an aggregate score for the entire image. This could be found by calculating a per-pixel Euclidean distance, then summing the distances. Or simply the sum of the overall per-channel score, or one of many other possibilities. If a consistent approach is to be taken with all focus measures, then many options become ruled out because of the nature of the different algorithms – summing the outputs from multiple wavelet transforms has little meaning. In support of this decision, Chern observed no significant difference in global maxima if different colour channels were used (ie red, green, blue or grey), and conluded that for most situations "greyscale focussing should be adequate".

Furthermore the visual system's response to colour and blur is more complex: Webster observed that "blurring only the light-dark variations in an image produced obvious changes in perceived image blur, yet when the same blur was applied only to the colour variations the image remained perceptually well-focused." [89, p113]. A similar discrepancy between monochromatic images along different colour axes was observed by Wuerger [84].

Thus, rather than using colour images, all images are reduced to a single channel representing luminance from RGB using the NTSC formula:

$$luminance = 0.2989 * R + 0.5870 * G + 0.1140 * B \tag{3.1}$$

3.4 Psychophysical assessment

Psychophysical measurements establish the perceptual behaviour of human senses. In this work, the perception of blur is being assessed. This is typically done by establishing the blur detection threshold (that is, the amount of additional blur that must be added to an unblurred image before it can be perceived), and the blur discrimination threshold (how much more blur must be added to an already blurred image before the extra can be perceived). In both these cases, the threshold is the extra amount of blur that is required to cause the subject to make the correct decision on a predetermined (ie criterion) proportion of trials. Typically the experiment starts with a large additional amount, and then reduces the additional blur until the subject is only able to make the right decision 82% of the time.

The experimental procedure for establishing these thresholds varies between the previous work in this field, though the different approaches are broadly similar. The key variations are:

3.4.1 Task

Establishing the blur threshold requires the subject to make a decision between stimuli, typically choosing between two or three stimuli. In a two-stimuli arrangement (called 2AFC - two alternative forced choice), the subject indicates which is least blurred. With three stimuli, two images are the same and the subject must indicate which of the three stimuli is the odd-one-out. In this work, 2AFC presentation was used.

3.4.2 Stimuli presentation

Independent of the number of stimuli is the choice of presentation used, either temporally or spatially separated. The spatially-separated 2AFC method shows all the stimuli at the same time, but spatially separated. For example, two images might be adjacent to one another, and the subject indicates whether it is the left or right image that is least blurred. The alternative option is to separate the stimuli temporally - that is, show one after the other, and the subject indicates whether the first or second image is least blurred. This latter method has the advantage that there is no need for the eye to saccade between the stimuli, though does prevent the subject from making repeated comparisons between the stimuli which could be done with a single-interval approach, albeit at the cost of longer experiments. These experiments used two images, temporally separated.

3.4.3 Presentation duration

Campbell showed that, when the subject had an unconstrained viewing time, discrimination thresholds were unaffected as pedestal magnitude changed. Only when the viewing time was constrained did the discrimination thresholds change with pedestal [42]. Thus, it is clear that presentation time must be limited. Investigating the opposite, the minimum presentation duration, Westheimer showed that performance plateaus when stimuli are presented for longer than 130ms [48]. So, a presentation duration longer than 130ms, but still of constrained duration, should be used. Typically, stimuli are displayed for 200-500ms (eg [163, 48, 164]), and a duration of 300ms (with a 500ms inter-stimulus interval) was found to be satisfactory in preliminary trials for these experiments, and was used in these experiments.

3.4.4 Screen

Over the years, the available technology for performing psychometric evaluations has changed, from 12-bit DEC minicomputers connected to an oscilloscope of known phosphor type (1970, [40]), through 35 mm film transparencies on a balsa-wood carrier driven by a servo motor [50]. Most recent trials use a conventional CRT screen connected to a general purpose computer, operating at a refresh rate of at least 75Hz.

To achieve linearity, the transfer function of the monitor is typically determined using a photometric sensor to measure the brightness whilst the computer is configured to display a particular intensity, and repeated at various intensities. From this, a look-up table is created to convert desired intensity into the pixel values required to achieve that intensity.

To maximise the number of distinctly achievable intensities, a video attenuator is connected between the computer and monitor which combines the individual red, green and blue signals in a predetermined ratio, and use the resultant signal to power each channel on the monitor [165]. That is, an arbitrary (R, G, B) pixel value will result in a precise grey-level being displayed on the monitor. The exact ratios within the attenuator are not of great importance as the overall transfer function also incorporates the output impedance of the computer's graphics card, slight discrepancies in the digital to analogue converters (DACs), and other confounding factors. Accordingly, the lookup table is constructed by treating the entire system as a black box with the input being the specified (R, G, B) value sent to the operating system, and the output being the brightness observed on the screen, to achieve linearity. This is the same approach as was used by Parraga et al [77].

It is also important to ensure that the stimulus is sufficiently bright for the cones within the retina to fire, as these are the neurons that dominate the fovea and provide the high acuity vision [17]. They start to fire when the illumination is above 1 cd/m^2 [166], and this is well beneath the range of a standard CRT monitor (up to approx 100 cd/m^2), meaning no special measures need taking.

Recent work, such as that by Karatzas has looked at methods for screen calibration that do not require specific hardware, such as a photometer or colorimeter [167]. It has interesting possible applications, such as for displaying museum artefacts over the internet in a manner that ensures visitors can see details of the artefacts in their true colours. However, psychometric experiments have yet to be conducted with this calibration strategy.

In summary, a CRT monitor connected to a computer via a video attenuator was used in this work.

3.4.5 Screen position

The stimulus can either be viewed with the fovea (that is, arranged such that the viewer looks directly at the stimulus, and for the stimulus to not extend more than the few degrees of angle that the fovea subtends), or shown in the periphery, requiring the subject to fixate on a target, and assess the stimulus with the periphery of their retina. As the natural way of looking at an object is to simply look at it, this foveal approach is what is used by the majority of previous work, and is used by the experiments in this work.

3.4.6 Viewing distance

Most papers use a viewing distance of between 1 and 2m, and ensure that the stimulus is confined in angular extent to be entirely visible in the fovea. For example, Weurger uses stimuli presented at a distance of approximately 1.16m on a 19" CRT monitor (with a visible screen area of 445mm diagonally). In this work, each stimulus was 256x256 pixels in size, and the screen was operating at a resolution of 1024x768. Thus, the stimulus was 89mm square, and at 1.16m this corresponds to a viewing angle of 4.4 degrees [84].

3.4.7 Viewing method

When viewing the stimulus, the observers might use both eyes or just one (having the other covered with an eye patch), and might be kept in a fixed location, such as by using a chin rest, or might be free to move around. Some previous work does

not describe the precise viewing method (eg [84]), and it is assumed in these cases that this unconstrained binocular vision method was used.

A preliminary trial showed no significant difference in psychophysical thresholds between monocular and binocular observations (see Appendix B). As such, and because each subject was required to complete a large number of trials, unconstrained binocular vision was used, as this is the natural way of using ones eyes, and likely to be the least tiring. A comfortable chair was used, at a fixed position, so as to help ensure the subject remained at a distance of 1.16m – the distance was reconfirmed at the start of each trial run.

3.4.8 Observers

Once the experimental arrangements are known, it is necessary to select subjects for participating in the trials. Frequently a small number of observers are used - for example, just two were used in [56,77], six in [84]. Other experiments have used a few subjects for all experiments and then validated the results on additional subjects, such as [79]. It is also common (eg [168]) to use both subjects who are aware of the purpose of the experiment, as well as those who are not ('naïve' observers). All observers should have good (or corrected) vision, and different methods have been used to assess this. For simplicity, subjects for these experiments were required to have had an eye test within the past 12 months, and to be wearing any prescribed correction. Subjects who said they were colour-blind were not used. Four observers were used in this work.

3.4.9 Number of trials and analysis of responses

The exact methodology for presenting stimuli, and for analysing the results, also varies between previous papers. There are two approaches typically employed - either using a simple staircase (analysed in [169]), or using a more intelligent stimulus selection strategy, such as QUEST [53]. A staircase approach means that a large additional-blur is displayed, such that the observer makes the correct decision as to which image is more blurred. Then, for as long as the observer makes the correct decision, the blur is reduced. Once the observer makes a mistake, the staircase 'reverses' and blur is increased until they once again make the correct decision. The final threshold is then taken as the average of the last few reversals after a fixed number of reversals have been recorded.

To reduce the number of decisions that the observer needs to make (and thus the time taken to perform the experiment), an estimate of the current most-likely

threshold can be made at each iteration, and used as the next stimulus. One such approach has been implemented as a reusable software library: QUEST. The details of implementing the QUEST model, and how its results are interpretted are described later (see Section 3.5.1).

3.4.10 Preparing the stimuli

Section 3.1 describes the sources of images used for these experiments. The psychophysical experiments apply varying amounts of mathematically added blur to a single, in-focus image, of each scene. To do this, a mathematical model of blur must be applied. Work by Bex [114] compares the perceptual response of applying blur via three routes: Varying α (the slope parameter, see Section 2.2.5), convolving with a 2D sinc function, and convolving with a 2D gaussian. It supports the results of Field and Brady [72], showing that varying α is not a satisfactory model of blur. Instead, blur must be applied by convolving the image with a model of the camera's optical system's point spread function. Pentland showed that the PSF of a camera's optical system can be approximated by a Gaussian kernel [19], and this is the method used in the majority of previous research.

Several of the previous experiments using a gaussian kernel do not explain how the additional blur is quantified. For example, Pääkkönen says "the blur width of an edge was specified by the standard deviation of the gaussian.", but plots his results using a scale of arc minutes [54]. This implies an equivalence whose justification or proof is not mentioned, but is assumed to be present in this, and other papers (eg [45, 56]). An alternative approach for applying image distortion is to measure the result in terms of root mean square (RMS) contrast distortion. This was done by Chandler et al [90] when assessing how observers respond to different types of distortion. In their experiment, a range of different distortions were applied to the original images such that each resultant image had a specific RMS contrast distortion.

As image processing affects the images appearance, there might be other artefacts that arise beyond the desired blurring which could be used by observers as a cue to the extent of the blur. The most significant of these confounding factors is contrast: As blur increases, the amount of contrast present in the image decreases. There are several strategies which could be employed to ensure that contrast cannot be depended upon when asessing blur – either randomising the contrast, or normalising it between images. Tolhurst et al's experiments used stimuli that "were constrained to have the same overall power and the same mean luminance; hence, they also had the same RMS contrast" [74]. Others have normalised the image post-distortion to

have a full dynamic range, or to preserve the mean brightness.

In this work, performance comparisons between humans and algorithms are being conducted. Providing that the same stimuli are used for both sets of observers, then the experiments are fair, regardless of the confounding factors that might be used to help observers discriminate blur. Secondly, images that are subjected to contrast randomisation or normalisation do not look as 'normal', and as this work is motivated by real-world applications, no contrast manipulation is performed on the stimuli.

For the experiments described in this work, convolution with gaussian is used, numerically quantified by the standard deviation of the function, with no post-processing or normalisation. The size of the gaussian kernel (in pixels, σ), is then mapped into an amount of blur (in arc minutes) using Equation 3.2:

$$blur_{arc\ minutes} = 60 \times \arctan\left(\frac{\sigma \times pixel\ width}{viewing\ distance}\right) \tag{3.2}$$

As filtering cannot work at the image boundaries, images are typically extended by means of reflection, replication, or periodic repetition. However, as in this work only a portion of the image is used following blurring, no image extension is used but instead the blurred image is cropped by the filter radius.

3.4.11 Summary

A variety of methods have been used in previous experiments, with a broadly similar approach. The methodology being followed here is a hybrid of those used by Wuerger [84], Morgan [163], Burr [56] and Kayargadde [59], and resulted in preliminary trials giving expected outcomes.

In summary, images from the Van Hateren image library were used. Blur was applied synthetically using a Gaussian kernel, and images were presented using temporally separated 2AFC. Images were presented for 300ms, with a inter-stimuli interval of 500ms, on a CRT monitor connected to a computer via a video attenuator to achieve better linearity. Four observers were used in these experiments, each looking directly at the stimulus with binocular vision from a distance of 1.16m, corresponding to a viewing angle of 4.4 degrees. QUEST was used to select candidate stimuli for presentation.

3.5 Mathematical tools

The operation of the experiments, and subsequent data analysis rely on a number of mathematical techniques. Some, such as ANOVA, are well known techniques applied in standard ways, but others deserve discussion, as they are either less familiar, or have been used in particular ways for this work:

3.5.1 QUEST

QUEST, introduced above, is "an adaptive psychometric procedure that places each trial at the current most probable Bayesian estimate of the threshold". That is, when trying to establish a threshold stimulus intensity, such that the observer can only make the correct decision between stimuli a certain percentage of the time, QUEST can recommend the optimal next intensity to measure, and thereby reduce the number of stimuli required to establish the threshold. It was proposed by Watson and Pelli in 1983 [170], and made more readily usable by being included in Pyschtoolbox [171,172], a library of useful functions for performing psychometric experiments in Matlab. It has been used in many previous experiments (eg [163, 168, 173]).

The QUEST model has several important parameters (additional explanation can be found in [174]):

Prior estimate This is the starting point for QUEST's estimates. This could be an accurate estimate (and thus potentially reduce the number of trials), or significantly away from the anticipated result, so as to ensure the observer has an unambiguous first trial. In preliminary experiments, it was found that observers required less reassurance when the first few trials were unambiguous, thus a large estimate of gaussian kernel's $\sigma = 0.5$ was used.

Standard deviation of prior estimate A large relative value was used (3, cf $\sigma = 0.5$), to reflect the fact that the actual anticipated result was considerably different to the prior estimate supplied to the model.

Threshold The threshold being measured, 82%.

Beta The steepness of the psychometric function. Beta of 3.5 was used.

Delta The proportion of trials where the observer makes a key-press error. Typically this is 0.01, though a value of 0.1 was used.

Gamma The fraction of trials that will generate an erroneous response with lowest intensity. Gamma of 0.5 was used.

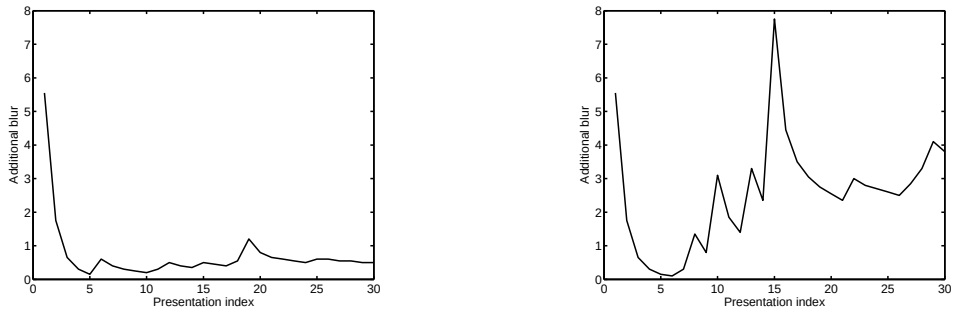
Separate to the model's parameters are considerations as to how the model itself should be used, and results selected. As the task, especially for observers who have never previously participated in psychometric experiments, is not one which is regularly performed in day-to-day life, a 'dummy' phase was employed. The decisions made by the subject in response to the first few stimuli were not fed back into the QUEST model, thereby helping to ensure that the observers were comfortable both with their physical environment, and the task at hand, and to do so without affecting the probability distribution function (PDF) within QUEST.

In addition to considering how the experiment starts, it is necessary to consider the termination criteria, and to select the results. Morgan et al [163] used a fixed number of trials (50); once these had been completed, they averaged the results of multiple experiments to determine the threshold ($\pm$ a confidence interval). Preliminary experiments for this work showed that multiple (interleaved) repeats of the same pedestal blur by the same observer did not yield the same threshold after 50 trials. And, after 100 trials, there is no significant improvement in inter-trial agreement. Accordingly, the results of multiple trials shorter trials were aggregated, as explained below.

To help minimise habituation and similar confounding factors that might be introduced if the same scene were viewed on every single stimulus, several experiments were interleaved during each session. To achieve this, a list of all experiments to be performed was prepared in software, and then the software ensured that five experiments were running concurrently. That is, when a particular condition was finished, the software pseudorandomly selected the next experiment to perform, and add that to the list of active experiments. At each iteration, the software selected an active experiment, and asked QUEST to suggest the threshold to assess. The order of presentation (that is, $b + \delta b$ then b or vice-versa) was determined pseudorandomly, and then the stimuli were presented.

Once the QUEST process has been followed, it is then necessary to select the results. Previous research has alluded to only some results being considered (for example, saying "Each estimate of threshold was based on at least three separate determinations (QUESTs) per measure" [56], whilst Simmers et al used "at least four separate determinations" [173]), though the methodology by which determinations are selected for inclusion in presented data is not discussed.

Preliminary trials showed that, in most cases, the QUEST determination does neatly converge, but that occasionally this does not happen. The reasons for a lack of convergence have not been explored, as it occurs on a minority of trials (less than 10% of trials). Figure 3.5 shows two trials for the same observer and the same



(a) Converging (b) Not converging

Figure 3.5: In most trials, QUEST's recommendations gradually converged, as can be seen in (a). However, in a small minority of trials, such convergence did not occur, and results similar to (b) were obtained. These examples were from the same observer and under the same conditions. A bootsrapping procedure was used to fit the data to psychometric functions, ensuring the observations made during non-convergent QUEST determinations were not discarded.

experimental conditions and stimuli, though only one of the trials converges.

In this work, QUEST is used to improve the speed with which data is collected, but the final threshold is not taken as the result of QUEST's convergence. That is, QUEST is used to direct the experiment to the next optimal measurement, rather than requiring the observer to respond to pairs of images at every point in the sample space. Once several determinations of QUEST have been performed, the results for each condition are then aggregated and tabulated, recording whether the subject responded correctly or not at each stimuli under test. Thus, for a given pedestal condition, there are a minimum of 120 data points, each comprising the additional blur being discriminated (δb), and whether or not that additional blur could be discriminated. These results were then subjected to a boot-strapping procedure to estimate the 82% threshold, and its 95% confidence interval. This was achieved by fitting the data to psychometric functions using $psignifit$, a software package which implements the maximum-likelihood method described by Wichmann and Hill [175, 176].

In summary, the method of using QUEST in this work is most similar to Burr et al's approach [56], but using bootstrapping to establish the psychometric thresholds rather than taking the result of each QUEST determination.

3.5.2 Computation and manipulation of the slope parameter

The statistics of natural scenes, and discovery that the amount of energy present at each frequency drops off as frequency increases, is introduced in Section 2.2.5. The method used in this work to calculate the value of α for a given image is based on Bex and Dakin's approach for computing the power present in each octave [150]. From these results, a linear best fit is performed to determine α.

To adjust α, each coefficient of the shifted fast Fourier transform (FFT) of the image is multiplied by a specific scaling factor. The matrix of scaling factors was determined experimentally to be a matrix of Euclidean distances from the centre, with each distance raised to a predetermined power, P:

$$P = \left(\alpha_{desired} - \alpha_{current}\right) / 0.995 \tag{3.3}$$

The DC component of the FFT was unchanged by setting its scaling factor to 1. The full code for changing an image's α, as a Matlab script, can be found in Appendix H.35.

3.5.3 Scoring focus measures

The method for scoring focus measures is introduced in Section 2.3.5, which explains the techniques used by other researchers. However, when focus measures were being scored, several issues arose.

Firstly, the 'accuracy' property is not defined if the focus measure gives multiple candidate images the same (highest) score. In this work, the best image is found by finding the mean image with the highest score, rounding down where necessary, and then the accuracy metric found by determining the difference between the measure-best and ground-truth.

Secondly, to assist in comparing scores between different scenes, the range value was computed as a percentage of the number of images, rather than an absolute number.

Thirdly, a score of zero was considered to be a minimum, even if there was only an increase in score on one side of the zero. Without this interpretation, the range of several focus measures is undefined (any focus measure without a false maximum will lack a minimum on either side of the primary peak, unless zero is considered to be a minimum).

The full implementation, as a Matlab script, can be found in Appendix H.34.

3.6 Data protection and ethics

Advice was sought from the university's Records Office who advised that the recording and analysis of anonymous data did not require adherence to the Data Protection Act 1998. Separately, the experiments in this work did not require approval by the university's Research Ethics Committee, as they comprise solely the use of behaviour observations and educational tests for which the participants cannot be identified nor would there be any consequences of disclosing their responses.

Chapter 4

Subjective experiments

The literature reviewed in Chapter 2 makes little reference to establishing the ground truth when comparing focus measures. Of the papers that do mention how it was established, all have used an 'experienced observer', or words to that effect – and there was no evidence of averaging or obtaining the consensus from multiple observers.

This chapter describes a series of experiments that were conducted to find human opinions about focus in a number of different scenes using a large number of observers. Two objectives were planned to establish peoples' opinions about focus. The first was to establishing the ground truth – which image was most in focus. The second was to establish whether a focus curve could be created from human opinions.

Whilst collecting the data for the first objective (Section 4.1), a trial was conducted to establish which of two software approaches was most appropriate for obtaining the necessary experimental results. It was found that a web-based experiment worked well, and this then used to collect all the data reported in this chapter.

The focus curve neded for the second objective is not a concept with which observers are familiar. As such, a series of tasks were given to the observers to try to indirectly capture the data necessary to plot a focus curve. These tasks are described in Section 4.2.

4.1 Ground truth

Mathematically, the ground truth can be precisely determined. For a simple optical system, this is calculated with the thin lens formula, where S_1 is the distance from the object to the lens, S_2 is the distance from lens to image, and f is the focal length of the lens, as shown in Figure 4.1.

(4.1)

Figure 4.1: Principle of the imaging provided by a convex lens [177]

However, given S_2 and a fixed optical system, there is only one solution for a single S_1. In real world scenes, depth is present, and so it is not possible to focus the entire scene at the same time, hence subjective opinions of a population of observers are required to esablish the 'best' focus distance.

It is clear, however, that for some scenes the best focus position may differ between viewers of the scene (eg Example Photograph A.1), or indeed may vary over time as an observer returns to the photo (eg Example Photograph A.2). But, whether there are inter-observer differences on less complex scenes has not been addressed by previous work. To assess this, multiple observers need to be shown a scene, and be permitted to change the focus until they have identified their preferred position – the 'best' focus distance.

For this experiment, simple scenes of domestic objects were photographed under artificial light using a tripod mounted Olympus 500-UZ. The camera was controlled with Pine Tree Computing's Camera Controller, and configured to take between 20 and 100 photographs at high resolution (for further details, see Section 3.1). Whilst the software controlling the camera was instructed to request equal changes in focus position from the camera, it is unclear whether the camera incorporates any feedback mechanism to confirm it is in the requested position. Nor is it known whether the camera's focus positions are equally spaced, and indeed what the mapping function is between requested position and resultant focal length of the lens. As such, the camera is used solely for acquisition of images with increasing focus distance, and no calibration or implied equivalent change between adjacent positions is present. This does not, however, affect the ability of humans (or focus measures) to make judgements on these scenes – decisions can be made between images without knowing their exact source.

The acquired photographs were then manually reviewed to ensure no object motion or other undesirable features were present. Once checked, they were down-sampled to 640x480 pixels for display, using the `resize` function in the open source ImageMagick software.

Using an existing library of calibrated images is discussed in Section 3.1. However, it is not possible to use these images for this experiment, as the are not available at a variety of focus distances. That is, whilst gaussian blur (shown to be a good approximation of optical blur) can be applied to these images, this can only be done globally. It is not possible to reconstruct the depth information from the photographed scenes, reduce the depth of field, and then produce multiple images from the scene with different focus distances, so these libraries cannot be used.

The next consideration is the graphical user interface (GUI) and presentation of images. Two approaches were evaluated and tested in pilot studies:

4.1.1 Desktop application

An application was written in Microsoft Visual Studio Express. This software platform was selected as it is relatively easy to develop powerful user interfaces that can run on other computers without requiring a complex configuration process. The initial user experience objective was to make the task feel like focussing a camera: That is, trying to minimise the digital feel by making the experience fluid, responsive and continuous. To assist this experience, a number of input devices were considered and the Griffin PowerMate selected. This is a USB-connected rotatable dial with a smooth motion. By virtue of its operation, just a single parameter can be adjusted, and thus is well suited to the task of 'focussing' through a pre-acquired set of images.

The most challenging (and important) part of the software development was to pre-load all the images into memory, so that the appropriate image could be displayed on screen as soon as the dial's motion determined that it was necessary. In addition, the transition between images needed to be performed in such a way that there was no flickering on the screen. By achieving these two aims, the experience of using the dial to select the most focused image from a series of images feels entirely natural, and not at all as if there is a computer 'in the way'.

A pilot trial involving two observers was performed. The observers were unaware of the purpose of the project, though were briefed on the task to perform. Both observers encountered difficulty with the task, finding it too time consuming and also appearing to forget the objective during the task, seeking clarification of the objective mid-task.

After analysis, the results from these two observers during the pilot study were considerably different to those obtained when testing and developing the experiment. On this basis, it was decided that a large number of observers should be used for the experiment.

So, despite the excellent experience available using Visual Studio in conjunction with the PowerMate experimental setup, the difficulty of recruiting a large body of observers to participate in the experiment at a fixed location meant that a different approach was pursued. The original intention to distribute the software to multiple computers for observers was rendered impractical given the dependence on the PowerMate – a piece of hardware not available on all computers – and so it was decided to develop a web-based implementation of the experiment.

4.1.2 Web application

Deploying an application over the internet is common practice for business, but less so in the field of subjective image assessment. By being available to potentially the entire planet, a diverse range of web browsers, computers and screens could access the experiment. Rather than try to recruit observers with particular equipment, it was decided to allow any observer and equipment combination, but to require them to complete a short questionnaire. This asked for self-declared answers about:

Age Free-text entry

Gender Choice of 'male' or 'female'

Uncorrected vision Choice of 'short-sighted', 'slight-short-sighted', 'normal', 'slight-long-sighted' or 'long-sighted'

Vision correction Choice of 'none', 'glasses' or 'contact lenses'

Colour blindness Choice of 'don't know', 'none', 'red-green', 'blue-yellow', 'other'

Figure 4.2: Griffin PowerMate: A USB device that can send different events to the computer when it is rotated or depressed

Screen Choice of 'don't know', 'CRT', or 'LCD'

Mother tongue Free-text entry

These answers were then available for analysing with the results, so as to investigate whether any particular answer caused a significant difference in the image selected. By being conducted in multiple, unsupervised, remote locations meant that the question answers could not be validated. Instead, it is assumed that people answered honestly.

To ensure the user experience was as responsive as possible, all the images (ie the images taken at each focus distance) of the scene being presented were preloaded before presentation started. This was achieved by using Secord's Image Preloader library [178]. Once all images were loaded, the progress bar was removed from the screen, and the relevant scene displayed. The observer could then browse through the focus distance by pressing the arrow keys on their keyboard. The JavaScript `EventListener` hooks for `keydown` were used to detect the key presses, then the `src` attribute of the displayed image was changed to be the new image to display. Because all images were preloaded, this transition occurred very rapidly, and without any perceived flicker, on all leading web browsers.

To reduce the impact of on-screen clutter acting as distractors from the task, the experiment was positioned centrally on a black background, occupying 640x480 pixels (approximately 50% of the screen area of a small monitor). The initial image displayed was clearly out of focus, so observers knew from the start of the task that focussing was required – a pilot trial placing the starting point at a random location left some observers confused when the initial image had been almost in focus, and they were unsure as to what they needed to do.

The experiment was advertised via various email mailing lists to students and friends. In total, 80 people started the experiment, though a few did not complete the entire experiment. Various screenshots from the web-implementation are shown in Figure 4.3, and the scenes used are shown in Figure 4.4.

The precise instructions given to the observers was: "In a moment, you will be shown a series of images. They will be slightly out of focus. Use the left and right arrow keys on your keyboard to change the focus of the image. When you have found the best image, press the enter key. Click here if you're unsure which keys these are." If the observer was unsure of the keys involved, then the link took them to a picture of a standard keyboard, with the relevant keys highlighted.

Performing experiments with unsupervised observers on equipment that has not been calibrated could cause misleading results. However, the author feels that the

participants who were selected for these experiments were likely to be cooperative, and unlikely to have motives for reporting inaccurate information or deliberately responding in detrimental ways. Furthermore, their demographic (being educated young professionals) is likely to have their computer well configured with reasonable colour reproduction and have a good quality monitor running at its native resolution. Such assumptions would need to be reviewed, and potentially additional investigation conducted, should a wider audience be used for future web-based experiments.

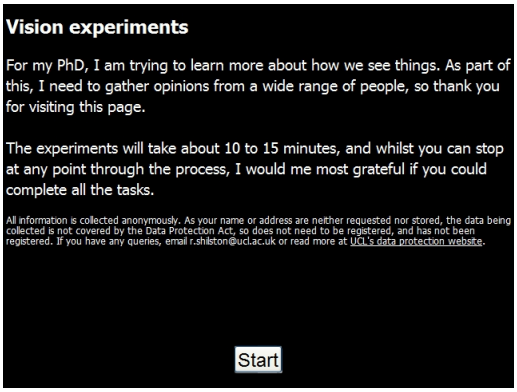
The web-server maintained a full log of the exact sequence of images shown to each observer, and the precise time at which it was displayed. No reference information or other feedback was provided to the observer, as these might have provided additional, non-visual information from which a decision could be made. Instead, the only source of information available to an observer was the desired image. From these logs, it is possible to see how one observer browsed through one of the scenes to find the image they thought was most in focus, as can be seen in Figure 4.5.

4.1.3 Results

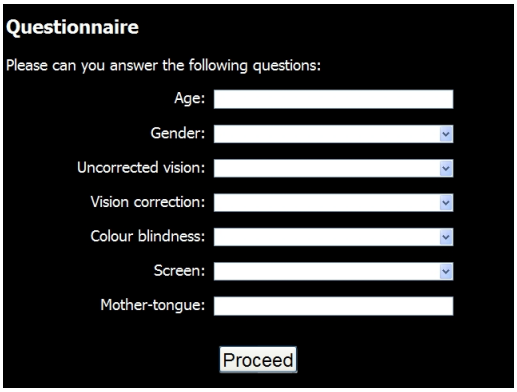
The questionnaire results shown in Figure 4.7 provide some information about the participants in the experiment. Figure 4.8 contains histograms showing the frequency with which each candidate image was selected as the 'best' version for that scene. There were anomalous results which have been identified and removed manually. These are shown on the graphs as grey bars, and most likely arose as people accidentally clicked through the experiment without actually performing it. Table 4.1 summarises the final results.

Scene	**Participants**	**Mode**	**Mean**	**Std dev**	**% selecting mode**
Chillis	74	20	19.85	0.86	46%
Coins	74	28	28.34	1.74	23%
Bolt	75	72	72.13	2.77	28%
Red onion	72	19	19.24	1.71	24%
Strawberries	75	13.5	13.44	1.49	25%*
Towel	68	33	32.13	2.31	19%

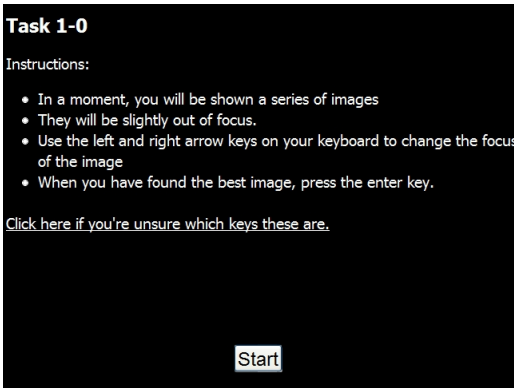
Table 4.1: Results for the 'best' experiment showing which images was chosen by the observers of the different scenes in the experiment. Note: The strawberries image produced a bimodal result – both images 13 and 14 were selected equally frequently by the observers as being most in focus. The reported percentage is for one of the modes.



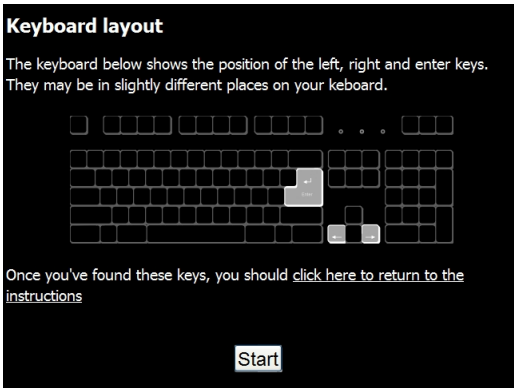
(a) Introduction



(b) Questionnaire



(c) Instructions

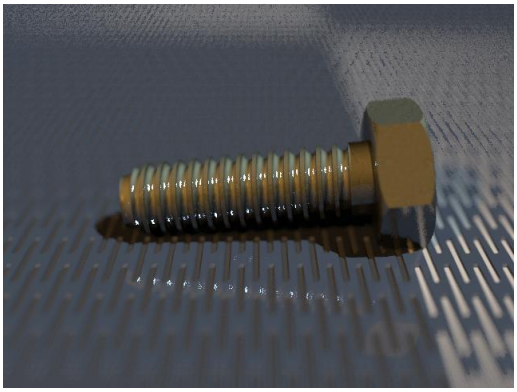


(d) Keyboard

Figure 4.3: Selected screenshots from the web implementation of the 'best' experiment.

(a) Chillis

(b) Coins



(c) Bolt

(d) Red onion

(e) Strawberries

(f) Towel

Figure 4.4: Scenes shown to observers during the 'best' experiment.

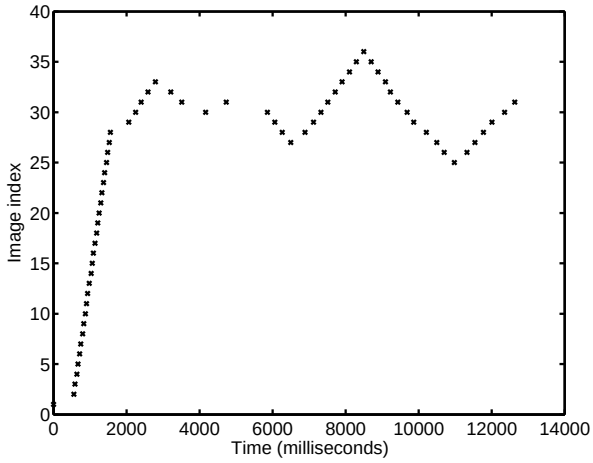


Figure 4.5: The journey one randomly selected observer made through the candidate images when selecting the most in-focus picture of coins.

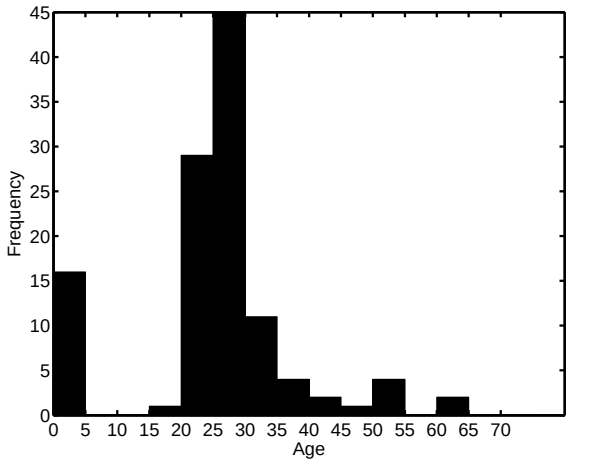
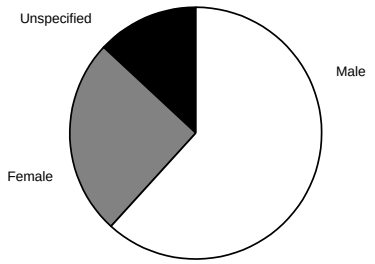
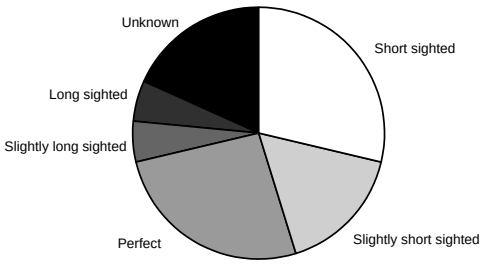


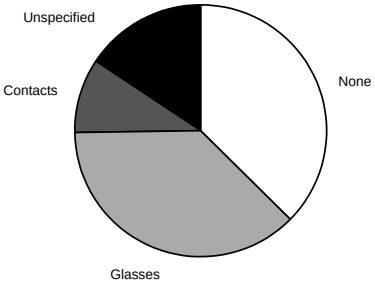
Figure 4.6: Histogram showing age of participants in the 'best' experiment. 16 participants did not specify their age, and are recorded in the 0-5 bin – there were no observers under 15 years of age.



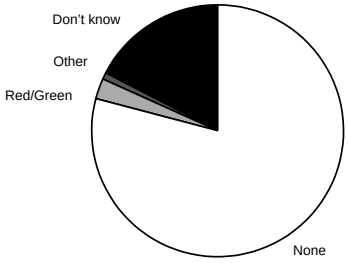
(a) Gender



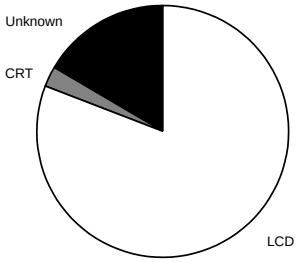
(b) Uncorrected vision



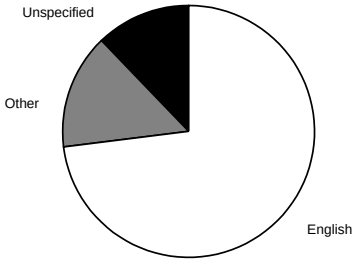
(c) Correction



(d) Colour blindness

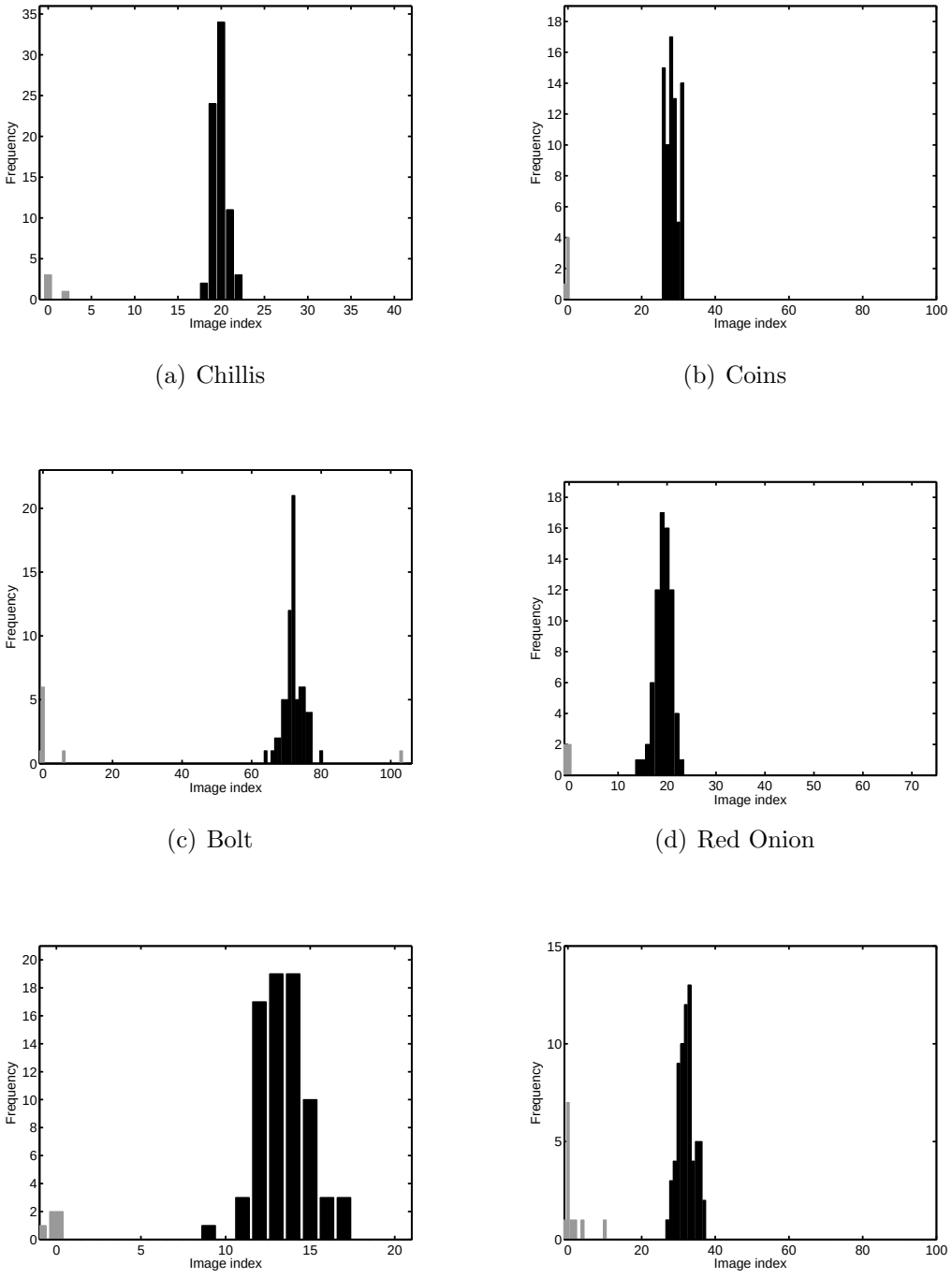


(e) Screen



(f) Mother tongue

Figure 4.7: Pie charts show the overall questionnaire results for participants of the web-based experiments.



(a) Chillis

(b) Coins

(c) Bolt

(d) Red Onion

(e) Strawberries

(f) Towel

Figure 4.8: The graphs show the frequency with which each candidate image was selected as being 'best' for the given scene. Bars in grey were identified as being anomalous and have been excluded from further analysis. They most likely resulted from observers proceeding to the next task in the sequence without selecting an optimal image.

4.1.4 Analysis

It was anticipated that observers would, to a certain extent, agree with one another, though the exact extent of agreement, and whether it would vary by environment or individual was unknown. The results in Figure 4.8 confirm that humans mostly agree with one another. Each scene resulted in a single group of selected images, which would be expected given that none of the scenes had multiple objects at different depths within them.

However, the observers were not unanimous in their opinions. To explore the reasons for the diverse answers, the results were processed using ANOVA as a linear model evaluated with the R statistical processing environment [179]. A model of all the individual parameters, as well as all combinations of pairs of parameters was constructed. The questionnaire results were collated slightly – as there were no responses to the 'What is your mother tongue?' question that occurred more than once (other than English), the model was built based on the answer to the question 'Is your mother tongue English?'. Similarly, 'age' was grouped both by decade, and separately to 'under 30' and 'over 30'. Thus, the full model used the following parameters, plus all combinations of two parameters:

- Over or under 30 (1 for under, 0 for over)
- Age (in decades, rounded down)
- Gender (1 for male, 2 for female)
- Screen type (1 for LCD, 2 for CRT)
- Colour blindness (1 for none, 2 for red/green, 3 for other)
- Correction (1 for none, 2 for glasses, 3 for contacts)
- Uncorrected vision (1 for short sighted through 5 for long sighted)
- English is mother tongue (1 for yes, 0 for no)

The full ANOVA tabulations, computed independently for each scene, are included in Appendix F. From these full tables, the most significant relationships were different between scenes:

Chillis $Pr(> F)$ for Gender was 0.096, and 0.065 for the compound variable DecadeAge and Correction, both of which are very weak effects.

Coins This had stronger correlations – $Pr(> F)$ was < 0.3 for both Young30:Correction and DecadeAge:Gender, whilst the value for Correction:Uncorrected was stronger still: $Pr(> F) = 0.0157$. There were no strong effects for single variables.

Bolt For this scene, the strongest effect was with Uncorrected vision (0.013), then Colourblind (0.019), followed by Young30:English (0.02856) and DecadeAge:Uncorrected (0.045)

Red onion Gender was the only strong effect, with $Pr(> F) = 0.0162$

Strawberries This scene had the strongest effect of all scenes, with $Pr(> F)$ for the compound variable Gender:Screen being 0.0001. There were weaker effects for DecadeAge:Gender, DecadeAge:Uncorrected and Gender:Correction.

Towel There were weak effects with Gender and DecadeAge:Correction

In summary, three of the scenes had no strong effects ($Pr(> F) < 0.05$) with any single variable. The single variables that did have strong effects were Uncorrected vision, Colourblindness and Gender, though these relationships were not present across multiple scenes. This suggests that whilst there were some effects, these were not widespread across the experiment. No correction for multiple comparisons was performed. Had such an approach been taken, it is anticipated that the individual effects appearing in the different analyses would have been suppressed. That is, there is not a subset of the observers who across all images made a particular preference towards focussing nearer or father away.

By reanalysing the data from one of the single-variable correlation (gender, whilst looking at the Red Onion scene), the impact of this variable can be seen. Figure 4.9 is of two histograms (normalised to the same area), for males and females, showing how the genders selected different images as their preferred 'best' image. The modal image indexes were 19 (male) and 20 (female), though the mean index was the other way around – 19.5 for male ($\sigma = 1.59$) and 18.5 for female ($\sigma = 1.84$). If the distribution were modelled as a normal distribution, then all of these figures are well within the 95% confidence interval (2σ).

No statistical analysis of the union of all scenes in this experiment was performed as doing so would require their results to be comparable. This is not possible, as there was no normalised focal position around which to equate the results – for this was the position being sought by the experiment, and because the differences in focus between adjacent images of the same scene are not necessarily equal.

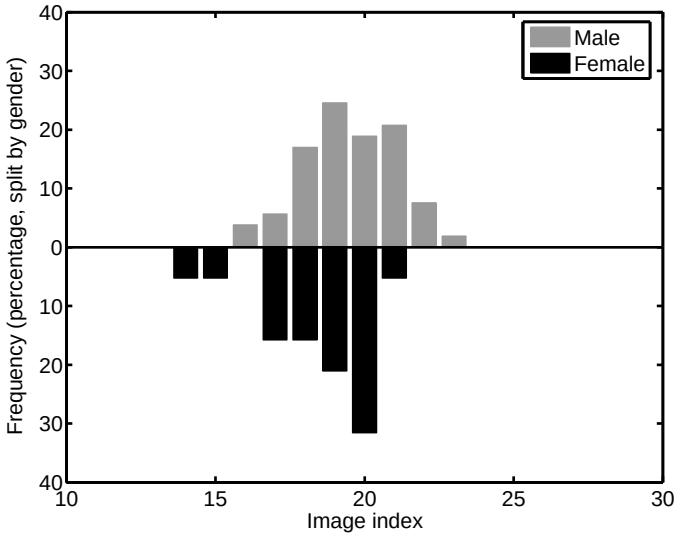


Figure 4.9: Comparison of male and female responses when asked to select the best image of the red onion scene

Instead, the aggregate human results could be considered to be a focus measure. A focus measure produces a score for each candidate image of a scene. The histogram showing the frequency with which humans selected each particular image could be considered a score. So, if three people thought image 10 was best, then image 10 receives a score of three, etc. This human focus measure can then be assessed using the methodology proposed in Section 2.3.5, to assess the width, range, and other properties of the ground truth's distribution. By definition, the average human response is the ground truth, so the accuracy is perfect. The other criteria are summarised in Table 4.2, with the score calculated using equation (2.8).

Scene	**Accuracy**	**Range***	**False maxima**	**Width**	**Noise level**	**Score**
Chillis	0	86%	0	1	1.460	0.321
Coins	0	97%	2	5	6.311	1.994
Bolt	0	93%	2	1	0.871	1.416
Red onion	0	86%	0	3	0.450	0.781
Strawberries	0	65%	1	3	1.324	0.723
Towel	0	92%	0	3	1.373	1.067

Table 4.2: Ranking human responses shows that humans were best (using the methodology introduced in Section 2.3.5) when viewing the chillis scene. Note: For clarity, the range figure has been expressed as a percentage of the available images outside the first minima on either side of the peak, as discussed in Section 3.5.3. Accuracy is zero, by definition, as these are human results being displayed.

These results, especially the very narrow width of the histograms at 50% amplitude, show that untrained humans are very much in agreement with each other when making focus decisions. There were false maxima present, but the largest of these was less than 40% of the peak amplitude, and none were wider than two images.

4.1.5 Conclusions

This experiment has shown that humans do agree with one another when focusing a scene on a computer screen. There are differences between subsets of the population on a per-experiment basis, but these have been shown to be small in comparison with a 95% confidence interval. That there are no common factors between scenes reinforces the interpretation that the per-experiment differences are not indicative of an underlying factor that does affect focus decisions.

The best quantitative comparison score was achieved when focussing a scene featuring some chillis. There were no significant differences between the observers' ability to focus on synthetic or photographed scenes, nor a difference between focusing on 2D or 3D objects. However, even in the scene with greatest agreement (chillis), only 46% of the observers selected the most popular image as their 'best', and significantly lower for the other scenes. Thus, for untrained observers, the opinion of just one observer is unlikely to be the consensual ground truth, and multiple observers opinions should be sought.

No justification of for the high level of agreement in the chillis scene can be made with certainty from these results. It is the author's opinion that there is, perhaps, a particularly attentive point in the scene which observers endeavoured to focus, and that there is not such a popular point in other scenes.

Having now established the ground truth for a number of scenes, Chapter 5 compares these results with existing focus measures.

4.2 Building a focus curve

The focus measures introduced in Section 2.3 each produce a single score for a single image. If multiple images of a scene are processed by the same measure, then a focus curve can be plotted, and the peak of that curve is the best focus position, as judged by that measure. The previous experiment asked human observers to manually select the best focus position, and then created a focus measure graph based on the frequency with which humans selected each particular image. These histograms are tall and thin in shape, which is expected as observers would not

select a clearly defocussed image as being 'best'. Thus, for many candidate images, the score was zero.

However, humans can discriminate between two out of focus images to determine which is sharpest, and thus there must be a measure that the brain is using. Previous work to determine this measure has not been found, though Fylan's proposal in 1998 to record visual evoked potentials (VEPs) might yield quantitative information about the measure used by the brain [34].

To determine a focus curve, one could present blurred images to the observer, and specify that their task be simply to score them (eg [58]). This opinion score would equate to the quality of the scene at each focus distance, and thus form a focus curve. However, observers are likely to simply line up the images in order of defocus and assign scores linearly on that basis, thereby simply producing a triangular function, and so is not a direct measure of the human focus opinion. So, several approaches were pursued to try to indirectly extract a human focus curve, without requiring the observer to retain a memory of multiple images:

4.2.1 Time to re-order

The first method attempted was to ask subjects to re-order photos of the same scene so that they were in the order of increasing focus. The hypothesis was that it would take subjects longer to reorder photos that had similar perceived focus quality (ie focus score), and so the overall focus curve could be plotted by using the response times as the gradient at each point. Specifically, the subject was shown on a computer screen three images of the same scene from random focus positions. Using the mouse, they could drag and drop the images to change their order. Once satisfied, the subject clicked a button to indicate they had finished. The subject was then shown another three photos from the same scene, but from different focus positions.

It was hoped that the reciprocal of the average response time (given the correct answer was received) for a triplet of source images centred at some focus point, p, could be used as the gradient of the focus curve at that point. Photos that were easy to order would have a low average response time, and so a steeper gradient, as there must be a large difference in perceived focus if observers could sort them quickly.

However, no meaningful conclusion could be drawn from the results, and this is likely to be for two reasons. Firstly, for the gradient to be known with confidence at any given point, then multiple subjects will have had to order a triplet of photos around that point, and such a large number of trials was not conducted. Secondly,

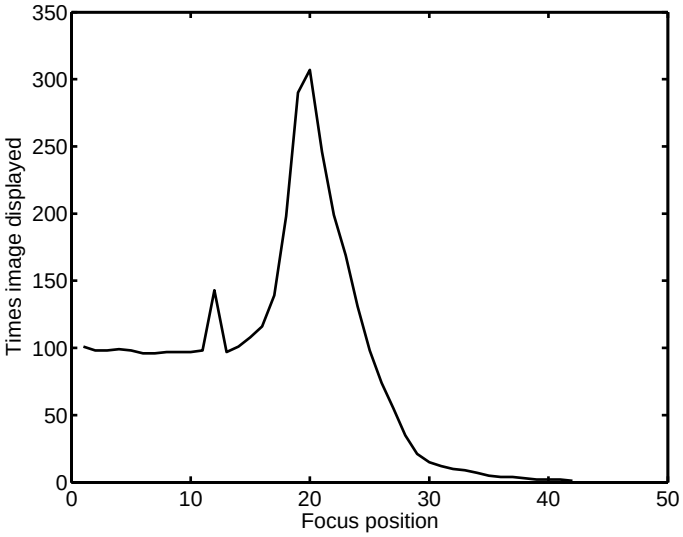


Figure 4.10: Popularity of images in the 'chillis' scene, given that each observer started at position 0.

and perhaps most significant, was that the response times did not vary significantly when the photos were "easy" or "hard" to order – most of the time was spent using the mouse to interact with the computer to record the subject's decision, rather than actually working out the correct order, so the experiment mainly recorded mouse dexterity rather than focus opinion.

4.2.2 Image viewing frequency

It was decided next to try to capture more instinctive information. The Griffen PowerMate was used to enable the subject to rapidly navigate through the images of the scene captured at different focus positions. The hypothesis was that subjects would more frequently look at images that were sharper in focus than those that were less focused.

The results, as reproduced in Figure 4.10, show that there is a peak in image popularity when the image is in focus. However, it is perhaps not justifiable to say that this represents the human focus curve, as the image on either side of the peak will always have to be seen frequently, as the subject will need to go past these to get to the peak.

4.2.3 Time spent on each image

To reduce the impact of the nearest-neighbour phenomenon, the same data was processed differently. The revised hypothesis was that the less in-focus an image, the less time a subject would spend looking at it. The results are shown in Figure 4.11. This shows that 99.8% of the images displayed were viewed for less than 125 milliseconds, and that those looked at for longer were significantly longer (typically a few seconds) – no image was looked at for more than 125 milliseconds excluding those exceeding 300 milliseconds. Therefore, the average time spent looking at each image, having excluded those times greater than 300 milliseconds, was plotted. This can be seen in Figure 4.12. The results are quite noisy, but do show a curve which matches with the expected observation near the peak. However, there is very little drop-off in viewing time beyond image 40, despite there being a continued deterioration in sharpness in this region which is obviously visible to human observers.

4.2.4 Blur equivalence

Rather than attempting to construct an entire focus curve, a new curve could be generated by anchoring one half of the curve to a known function, and asking observers to find images that are equivalently blurred. Having established the ground truth (as described above), the images were partitioned into two groups. The first group containing only images focussed further away than the best image, whilst the second group containing the images whose focal distance is nearer than the best image. The images in the first set were manually assigned a linear score, between 0 (for the least focussed image) to 1 (the most focussed).

Observers were presented with a screen divided in two. The left side of the screen showed a static image selected from the first set of images. The right side of the screen could be controlled by the observer to vary the image displayed, but restricted to only show images from the second set. The observers were then asked to select the image on the right that they felt was equivalently blurred to the image on the left. Figure 4.13 shows the user-interface for this experiment.

Whilst this is a straight forward task for flat scenes, such as the coins or towels scenes, it is harder for an observer looking at a scene with depth. This is because some points of the target scene might be in focus, yet the overall feel for the scene's focus needs to be assessed and reproduced on the matching scene. 84 observers participated in this experiment, all using the red onion scene. Each observer had one attempt on each of four stimuli images. Their results are shown in Figure 4.14.

The results, as with the 'best' experiment, show a tight grouping of observer

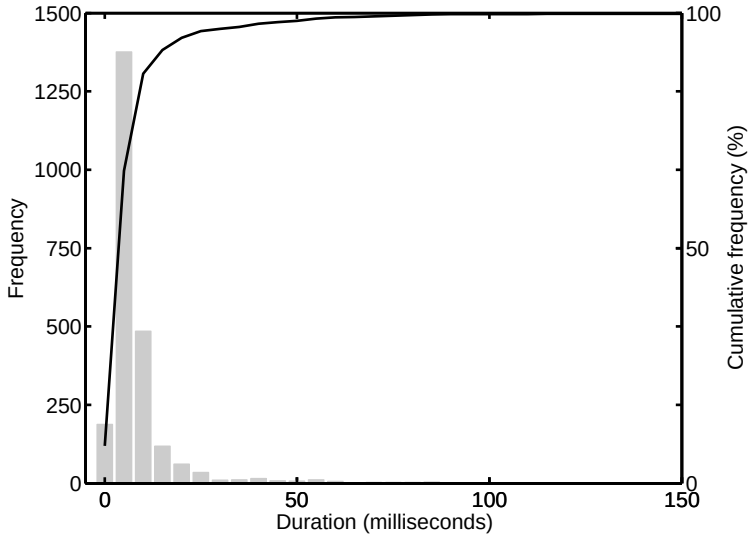


Figure 4.11: Distribution of viewing times when browsing the 'chillis' scene using the Griffen Powermate. 99.8% of the images viewed were shown for less than 125 milliseconds. The grey bars show the distribution of viewing times, and the black line shows cumulative percentage.

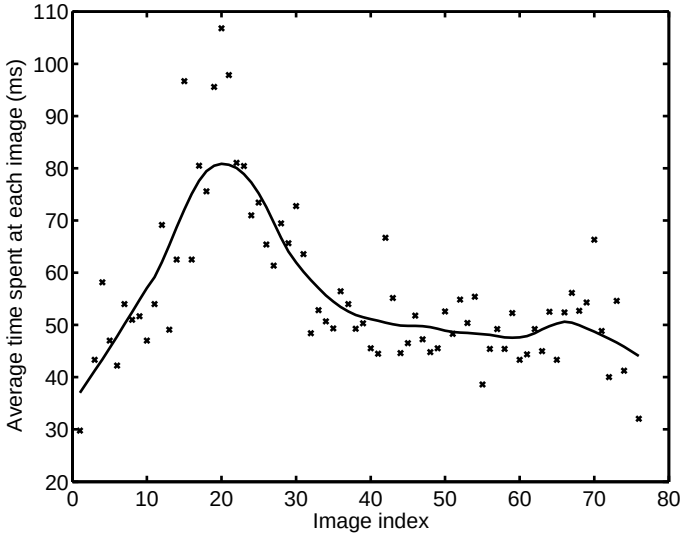


Figure 4.12: Average time spent looking at each position, given that the time was less than 300 milliseconds. A trend line has been added using the Lowess method, with a window size of 20 points.

Figure 4.13: User-interface for the 'equivalent' experiment. The left side of the screen showed a target image, and observers were required to use the arrow keys on their keyboard to control the image on the other side and to select the equivalently blurred image.

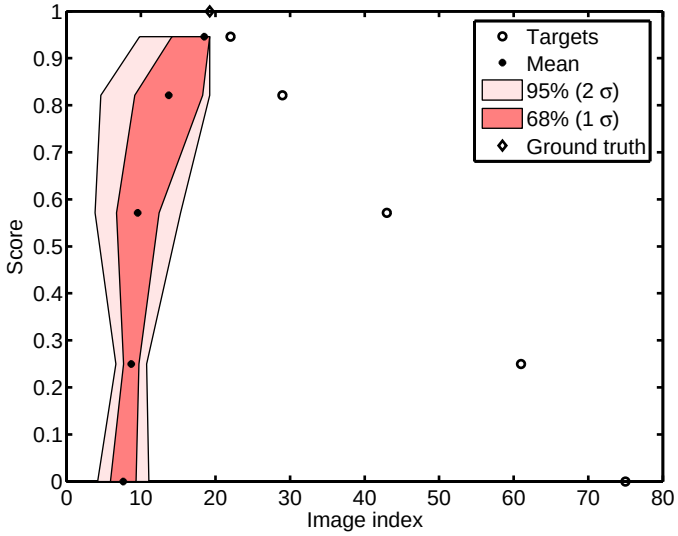


Figure 4.14: Results for 'equivalent' experiment: 84 observers were shown target images from the red onion scene, whose score had been linearly assigned, and asked to select the image from the otherside of the 'best' position such that it was blurred to an equivalent amount as the target image. The mean responses, together with confidence intervals, can be seen on the left side of this graph. For each score the horizontal width of the shaded area shows the range of opinions for the image at that score, .

opinion. With a normal distribution, 95% of the results lie within two standard-deviations of the mean. This region, shown on the graph, is relatively narrow for the majority of the curve, encompassing between four and eleven images when the target is most blurred. The results show a similarly steep drop-off rate of score with focus position as has been seen in the earlier approaches to producing a focus curve for human opinions. However, within the 95% confidence region, the same image (number 10) could be the equivalently blurred version of all of the five different targets. Therefore, whilst the expected trend is present, and the sample size is relatively large (84 observers), the data is also consistent with no trend being present.

Observers reported that it was frustrating (and thus hard) to try to find the equivalently blurred image when part of the scene was in focus. By the experiment's design, the same portion of image would never be in focus in the candidate set of images, and this was the underlying issue. This difficulty with images that were partially in focus is the most likely explanation of the larger standard deviation when closer to the ground truth of the 'best' image.

4.2.5 Half-way focused

A final approach was pursued wherein human results were used to plot the entire focus curve, and was designed following observer feedback on the equivalence experiment. Rather than selecting an equivalently blurred image, observers were asked to select an image that was equally defocussed with respect to two stimuli (ie halfway defocussed between two target images). This is also known as 'subjective bisection'. Using this approach, each trial was restrained to being on just one side of the ground truth, and so the difficulty reported by observers in Section 4.2.4 eliminated.

By manually assigning the score 0 to the least focussed image at either extreme (positions A and B in Figure 4.15), and 1 to the in-focus image (position C), the hypothesis was that the middle of the graph could be populated by asking observers to identify the image whose score was precisely 0.5, from either side of the in-focus position (positions D and E). Then, by repeating the process, the quartile values (positions g, h, i and j) could be determined.

A g D h **C** i E j **B**

Figure 4.15: Establishing the half-way defocused images: Images A, B and C were predetermined, and from those, observers selected positions D and E, then g, h, i and j.

Figure 4.16: User-interface for the 'halfway' experiment. The left and right sides of the screen showed target images, and observers were required to use the arrow keys on their keyboard to control the image in the centre to select the image that was half way between the target images.

To help reduce any confounding effects that might arise, each position on the curve was presented to the observers twice, once with either arrangement of targets. So, the observer had two attempts at identifying image D; once with A on the left and C on the right, and vice versa. Figure 4.16 shows the user interface for this experiment.

72 observers participated in this experiment, whose results are shown in Figure 4.17. As with the previous experiments, it shows that observers are in agreement with each other. There were a few outlying data points, which were not manually removed, but by virtue of the large number of observers, these are outside the confidence intervals calculated from the standard deviation. Most importantly, the results show that this experimental methodology can be used to produce a focus curve.

4.3 Conclusions

The experiments conducted have provided answers to the first few key questions that arose from the thesis statement. Across a number of different scenes, comprising both natural photographs and synthetically generated images, it has been shown that there is agreement between observers as to which is the most in-focus image. Most informatively, the agreement is not unanimous, and the ground-truth image was never selected by the absolute majority of observers. This suggests that future work investigating focus should not rely upon a single observer when looking

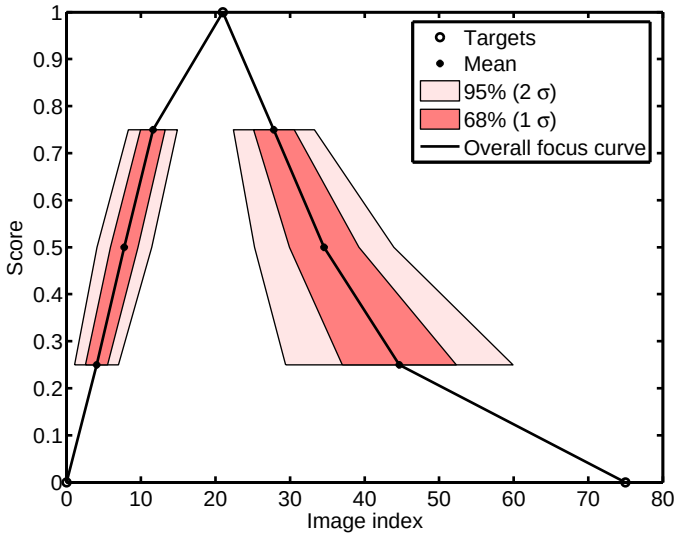


Figure 4.17: Results for 'halfway' experiment: 72 observers were shown two target images from the red onion scene, and asked to select the mid-point in terms of focus. This was repeated to find the scores at 0.25, 0.5 and 0.75. Their mean responses, together with confidence intervals are plotted.

at general purpose images, but instead must gather the opinions of a number of observers.

Reassuringly, for the scenes used in these experiments, there was no distinguishable subsets of the observer population which selected different ground-truths. There were some factors which had weak correlations with the selected images, but these factors did not persist across scenes. This suggests that, regardless of gender, age, vision or screen type, observers make equally accurate decisions when selecting the ground truth. That said, the population of observers was skewed towards those people known by the author, and thus are more likely to have undertaken higher education and be computer literate. Whether these variables cause an influence on the results is unknown, but in the author's opinion, unlikely.

Several approaches were pursued to try to construct a human focus curve. Data analysis and observer feedback led to the gradual refinement of the methodology, until a working approach was found. The methodology for establishing a focus curve minimises the amount of specified data, and is completely observer-generated. No time limits are imposed, and observers are free to browse through the candidate images as many times as they desire, thereby improving the accuracy of their opinion.

The next step is to compare these ground truths established from a large body of observers with the behaviour of the various focus measures proposed in the literature,

and this is investigated in the following chapter.

Chapter 5

Performance of focus measures

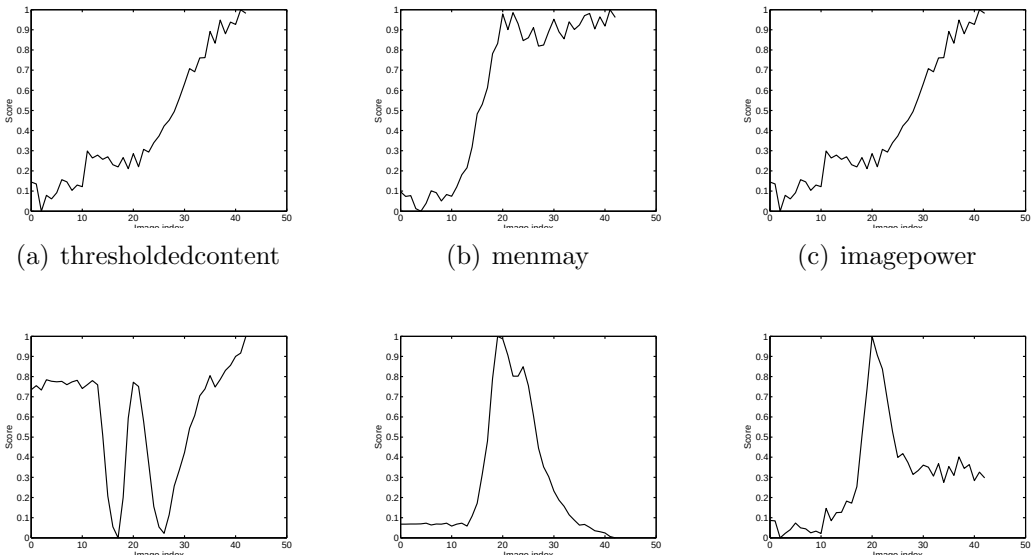
The results generated in the previous chapter provide the ground truth for a number of scenes. The most in-focus image of each scene has been identified, and a focus curve has been generated. This chapter shows the results of processing the same scenes with the focus measures discovered in the literature and created during the course of this research.

A focus measure is a mathematical function which processes a number of images and determines which image is most in focus. All the measures reviewed in Sections 2.3 and 3.2 process one image at a time, giving each image a score. The image of the scene with the highest score is the most in-focus image.

Section 2.3.5 describes how previous researchers have compared focus measures, by giving them a quantitative score according to their performance. The score is derived from a number of individual properties measured from the focus curve. Whilst this score is of merit, some of the underlying properties measured have more importance than others – for example, a high-scoring measure might still be inaccurate. A focus measure should be able to identify the same image as being maximally in-focus as that identified by human observers, and this accuracy should be of greater importance than, say, the width of the curve at 50% height. As such, consideration of the focus measures' accuracy is given special attention:

5.1 Accuracy

For each scene shown to the human observers, the same images were prepared for mathematical assessment by a range of focus measures; the same resolution (640x480 pixels) was used, though they were reduced to grayscale from RGB colour, as the focus measures process a single luminance channel, rather than full colour information. This colour reduction was performed by extracting the luminance information



(a) thresholdedcontent (b) menmay (c) imagepower

(d) phasecongruence (e) cranepeak (f) tenengrad

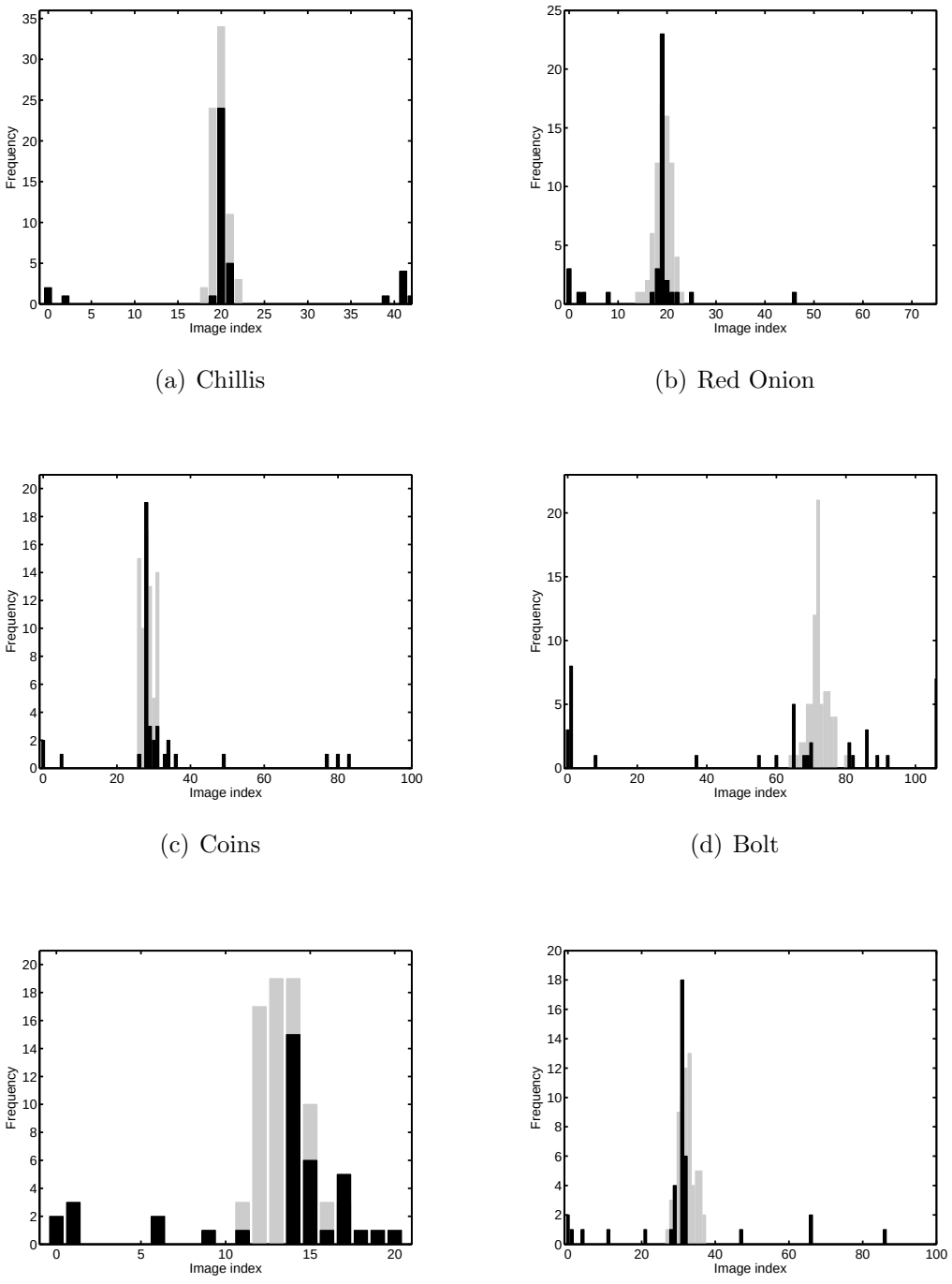
Figure 5.1: Score plotted against focus distance for a range of focus measures using the 'chillis1' sample scene. Four measures fail to form a global peak near the human observers, who selected image 20 as most in focus.

from the colour image, using the NTSC luminance formula (Equation 3.1). Once the scores were computed, they were normalised such that, for each measure, the range of scores given to each scene was 0–1.

The focus curves for all measures were plotted, then these graphs examined by eye, allowing several initial qualitative observations to be made: Firstly, certain focus measures fail to produce a peak in score anywhere in the middle of the set of images. Examples of this behaviour include the *thresholdedcontent*, *menmay* and *imagepower* measures. Secondly, the *phasecongruence* measure, whilst it does have a maximum in the right place, this is not the global peak for the measure. Figure 5.1 shows these four focus curves, together with those for *cranepeak* and *tenengrad*, two measures which performed well. The full set of focus curves for the 'chillis' scene are included in Figures 5.5 through 5.7 (pages 116 – 118).

Results from Section 4.1.3 show that the modal response of human observers viewing the chillis scene was to select image 20 as the most in focus, with a 95% confidence interval of ± 1.72, but only 80% of the measures selected images 18-22 as their peak. This is illustrated by Figure 5.2, a series of histograms showing the distribution of the best image selected by each measure across the selection of scenes as used by the human observers.

Indeed, by considering the accuracy of the measures when assessing all six scenes under test, it is possible to aggregate the results and establish which measures are



(a) Chillis

(b) Red Onion

(c) Coins

(d) Bolt

(e) Strawberries

(f) Towel

Figure 5.2: The graphs show the frequency with which each candidate image was selected as being 'best' for the given scene by individual focus measures. Bars in grey show the human results for the same scene, as Figure 4.8. It is not possible to determine which is the best measure by examination of these frequency charts. What is shown is that (for most scenes) the image most selected as 'best' by observers is also the modal image selected by focus measures. An assessment of the individual performance of the focus measures is described in the following section.

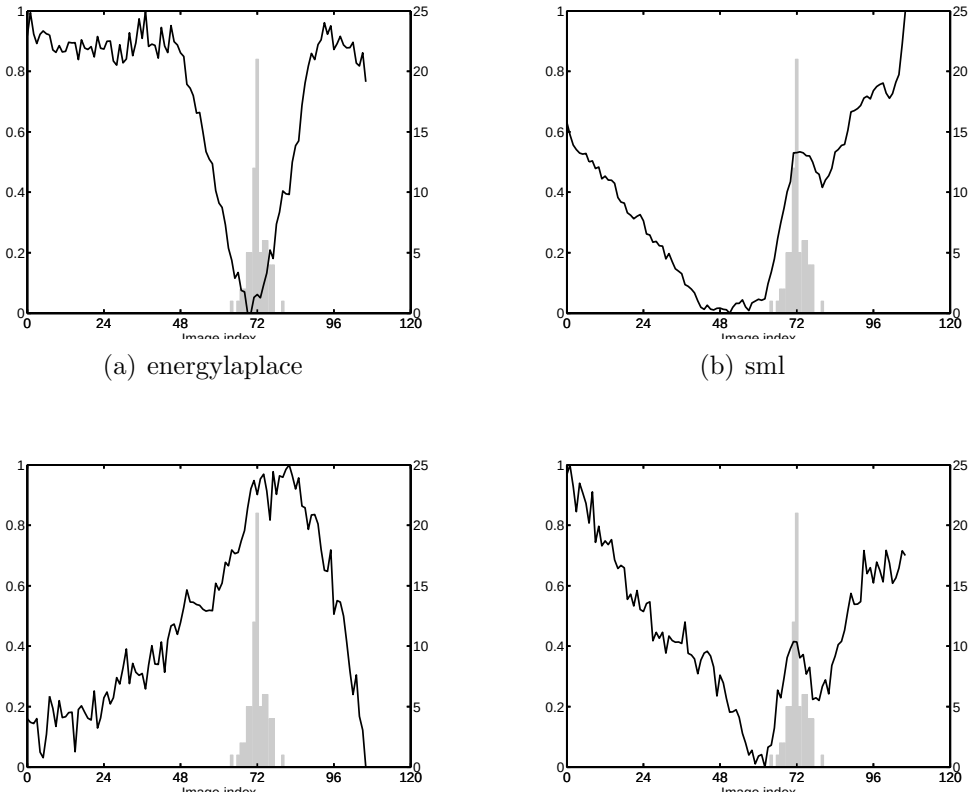
reliable. For a given scene, each measure was classified as either perfectly agreeing with the modal human response, falling within the 95% confidence intervals of the human result, or failing to match human observers. Starting with the scene with the fewest number of measures that failed to match human observers, the measures that succeeded (either because they agreed, or were within the 95% intervals) in a given scene were then assessed against the next scene, and the process repeated. These results are shown in Table 5.1:

Scene	**Perfect agreement**	**Within 95%**	**Failed**	**Not considered**
Chillis	28	8	9	n/a
Red Onion	27	7	2	9
Coin	23	8	3	11
Towel	0	27	4	14
Strawberries	0	20	7	18
Bolt	0	2	18	25

Table 5.1: Whilst most focus measures do well for some scenes, the subset that perform well across a number of scenes is far lower. The scenes in this table are sorted such that the fewest measures are eliminated at each stage.

This shows that no single focus measure selected the same best image as the human consensus every time. And, over all the scenes, only two measures came close to matching the ground truth. They were *cranepeak* and *tenengrad*, two of the earliest proposed focus measures (1966 and 1970 respectively). However, it is worth investigating the final scene in detail. Prior to the measures being tested against the bolt scene, there were still 20 measures that had performed satisfactorily in the first five scenes (*autocorrelation*, *brennergradient*, *cranepeak*, *cranesum*, *energylaplace*, *energylaplace5a*, *energylaplace5b*, *energylaplace5c*, *laplace*, *smd*, *sml*, *squaredgradient*, *tenengrad*, *thresholdedabsolutegradient*, *triakis11s*, *va*, *voll4*, *waveletw1*, *waveletw2* and *waveletw3*). That almost all of these successful measures were eliminated in the bolt scene suggests that there must be something particularly difficult for the focus measures within it. By looking at a few of the measures that failed on the bolt scene, it is apparent that they are scoring some candidate images very highly, and that these scenes are a long way from the ground truth. Four particular measures highlight the difficulties encountered whilst assessing this scene, and they are shown in Figure 5.3.

Firstly, *energylaplace* completely fails – it indicates that the frame contains least energy (and thus lowest score) when the scene is actually in focus. Secondly, both *sml* and *waveletw1* report a minima around image 50–60, before climbing again. Both show a local maxima for the image most frequently selected by the observers, but this is small in relation to the measures' overall scores. Finally, *triakis11s* does



(a) energylaplace (b) sml

(c) triakis11s (d) waveletw1

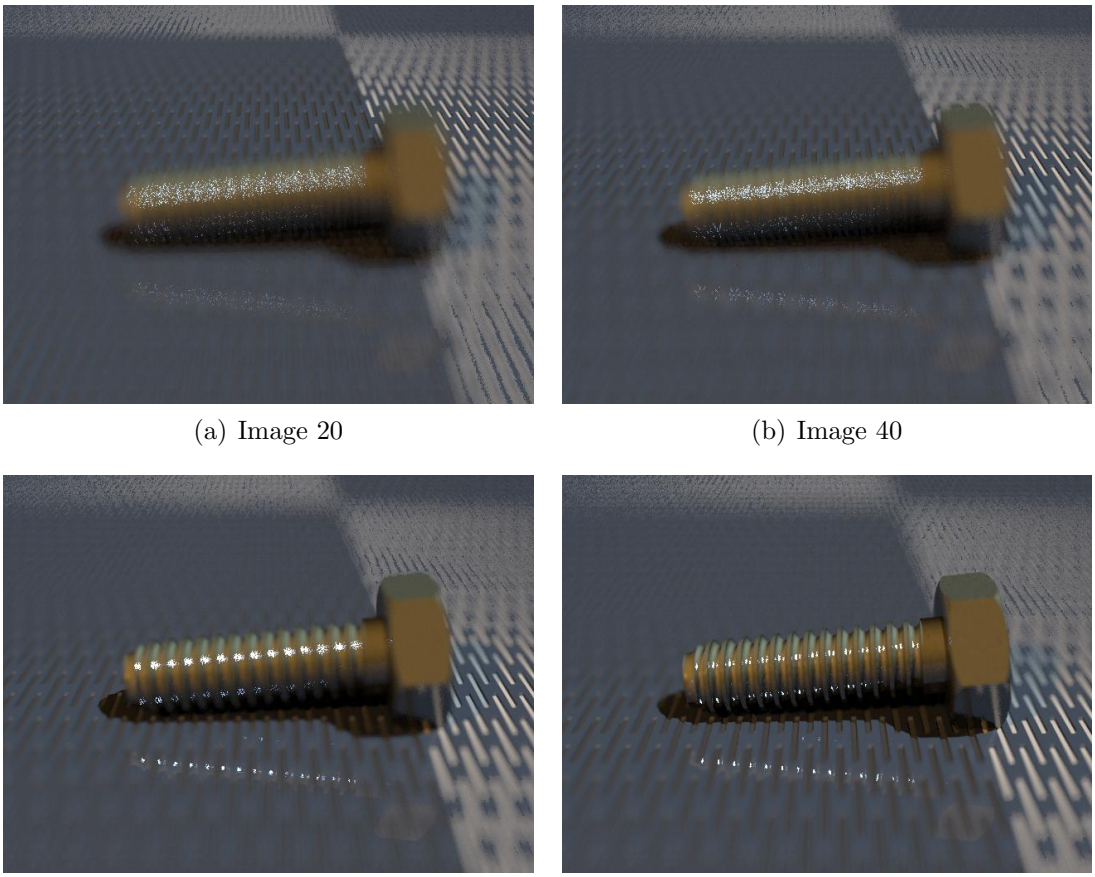
Figure 5.3: Score plotted against focus distance for a few of the focus measures using the 'bolt' sample scene. The black line indicates the score given by the relevant focus measure, whilst the frequency with which each image was selected by human observers as being the best is shown as a grey histogram.

significantly better, but its peak is well outside the human responses, and so is insufficiently accurate.

Unlike all the other scenes, which were captured photographically, the bolt scene was artificially rendered using ray tracing by Povray (see Section 3.1.3). This did not prevent humans from selecting their preferred best image, but did prevent the mathematical focus measures from making the correct decision.

The nature of ray tracing is to render the image by considering one pixel at a time, and trace it through the three-dimensional model of the scene until its colour and brightness can be characterised. Because the PSF of an optical system is approximately gaussian in shape, and the gaussian distribution has a wide extent, accurate ray tracing with optical blur takes a very long time. For the bolt scene, Povray was instructed to continue rendering until 250 rays had been processed per pixel, or the pixel's value was over 95% certain, whichever was sooner (Povray took approximately 90 minutes to render each frame when using these parameters). However, it is clear from Figure 5.4 that, for the images identified from the local minima and maximas in Figure 5.3, there is speckling and dithering present in these rendered scenes.

Thus, it appears that most measures are susceptible to dithering, a technique typically used to produce visually appealing images even when the absolute number of colours present in the image is low. Significant work has been done to improve the way dithering is performed so that it looks best under a model of human perception (eg [180]), but this can only be done once the full image is known. With ray tracing, dithering is an artefact of the uncertainty in the image that has been produced so far, rather than a mechanism for coping with fewer colours. Indeed, the dithering apparent in these images is very like that shown in Figure 2.17 (page 57), an example of compressive imaging, where the image quality slowly improves as ever more data points are acquired. So, whilst humans can apply top-down knowledge, and know that they are looking at a bolt (or letter R in the compressive imaging example), algorithms that attempt to determine the scene's focus on a per pixel basis are greatly hindered by the high contrast between adjacent pixels, and struggle to obtain a valid measure in such conditions.



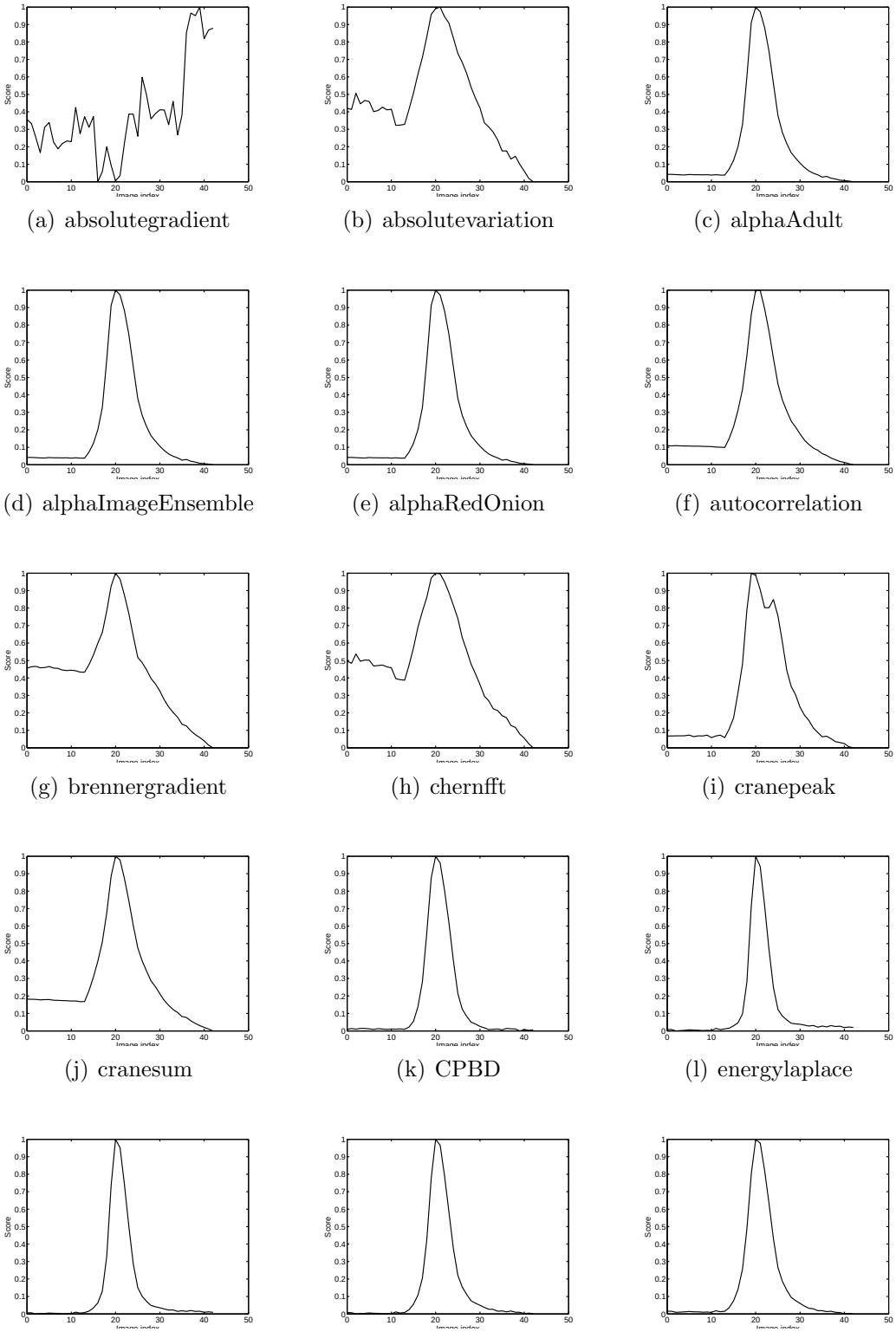
(a) Image 20

(b) Image 40

(c) Image 60

(d) Image 70

Figure 5.4: The bolt scene was rendered by Povray with a number of different camera distances. However, rendering a scene exhibiting focal blur with a ray tracing algorithm takes a very long time. Even after 90 minutes rendering, there are artefacts present which confused many of the focus measures.



(a) absolutegradient (b) absolutevariation (c) alphaAdult

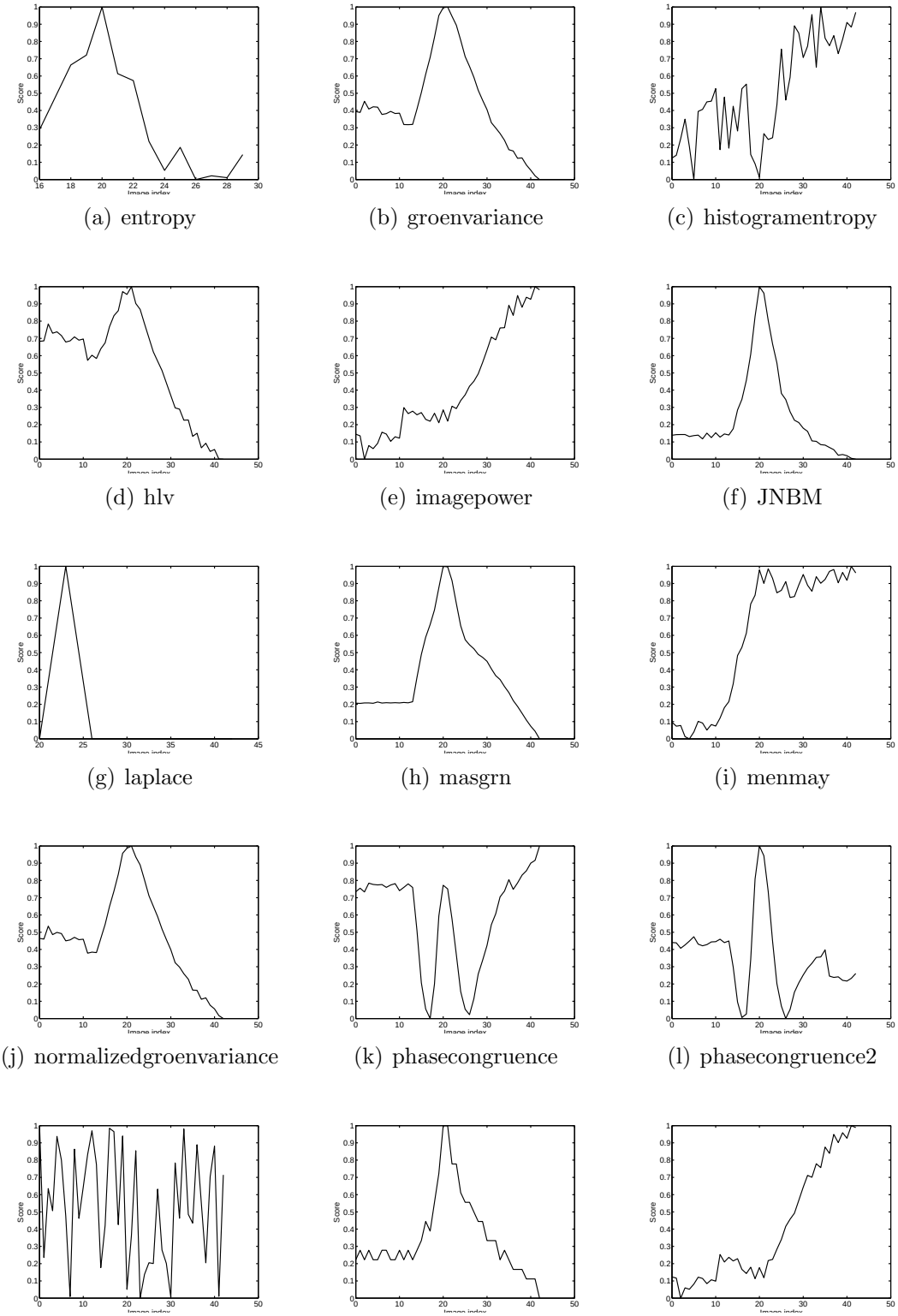
(d) alphaImageEnsemble (e) alphaRedOnion (f) autocorrelation

(g) brennergradient (h) chernfft (i) cranepeak

(j) cranesum (k) CPBD (l) energylaplace

(m) energylaplace5a (n) energylaplace5b (o) energylaplace5c

Figure 5.5: Score plotted against focus distance for a range of focus measures using the 'chillis1' sample scene



(a) entropy (b) groenvariance (c) histogramentropy

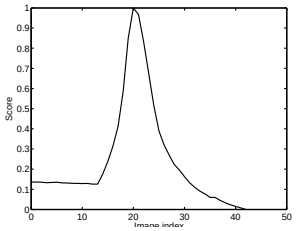
(d) hlv (e) imagepower (f) JNBM

(g) laplace (h) masgrn (i) menmay

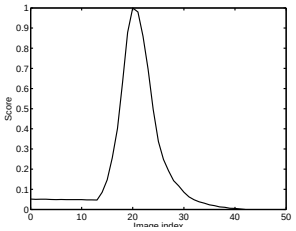
(j) normalizedgroenvariance (k) phasecongruence (l) phasecongruence2

(m) randomnumber (n) range (o) rawlaplace

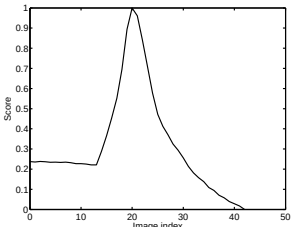
Figure 5.6: Score plotted against focus distance for a range of focus measures using the 'chillis1' sample scene



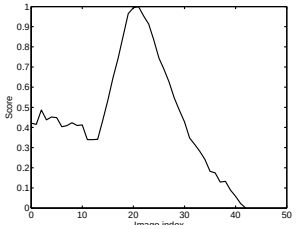
(a) smd



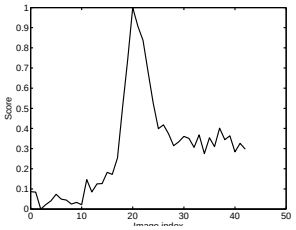
(b) sml



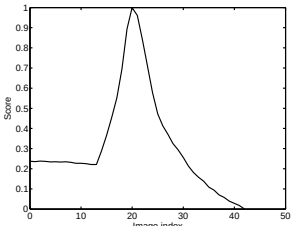
(c) squaredgradient



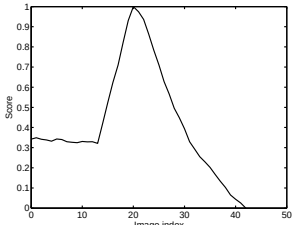
(d) standarddeviationbased-autocorrelation



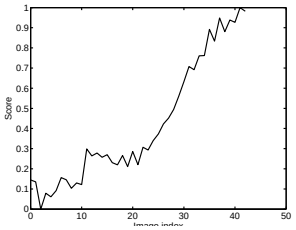
(e) tenengrad



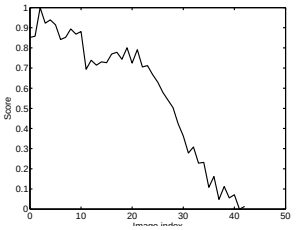
(f) thresholdedabsolutegradient



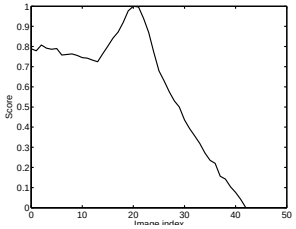
(g) triakis11s



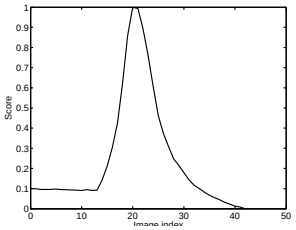
(h) thresholdedcontent



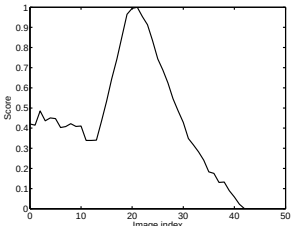
(i) thresholdedpixelcount



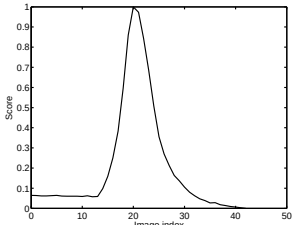
(j) va



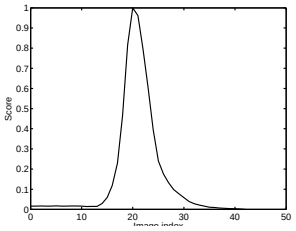
(k) voll4



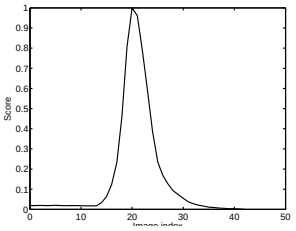
(l) voll5



(m) waveletw1



(n) waveletw2



(o) waveletw3

Figure 5.7: Score plotted against focus distance for a range of focus measures using the 'chillis1' sample scene

5.2 Quantitative scoring

Beyond the accuracy of the individual measures, discussed in the previous section, there are other properties of the focus curve that can be quantified and scored using the methodology introduced in Section 2.3.5, and specifically using Equation (2.8). To recap, the score is the Euclidean distance between the characteristics of the focus curve (including the location of the peak, sharpness, and number of false maxima) and those of an ideal response – a low score is better. All the focus measures were scored for each scene using this method, using the modal human response as the ground truth for assessing the accuracy. The full results for all scenes are included in Appendix E. These have been summarised across the scenes in Table 5.2.

Overall, the measures had a considerable number of false maxima, and thus a low range. This means that any control system using these focusing measures might need to average the values from adjacent points, or do a more wide-ranging search for the maxima to ensure that it did not settle upon an undesirable local maximum within the curve.

The measures that ranked highly in this quantitative comparison (*cranesum* and *autocorrelation*) are not those which performed best in the earlier assessment of accuracy. Thus, they have attributes upon which the existing comparison methodology places importance, such as a wide range and few false maxima, but do not have the highest overall accuracy; both perform very poorly on the bolt scene.

If the bolt scene is excluded from this quantitative ranking, then the best measure is *waveletw3*, followed by *waveletw2*, *energylaplace5c*, *autocorrelation* down to *laplace* ranked fifth. The two measures identified as being most accurate in the previous section are then ranked 21st and 23rd (*cranepeak* and *tenengrad*), rather than 32nd and 27th respectively if all scenes are considered.

Previous surveys of focus measures show a certain degree of similarity to one-another, with Santos, Sun and Groen's surveys all indicating that the *normalised-groenvariance* measure was amongst the best few measures. However, in these experiments, *normalisedgroenvariance* measure has not performed as well. It was a long way from the ground truth position in the strawberries scene, and suffered from having a wide peak, and thus scoring badly on the 'width' criteria.

In addition to not ranking *normalisedgroenvariance* highly the results here have other differences with previous surveys. To a certain extent, this is to be expected: As previous surveys have not examined as many measures, the lack of a match between the best algorithms found by them, and the results here, are understandable. For example, the best measure found by Santos was *voll4* and *voll5*. However, none of the measures performing better in the results presented here were considered by

Measure	Chillis	Coins	Bolt	Red onion	Strawberries	Towel	Total	Rank
absolutegradient	42	31	38	36	43	35	225	42
absolutevariation	25	12	11	35	30	1	114	19
alphaAdult	15	18	16	6	11	30	96	12
alphaImageEnsemble	17	11	20	6	12	27	93	9
alphaRedOnion	15	22	16	6	38	37	134	28
autocorrelation	4	8	36	5	10	9	72	3
brennergradient	7	17	23	30	25	16	118	22
chernfft	24	5	13	34	37	8	121	24
cranepeak	30	33	24	20	24	28	159	32
cranesum	3	4	19	16	18	7	67	1
CPBD	28	35	22	18	14	32	149	30
energylaplace	27	27	31	13	1	20	119	23
energylaplace5a	26	26	34	14	2	19	121	24
energylaplace5b	23	25	32	15	3	18	116	20
energylaplace5c	19	23	27	10	7	17	103	18
entropy	43	44	44	44	44	42	261	44
groenvariance	21	10	10	24	29	2	96	12
histogramentropy	44	42	26	41	39	36	228	43
hlv	35	32	1	42	22	43	175	33
imagepower	39	36	8	32	41	33	189	34
JNBM	18	30	39	19	20	31	157	31
laplace	14	7	33	11	4	26	95	11
masgrn	12	20	28	17	26	23	126	26
menmay	37	40	12	39	40	44	212	41
normalizedgroenvariance	31	16	7	31	28	3	116	20
phasecongruence	41	43	25	25	21	40	195	36
phasecongruence2	33	39	14	38	33	39	196	37
randomnumber	45	45	45	45	45	45	270	45
range	29	34	43	37	34	29	206	39
rawlaplace	38	38	18	43	36	38	211	40
smd	2	3	37	12	8	6	68	2
sml	6	28	21	2	15	22	94	10
squaredgradient	9	1	29	22	16	13	90	7
standarddeviationbasedautocorrelation	20	15	4	27	31	5	102	16
tenengrad	34	29	2	21	19	25	130	27
thresholdedabsolutegradient	9	1	29	22	16	13	90	7
triakis11s	11	13	6	28	23	21	102	16
thresholdedcontent	39	36	8	32	41	33	189	34
thresholdedpixelcount	36	41	15	40	27	41	200	38
va	32	19	5	29	35	24	144	29
voll4	1	6	35	9	13	11	75	4
voll5	22	14	3	26	32	4	101	15
waveletw1	13	9	40	1	9	12	84	5
waveletw2	8	24	42	4	6	15	99	14
waveletw3	5	21	41	3	5	10	85	6

Table 5.2: Overall quantitative scores for all measures, together with an overall rank of their aggregate performance. Note that the measures that best matched human results (*cranepeak* and *tenengrad*) and ranked relatively low by this quantitative assessment.

Santos, and thus their conclusion matches these results.

At the other extreme, Sun reported that the wavelet measures performed poorly, and that the best was *normalisedgroenvariance* followed by *standarddeviationbased-autocorrelation*, results which are quite contradictory to these presented here. One possible explanation is the difference in subject matter – Sun was considering images derived from microscopy rather than general purpose scenes.

Discrimination between focus measures by Firestone et al was mainly a commentary on the width at 50%, as their results showed few false positives and high accuracy. That these parameters of the focus curve differed in the results presented here is most likely due to the higher resolution of images used here than by Firestone.

In conclusion, quantitative results have been obtained, but as with previous surveys, they do not point to a single measure as being a clear winner. *Cranesum* and *smd* have achieved the highest rank in these scenes, but they are not the most accurate measures, nor have they been identified by previous surveys as being especially highly performing.

5.3 Shape

One further analytical technique that can be applied to the focus curves is to see how they compare with the human-derived focus curves. By mapping the near and far sides of the focus curves to each run from 0 to 1, the curve can be compared with the human results determined in Section 4.2 and shown in Figure 4.17 in the red onion scene. Using only those measures whose peak was within the 95% confidence interval of the human opinion, Table 5.3 shows the defocussed image identified by each measure as being the appropriate image for each position. The columns labelled g, D, h, i, E and j, the same references as in Figure 4.15, represent the positions in the focus curve where the score is 25%, 50% and 75% on the ascending side of the curve, and the converse on the curve's descent.

Eleven of the measures fit within the human results. Of those that do not fully match, most match the descent (far side) of the curve. There is not a measure which matches the ascent but not the descent. Only a few measures are fail to match at all: *Thresholdedpixelcount* and *phasecongruence2* both wildly mismatch, and the *energy-laplace* series of measures have a narrower peak than the human results. However, the progression from the *5a* to *5c* variants of this family of measures does cause the curve to widen and then start matching human results.

Measure	**G (25%)**	**D (50%)**	**H (75%)**	**I (75%)**	**E (50%)**	**J (25%)**	**Comment**
Humans	*1.1-7.0*	*4.1-11.4*	*8.3-14.9*	*22.4-33.3*	*25.2-43.9*	*29.4-59.9*	*95% confidence intervals*
absolutevariation	3	6	11	26	32	43	Fits
alphaAdult	13	15	17	25	27	33	Fits right side
alphaImageEnsemble	13	15	17	25	27	33	Fits right side
alphaRedOnion	13	15	17	25	27	33	Fits right side
autocorrelation	11	14	17	25	28	35	Fits right side
brennergradient	2	4	13	24	26	33	Fits
chernfft	3	6	11	25	33	43	Fits
cranesum	7	12	16	25	29	38	Fits right side
CPBD	14	16	17	24	26	29	Fits right side
energylaplace	16	17	19	23	24	27	Too narrow
energylaplace5a	15	17	18	23	25	27	Wider, but still too narrow
energylaplace5b	14	16	18	23	25	31	Wider, but still too narrow
energylaplace5c	13	16	18	24	26	32	Fits right side
groenvariance	3	6	11	26	32	47	Fits
JNBM	11	15	17	24	27	33	Fits right side
laplace	14	16	18	23	25	30	Fits right side
masgrn	3	13	17	24	26	31	Fits right side
normalizedgroenvariance	3	6	11	26	31	43	Fits
phasecongruence2	14	16	18	24	25	77	Not a fit
smd	8	14	17	24	26	34	Fits right side
sml	13	14	17	24	28	34	Fits right side
squaredgradient	5	8	15	24	27	39	Fits
standarddeviationbasedautocorrelation	3	6	10	27	33	47	Fits
thresholdedabsolutegradient	5	8	15	24	27	39	Fits
triakis11s	4	8	14	24	29	40	Fits
thresholdedpixelcount	4	7	11	62	77	77	Not a fit
va	2	4	12	24	27	36	Fits
voll4	10	14	17	25	29	37	Fits right side
voll5	3	6	11	27	33	48	Fits
waveletw1	12	15	17	24	27	33	Fits right side
waveletw2	14	16	18	23	25	29	Not a fit
waveletw3	14	16	18	23	25	29	Not a fit

Table 5.3: Focus measures scores were mapped such that they went to zero at each extreme. Their images for the 25, 50 and 75% score values were then determined and compared with the human results (Figure 4.17). Most measures match the human results.

5.4 Conclusions

The methodology proposed by earlier work has been used to evaluate all focus measures on the same scenes as the human observers. The best measures, by this quantitative assessment, are *cranesum* and *smd*. Even the best measure was relatively poor on some scenes, getting ranked as low as 19th in the bolt scene, suggesting that there is an opportunity for a better focus measure to be developed.

If accuracy is solely considered, then no measures perfectly agree with humans, and only two (*cranepeak* and *tenengrad*) even come close to matching human results in all scenes. One particular scene was especially troublesome for the focus measures which appear unable to handle images acquired via iterative methods such as ray tracing or compressive imaging, due to the dithered nature of the image as it is being constructed.

Beyond quantitative comparisons, the shape of the focus curves have also been compared with human results. This has shown that eleven of the measures being evaluated fit within the 95% confidence intervals of the human results.

The most promising focus measures are as tabulated below in Table 5.4. No single measure has performed well in all the tests. Indeed, it can also be seen that the highest scoring measures have been amongst the worst in other tests.

Overall, the results from this chapter show that a significant subset of the measures under test fall within the range of likely human opinion, though none perfectly agree in all scenes or with all tests. To refine the evaluation further, additional human-produced information is required, and this is described in the following chapter.

Measure	Accuracy	Score rank	Shape
tenengrad	Best	27	-
cranepeak	Best	32	-
smd	Good	2	Half-fit
autocorrelation	Good	3	Half-fit
voll4	Good	4	Half-fit
waveletw1	Good	5	Half-fit
waveletw3	Good	6	No fit
squaredgradient	Good	7	Fit
thresholdedabsolutegradient	Good	7	Fit
sml	Good	10	Half-fit
laplace	Good	11	Half-fit
waveletw2	Good	14	No fit
triakis11s	Good	16	Fit
energylaplace5c	Good	18	Half-fit
energylaplace5b	Good	20	No fit
brennergradient	Good	22	Fit
energylaplace	Good	23	No fit
energylaplace5a	Good	24	No fit
va	Good	29	Fit
cranesum	Poor	1	-
alphaImageEnsemble	Poor	9	Half-fit
groenvariance	Poor	12	Fit
absolutevariation	Poor	19	Fit

Table 5.4: Summary of top performing focus measures by the different assessment criteria used described in this chapter. Only measures that were found to be accurate, or scored in the top ten, or were a reasonable match to the focus curve shape have been included; many measures have not performed sufficiently well to be listed. The measures are sorted first by accuracy, then score, then shape. Note that no measure scores well and has high accuracy and a good fit to the human focus curve.

Chapter 6

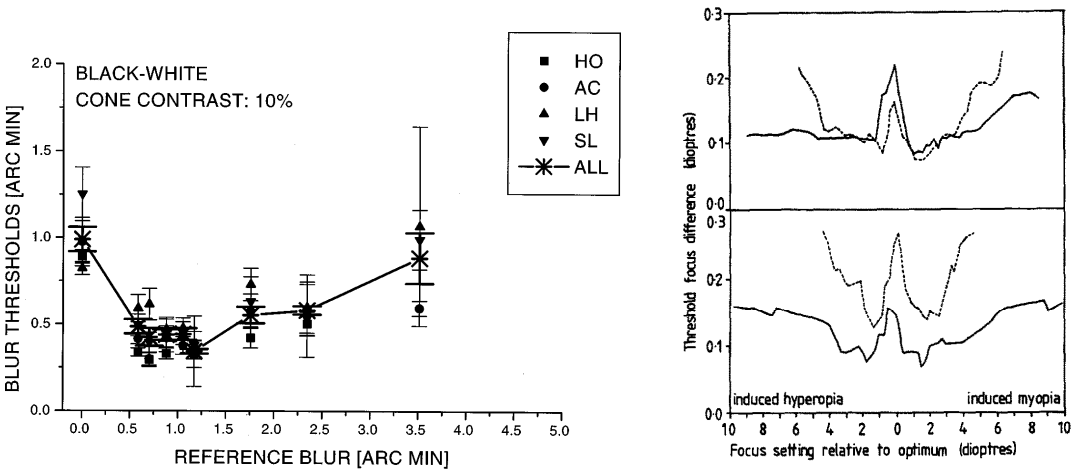
Psychophysical experiments

The results achieved from the experiments of the previous chapters have shown that some focus measures described in the literature are close to the results seen from a large body of human observers, and certainly within a 95% confidence interval. Others have performed highly in the quantitative assessment but do not match human results. Continuing the theme of previous chapters, further human results, collected via a different approach, need to be obtained in order to narrow down the focus measure that most closely matches humans.

The experiments in previous chapters allowed the subjects to spend an unconstrained amount of time considering their answer. This enabled them to think about the content of the images, scrutinising whichever aspects attracted their attention. The results show that people agree with each other, and no populations have been found which consistently provide different answers. However, there is no strong correlation between the human opinion, and any particular focus metric. This chapter describes a different set of experiments which approach the concept of focus from the other side – rather than asking peoples' considered opinions, they attempt to capture normal peoples' underlying perceptual response.

Experiments described in the literature show how human sensitivity to blur, both in terms of detection and discrimination, can be measured using carefully controlled experimental methodologies and more constrained viewing conditions than used thus far in this research. The results published by other researchers, however, have been from experiments using relatively simple stimuli, such as a simple step edge [45] in white light, and under different monochromatic luminances [84]. Experiments involving real scenes have been limited to Walsh, using observers with anaesthetised eyes and the observer method, showing peak discrimination ability with a pedestal of approximately 2D [52].

This chapter describes experiments conducted with multiple subjects following



(a) (b)

Figure 6.1: Graphs showing the results of similar experiments by other researchers. (a) Blur discrimination, with stars showing blur thresholds averaged over all observers, on a simple grating stimuli [181]. (b) The solid line shows the threshold for detection of focus curve on a street scene; the dashed line shows the corresponding results from a simple 5.1 c/deg grating. The lower figure shows the results under monochromatic green light, and the upper figure under white illumination [52]

.

a methodology similar to that used by Wuerger [84] and Burr [56], and measures the blur detection and discrimination thresholds with various amounts of baseline (pedestal) blur, using natural scenes from the Van Hateren image library. It was anticipated that the results from such experiments would be qualitatively the same as published results from simpler stimuli, though unlikely to show exactly the same thresholds, due to the complexity of the stimuli. Such a quantitative difference has been reported before – Walsh's experiments showed that for a complex stimuli, the discrimination threshold at the point of optimum discriminability was greater than for simple sinusoidal gratings. Figure 6.1 shows the results from these earlier experiments.

Walsh showed that with grating stimuli (see Figure 2.9), discrimination was best at approximately 2D of defocus, where changes of approximately 0.15D could be discriminated. However, for his complex stimuli (the street scene), the best discrimination performance was at 1.5D, where the threshold (0.1D) was smaller than that at which observers could discriminate blur in the grating stimulii. That is, blur discrimination was better in the more complex scene.

6.1 Method

In summary, the experiment involves showing observers a pair of images and asking them which is most in focus. The images are changed, and the process repeated. By refining the pairs of images displayed, the experiment determines the threshold at which the subject can only correctly discriminate the most in-focus image 82% of the time. Similar experiments conducted by previous researchers have shown that discrimination improves as a small amount of blur is added, but this trend reverses as more blur is added. This dipper response is described in more detail in Section 2.2.

All experiments were conducted by the author. Additional subjects naive to the purpose of the experiment and unfamiliar with the concept of the 'dipper' response and psychophysics. Subjects were not colour blind, and required to self-declare themselves as having had an eye test within the past 12 months (wearing any correction as deemed necessary by their optician).

The stimuli were presented centrally on a Dell P992 CRT monitor connected to an ATI Radeon 9800 graphics card, via a video attenuator (as described by Pelli [165]) which enables pseudo 12-bit control of pixel luminance, and so permitting the compensation of nonlinearites in the presentation whilst retaining the ability to display a full precise range of 256 grey levels. The monitor was switched on at least two minutes before any experiments or calibration were performed, to minimise any artefacts whilst stabilising. The equipment was calibrated linearly through its luminance range by using a Minolta CS100 photometer, and applying compensation by the use of a look up table.

The screen operated at resolution of 1024x768 pixels, with a refresh rate of 100Hz. Its visible area was 36.5cm by 27.4cm, and stimuli were 256x256 pixels in size (8.7cm by 8.7cm). The part of the screen not displaying the stimuli had a mean luminance of 44 cd/m^2. Peak luminance was 112 cd/m^2, and the darkest part of the screen 0.5 cd/m^2. The stimuli did not have the same mean intensity, and so the mean luminance of the stimuli could differ from the peripherary of the screen. Low ambient illumination was used, such that the screen was the brightest object in the room.

Subjects were seated in a comfortable office chair in a fixed location so that they were 116cm from the screen, equating to the stimuli subtending 4.3°. No chin rests or other guides or restraints were used. The experiments were divided into intervals of approximately ten minutes duration, between which subjects were encouraged to relax, stretch their legs, and have refreshments if they desired. When no stimulus was displayed on the screen, a small cross was shown at the centre of the screen

to provide a fixation target, helping the observers remain comfortable and ready to observe the stimulus.

Images from the scenes used in the previous experiments (such as the chillis or strawberries) are not suitable for these experiments. This is because they are not uniformly in focus, and so observers might make decisions based on the underlying blur present in the image (as a result of the 3D nature of the scene), rather than the additional blur added as part of this experiment. Secondly, whilst the individual images were acquired in sequence, no calibration to establish the precise change in either focal distance or optical properties of the lens assembly was conducted. This means that it is not possible to say that the additional blur between image 1 and 2 is equal to that between image 50 and 51. Instead, two scenes, both from the Van Hateren image set (numbers 0005 and 1342, see Figure 3.4) were used for the experiments in this chapter. Similarly, the Van Hateren images could not be have been used in previous experiments, as they are uniformly in focus, and any blur that is applied will appear to defocus them in the same direction. That is, there is no depth information available with which to artificially reduce the depth of field and then display the scene with a variety of focal positions, each with a narrow depth of field.

The experiment required observers to discriminate between two images temporally separated. The original Van Hateren image was blurred to the baseline condition, $\sigma = b$, to create the first stimuli. The second stimuli was created by blurring the original image to $\sigma = b + \delta b$, where b represents the pedestal blur extent, and δb the amount of additional blur being discriminated. Blur was added by convolving the image with a gaussian kernel whose standard deviation, σ, was set to the desired blur. From these blurred images, a randomly selected square subset of 256x256 pixels was extracted. The computer determined, randomly, which order to display the two images. The QUEST [170] procedure was used to find the thresholds. It builds a model of responses received, and uses them to place the next value of δb at the mode of its current posteria probability distribution function – that is, at its current threshold estimate. Both stimuli were displayed centrally on the screen, in the same position, for 300ms separated by a 500ms interval during which the fixation target was displayed.

Abrupt and artefact-free stimuli change, as well as accurate presentation duration, was possible by using the timing and graphics functions present in the Psychophysics toolbox for Matlab [171]. This software library provides easy access to OpenGL functions, which include the ability to prepare the stimuli in such a way that it can be applied to the screen during a vertical retrace (the `flip` command).

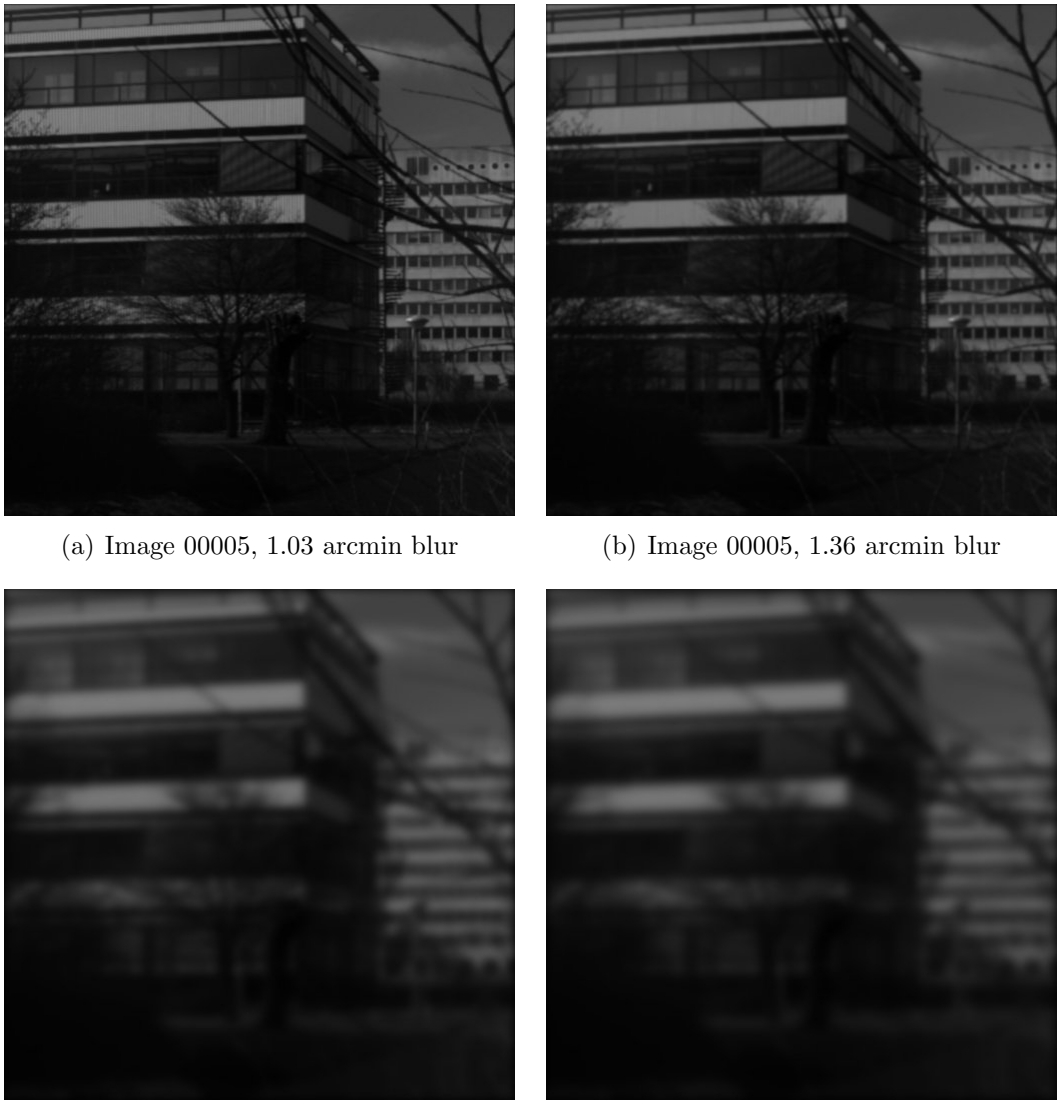
Such a technique means that no ripping or tearing of the image is visible on the screen; artefacts which would be present if the stimuli were directly copied into the active framebuffer. Separately, because the onset happens at an exact time (coincident with a particular screen refresh cycle), the precise onset and duration of stimuli presentation can be controlled.

The observers were instructed to record their opinion as to whether the first (or second) image was most in focus after the pair of images had been displayed, by pressing one of two buttons, and no time restriction was placed on their responses. Audio feedback was provided to indicate whether the choice was correct. To ensure that the observer was familiar with the process, and to help them settle into the experiment, the first five responses for each condition were not used to refine the QUEST model, and instead were discarded. These initial images also had a large δb to help observers. Figures 6.2 and 6.3 shows a few typical stimuli that would be encountered by an observer during a session.

Once the initial practice phase was completed, the observers' response was compared with the truth known by the software (for which there was no ambiguity as blur was applied mathmatically). The QUEST model was then updated with the knowledge of whether the observer correctly discriminated the additional blur, δb.

Pilot trials were conducted to determine indicative responses, which were used to select the pedestal values, as well as to establish the typical rate at which observers completed the experiment. Based on these trials, eleven pedestal values were selected to be equally spaced (in log-space) between a gaussian kernel size of $\sigma = 10^{-1}$ and $\sigma = 10^{2.1667}$. This equates to a blur of between 0.1 and 151 arc minutes at the specified viewing distance. An experimental condition comprised the pedestal value and scene. Each observer was measured multiple times for a given condition.

To minimise habituation and help reduce observer fatigue, up to five different conditions were interleaved and each experimental session lasted approximately ten minutes. That is, when the experiment started, five conditions were selected, and presented randomly to the observer. At the completion of the QUEST procedure for a given condition, the software selected a new condition to start, providing no more than eight minutes had elapsed since the start of the session. This meant that observers completed each session after approximately 8-12 minutes, and that at the end of a session there were no incomplete QUESTs. After a short break, the observer started a new session, until all conditions were completed. For each condition, an absolute minimum of 105 observations were recorded (three repetitions per condition, with a minimum of 35 samples per run). Typically, around 160 observations per condition were actually recorded, as the QUEST procedure only



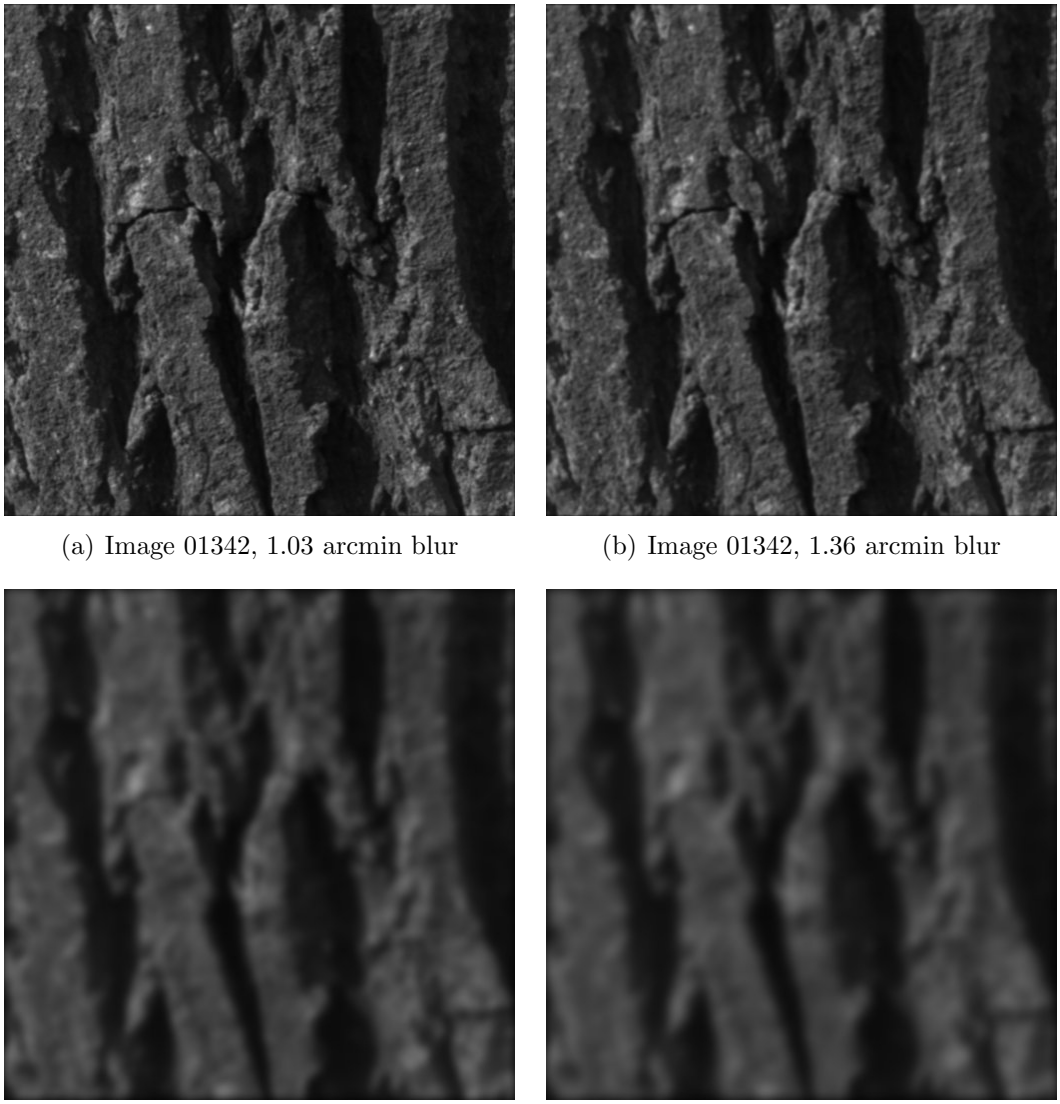
(a) Image 00005, 1.03 arcmin blur

(b) Image 00005, 1.36 arcmin blur

(c) Image 00005, 4.78 arcmin blur

(d) Image 00005, 5.53 arcmin blur

Figure 6.2: Typical stimuli shown to observers during the experiment described in Section 6.1. This is image 00005 from the Van Hateren set. These particular stimuli are shown as they demonstrate the discrimination thresholds determined for observer RAU. The images on the left are the pedestal images, and the images on the right show the amount of extra blur that needed to be added before the observer could discriminate 82% of the time under the experimental conditions. The results are introduced in Section 6.2. The slight darkening at the peripherary of each image is an artefact of the application of blur, as described in Section 3.4.10. These edges were never shown to the observer, as in each trial a random sub-portion of the image was selected, avoiding the edges. These images appear significantly darker in print – they are intended to be viewed on a linearly calibrated monitor.



(a) Image 01342, 1.03 arcmin blur

(b) Image 01342, 1.36 arcmin blur

(c) Image 01342, 4.78 arcmin blur

(d) Image 01342, 6.00 arcmin blur

Figure 6.3: Typical stimuli shown to observers during the experiment described in Section 6.1. This is image 01342 from the Van Hateren set. These particular stimuli are shown as they demonstrate the discrimination thresholds determined for observer RAU. The images on the left are the pedestal images, and the images on the right show the amount of extra blur that needed to be added before the observer could discriminate 82% of the time under the experimental conditions. The results are introduced in Section 6.2. The slight darkening at the peripherary of each image is an artefact of the application of blur, as described in Section 3.4.10. These edges were never shown to the observer, as in each trial a random sub-portion of the image was selected, avoiding the edges. These images appear significantly darker in print – they are intended to be viewed on a linearly calibrated monitor.

terminated after the earlier of 20 reversals or 75 stimulii responses being recorded.

6.2 Results

Four people, aged approximately 25-30, participated in the experiment – the author and three observers. Each subject performed multiple QUEST convergences for each pedestal condition – that is, during the experiment QUEST selected the candidate δb, refining its estimate with each response. For the purposes of analysis, the individual responses (a combination of δb, and whether the observer could discriminate at that point) were collated and analysed with a bootstrapping procedure to establish the 82% threshold and the 95% confidence intervals of that discrimination threshold. This was achieved by fitting a psychometric function using *psignifit*, a software package which implements the maximum-likelihood method described by Wichmann and Hill [175, 176], to the data.

The discrimination threshold, and its confidence intervals, are shown in Figures 6.4 and 6.5. Overlaid on the raw thresholds are the results of fitting the data to a contrast response function [182, p222] of the form in Equation 6.1, a hyperbolic function, using the the `rvc` and `fitit` functions from the Matteobox Matlab toolbox [183].

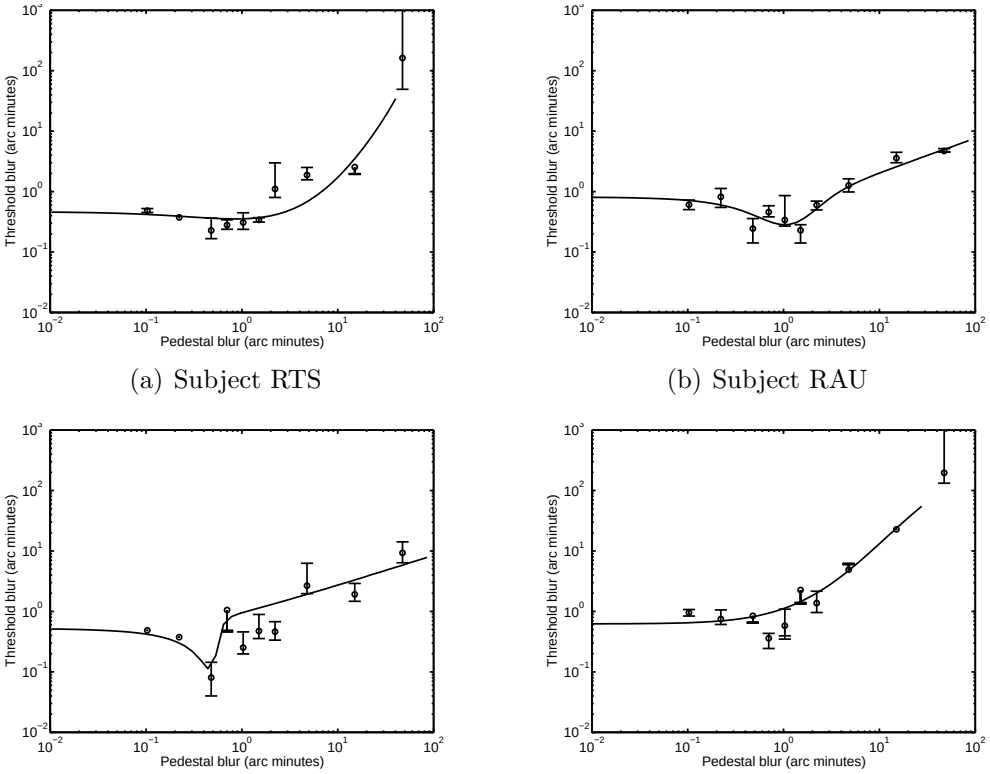
$$R(c) = k\frac{c^{m+n}}{\sigma^m + c^m} + ac \tag{6.1}$$

Representative images for one particular observer, RAU, and their blur discrimination thresholds are shown in Figures 6.2 and 6.3. These figures show the images from the pedestal position, and the additional blur that needed to be added for the observer to correctly discriminate 82% of the time (under experimental conditions).

As was expected, and previously observed in simpler stimuli, dip shaped responses were obtained for many observer/image combinations, though this was not universally observed in these experiments.

All but one observer/image combination show that discrimination is better when $b = 1$ arcmin than at $b = 0.1$ arcmin. By inspection, discrimination improves by approximately $\delta b = 0.5$ arcmin. Further, all observers show a continued reduction in discrimination performance beyond a pedestal condition of $b = 1$ arcmin of blur. One observer, NCA, shows the least dipper-like response – part of this is a result of a poor fit by the contrast function to the underlying data (see Figure 6.4(d)), but this does not explain the pattern of underlying results in Figure 6.5(d), for which no explanation can be found.

As the experimental conditions were the same for all observers, the presence of



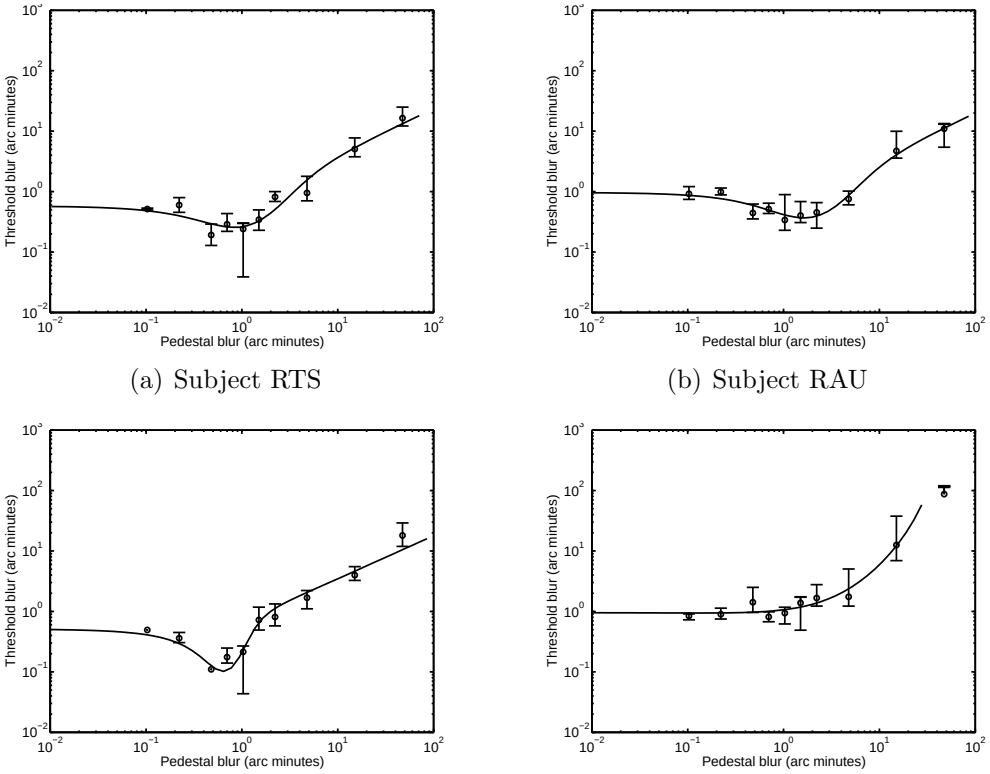
(a) Subject RTS

(b) Subject RAU

(c) Subject CAS

(d) Subject NCA

Figure 6.4: Blur discrimination thresholds as a function of pedestal blur, for observers viewing image 01342 from the Van Hateren library. Each estimate of threshold was based on at least three separate determinations (QUESTs) per measure, followed by a bootstrapping procedure to establish the 82% threshold and confidence intervals. This is overlaid with the best fit to a contrast response function (Equation 6.1) computed with a least squares fitting.



(a) Subject RTS

(b) Subject RAU

(c) Subject CAS

(d) Subject NCA

Figure 6.5: Blur discrimination thresholds as a function of pedestal blur, for observers viewing image 00005 from the Van Hateren library. Each estimate of threshold was based on at least three separate determinations (QUESTs) per measure, followed by a bootstrapping procedure to establish the 82% threshold and confidence intervals. This is overlaid with the best fit to a contrast response function (Equation 6.1) computed with a least squares fitting.

these different results suggests that some confounding effect beyond the immediate environment was likely to have been present. Alternative hypotheses, such as the 'dip' response not being present with certain subjects, whilst possible, are unlikely given the absence of such results in previous research. Even discounting these unusual responses, a greater variability was found in these results overall than has been published by previous researchers. This may be a result of the differences in result selection methodology, discussed in Section 3.5.1 – in these experiments, all data from all conditions presented to the observers is included; no data has been excluded from this analysis.

Certain data points do not have an upper or lower confidence interval plotted on the graph. This is an artifact of the bootstrap fitting routine which has been unable to determine the confidence interval at that point. It is anticipated that this could be addressed by discarding these results and re-measuring at the relevant conditions.

6.3 Analysis

Qualitatively, the results presented here agree with Watt and Morgan's findings for simple edge blur: "blur comparison is most precise at some non-zero criterion [pedestal] blur ... showing a decrease in threshold as criterion blur is increased from zero to an optimum level, beyond which the threshold rises rapidly".

By inspection, the measured discrimination thresholds do appear to vary more between pedestals (as expected), than to vary with other parameters. To confirm this, an analysis of variance (ANOVA) was carried out of the blur discrimination using pedestal, image and observer, and their combinations, as factors. This is shown in Table 6.1, from which it is confirmed that the blur discrimination is not correlated with observers, but highly dependent ($p < 0.001$) on the pedestal condition.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Participant	1	205.22	205.22	0.30	0.5840
Image	1	2073.96	2073.96	3.05	0.0843
Pedestal	1	26667.26	26667.26	39.28	0.0000
Participant:Image	1	18.13	18.13	0.03	0.8706
Participant:Pedestal	1	121.15	121.15	0.18	0.6738
Image:Pedestal	1	5580.33	5580.33	8.22	0.0053
Residuals	81	54991.97	678.91		

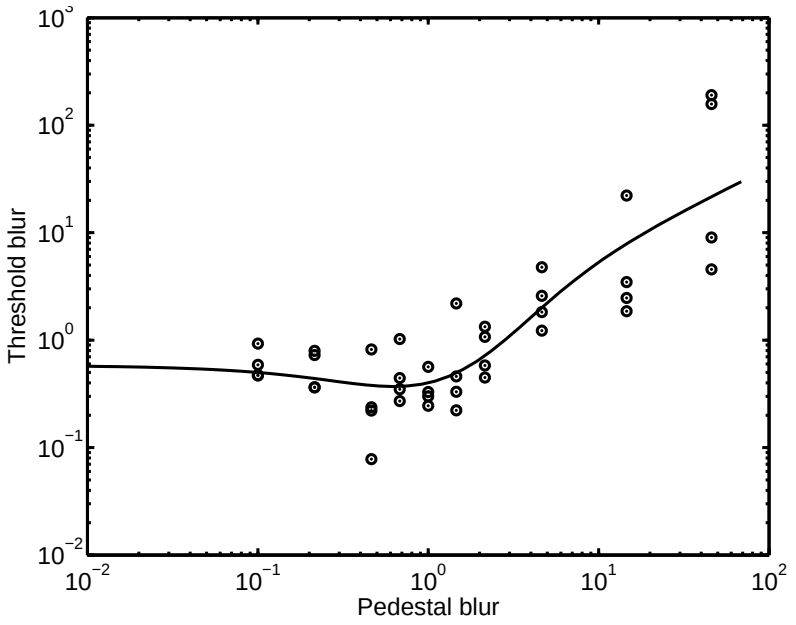
Table 6.1: Anova for blur discrimination

Subject	Image	Detection	Opt. pedestal (b)	Opt. discrim. (δb)
NCA	01342	0.96	0.70	0.36
NCA	00005	0.84	0.70	0.81
CAS	01342	0.49	0.48	0.08
CAS	00005	0.49	0.48	0.11
RAU	01342	0.61	1.50	0.23
RAU	00005	0.92	1.03	0.34
RTS	01342	0.48	0.48	0.23
RTS	00005	0.51	0.48	0.19
Mean	01342	0.65	0.81	0.23
Mean	00005	0.71	0.69	0.37
Watt [45]	–	0.3	1-2.5	0.15-0.20
Wuerger [181]	–	1	0.5-1	0.3-0.5
Walsh [52]	Street	0.22D	1.5D	0.1D
Walsh [52]	Grating	0.16D	2D	0.15D

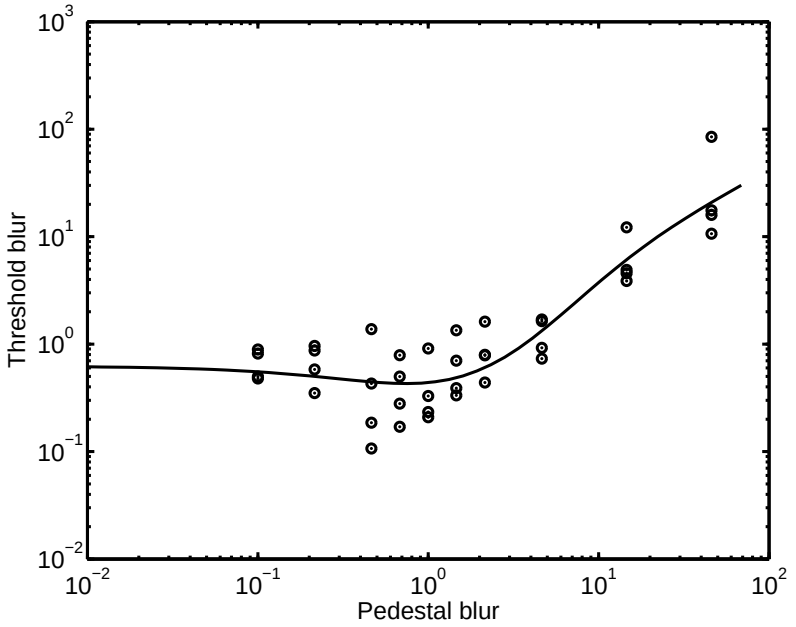
Table 6.2: Various key properties of the psychophysical responses were calculated from the bootstrapped results: Detection threshold, the point of best discrimination (in terms of pedestal value and actual discriminable blur at the optimum pedestal blur). All figures without units are in arc minutes. Equivalent results from previous papers (Watt and Morgan [45] and Wuerger [181]) are included either using figures from their respective texts, or reading from graphs. Walsh et al's results, whilst not using the same units, are included here for comparison. Results from Burr et al are not included as they relate to moving edges rather than static stimuli.

There is a smaller, second-order correlation with image and pedestal. This secondary relationship is similar to the observation made by Walsh that discrimination thresholds vary with scene complexity. In this experiment the building scene (00005) is more recognisable as a real-world scene by observers than the tree bark in the second experiment, which when cropped and blurred as in the experimental conditions, tends to look a little abstract.

Overall results are presented in Figure 6.6, which shows the discrimination threshold determined for each observer and condition, together with a best-fit to these data points. This shows a clear dip, followed by a monotonic increase in discrimination with pedestal condition. Table 6.2 shows how these figures compare with the individual responses, and that they are in agreement with observations by Wuerger and similar to Watt and Morgan's seminal results. The differences between these results and those of other researchers is probably related to the extent of the methodological (or stimuli) differences.



(a) Image 01342



(b) Image 00005

Figure 6.6: Overall psychophysical results: The points on these figures are the 82% discrimination thresholds determined from each observer at each pedestal condition. Overlaid is the best fit to the contrast response function in Equation (6.1).

6.4 Conclusions

The results show that blur discrimination sensitivity when viewing natural scenes is optimal when a sharp image has been slightly blurred, rather than viewing the original sharp image. These findings are in agreement with previously reported blur discrimination thresholds for luminance edges, and extend the results to natural scenes.

The peak discrimination ability occurs in natural scenes at around $b = 0.75$ arcmins, in the middle of the range of values reported by Wuerger. Discrimination performance with natural scenes slightly exceed those reported in simpler stimuli reported by Wuerger (δb was measured to be between 0.1 and 0.4 arc minutes in all but one observer/scene combination, a little better than the range of 0.3-0.5 previously reported).

In conclusion, previously reported results in blur perception have been shown to be present in natural scenes. If the focus measures discussed in previous chapters are good models of human blur perception, then they should reproduce the results achieved in this chapter. This is investigated in the following chapter.

Chapter 7

Model observers

In the preceeding chapters, various sets of data have been acquired from human observers, such that human ground truth can be established. Chapter 5 showed that over all the scenes, only two focus measures are within the 95% confidence interval of human opinions when identifying which image of a scene is best in focus, but many more agreed when just one scene was excluded from the analysis. Separately, ten measures were close to the focus curve determined from human observations. Chapter 6 describes further experiments which constrain the task and viewing conditions such that aspects of the underlying psychophysical response can be measured. These techniques established the threshold for blur discrimination at a number of different baseline conditions.

This chapter describes how focus measures are used as observers to replicate the experiments described in Chapter 6 which were performed with human observers. This will provide further criteria for selecting a focus measure which is a good match to human observations, for it is hypothesised that the focus measure which most accurately reproduces human results should reproduce the 'dipper' responses found in the previous chapter.

To use focus measures as observers, a computer is used in place of the human observer in the experiments described in the previous chapter. However, there are several possible ways of doing this.

7.1 Stimuli and presentation

Firstly, as the human observers viewed the candidate images on a computer screen, then perhaps the computer should perform the same task – viewing the stimuli via an analog presentation and subsequent acquisition (eg via a camera). However, such an arrangement would require significant calibration to parameterise the luminance

transfer function and PSF of the camera and its sensor.

The alternative is to perform the observations entirely within the computer, and have no visual presentation but simply a digital transfer of the stimuli from the software preparing it across to the program used to make a decision as to which of the stimuli is most in focus.

Given that the human observers were using a linearly calibrated monitor to ensure they were viewing the stimuli in the way in which it was intended by the software, the closest parallel for the model observer is the second of the options above – direct digital transfer of the image.

Beyond the image transfer, there are several other aspects of stimuli presentation which apply to human observers and which are not modelled in this work. Firstly, to minimise the confounding effect of higher level discrimination, presentation duration is constrained with human observers. Secondly, stimuli were presented to humans sequentially (ie temporally rather than spatially separated), such that no concurrent comparison could be performed. Neither of these restrictions are necessary for the computer models as each focus measure has knowledge of only one stimulus at a time, and makes no attempt to understand the content.

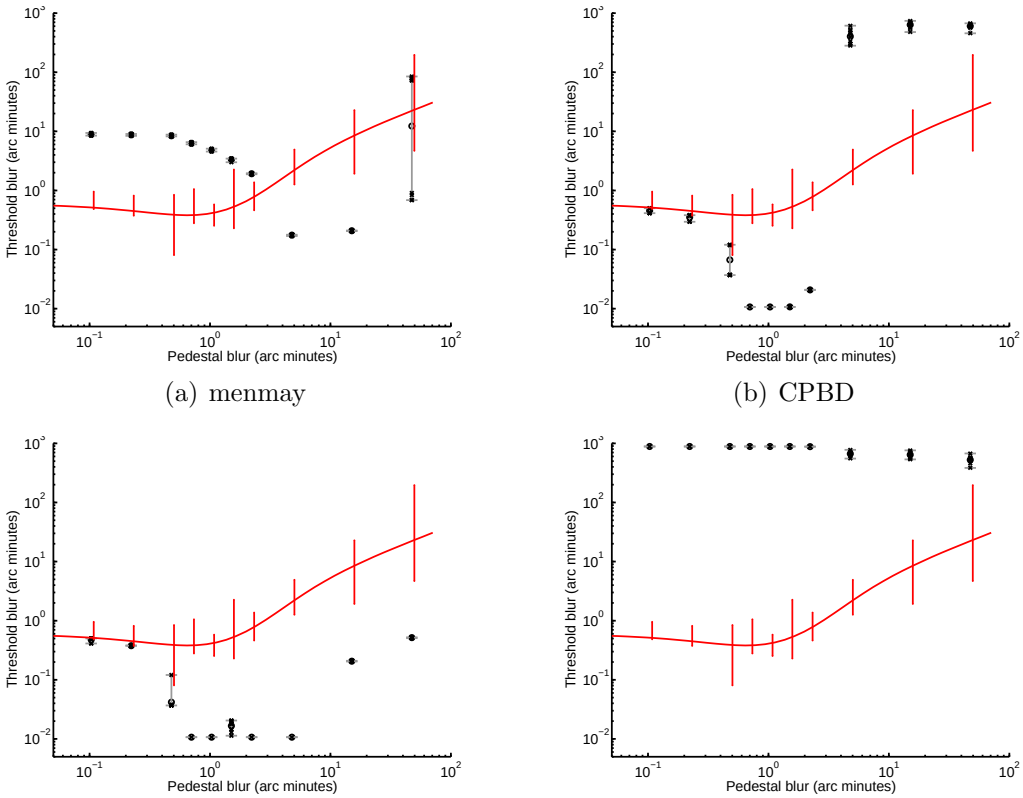
7.2 Preliminary results

The focus measures were implemented into the software described in Section 6.1 by replacing the request for user-input with a simple magnitude comparison between the scores obtained by processing each candidate image pair with the measure under test. Using this approach, psychophysical-type discrimination curves were obtained for each measure, examples of which are shown in Figure 7.1.

Broadly, the results were found to fall into four groups. Firstly, for some measures, the observed discrimination was far from the expected dip shape. For example, *menmay* showed far better discrimination with very large amounts of blur than with sharp images, the opposite of the human observations. This indicates that these measures are not good at modeling human blur discrimination.

A second group showed an improvement in discrimination as pedestal blur increased. This improvement plateaued at a floor, showing that discrimination is never better than a threshold of approximately $\delta b = 0.02$ arc minutes, beyond which discrimination deteriorated as pedestal blur continued to increase. This low ability to discriminate between highly blurred stimuli was far worse than human observers. *CPBD* is one such example of a measure exhibiting this behaviour.

The third group (comprising the majority of measures) showed a dip response,



(a) menmay (b) CPBD

(c) groenvariance (d) thresholdedpixelcount

Figure 7.1: Discrimination curves generated using various focus measures as observers. The results in black are from five QUEST convergences per pedestal blur for the model observers considering Van Hateren image 01342. Grey lines highlight the range between maximum and minimum convergence points. Human results for the same image are shown in red. The curve is the best fit to the bootstrapped human results, the range of which is shown as red vertical lines. These human results are those reported earlier in Figure 6.6.

with peak discrimination found in approximately the same conditions as for human observers, but with discrimination in all conditions being equal to, or better than humans. Again, many of the measures in this group exhibit a discrimination floor at $\delta b = 0.02$ arc minutes. Finally, some measures, including *thresholdedpixelcount* and *randomnumber* failed to produce any dip – discrimination performance in all conditions was equally poor.

Overall, these responses indicate discrimination thresholds considerably different to those determined from human observations with few exhibiting dips, and most showing a significantly greater ability to discriminate blur than humans – the best human discrimination is approximately $\delta b = 0.1$ arc minutes.

Reviewing the underlying scores upon which the focus measures were compared in such minima showed that the absolute difference between the scores was typically very very small in comparison to the scores themselves. The overly precise ability of the computer algorithm to make such decisions is a result of the direct digital transfer of the image from the generation to assessment functions of the software. Combined with Quest's ability to refine the stimuli to determine the additional stimuli required to achieve 82% meant that the determined threshold was more a result of the finite numerical precision in the calculations, than an underlying feature of the measure under test.

That the floor was present at approximately $\delta b = 0.02$ is an artefact caused by the manner in which blur is being applied to the candidate images. The images are blurred by convolving the image with a specific kernel. As described in Section 6.1, this is a gaussian kernel whose standard deviation is set to the desired blur. For large blur extents this convolution takes a few seconds. So, to maximise the performance of the software, thus minimising inter-stimuli delay for human observers (and eliminate any cues that might be inferred by such a delay), stimuli were pre-computed with blur extents quantised at approximately one tenth of the peak discrimination observed in preliminary trials.

Such a step size is far too large to fully explore the ultimate performance of the focus measures to make discrimination. Indeed, whilst the group mean peak discrimination for human observers was found to be $\delta b = 0.23$ arc minutes, focus measures can discriminate minute changes in blur. Many of the measures can correctly discriminate when δb is 1000 times smaller than the peak human ability.

Overall, this means that focus measures can significantly out-perform human observers. Thus, to use them to model humans and obtain dipper responses matching human results, a certain amount of performance degradation must be added.

7.3 Decision noise

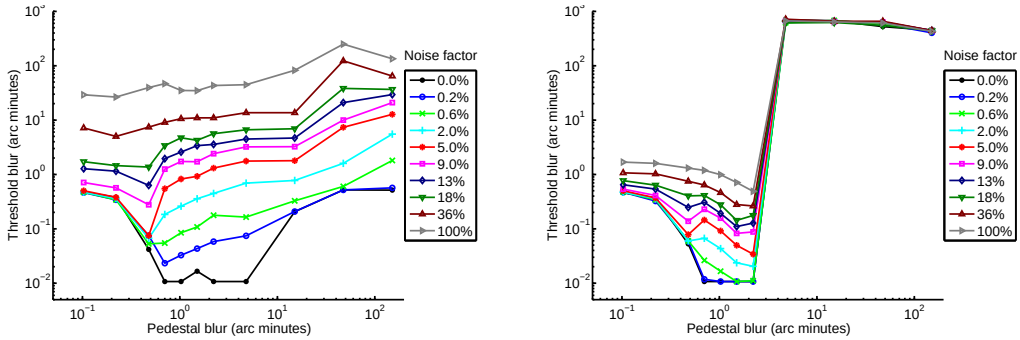
By adding noise to the decision previously made simply on the basis of a magnitude comparison of the raw scores, a degree of filtering is applied to the decision. The effect of this is that only large differences in score will cause the same decision to be made reliably for the same pair stimuli, thus reducing the performance of the focus measures. The concept of decision noise was proposed by Mueller and Weidemann [184] who observe that: "decision noise has frequently been ignored because it often cannot be separated from perceptual noise ... underestimating the level of perceptual sensitivity".

There are several plausible explanations for decision noise in human observers in the experiments in previous chapters. Human observers (by definition) need to view the stimuli with their eyes, whose state is not permanently fixed. Thus, small changes in its physical condition (such as accommodation or fixation), as well as microscopic fluctuations within the eye, would be likely to effect slight changes to the resultant image projected onto the retina. Such small changes might propagate through the judgement process to affect the decision. That is, whilst the differences occur earlier than the actual decision, they might be modelled as decision noise given fixed stimuli.

A second explanation centres around the nature of the stimuli presentation. Consecutive stimuli presentation depends upon the observer retaining a quantitative impression of the first scene such that it can be compared to the second. It is plausible that this degrades with time, and thus the decision accuracy might be related to the magnitude of the difference, and inversely related to the time between presentation and decision being made.

Beyond physical differences, human observers may also be influenced by other higher level factors. For example, prior experience and personal interests could affect scene interpretation. Similarly distraction is likely to affect an observer's decision, be that distraction in the form of regions of high visual attention, or others peripheral to the stimuli such as fatigue.

Nevertheless, regardless of the biological justification for adding decision noise, it solves an implementation problem where the focus measures are extremely sensitive to minute changes in input, thus producing discrimination thresholds that are a function of the numerical accuracy of Matlab and the measures' implementation, rather than their actual performance. To simulate decision noise, the scores produced by the focus measure under test were scaled by an arbitrary amount of normally-distributed random noise ($\sigma = 1$; using Matlab's `randn` function) such that:



(a) groenvariance, 01342 (b) sml, 01342

Figure 7.2: Discrimination curves generated using the *groenvariance* and *sml* measures with a selected range of different noise factors. The lines on this figure are straight lines between the data points to help the observer see the data sets, and are not meant to suggest a linear relationship inter-data point. The noise factor is the amount of multiplicative noise added.

$$score = score * (1 + noisefactor * randn) \tag{7.1}$$

7.4 The impact of noise

To evaluate the impact of this noise, two focus measures were considered, *groenvariance* and *energylaplace5b*, both measures which have performed well over the tests described in earlier chapters. Whilst uing these measures as observers, a variety of noise factors were added, between 10^{-4} and 1, to examine how the discrimination curve changes with the addition of noise. The impact of increasing this decision noise is shown in Figure 7.2.

Taking *groenvariance*, shown in Figure 7.2(a), the lowest line corresponds to the situation when no noise is present, where the measured discrimination performance is limited by the implementation. This plateaued minimum disappears once even a small amount of noise is added, whereupon, the expected dip is present. However, with the lowest added noise, the discrimination performance with high blur extents is approximately the same as with no blur, which is not the case with human observers. Only with additional decision noise does this measure start to exhibit a dipper shape with similar attributes to human observers. As noise rises about 5%, the measure's performance reduces in the conditions with low blur. Beyond 15% noise, the dipper shape itself starts to disappear.

In contrast, *sml* (Figure 7.2(b)), shows more clear defined dipper shapes over

a wider range of noise factors, but has equal performance when the pedestal blur extent exceeds approximately 5 arc minutes. This is in contrast with human observers whose discrimination performance continues to deteriorate as blur increases, suggesting that *sml* is not a good model of human blur discrimination.

To quantify the impact of noise, several metrics have been determined numerically from the *groenvariance* data shown in Figure 7.2, and are tabulated in Table 7.1, which also include equivalent values determined from human observations as reported in the previous chapter. Specifically, the *detection threshold* is defined as the discrimination when the smallest amount of blur is present. The *optimal pedestal* is the point with the best discrimination at which point the discriminable blur is the *optimal discrimination*. The *exponent* is the value of α for a line of the form $y = mx^{\alpha}$ between the trough of the dip and the discrimination threshold for the largest pedestal measured. (N.B. If the trough was a plateaued minimum, the condition with greatest pedestal blur was used.) The resultant equation for determining α from these two points, (x_1, y_1) and (x_2, y_2) is:

$$\alpha = \frac{log\,(y_1/y_2)}{log(x_1/x_2)} \tag{7.2}$$

These four characteristics can be plotted against noise factor to see more clearly how each varies with noise. Figure 7.3 shows that there is no amount of noise which causes *sml* to coincide with the human mean – at no point are all the data points close to the human results shown as horizontal lines. However, with approximately 15% noise ($nf \approx 10^{-0.8}$), *groenvariance* does coincide.

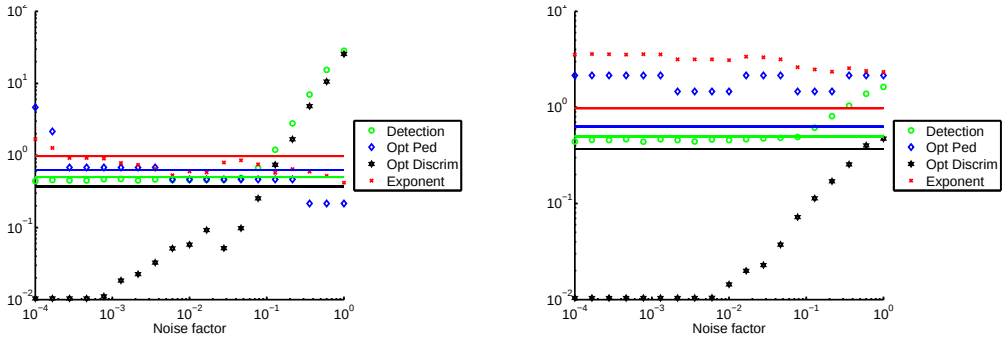
Using a Euclidean distance metric, the noise factor closest to human opinions for each focus measure was computed. That is, for each measure and at each noise factor, a similarity metric was computed by comparing the measures' attributes against human results. So, for the Van Hateren image 01342, the similarity measure is:

$$similarity = \sqrt{(D - 0.50)^2 + (OP - 0.63)^2 + (OD - 0.37)^2 + (E - 0.98)^2} \tag{7.3}$$

where equal weight was given to D the detectable blur, OP the pedestal blur at the point of best performance at which $\delta b = OD$ blur can be discriminated, and E the exponent of the curve beyond point OP. The numeric constants are the values obtained from human observations by inspection of Figure 6.6. An equivalent equation was used for assessing measures' similarity to human results with the other scene, Van Hateren image 00005.

Noise	**Detection**	**Opt. pedestal** (b)	**Opt. discrim.** (δb)	**Exponent** (α)
0.00%	0.464	4.64	0.0104	1.69
0.01%	0.44	4.64	0.0104	1.69
0.02%	0.456	2.15	0.0104	1.27
0.03%	0.448	0.68	0.0104	0.921
0.05%	0.45	0.68	0.0104	0.921
0.08%	0.466	0.68	0.011	0.905
0.13%	0.472	0.68	0.0183	0.787
0.22%	0.449	0.68	0.0225	0.738
0.36%	0.463	0.68	0.0327	0.65
0.60%	0.456	0.464	0.0513	0.53
1.00%	0.457	0.464	0.0577	0.602
1.67%	0.459	0.464	0.0921	0.582
2.78%	0.458	0.464	0.0516	0.798
4.64%	0.487	0.464	0.0979	0.856
7.74%	0.672	0.464	0.254	0.75
12.9%	1.2	0.464	0.742	0.572
21.5%	2.79	0.464	1.68	0.647
35.9%	6.97	0.215	4.85	0.597
59.9%	15.4	0.215	10.6	0.52
100%	28.3	0.215	25.6	0.419
Human mean	0.50	0.63	0.37	0.98

Table 7.1: Discrimination curves were generated for one focus measure, *groenvariance*, and their key attributes determined. The exponent is the value of α for a straight line of the form $y = mx^{\alpha}$ between the trough of the dip and the discrimination threshold for the largest pedestal measured. These results are for Van Hateren image 01342.



(a) groenvariance, 01342 (b) sml, 01342

Figure 7.3: Parameters which characterise the discrimination curves were determined by analysis of the various curves generated using the *groenvariance* and *energylaplace5b* measures with a number of different noise factors. The horizontal lines on this figure indicate the results for human observers from Figure 6.6. The vertical scale is δb for the detection and discrimination datasets, b for the optimal pedestal, and the value of α for the exponent of the curve. These graphs show that there is no amount of noise which causes all measured parameters to coincide with human observations for *sml*, but with 15% noise ($nf \approx 10^{-0.8}$), *groenvariance* does coincide. Results for Van Hateren 00005 were very similar.

An alternative method for assessing similarity between human observers and machine results, the Mean square error (MSE), was also computed. As with the analysis described earlier in this chapter, the average human results from Figure 6.6 were taken and compared with each focus measure and noise factor combination.

7.5 Results

Using both the proposed similarity metric, and computing the MSE, each measure was scored with each different noise factor. Thus, for each scene, a total of 910 results were ranked by the two metrics. These scores are summarised in Table 7.2 which shows the top combinations of measure and noise factor in terms of similarity to human observations under these metrics.

Several interesting observations can be made from this data: Firstly, some measures appear multiple times in this list, which indicates that they remain very similar to human opinions across a range of decision noise, and thus perhaps are more robust. Secondly, no measure scores highly across all ranks. With 4.6% noise, *range* has the lowest average rank, though *cranesum* is an excellent fit to all but the similarity metric for Van Hateren 00005. Overall, *cranesum*, *range* and *smd* all appear four times in the list of top performing measures.

In summary, using focus measures as observers does produce the expected dip-

shaped discrimination curves, but decision noise needs to be added for these curves to be close to human results. No single value of decision noise produces the best performance from all measures. Comprehensive results are included in Appendix G on page 208 which shows the model observer results for all focus measures operating with the noise factor causing it to be most similar to human results.

7.6 Conclusions

All the focus measures from previous chapters were used as observers within the blur discrimination experiment described in Chapter 6. Using this approach, psychophysical discrimination curves were generated, which showed that not all measures yielded a dip-like response, the response that was desired given the human results found in this work and work by others.

When decision noise was introduced, the discrimination curve changed shape. With certain amounts of noise, discrimination curves for far more measures were of

Measure	**Noise factor**	**R1**	**R2**	**R3**	**R4**	**Overall**
cranesum	7.7%	4	1	8	184	1
cranesum	12.9%	1	3	2	140	1
range	4.6%	3	30	1	10	1
range	7.7%	9	56	92	1	1
smd	12.9%	2	2	95	229	2
voll4	12.9%	36	20	66	2	2
chernfft	7.7%	13	53	4	3	3
range	2.8%	8	35	3	82	3
groenvariance	7.7%	11	37	15	4	4
smd	7.7%	5	4	20	249	4
cranesum	21.5%	7	32	87	5	5
smd	2.8%	93	57	5	17	5
triakis11s	1.7%	28	5	424	367	5
absolutevariation	2.8%	16	27	14	6	6
cranesum	4.6%	24	6	17	166	6
smd	21.5%	6	28	156	205	6
voll4	21.5%	95	69	6	87	6
normalizedgroenvariance	7.7%	10	36	27	7	7
range	0.4%	209	163	7	18	7
triakis11s	2.8%	15	7	457	423	7

Table 7.2: The combinations of focus measure and noise factor in terms of similarity to human observations. Ranks 1 and 2 use MSE, whilst the similarity measure as proposed in Eq (7.3) is used for Ranks 3 and 4. Ranks 1 and 3 are for Van Hateren image 01342, with image 00005's scores being shown in Ranks 2 and 4. The data is sorted such that all top rated measures are shown first, then second rated etc. Note that many measures appear multiple times, demonstrating their high performance across a range of noise factors.

the expected dip shape. Several plausible explanations were offered to justify the addition of noise.

To rank the ability of focus measures to match human results, several metrics were evaluated, including a new similarity measure proposed in this work compares four parameters of each measure's discrimination curve with the mean human results from the previous chapter, and produces a single similarity score.

Using this similarity metric, and by computing the MSE against human results, the performance of each focus measure in conjunction with a range of noise factors was assessed. This showed that there was not a single universal amount of noise that, when added to all measures, yielded the most similar results to humans. Instead, a specific noise-factor needed to be selected for each individual measure, which in this work was made from a finite set of noise-factors. That is, whilst a good noise factor was selected, no optimisation was performed.

Overall, the measures found to be most similar to human results in these experiments are *cranesum*, *smd* and *range*. The first two match human results when between 5% and 20% decision noise is added, whilst *range* matches results when between 0.4% and 8% noise is added. That each performs well over a range of noise factors indicates that they are not especially sensitive to this parameter.

Optimisation of the noise factor for each measure could be pursued, but given the wide range of human results in the previous chapter, and the robustness of the highly ranked focus, the precision with which the noise factor should be determined is unclear. Secondly, the optimisation would be very computationally expensive. It takes between one minute and ten hours to process each noise factor / measure combination, depending on the measure. But, it must be re-stated that at no point has computation time been used to judge measures, especially as one of the slowest (*triakis11s*) was explicitly designed for dedicated hardware that is not present on the general purpose computers used for processing in these experiments.

Chapter 8

Conclusions and future work

In the preceding chapters, various sets of data have been acquired from human observers. This has captured the ground truth across a number of scenes, an indication of the shape of the human focus curve, and psychophysical blur discrimination thresholds. These have been compared with a large number of focus measures discovered in the literature and created during the course of this work.

Before this data was collected and analysed, existing methodologies were reviewed. The primary approaches taken in this work are broadly similar to those existing methodologies described in the literature: Focus measures were scored used the quantitative metric described by Groen, blur was computationally applied by convolution with a gaussian kernel, and psychophysical blur-discrimination thresholds were measured with a 2AFC experimental procedure with QUEST providing estimates for each trial. However, some of the methodologies made assumptions that warranted exploration, and those were the starting points of this work.

Existing quantitative comparisons of focus measures used a ground truth which was established by using a 'trained observer', but did not comment on the reliability or repeatability of this observation. The first experiment in this work used a large number of observers to explore the nature of the ground truth. The results of this experiment showed that humans are broadly consensual in their selection of the 'best' image, though there is a small spread of opinions. No subset of the population could be found that made a different decision.

This ground truth was then used to score a large number of focus measures. Existing measures were collated from an extensive body of published work and implemented to a common software interface, alongside new measures created during this work. This permitted one of the largest comparisons between focus measures to be performed – over double the number of measures than compared in previous review papers. Chapter 5 showed that over all the scenes, only two focus measures

are within the 95% confidence interval of human opinions when identifying which image of a scene is best in-focus. Many more were within 95% for just five of the six scenes considered.

Beyond establishing the ground truth, various methods (such as finding blur-equivalence for candidate images) were employed to establish a focus curve from human opinions. The measures that were accurate in their agreement with the human-selected 'best' were not the same as those which fitted within this focus curve measured from human opinions.

Psychometric blur discrimination thresholds were measured in naïve observers in two calibrated natural images. Dipper-shaped responses, as had been reported by other researchers looking at simple stimuli, were shown to be present in these natural scenes. However, the confidence intervals were wider than reported in the literature. It is postulated that this may be due to the more complex nature of the stimuli than those used by other researchers, and also due to top-down influences that are more likely to be present when viewing natural scenes than viewing synthetic images.

Focus measures were then used as model observers to see if any measure could reproduce the blur discrimination thresholds obtained from human observers. None of the measures evaluated could do this until an amount of decision noise was added. No single amount of noise could be added which universally made focus measures agree with humans. Using a new similarity metric, these discrimination thresholds established with these observers were compared with human results, so as to establish which measures were closest to human results.

Considering the actual results of the various experiments, Table 8.1 aggregates the data from throughout the thesis, and shows the focus measures that performed well in each assessment. Many measures (such as *menmay* and *thesholdedpixelcount*) failed to perform well under any of the experiments. These measures rely on global image properties that are not sufficiently prominent in the best image, either because of pre-determined thresholds, or their basis on statistics that are not necessary to be present for an image to be in-focus. Those that performed best are measures whose score depends on small features; where the underlying mathematical function computes a score for each pixel in the image based upon comparisons with nearby pixels, and then produces an overall score by some form of aggregation. So, whilst there was no single focus measure that performed especially well in all scenarios, those that did perform well were those which process pixels locally in the spatial domain.

The measures that best matched the modal human opinion of most in-focus image when considering a range of scenes were *tenengrad* and *cranepeak*. Accordingly,

Measure	Core function	Accuracy	R1	R2	Shape
tenengrad	Convolution	Best	27	15	-
cranepeak	2D gradient	Best	32	16	-
smd	2D gradient	Good	2	4	Half-fit
autocorrelation	Autocorrelation	Good	3	13	Half-fit
voll4	Autocorrelation	Good	4	3	Half-fit
waveletw1	Wavelet transform	Good	5	20	Half-fit
waveletw3	Wavelet transform	Good	6	27	No fit
squaredgradient	1D gradient	Good	7	29	Fit
thresholdedabsolute-gradient	1D gradient	Good	7	28	Fit
sml	Convolution	Good	10	36	Half-fit
laplace	Convolution	Good	11	18	Half-fit
waveletw2	Wavelet transform	Good	14	24	No fit
triakis11s	Voxel statistic	Good	16	7	Fit
energylaplace5c	Convolution	Good	18	21	Half-fit
energylaplace5b	Convolution	Good	20	22	No fit
brennergradient	1D gradient	Good	22	23	Fit
energylaplace	Convolution	Good	23	26	No fit
energylaplace5a	Convolution	Good	24	25	No fit
va	Attention	Good	29	17	Fit
cranesum	2D gradient	Poor	1	1	Half-fit
alphaImageEnsemble	Spectrum statistic	Poor	9	37	Half-fit
groenvariance	Statistical	Poor	12	6	Fit
absolutevariation	Statistical	Poor	19	8	Fit
vol5	Autocorrelation	Poor	15	12	Fit
hlv	Histogram	Poor	33	33	-
standarddeviation-basedautocorrelation	Autocorrelation	Poor	16	14	Fits
range	Statistical	Poor	39	2	-
chernfft	Spectrum statistic	Poor	24	5	-
normalizedgroenvariance	Statistical	Poor	20	9	-

Table 8.1: Summary of top performing focus measures by the different assessment criteria. R1 is the rank of the measure's focus curve (see Figure 5.2 on page 120). R2 is the rank of the measure's ability to match human psychophysical results, as discussed in Section 7.5. The measures are sorted first by accuracy, then R1 – low numbers are better. Overall, no single measure performed well in all experiments.

these are likely to be the most appropriate measures of those reviewed to use in a general purpose auto-focus application. Many other measures are accurate when considering conventional photographs, but are sensitive to the artefacts present in new imaging techniques, be that images created by ray tracing, or acquired with compressive imaging.

These two best measures scored very poorly when assessed with a frequently used focus scoring metric, suggesting that this existing methodology is not necessarily the best way of identifying good focus measures. The poor score was due to nearby false maxima on their overall focus curve. Such a deficit could be resolved by using these measures for providing fine focus control having first found the approximate global maxima using another measure that has fewer erroneous maxima, such as *smd* or *cranesum*. An alternative solution would be to ensure the peak-finding logic was robust when false maxima are present.

Blur discrimination measurements for naïve observers viewing calibrated natural scenes were within the range of results found by other researchers. The peak discrimination ability was found at $b = 0.37$ arc minutes, where discrimination of $\delta b = 0.63$ arc minutes was possible. The closest measures to reproducing these results were *cranesum*, *range* and *smd*. However, for this to happen, a degree of noise needed to be added to the model.

8.1 Review of key research questions

This thesis suggested that it is possible to construct an algorithm that can accurately replicate the results of perceptual and subjective experiments, and several key research questions were proposed to help support this thesis. These are now considered in turn to establish the extent to which the results described answers them.

1. **Can ground truth data be obtained that is suitable for testing focus measures?**
   With a large population of observers, it was found that humans are broadly consensual when asked to identify the 'best' image. No subset of the observers was found to make a particular preference, across all scenes, that was different to the modal result. It was possible to establish quantitative data against which focus measures could be compared.

2. **How well do focus measures perform when compared against the ground truth?**

A large number of focus measures were implemented from the literature, and others devised during the course of this research (as listed on page 45). These were all assessed with the existing methodology used in previous reviews [97, 101, 96]. However, the lack of weight placed on the accuracy aspect of this scoring metric perhaps reduces its value, and it is suggested that other approaches might be better suited for comparing focus measures depending on the intended application – some measures were found to be good at matching the ground truth, whilst having a poor score under the existing methodology. A class of images was identified which hindered the ability of focus measures to identify the 'best' image, whilst not hindering human observers.

3. **How can human blur opinions and perception be measured?**
   Various approaches were attempted to gather information from observers which could be used to construct a focus curve. Based on feedback from the experiments' participants, the most comfortable task involved finding the image that represented the mid-point (in terms of blur) between two target images. By exploring successive mid-points, a focus curve was constructed. In addition, a standard 2AFC blur discrimination experiment was conducted to measure blur discrimination in natural scenes under natural PSF. Blur discrimination thresholds were measured in four human observers, and found to be of the expected shape. Confidence intervals were larger than reported in previous work, and it is suggested that this is due to the complexity of the stimuli.

4. **Can human results be compared with those from focus measures?**
   Throughout this work, the results obtained from human observers were compared with those produced by the individual focus measures. The ground-truth focus results and human-derived focus curve were compared with the measures to identify those that best reproduced the human results. The blur-discrimination thresholds of the focus measures was established by using each measure as an observer. When a degree of noise was added, many of the measures produced a dip-shaped response typical of human psychophysical discrimination. Noise was varied to match the results of these observers with those of human observers. No single measure was able to reproduce human results in all experiments.

The thesis statement is "It is hypothesised that it is possible to construct an algorithm that accurately replicates human perceptual and subjective experimental results". Having evaluated 46 measures, the measure identified as best meeting this

hypothesis is *smd*. No measure to-date exactly reproduces psychophysical results and at the same time performs accurately. Instead, different algorithms have been shown to excel at particular tasks, with the best measures being those making judgements based on small features in the spatial domain.

8.2 Significant findings

The work described in this thesis leads to several significant findings:

- Given a set of images taken at different focal distances, the 'best' in terms of human opinion is not unanimous, and should be measured per scene. Specifically, the opinion of a single observer should not be used when establishing ground truth, but that modal response from a number of observers is preferable.

- Overall, focus measures do work for most scenes, but not for scenes generated using PovRay. These are artificial, but likely to be representative of the challenges that might be encountered in the future when using compressive imaging. It seems necessary that a new approach to focus will be necessary to process images generated with these latest image acquisition techniques. This reinforces Vollath's statement that "no proof has so far been provided that any autofocus technique can operate reliably for every type of image" [105].

- The existing methodology for comparing focus measures arbitrarily puts equal importance on a number of criteria, meaning some highly accurate measures perform poorly. A different framework for comparing focus measures is introduced, which discards focus measures which fall outside the 95% confidence intervals of human observations.

- Blur discrimination thresholds have been measured in natural scenes subjected to computationally blurring using a gaussian kernel, and are shown to be in line with those found in simple stimuli. The confidence intervals on these thresholds measured in natural scenes are wider than those in simple stimuli, and it is suggested that this difference is due to top-down influences and merits further study.

- Focus measures had not previously been used as observers to establish their blur discrimination thresholds. The results reported here show that only some are able to reproduce human discrimination behaviour, and that to do this decision noise must be introduced.

8.3 Future work

The conclusions drawn from this work point towards a number of future research questions which could be investigated. Firstly, it has been shown that dithering affects the performance of focus measures. Are there other image artefacts which do not affect human observers' ability, but which cause difficulties for focus measures? Are there focus measures which are immune to the artefacts likely to be present in future novel imaging devices, and are these a better match to human results?

Further investigations could be conducted to understand what features of scenes affect peoples' performance when assessing blur. The use of eye-tracking equipment would expose fixation points that might help identify the salient features. This could be compared with a 'bubbles' approach to identify the regions of importance when discriminating blur [185], which could be performed by both human observers and candidate focus measures. Inter-observer behaviour could also be examined to establish whether blur-perception is similar amongst all observers, or whether top-down influences can affect blur discrimination thresholds – do people perform differently looking at a lion than when they are looking at a tree?

Preliminary results measuring VEPs were reported in 1998. Since then the technologies available to monitor activity within the human body have become ever more powerful. It might be possible to make measurements of the blur perception and track how that influences the ciliary muscles as the subject is shown a variety of stimuli. Such measurements could also be made in conjunction with eye tracking devices to further understand what aspects of images affect accommodation, from which an improved model of accommodation could be devised.

Finally, a valuable contribution of this work is the implementation of a large number of focus measures. To assist others in future work, these have been shared online as Matlab files, together with the images that were acquired during this work [186].

Bibliography

[1] R. T. Shilston and F. W. M. Stentiford, "Method for focus control," US Patent Application 20090310011, June 2007.

[2] B. T. Wagner and D. Kline, "The bases of colour vision," `http://www.psych.ucalgary.ca/pace/va-lab/Brian/history.htm`, 2002.

[3] P. Findlen and R. Bence, "A history of the eye," `http://www.stanford.edu/class/history13/earlysciencelab/body/eyespages/eye.html`, 1999.

[4] E. G. Boring, *Sensation and perception in the history of experimental psychology*, New York : Appleton-Century-Crofts, 1942.

[5] eduMedia, "Focusing via visual accommodation," `http://www.edumedia-sciences.com/en/a306-focusing-via-visual-accommodation`, Nov 2010.

[6] S. J. Judge and D. I. Flitcroft, *Nervous Control of the Eye*, chapter Control of Accommodation, pp. 93–115, Harwood Academic Publishers, 2000.

[7] E. F. Fincham, "The accommodation reflex and its stimulus," *British Journal of Opthalmology*, vol. 35, pp. 381–393, 1951.

[8] E. A. Walkton Ball, "A study in consensual accommodation," *American Journal of Optometry and Archives of American Academy of Optometry*, vol. 29, no. 11, pp. 561–574, 1952.

[9] D. I. Flitcroft, S. J. Judge, and J. W. Morley, "Binocular interactions in accommodation control: effects of anisometropic stimuli," *Journal of Neuroscience*, vol. 12, no. 1, pp. 188–203, 1992.

[10] Marran L. and Schor C.M., "Lens induced aniso-accommodation," *Vision Research*, vol. 38, pp. 3601–3619(19), November 1998.

[11] P. B. Kruger and J. Pola, "Stimuli for accommodation: blur, chromatic aberration and size," *Vision Research*, vol. 26, pp. 957–971, 1986.

[12] W. Taylor and H. W. Lee, "The development of the photographic lens," *Proceedings of the Physical Society*, vol. 47, no. 3, pp. 502–518, 1935.

[13] P. B. Kruger, S. Mathews, M. Katz, K. R. Addarwala, and S. Nowbotsing, "Accommodation without feedback suggests directional signals specify ocular focus," *Vision Research*, vol. 37, no. 18, pp. 2511–2526, 1997.

[14] T. Niemann, "Ca_tca00.jpg," `http://wiki.panotools.org/Chromatic_aberration`, July 2006, [accessed 13-October-2008].

[15] H. D. Crane, "A theoretical analysis of the visual accommodation system in humans," Tech. Rep. NASA CR-606, NASA - Ames Research Centre, Sep 1966.

[16] A. Hong-Chen, "Is there any difference in using blur as a stimulus for accommodation between emmetropes and myopes?," *Journal Documenta Ophthalmologica*, vol. 105, no. 1, pp. 33–39, July 2002.

[17] F. W. Campbell, "The minimum quantity of light required to elicit the accommodation reflex in man," *J Physiol*, vol. 123, no. 2, pp. 357–366, 1954.

[18] Oxford University Press, "focus, n.," The Oxford English Dictionary, 1989.

[19] A. P. Pentland, "A new sense for depth of field," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 4, pp. 523–531, 1987.

[20] D. H. Fender, "Control mechanisms of the eye," *Scientific American*, vol. 211, pp. 24–33, July 1964.

[21] M. Concetta Morrone and D. C. Burr, "Feature detection in human vision: A phase-dependent energy model," *Proceedings of the Royal Society of London. Series B, Biological Sciences*, vol. 235, no. 1280, pp. 221–245, 1988.

[22] P. Kovesi, "Image features from phase congruency," *Videre: Journal of Computer Vision Research*, vol. 1, no. 3, pp. 2–26, 1999.

[23] Z. Wang, E. P. Simoncelli, and H. Hughes, "Local phase coherence and the perception of blur," in *Advanced Neural Information Processing Systems (NIPS03*. 2004, pp. 786–792, MIT Press.

[24] A. Troelstra, B. L. Zuber, D. Miller, and L. W. Stark, "Accommodative tracking: A trial-and-error function," *Vision Research*, vol. 4, pp. 585–594, 1964.

[25] L. Stark and Y. Takahashi, "Absence of an odd-error signal mechanism in human accommodation," *IEE Transactions on Bio-medical Engineering*, vol. 12, no. 3/4, pp. 138–146, 1965.

[26] S. Philips and L. Stark, "Blur: A sufficient accommodative stimulus," *Documenta Ophthahnologica*, vol. 1, no. 43, pp. 65–89, 1977.

[27] J.A. Marshall, C.A. Burbeck, D. Ariely, J.P. Rolland, and K.E. Martin, "Occlusion edge blur: A cue to relative visual depth," April 1996.

[28] G. Mather, "Image Blur as a Pictorial Depth Cue," *Royal Society of London Proceedings Series B*, vol. 263, pp. 169–172, Feb. 1996.

[29] W. H Ittelson and A. Ames Jr, "Accommodation, convergence, and their relation to apparent distance," *The Journal of Psychology*, vol. 30, pp. 43–62, 1950.

[30] J. N. Trachtman, V. Giambalvo, and J. Feldman, "Biofeedback of accommodation to reduce functional myopia.," *Biofeedback and self-regulation*, vol. 6, no. 4, pp. 547–562, 1981.

[31] P. D. R. Gamlin, *Nervous Control of the Eye*, chapter Functions of the Edinger-Westphal Nucleus, pp. 117–154, Harwood Academic Publishers, 2000.

[32] T. Sugimoto, K. Itoh, and N. Mizuno, "Direct projections from the ediger-westphal nucleus to the cerebellum and spinal cord in the cat: An hrp study," *Neuroscience Letters*, vol. 9, no. 1, pp. 17 – 22, 1978.

[33] G. K. Hung, K.J. Ciuffreda, and B.C. Jiang, *Models of the visual system*, chapter Models of accommodation, pp. 287–340, Kluwer Academic Publishers/Plenum Publishers, 2002.

[34] F. Fylan and M. G. A. Thomson, "Visual evoked responses to blur in natural scenes," in *Perception 27 ECVP Abstract Supplement*, 1998.

[35] S. M. Shinners, *Modern control system theory and design*, J. Wiley, 1998.

[36] B. Winn, J. R. Pugh, B. Gilmartin, and H. Owens, "Arterial pulse modulates steady-state ocular accommodation," *Current Eye Research*, vol. 9, pp. 971–975, October 1990.

[37] W. N. Charman and G. Heron, "Fluctuations in accommodation: a review," *Ophthalmic and Physiological Optics*, vol. 8, no. 2, pp. 153–164, 1988.

[38] F. M. Toates, "A model of accommodation," *Vision Research*, vol. 10, pp. 1069–1076, 1970.

[39] L. Stark, Y. Takahashi, and G. Zames, "Nonlinear servoanalysis of human lens accommodation," *IEEE Transactions on Systems Science and Cybernetics*, vol. 1, no. 1, pp. 75–83, 1965.

[40] M. Khosroyani and G. K. Hung, "A dual-mode dynamic model of the human accommodation system," *Bulletin of Mathematical Biology*, vol. 64, pp. 285–299, 2002.

[41] A. S. Eadie and P. J. Carlin, "Evolution of control system models of ocular accommodation, vergence and their interaction," *Medical and Biological Engineering and Computing*, vol. 33, pp. 517–524, 1995.

[42] F. W. Campbell and J. J. Nachmias, "Spatial-frequency discrimination in human vision," *Journal of the Optical Society of America (1917-1983)*, vol. 60, no. 4, pp. 555–559, Apr. 1970.

[43] D. Marr and E. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society of London: Series B, Biological Sciences*, vol. 207, no. 1167, pp. 187–217, Feb. 1980.

[44] J. R. Hamerly and C. A. Dvorak, "Detection and discrimination of blur in edges and lines," *J. Opt. Soc. Am.*, vol. 71, no. 4, pp. 448–452, 1981.

[45] R. J. Watt and M. J. Morgan, "The recognition and representation of edge blur: Evidence for spatial primitives in human vision," *Vision Research*, vol. 23, no. 12, pp. 1465–1477, 1983.

[46] D. M. Levi and S. A. Klein, "Equivalent intrinsic blur in spatial vision," *Vision Research*, vol. 30, no. 12, pp. 1971–1993, 1990.

[47] D. M. Levi and S. A. Klein, "Equivalent intrinsic blur in amblyopia," *Vision Research*, vol. 30, no. 12, pp. 1995–2022, 1990.

[48] G. Westheimer, "Sharpness discrimination for foveal targets," *Journal of the Optical Society of America A*, vol. 8, no. 4, pp. 681–685, 1991.

[49] G. Mather, "The use of image blur as a depth cue," *Perception*, vol. 26, pp. 1147–1158, 1997.

[50] G. Mather and D. R. Smith, "Depth cue integration: stereopsis and image blur," *Vision Res.*, vol. 40, pp. 3501–3506, 2000.

[51] R. J. Jacobs, G. Smith, and C. D. Chan, "Effect of defocus on blur thresholds and on thresholds of perceived change in blur – comparison of source and observer methods," *Optometry and Vision Science*, vol. 66, pp. 545–553, 1989.

[52] G. Walsh and W. N. Charman, "Visual sensitivity to temporal change in focus and its relevance to the accommodation response," *Vision Research*, vol. 28, no. 11, pp. 1207–1221, 1988.

[53] D. I. Flitcroft, "Accommodation and flicker: evidence of a role for temporal cues in accommodation control?," *Ophthalmic and Physiological Optics*, vol. 11, no. 1, pp. 81–90, 1991.

[54] A. K. Pääkkönen and M. J. Morgan, "Effects of motion on blur discrimination," *Journal of the Optical Society of America A*, vol. 11, pp. 992–1002, Mar. 1994.

[55] S. T. Hammett, "Motion blur and motion sharpening in the human visual system," *Vision Research*, vol. 37, no. 18, pp. 2505–2510, 1997.

[56] D. C. Burr and M. J. Morgan, "Motion deblurring in human vision," *Royal Society of London Proceedings Series B*, vol. 264, pp. 431–436, Mar. 1997.

[57] V. Kayargadde and J. B. Martens, "Estimation of edge parameters and image blur using polynomial transforms," *CVGIP: Graphical Models and Image Processing*, vol. 56, no. 6, pp. 442–461, 1994.

[58] V. Kayargadde and J. B. Martens, "Perceptual characterization of images degraded by blur and noise: model," *J. Opt. Soc. Am. A*, vol. 13, no. 6, pp. 1178, 1996.

[59] V. Kayargadde and J. B. Martens, "Estimation of perceived image blur using edge features," *International Journal of Imaging Systems and Technology*, vol. 7, no. 2, pp. 102–109, 1996.

[60] E. Peli, "Feature detection algorithm based on a visual system model," *Proceedings of the IEEE*, vol. 90, no. 1, pp. 78–93, January 2002.

[61] R. Ferzli and L. J. Karam, "A human visual system-based model for blur/sharpness perception," in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, 2006.

[62] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of Just Noticeable Blur (JNB)," *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 717–728, April 2009.

[63] N. Narvekar and L. Karam, "An improved no-reference sharpness metric based on the probability of blur detection," in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, January 2010.

[64] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proc. IEEE International Conference on Image Processing*, 2002, vol. 3, pp. 57–60.

[65] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: application to JPEG2000," *Signal Processing: Image Communication*, vol. 19, pp. 163–172, 2004.

[66] F. W. Campbell, "The physics of visual perception," *Royal Society of London Philosophical Transactions Series B*, vol. 290, pp. 5–9, July 1980.

[67] S. A. Klein, "Channels in the visual nervous system: Neurophysiology, psychophysics and models," in *Channels: Bandwidth, channel independence, detection vs discrimination*, B. Blum, Ed. 1992, pp. 11–27, London: Freund.

[68] V. A. Billock, "Neural acclimation to 1/f spatial frequency spectra in natural images transduced by the human visual system," *Physica D: Nonlinear Phenomena*, vol. 137, no. 3-4, pp. 379 – 391, 2000.

[69] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am. A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[70] C. R. Carlson, "Thresholds for perceived image sharpness," *Photographic science and engineering*, 1978.

[71] D. C. Knill, D. Field, and D. Kersten, "Human discrimination of fractal images," *Journal of the Optical Society of America A*, vol. 7, no. 6, pp. 1113–1123, 1990.

[72] D. J. Field and N. Brady, "Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes," *Vision Research*, vol. 37, no. 23, pp. 3367–3383, 1997.

[73] Y. Tadmor and D. J. Tolhurst, "Discrimination of changes in the second-order statistics of natural and synthetic images," *Vision Research*, vol. 34, no. 4, pp. 541–554, 1994.

[74] D. J. Tolhurst and Y. Tadmor, "Discrimination of changes in the slopes of the amplitude spectra of natural images: band-limited contrast and psychometric functions," *Perception*, vol. 26, no. 8, pp. 1011–1025, 1997.

[75] Ian BC North on flickr, "Blinds," `http://www.flickr.com/photos/91997797@N00/4057584742/`, October 2009.

[76] larryvincent on flickr, "Large blinds," `http://www.flickr.com/photos/larryvincent/3122806255/sizes/l/`, December 2008.

[77] C. A. Párraga and D. J. Tolhurst, "The effect of contrast randomisation on the discrimination of changes in the slopes of the amplitude spectra of natural scenes," *Perception*, vol. 29, pp. 1101–1116, 2000.

[78] M. G. A. Thomson and D. H. Foster, "Role of second- and third-order statistics in the discriminability of natural images," *J. Opt. Soc. Am. A*, vol. 14, no. 9, pp. 2081–2090, 1997.

[79] C. A. Párraga, T. Troscianko, and D. J. Tolhurst, "The effects of amplitude-spectrum statistics on foveal and peripheral discrimination of changes in natural images, and a multi-resolution model," *Vision Research*, vol. 45, no. 25-26, pp. 3145 – 3168, 2005.

[80] M. P. S. To, I. D. Gilchrist, T. Troscianko, J. S. B. Kho, and D. J. Tolhurst, "Perception of differences in natural-image stimuli: Why is peripheral viewing poorer than foveal?," *ACM Trans. Appl. Percept.*, vol. 6, no. 4, pp. 26, 2009.

[81] B. Wang and K. J. Ciuffreda, "Foveal blur discrimination of the human eye," *Opthal. Physiol. Opt*, vol. 25, pp. 45–51, 2005.

[82] B. Wang and K. J. Ciuffreda, "Blur discrimination of the human eye in the near retinal periphery," *Optometry and Vision Science*, vol. 82, no. 1, pp. 52–58, January 2005.

[83] G. Westheimer, S. Brincat, and C. Wehrhahn, "Contrast dependency of foveal spatial functions: orientation, vernier, separation, blur and displacement discrimination and the tilt and Poggendorff illusions," *Vision Research*, vol. 39, pp. 1631–1639, 1999.

[84] S. M. Wuerger, H. Owens, and S. Westland, "Blur tolerance in different colour directions," in *International Conference on Color in Graphics and Image Processing – CGIP2000*, 2000.

[85] M. A. Webster, S. M. Webster, J. MacDonald, and S. R. Bahradwadj, "Adaptation to blur," *Human Vision and Electronic Imaging VI*, vol. 4299, no. 1, pp. 69–78, 2001.

[86] S. M. Webster, M. A. Webster, and J. Taylor, "Simultaneous blur contrast," *Human Vision and Electronic Imaging VI*, vol. 4299, no. 1, pp. 414–422, 2001.

[87] M. A. Webster, M. A. Georgeson, and S. M. Webster, "Neural adjustments to image blur," *Nature Neuroscience*, vol. 5, no. 9, pp. 839–840, 2002.

[88] A. C. Bilson, Y. Mizokami, and M. A. Webster, "Visual adjustments to temporal blur," *J. Opt. Soc. Am. A*, vol. 22, no. 10, pp. 2281–2288, 2005.

[89] M. A. Webster, Y. Mizokami, L. A. Svec, and S. L. Elliott, "Neural adjustments to chromatic blur," *Spatial Vision*, vol. 19, no. 2-4, pp. 111–132, 2006.

[90] D. M. Chandler, K. H. Lim, and S. S. Hemami, "Effects of spatial correlations and global precedence on the visual fidelity of distorted images," in *Human Vision and Electronic Imaging XI. Edited by Rogowitz, Bernice E.; Pappas, Thrasyvoulos N.; Daly, Scott J. Proceedings of the SPIE, Volume 6057, pp. 131-145 (2006).*, B. E. Rogowitz, T. N. Pappas, and S. J. Daly, Eds., Feb. 2006, vol. 6057 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pp. 131–145.

[91] E. Vansteenkiste, D. Van der Weken, W. Philips, and E. Kerre, *Perceived Image Quality Measurement of State-of-the-Art Noise Reduction Schemes*, pp. 114–126, Springer Berlin / Heidelberg, 2006.

[92] A. Erteza, "Sharpness index and its application to focus control," *Appl. Opt.*, vol. 15, no. 4, pp. 877–881, 1976.

[93] R. A. Muller and A. Buffington, "Real-time correction of atmospherically degraded telescope images through image sharpening," *J. Opt. Soc. Am.*, vol. 64, no. 9, pp. 1200–1210, 1974.

[94] N. K. Chern, P. A. Neow, and Jr. Ang, M.H., "Practical issues in pixel-based autofocusing for machine vision," *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, vol. 3, pp. 2791 – 2796 vol.3, 2001.

[95] D. Vollath, "The influence of the scene and of noise on automatic focussing," *Journal of Microscopy*, vol. 151, no. S, pp. 133–146, 1988.

[96] F. C. Groen, I. T. Young, and G. Ligthart, "A comparison of different focus functions for use in autofocus algorithms," *Cytometry*, vol. 6, no. 2, pp. 81–91, 1985.

[97] Y. Sun, S. Duthaler, and B. Nelson, "Autofocusing in computer microscopy: Selecting the optimal focus algorithm," *Microscopy Research and Technique*, vol. 65, no. 3, pp. 139–149, 2004.

[98] E. Krotkov, "Focusing," *Intl. Journal of Computer Vision*, vol. 1, no. 3, pp. 223–237, October 1987, Reprinted in Physics-Based Vision, Jones and Bartlett, 1992.

[99] A. Santos, C. Ortiz de Solorzano, J. J. Vaquero, J. M. Pena, N. Malpica, and F. del Pozo, "Evaluation of autofocus functions in molecular cytogenetic analysis," *Journal of Microscopy*, vol. 188, no. 3, pp. 264–272, 1997.

[100] M. L. Mendelsohn and B. H. Mayall, "Computer-oriented analysis of human chromosomes–iii. focus," *Computers in Biology and Medicine*, vol. 2, no. 2, pp. 137 – 138, IN9–IN10, 139–145, 146–150, 1972, Chromosome Analysis.

[101] L. Firestone, K. Cook, K. Culp, N. Talsania, and K. Preston Jr, "Comparison of autofocus methods for automated microscopy," *Cytometry*, vol. 12, no. 3, pp. 195–206, 1991.

[102] Carl Zeiss, "Method of and device for the automatic focusing of microscopes," US Patent 3721759, March 1973.

[103] D. C. Mason and D. K. Green, "Automatic focusing of a computer-controlled microscope," *Biomedical Engineering, IEEE Transactions on*, vol. BME-22, no. 4, pp. 312 –317, July 1975.

[104] R. A. Jarvis, "Focus optimization criteria for computer image processing," *Microscope*, vol. 24, no. 2, pp. 163–180, 1976.

[105] D. Vollath, "Automatic focusing by correlative methods," *Journal of Microscopy*, vol. 147, no. 3, pp. 279–288, 1987.

[106] S. K. Nayar, "Shape from focus," M.S. thesis, Carnegie Mellon University, 1989.

[107] E. Peli, "Contrast in complex images," *J. Opt. Soc. Am. A*, vol. 7, no. 10, pp. 2032–2040, Oct 1990.

[108] G. Yang and B. J. Nelson, "Wavelet-based autofocusing and unsupervised segmentation of microscopic images," *Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, vol. 3, pp. 2143 – 2148 vol.3, oct. 2003.

[109] J Caviedes and F. Oberti, "A new sharpness metric based on local kurtosis, edge and energy information," *Signal Processing: Image Communication*, vol. 19, pp. 147–161, 2004.

[110] R. Shilston and F. W. M. Stentiford, "An attention based focus control system," in *Image Processing, 2006 IEEE International Conference on, Vol., Iss.*, October 2006, pp. 425–428.

[111] M. Subbarao, T. Choi, and A. Nikzad, "Focusing techniques," *Journal of Optical Engineering*, vol. 32, pp. 2824–2836, 1993.

[112] AccelerEyes, "Success stories for jacket," `http://www.accelereyes.com/successstories`, March 2010.

[113] R. J. Watt and M. J. Morgan, "A theory of the primitive spatial code in human vision," *Vision Research*, vol. 25, pp. 1661–1674, 1985.

[114] P. J. Bex and S. Murray, "Blur perception and discrimination in naturally-contoured images," in *Association for Research in Vision and Opthalmology 2010 conference: For sight – The future of eye vision and research.*, 2010.

[115] S. Murray and P. J. Bex, "Perceived blur in naturally-contoured images depends on phase," *Frontiers in Psychology*, vol. 1, no. 0, Jun 2010.

[116] A. Roorda, F. Romero-Borja, III W. Donnelly, H. Queener, T. Hebert, and M. Campbell, "Adaptive optics scanning laser ophthalmoscopy," *Opt. Express*, vol. 10, no. 9, pp. 405–412, May 2002.

[117] R. Navarro, E. Moreno, and C. Dorronsoro, "Monochromatic aberrations and point-spread functions of the human eye across the visual field," *J. Opt. Soc. Am. A*, vol. 15, no. 9, pp. 2522–2529, 1998.

[118] G. L. Ligthart and Groen F. C. A., "A comparison of different autofocus algorithms," in *International Conference on Pattern Recognition*, 1982, pp. 597–600.

[119] R. Leggat, "A history of photography from its beginnings till the 1920s," `http://www.rleggat.com/photohistory/`, 1995, [accessed 02-May-2007].

[120] P. Greenspun, "History of photography timeline," `http://photo.net/history/timeline`, January 2007, [accessed 02-May-2007].

[121] M. Minsky, "Memoir on inventing the confocal scanning microscope," *Scanning*, vol. 10, pp. 128–138, 1988.

[122] L. Byoung-kwon and K. Hyon-soo, "Focusing method for digital photographing apparatus," US Patent Application 2005185082, August 2005.

[123] Canon, "Canon launches first powershot a-series camera with optical image stabilization," Press Release, August 2006, [accessed 06-January-2007].

[124] Canon, "Digital ixus i7 zoom," `http://www.canon-europe.com/For_Home/Product_Finder/Cameras/Digital_Camera/IXUS/Digital_IXUS_I7_zoom/index.asp?ComponentID=393526\&SourcePageID=231640`, 2006, [accessed 06-January-2007].

[125] Canon, "Digital ixus i7 advanced camera user guide," 2006.

[126] Nikon Corporation, "Nikon face-priority AF," `http://www.nikonimaging.com/global/news/2005/pdf/FacepriorityAFnr.pdf`, February 2005.

[127] Steve's Digicam Online Inc, "Canon EOS Digital Rebel XSi," `http://www.steves-digicams.com/2008_reviews/canon_rebel_xsi.html`, May 2008.

[128] P. Porntrakoon, D. Kesrarat, and J. Daengdej, "A model for auto-focus using location detection based on sound localization," *Applied Simulation and Modelling, Proceedings of*, June 2004.

[129] H. Jung, J. Bae, S. Suh, and H. Yi, "Use of TRIZ to develop a novel auto-focus camera module," *Triz Journal*, August 2006.

[130] M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "An architecture for compressive imaging," in *Image Processing, 2006 IEEE International Conference on*, Oct. 2006, pp. 1273–1276.

[131] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," *Computer Graphics*, vol. 31, no. Annual Conference Series, pp. 369–378, 1997.

[132] Fujifilm, "Further progress of the Super CCD," `ttp://www.fujifilmholdings.com/en/pdf/investors/annual_report/ff_ar_2003_part_004.pdf`, March 2003.

[133] P. Askey, "Fujifilm FinePix S3 Pro review," `http://www.dpreview.com/reviews/fujifilms3pro`, March 2005.

[134] Hewlett Packard, "HP Adaptive Lighting technology," Tech. Rep., Hewlett-Packard, 2004.

[135] Hewlett Packard, "HP Photosmart digital cameras HP Real Life technologies," Tech. Rep., Hewlett-Packard, 2004.

[136] R. Ng, M. Levoy, M. Brdif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Tech. Rep. CTSR 2005-02, Stanford, 2005.

[137] A. Isaksen, L. McMillan, and S. J. Gortler, "Dynamically reparameterized light fields," in *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 2000, pp. 297–306, ACM Press/Addison-Wesley Publishing Co.

[138] Y. Xiong and S. Shafer, "Depth from focusing and defocusing," Tech. Rep. CMU-RI-TR-93-07, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, March 1993.

[139] P. Favaro, H. Jin, A. Yezzi, and S. Soatto, "A variational approach to shape from defocus," *Proc. of the European Conference on Computer Vision*, May 2002.

[140] M. A. Georgeson, K. A. May, T. C. A. Freeman, and G. S. Hesse, "From filters to features: Scalespace analysis of edge and blur coding in human vision," *Journal of Vision*, vol. 7, no. 13, pp. –, 2007.

[141] E. Renrich and R. Caruana, "Infinite depth of field methodology," http://www.cs.cornell.edu/~erenrich/dof/, 2003.

[142] S. Ball, "Deep focus," http://www.dipteristsforum.org.uk/deepfocus/DeepFocus.pdf, 2004.

[143] M. Brady and G. E. Legge, "Camera calibration for natural image studies and vision research," *J. Opt. Soc. Am. A*, vol. 26, no. 1, pp. 30–42, 2009.

[144] FlashPoint Technology, "Scripting manual for DigitaScript 1.5," Tech. Rep. 6-2120-0001-01 v1.5.2, FlashPoint Technology, 2000.

[145] R. Willson and S. Shafer, "Precision imaging and control for machine vision research at carnegie mellon university," Tech. Rep. CMU-CS-92-118, Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, March 1992.

[146] J. V. Piqueres, "Still life with bolts," http://www.ignorancia.org/en/index.php?page=Bolts_still, 2002.

[147] POV-Ray, "Pov-ray hall of fame," http://hof.povray.org/, March 2010.

[148] D. J. Tolhurst and Y. Tadmor, "Band-limited contrast in natural images explains the detectability of changes in the amplitude spectra," *Vision Research*, vol. 37, no. 23, pp. 3203 – 3215, 1997.

[149] P. J. Bex, I. Mareschal, and S. C. Dakin, "Contrast gain control in natural scenes," *Journal of Vision*, vol. 7, no. 11, pp. 1–12, 8 2007.

[150] P. J. Bex, S. G. Solomon, and S. C. Dakin, "Contrast sensitivity in natural scenes depends on edge as well as spatial frequency structure," *J. Vis.*, vol. 9, no. 10, pp. 1–19, 9 2009.

[151] A. Olmos and F. Kingdom, "McGill calibrated colour image database," http://tabby.vision.mcgill.ca, 2004.

[152] J. H. van Hateren and A. van der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," *Proceedings: Biological Sciences*, vol. 265, no. 1394, pp. 359–366, Mar 1998.

[153] J. F. Brenner, B. S. Dew, J. B. Horton, T. King, P. W. Neurath, and W. D. Selles, "An automated microscope for cytologic research a preliminary evaluation," *J. Histochem. Cytochem.*, vol. 24, no. 1, pp. 100–111, 1976.

[154] Wikipedia, "List of color spaces and their uses — Wikipedia, the free encyclopedia," 2006, [accessed 20-November-2006].

[155] C. Poynton, "Frequently asked questions about colour," `http://www.poynton.com/PDFs/ColorFAQ.pdf`, 1997.

[156] M. Stokes, M. Anderson, S. Chandrasekar, and R. Motta, "A standard default color space for the internet - sRGB," Tech. Rep., HP and Microsoft, November 1996.

[157] Adobe Systems Inc, "Adobe RGB (1998) color image encoding," Tech. Rep., Adobe, May 2005.

[158] Y. K. Lee, "Comparison of CIELAB dE* and CIEDE2000 color-differences after polymerization and thermocycling of resin composites," *Dental Materials*, vol. 21, no. 7, pp. 678–682, July 2005.

[159] C. H. Chou, K. C. Liu, and C. S. Lin, "Perceptually optimized JPEG2000 coder based on CIEDE2000 color difference equation," *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 3, pp. 1184–1187, September 2005.

[160] M. W. Schwarz, W. B. Cowan, and J. C. Beatty, "An experimental comparison of RGB, YIQ, LAB, HSV, and opponent color models," *ACM Trans. Graph.*, vol. 6, no. 2, pp. 123–158, 1987.

[161] G. Paschos, "Perceptually uniform color spaces for color texture analysis: an empirical evaluation.," *IEEE Transactions on Image Processing*, vol. 10, no. 6, pp. 932–937, 2001.

[162] J. McNames, "An effective color scale for simultaneous color and gray-scale publications," *IEEE Signal Processing Magazine*, vol. 23, no. 1, pp. 82–96, Jan 2006.

[163] M. Morgan, C. Chubb, and J. A. Solomon, "A dipper function for texture discrimination based on orientation variance," *Journal of Vision*, vol. 8, no. 11, pp. –, 2008.

[164] G. Mather and D. R. R. Smith, "Blur discrimination and its relation to blur-mediated depth perception," *Perception*, vol. 31, no. 10, pp. 1211 – 1219, 2002.

[165] D. G. Pelli and L. Zhang, "Accurate control of contrast on microcomputer displays," *Vision Research*, vol. 31, no. 7-8, pp. 1337 – 1350, 1991.

[166] Wikipedia, "Scotopic vision — wikipedia, the free encyclopedia," 2010, [Online; accessed 22-June-2010].

[167] D. Karatzas and S. M. Wuerger, "A hardware-independent colour calibration technique," *Annals of the British Machine Vision Association*, vol. 2007, no. 3, pp. 1–11, Jan. 2007.

[168] J. A. Solomon, "Contrast discrimination: Second responses reveal the relationship between the mean and variance of visual signals," *Vision Research*, vol. 47, no. 26, pp. 3247 – 3258, 2007.

[169] M. A. Garcia-Perez, "Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties," *Vision Research*, vol. 38, no. 12, pp. 1861 – 1881, 1998.

[170] A. B. Watson and D. G. Pelli, "QUEST: a Bayesian adaptive psychometric method.," *Perception and Psychophysics*, vol. 33, no. 2, pp. 113–120, February 1983.

[171] D. H. Brainard, "The psychophysics toolbox," *Spatial Vision*, vol. 10, pp. 433–436(4), 1997.

[172] D. G. Pelli, "The videotoolbox software for visual psychophysics: transforming numbers into movies," *Spatial Vision*, vol. 10, pp. 437–442(6), 1997.

[173] A. J. Simmers, P. J. Bex, and R. F. Hess, "Perceived Blur in Amblyopia," *Investigative Ophthalmology and Visual Science*, vol. 44, no. 3, pp. 1395–1400, 2003.

[174] I. Fine and A. Rokem, "Psychtoolbox tutorial," Tech. Rep., Silver lab, University of California, Berkeley, 2007.

[175] F. A. Wichmann and N. J Hill, "The psychometric function: I. Fitting, sampling and goodness-of-fit," *Perception and Psychophysics*, vol. 63, no. 8, pp. 1293–1313, 2001.

[176] F. A. Wichmann and N. J Hill, "The psychometric function: II. Bootstrap-based confidence intervals and sampling," *Perception and Psychophysics*, vol. 63, no. 8, pp. 1314–1329, 2001.

[177] DrBob, "Principle of the imaging provided by the convex lens," http://commons.wikimedia.org/wiki/File:Lens3.svg, March 2006, [accessed 01-June-2010].

[178] R. J. Secord, "Image preloader with progress bar," Online tutorial, http://www.knowledgesutra.com/forums/topic/21316-image-preloader-with-progress-bar-status/, 2005.

[179] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2008, ISBN 3-900051-07-0.

[180] J. M. Buhmann, D. W. Fellner, M. Held, J. Ketterer, and J. Puzicha, "Dithered color quantization," *Computer Graphics Forum*, vol. 17, no. 3, pp. 219–231, 1998.

[181] S. M. Wuerger, H. Owens, and S. Westland, "Blur tolerance for luminance and chromatic stimuli," *Journal of the Optical Socity of America A*, vol. 18, no. 6, pp. 1231–1239, June 2000.

[182] D. G. Albrecht and D. B. Hamilton, "Striate cortex of monkey and cat: contrast response function," *J Neurophysiol*, vol. 48, no. 1, pp. 217–237, 1982.

[183] M. Carandini, "Matteobox toolbox," 2001.

[184] S. T. Mueller and C. T. Weidemann, "Decision noise: An explanation for observed violations of signal detection theory," *Psychonomic Bulletin and Review*, vol. 15, no. 3, pp. 465–494, 2008.

[185] F. Gosselin and P. G. Schyns, "Bubbles: a technique to reveal the use of information in recognition tasks," *Vision Research*, vol. 41, no. 17, pp. 2261 – 2271, 2001.

[186] R. T. Shilston, "Matlab implementation of focus measures," http://shilston.com/focusmeasures, September 2011.

[187] R. Ferzli, L. J. Karam, and J. Caviedes, "A robust image sharpness metric based on kurtosis measurement of wavelet coefficients," in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, 2005.

[188] R. Ferzli and L. J. Karam, "No-reference objective wavelet based noise immune image sharpness metric," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, September 2005, vol. 1, pp. I – 405–8.

[189] Sony, "EVI D30/31 command list," `http://bssc.sel.sony.com/Professional/docs/manuals/evid30commandlist1-21.pdf`, 1999.

[190] University of Southern California iLab, "iLab Neuromorphic Vision C++ Toolkit," `http://ilab.usc.edu/toolkit/`, Jan 2006.

[191] BBC, "Mobile phone sales start to slow," `http://news.bbc.co.uk/1/hi/uk_politics/5403588.stm`, Oct 2006.

[192] K. Behr, "'Same-as-Difference': Narrative transformations and intersecting cultures in Harry Potter," *Journal of Narrative Theory*, vol. 35, no. 1, pp. 112–132, 2005.

[193] M. Bollmann, R. Hoischen, and B. Mertsching, "Integration of static and dynamic scene features guiding visual attention," in *Mustererkennung 1997, 19. DAGM-Symposium*, London, UK, 1997, pp. 483–490, Springer-Verlag.

[194] A. P. Bradley and F. W. M. Stentiford, "JPEG2000 and region of interest coding," *Digital Image Computing Techniques and Applications, Australia, Proceedings of*, Jan 2002.

[195] C. Bundesen, "A theory of visual attention," *Psychological Review*, vol. 97, no. 4, pp. 523–547, 1990.

[196] W. H. Cheng, W. T. Chu, and J. L. Wu, "A visual attention based region-of-interest determination framework for video sequences," *IEICE Trans Inf Syst*, vol. E88-D, no. 7, pp. 1578–1586, 2005.

[197] S. C. Cheung, "Visualizing marriage in Hong Kong," *Visual Anthropology*, vol. 19, no. 1, pp. 21–37, 2006.

[198] T. N. Cornsweet and H. D. Crane, "Experimental study of visual accommodation," Tech. Rep. NASA CR-2007, NASA - Ames Research Centre, March 1972.

[199] S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman, "Indexing by latent semantic analysis," *Journal of the American Society of Information Science*, vol. 41, no. 6, pp. 391–407, 1990.

[200] A. D. Doulamis, N. Doulamis, G. Akrivas, and S. Kollias, "Non-sequential video content representation using temporal variation of feature vectors," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 3, pp. 758–768, 2000.

[201] S. M. Ebenholtz, *Oculomotor systems and perception*, Cambridge University Press, 2001.

[202] M. Fairchild, "Refinement of the RLAB color space," *Color Res. Appl*, vol. 21, pp. 338–346, 1996.

[203] A. Mfit Ferman and A. Murat Tekalp, "Two-stage hierarchical video summary extraction to match low-level user browsing preferences," *Multimedia, IEEE Transactions on*, vol. 5, pp. 244–256, June 2003.

[204] D. M. Frohlich, *Audiophotography: Bringing photos to life with sounds*, Kluwer, 2004.

[205] P. Getreuer, "Matlab CIE functions," `http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=7744`, 2005.

[206] J. Geusebroek, F. Cornelissen, A. Smeulders, and H. Geerts, "Robust autofocusing in microscopy," *Cytometry*, vol. 39, no. 1, pp. 1–9, 2000.

[207] gi@como on flickr, "A good friend," `http://www.flickr.com/photos/49784886@N00/441230864/`, March 2007.

[208] Y. Gong and X. Liu, "Video summarization with minimal visual content redundancies.," in *ICIP (3)*, 2001, pp. 362–365.

[209] Y. Gong and X. Liu, "Summarizing video by minimizing visual content redundancies.," in *ICME*, 2001.

[210] Y. Gong and X. Liu, "Generating optimal video summaries," in *IEEE International Conference on Multimedia and Expo (III)*, 2000, pp. 1559–1562.

[211] Y. Gong and X. Liu, "Video summarization using singular value decomposition," in *Proc. IEEE Intl. Conf. on Computer Vision and Pattern Recoginition*, June 2000, pp. 174–180.

[212] P. J. Green, "Colorimetry and colour difference," in *Colour Engineering*, P. J. Green and L. W. MacDonald, Eds. Wiley, 2002.

[213] A. Hanjalic and H. J. Zhang, "An integrated scheme for automated video abstraction based on unsupervised cluster-validity analysis," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 9, no. 8, pp. 1280–1289, 1999.

[214] D. Harper, "Talking about pictures: a case for photo elicitation," *Visual Studies*, vol. 17, no. 1, pp. 13–26, April 2002.

[215] L. He, E. Sanocki, A. Gupta, and J. Grudin, "Auto-summarization of audio-video presentations," in *MULTIMEDIA '99: Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, New York, NY, USA, 1999, pp. 489–498, ACM Press.

[216] L. Itti and C. Koch, "Computational modeling of visual attention," *Nature Reviews: Neuroscience*, vol. 2, pp. 2–11, March 2001.

[217] L. Itti and C. Koch, "Feature combination strategies for saliency-based visual attention systems," *Journal of Electronic Imaging*, vol. 10, no. 1, pp. 161–169, January 2001.

[218] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision Research*, vol. 40, no. 10-12, pp. 1489–1506, May 2000.

[219] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[220] G. M. Johnson and M. D. Fairchild, "A top down description of S-CIELAB and CIEDE2000," *Color Research and Application*, vol. 28, pp. 425–435, 2003.

[221] M. M. Jose, "MPEG-7 overview," Tech. Rep. ISO/IEC JTCI/SC/29/WG11 N6828, ISO, 2004.

[222] jpstanley on flickr, "Moon and elnath," `http://www.flickr.com/photos/jpstanley/138038291/`, April 2006.

[223] G. King, *Say 'cheese'! : the snapshot as art and social history*, Collins, 1986.

[224] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry.," *Hum Neurobiol*, vol. 4, no. 4, pp. 219–227, 1985.

[225] J. C. Kotulak and C. M. Schor, "The accommodative response to subthreshold blur and to perceptual fading during the troxler phenomenon," *Perception*, vol. 15, pp. 7–15, 1986.

[226] R. G. Kuehni, "Color space and its divisions," *Color Research & Application*, vol. 26, no. 3, pp. 209–222, 2001.

[227] T. Kurita and T. Kato, "Learning of personal visual impression for image database systems," in *in Proceedings of the Second International Conference on Document Analysis and Recognition*, 1993, pp. 547–552.

[228] S. Lee, S. Lee, and D. Chen, "Automatic video summary and description," in *VISUAL '00: Proceedings of the 4th International Conference on Advances in Visual Information Systems*, London, UK, 2000, pp. 37–48, Springer-Verlag.

[229] J. E. Lewis and L. Maler, "Blurring of the senses: common cues for distance perception in diverse sensory systems," *Neuroscience*, vol. 114, no. 1, pp. 19–22, September 2002.

[230] J. Li, "Autofocus searching algorithm considering human visual system limitations," *Optical Engineering*, vol. 44, no. 11, pp. 113201, 2005.

[231] Y. Li, T. Zhang, and D. Tretter, "An overview of video abstraction techniques," Tech. Rep. HPL-2001-191, HP Laboratories, Palo Alto, July 2001.

[232] R. Lienhart, S. Pfeiffer, and W. Effelsberg, "Video abstracting," *Communications of the ACM*, vol. 40, no. 12, pp. 54–62, 1997.

[233] M. R. Luo, C. Mincheq, P. Kenyon, and G. Cui, "Verification of CIEDE2000 using industrial data," in *AIC 2004 Color and Paints, Interim Meeting of the International Color Association*, 2004, pp. 97–102.

[234] Y. Ma, L. Lu, H. Zhang, and M. Li, "A user attention model for video summarization," in *MULTIMEDIA '02: Proceedings of the tenth ACM international conference on Multimedia*, New York, NY, USA, 2002, pp. 533–542, ACM Press.

[235] Y. F. Ma, L. Lu, H. J. Zhang, and M. Li, "A user attention model for video summarization," in *MULTIMEDIA '02: Proceedings of the tenth ACM international conference on Multimedia*, New York, NY, USA, 2002, pp. 533–542, ACM Press.

[236] Y. F. Ma and H. J. Zhang, "A model of motion attention for video skimming," in *Proceedings of the International Conference on Image Processing*, 2002, pp. 129–132.

[237] D. L. MacAdam, "Visual sensitivities to color differences in daylight," *Journal of the Optical Society of America*, vol. 42, no. 5, pp. 247–274, May 1942.

[238] C. McCreery, "Analysis of variance," Tech. Rep. 2007-3, Oxford Forum, 2007.

[239] J. Meyer, "The future of digital imaging - high dynamic range photography (HDR)," `http://www.cybergrain.com/tech/hdr`, 2004.

[240] A. Misawa, "Focusing apparatus for adjusting focus of an optical instrument," US Patent 6,864,474, Feb 2002.

[241] J. Nachmias and R. V. Sansbury, "Grating contrast: Discrimination may be better than detection," *Vision Research*, vol. 14, no. 10, pp. 1039 – 1042, 1974.

[242] J. Nam and A. H. Tewfik, "Dynamic video summarization and visualization," in *MULTIMEDIA '99: Proceedings of the seventh ACM international conference on Multimedia (Part 2)*, New York, NY, USA, 1999, pp. 53–56, ACM Press.

[243] J. Nam and A. H. Tewfik, "Video abstract of video," in *Multimedia Signal Processing, 1999 IEEE 3rd Workshop on*, 1999, pp. 117–122.

[244] Netscape, "Colors," `http://wp.netscape.com/home/bg/colorindex.html`, 1995.

[245] H. Nishiyama, S. Kin, T. Yokoyama, and Y. Matsushita, "An image retrieval system considering subjective perception," in *CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems*, New York, NY, USA, 1994, pp. 30–36, ACM.

[246] N. Ohta and A. R. Robertson, *Colorimetry: Fundamentals and Applications*, Wiley, 2005.

[247] W. M. Osberger, "Visual attention model," US Patent 6,670,963, December 2003.

[248] O. K. Oyekoya and F. W. Stentiford, "Eye tracking – a new interface for visual exploration," *BT Technology Journal*, vol. 24, no. 3, pp. 57–66, 2006.

[249] D. Parkhurst, K. Law, and E. Niebur, "Modeling the role of salience in the allocation of overt visual attention," *Vision Research*, vol. 42, pp. 107–123, 2002.

[250] M. R. Pointer and G. G. Attridge, "Some aspects of the visual scaling of large colour differences," *Color Research and Application*, vol. 22, no. 5, pp. 298–307, 1997.

[251] S. V. Porter, M. Mirmehdi, and B. T. Thomas, "A shortest path representation for video summarisation," in *Proceedings of the 12th International Conference on Image Analysis and Processing*. September 2003, pp. 460–465, IEEE Computer Society.

[252] Oxford University Press, "picture, n.," The Oxford English Dictionary, 1989.

[253] M. Ronnier-Luo, "Colour difference formulae: Past, present and future," in *ISCC, Ottowa*, May 2006.

[254] C. Ronse, "The phase congruence model for edge detection in multi-dimensional pictures," Tech. Rep. 97/16, Universit Louis Pasteur, Strasbourg, France, 1997.

[255] J. K. Rowling, *Harry Potter and the Philosopher's Stone*, Bloomsbury, London, 1997.

[256] D. M. Russell, "A design pattern-based video summarization technique: Moving from low-level signals to high-level structure.," in *HICSS*, 2000.

[257] M. Russell, "Using eye-tracking data to understand first impressions of a website," *Usability News*, vol. 7, no. 1, pp. 1–14, February 2005.

[258] I. A. Rybak, V. I. Gusakova, A. V. Golovan, L. N. Podladchikova, and N. A. Shevtsova, "A model of attention-guided visual perception and recognition," *Vision Research*, vol. 38, pp. 2387–2400, 1998.

[259] R. Scruton, "Photography and representation," *Critical Inquiry*, vol. 7, no. 3, pp. 577–603, 1981.

[260] G. Sharma, W. Wu, and E. N. Dalal, "The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," *Color Research and Application*, vol. 30, no. 1, pp. 21–30, February 2005.

[261] R. T. Shilston, "Visual attention: An identification of new applications, and an implementation on the TMS320DM642," M.S. thesis, University of Bristol, 2005.

[262] M. Smith and T. Kanade, "Video skimming and characterization through the combination of image and language understanding," in *Proceedings of the 1998 IEEE International Workshop on Content-Based Access of Image and Video Databases*, January 1998, pp. 61 – 70.

[263] J. A. Solomon, "The history of dipper functions," *Attention, Perception and Psychophysics*, vol. 71, no. 3, pp. 435–443, 2009.

[264] E. Spiekermann and E. M. Gringer, *Stop stealing sheep & find out how type works*, Adobe Press, 1993.

[265] S. Srinivasan, D. B. Ponceleon, A. Amir, and D. Petkovic, ""what is in that video anyway?" in search of better browsing.," in *ICMCS, Vol. 1*, 1999, pp. 388–393.

[266] F. W. M. Stentiford, "Attention based auto image cropping," in *Workshop on Computational Attention and Applications, Bielefeld*, March 2007.

[267] F. W. M. Stentiford, "An estimator for visual attention through competitive novelty with application to image compression," in *Picture Coding Symposium*, April 2001, pp. 101–104.

[268] F. W. M. Stentiford, "An evolutionary programming approach to the simulation of visual attention," in *Proceedings of the 2001 Congress on Evolutionary Computation CEC2001*, COEX, World Trade Center, 159 Samseong-dong, Gangnam-gu, Seoul, Korea, May 2001, pp. 851–858, IEEE Press.

[269] B. Suh, H. Ling, B. B. Bederson, and D. W. Jacobs, "Automatic thumbnail cropping and its effectiveness," in *UIST '03: Proceedings of the 16th annual ACM symposium on User interface software and technology*, New York, NY, USA, 2003, pp. 95–104, ACM Press.

[270] J. M. Thorne, "Supporting informal communication and closeness through video snapshots," in *London Communications Symposium 2005, September 2005, London, England*, 2005.

[271] C. Toklu, S. P. Liou, and M. Das, "Video abstract: A hybrid approach to generate semantically meaningful video summaries.," in *IEEE International Conference on Multimedia and Expo (III)*, 2000, pp. 1333–1336.

[272] B. L. Tseng, C. Lin, and J. R. Smith, "Using MPEG-7 and MPEG-21 for personalizing video," *IEEE Multimedia*, vol. 11, no. 1, pp. 42–52, 2004.

[273] B. L. Tseng, C. Y. Lin, and J. R. Smith, "Video summarization and personalization for pervasive mobile devices," in *Proc. SPIE Vol. 4676, p. 359-370, Storage and Retrieval for Media Databases 2002, Minerva M. Yeung; Chung-Sheng Li; Rainer W. Lienhart; Eds.*, M. M. Yeung, C.-S. Li, and R. W. Lienhart, Eds., Dec. 2001, pp. 359–370.

[274] J. K. Tsotsos, "On attention, saliency, recognition and feature binding," in *CCNCV keynote presentation, Bielefeld, Germany*, 2007.

[275] J. K. Tsotsos, S. M. Culhane, W. Y. K. Wai, Y. Lai, N. Davis, and F. Nuflo, "Modeling visual attention via selective tuning," *Artificial Intelligence*, vol. 78, no. 1-2, pp. 507–545, 1995.

[276] S. Uchihashi, J. Foote, A. Girgensohn, and J. Boreczky, "Video manga: generating semantically meaningful video summaries," in *MULTIMEDIA '99: Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, New York, NY, USA, 1999, pp. 383–392, ACM Press.

[277] K. N. Walker, T. F. Cootes, and C. J. Taylor, "Locating salient object features," *British Machine Vision Conference*, vol. 2, pp. 557–566, 1998.

[278] Y. Wang, P. Zhao, D. Zhang, M. Li, and H. Zhang, "Myvideos: a system for home video management," in *MULTIMEDIA '02: Proceedings of the tenth ACM international conference on Multimedia*, New York, NY, USA, 2002, pp. 412–413, ACM Press.

[279] Wikipedia, "Helvetica — wikipedia, the free encyclopedia," 2008, [accessed 2-September-2008].

[280] W. D. Wright, "The sensitivity of the eye to small colour differences," *The Proceedings of the Physical Society*, vol. 53, no. 296, pp. 93–112, March 1941.

[281] Y. Zhuang, Y. Rui, T. S. Huang, and S. Mehrotra, "Adaptive key frame extraction using unsupervised clustering.," in *ICIP (1)*, 1998, pp. 866–870.

[282] T. T. Yeo, S. H. Ong, Jayasooriah, and R. Sinniah, "Autofocusing for tissue microscopy," *Image and Vision Computing*, vol. 11, no. 10, pp. 629 – 639, 1993.

Appendices

Appendix A

Example photographs

Photographs are taken for many purposes, but there are some specific situations and stories reported to the author that demonstrate the challenges affecting any automatic processes used to take them. This appendix includes a such few photographs, together with a description of their meaning and issues they provoke. These help to explain why photography is so subjective, and that there are many issues that remain present and ripe for further investigation.

Figure A.1: *Land and water - Lasalle* This photo, from the McGill library [151], is an excellent photo with which to demonstrate how different people might want to focus at different distances, but retaining the same composition. Such top-down influence is likely to be present in many photos that are taken – in this scene, a geologist, botanist and ornithologist would be likely to focus differently.

Figure A.2: *An old picture of Boyton School* This photo, from the Boyton historial archive, shows children sitting on a wall in front of the school house. Recently, this photo was examined to try to judge when certain portions of the wall on which the children are sat was rebuilt. This demonstrates that photographs are sometimes used in ways that the photographer would not, or could not, anticipate.
A framework for examining such varying usage of a photos, discussing in more detail the relationship between photographer, subject, audience and photograph, and how that evolves over time, is described by Frohlich [204, Figure 3.5].

Appendix B

Preliminary experiments

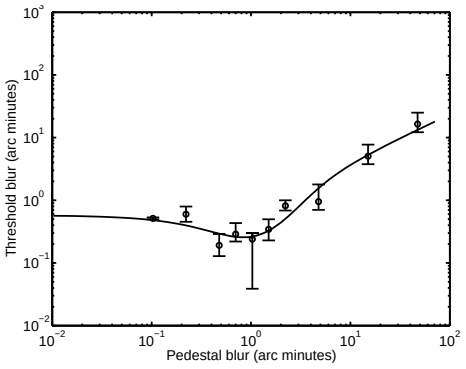
Section 3.4 (page 74) describes other researchers' methodologies for exploring human blur discrimination using controlled viewing conditions. These previous experiments used a variety of approaches, with experimental differences in aspects of stimuli presentation, observation and analysis. Several aspects, such as presentation duration, were assessed in preliminary trials and existing published approaches found to be satisfactory. Contrast normalisation was considered (see page 79), but for the experiments in this work was not employed, to ensure that stimuli looked like natural images, rather than being artificially distorted.

One aspect of the methodology, however, that was less clear in the existing literature is the viewing method. Whilst it was assumed that unconstrained binocular viewing was used in previous experiments in the literature, it was felt necessary to conduct preliminary experiments to measure the discrimination response in an observer when viewing the stimuli monocularly as well as binocularly to confirm whether blur discrimination was affected by viewing conditions.

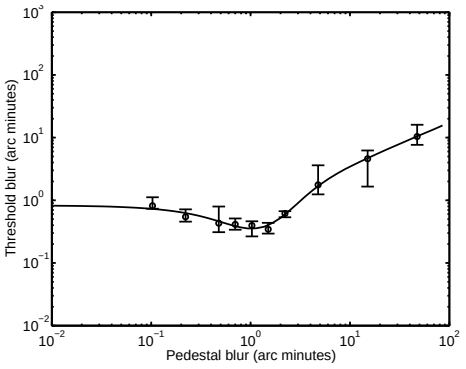
To assess any difference in discrimination when viewing monocularly, one observer performed the experiments described in Chapter 6 both with binocular vision, and also with an eye patch covering one eye. Apart from this patch, the experimental procedure was identical. The results of this trial are shown in Figure B.1.

What is apparent from these results is that the same underlying dipper response is present with both viewing methods. Quantitatively, the results are similar, though there is slightly better discrimination performance with binocular vision than monocular viewing across all pedestal conditions. The cause of this slight change in discrimination was not investigated, though from observer feedback it is possible that monocular viewing is more tiring, and this reduced the ability to discriminate between stimuli.

On the basis of these results, and similar ones for the second Van Hateren scene,

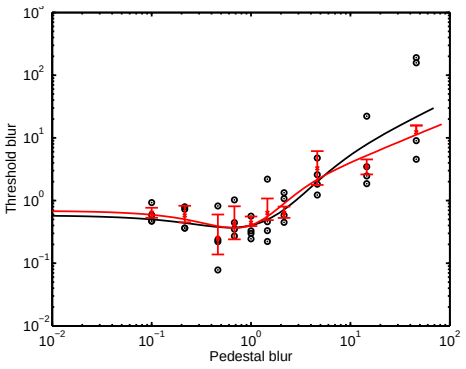


(a) Image 00005, binocular

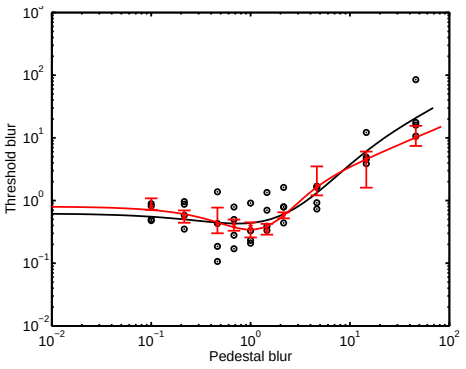


(b) Image 00005, monocular

Figure B.1: Blur discrimination thresholds as a function of pedestal blur, for one observer (RTS) viewing Van Hateren image 00005 both binocularly and monocularly. Each estimate of threshold was based on at least three separate determinations (QUESTs) per measure, followed by a bootstrapping procedure to establish the 82% threshold and confidence intervals. This is overlaid with the best fit to a contrast response function (Equation 6.1) computed with a least squares fitting.



(a) Image 01342



(b) Image 00005

Figure B.2: Blur discrimination thresholds as a function of pedestal blur, for one observer (RTS) viewing Van Hateren image 00005 and 01342 monocularly overlaid on the results of multiple binocular observers (from Figure 6.6). Monocular results, in red, show the estimate of thresholds based on at least three separate determinations (QUESTs) per measure, followed by a bootstrapping procedure to establish the 82% threshold and confidence intervals. This is overlaid with the best fit to a contrast response function (Equation 6.1) computed with a least squares fitting. Binocular results, in black, show the mean of each individual observers' results after applying a bootstrapping procedure.

it was decided that no marked difference was present between viewing techniques. Thus, just as when considering contrast where emphasis was placed on the similarity with normal viewing behaviour, it was decided to perform all experiments with binocular viewing.

Once the full experimental data had been collected, it was possible to validate this response by demonstrating that the preliminary results for binocular viewing fit within the range of binocular results found in the reported experiments. Figure B.2 overlays the monocular results on the binocular human results reported in Figure 6.6. The black line (the best fit to the mean binocular observations), falls within the red error bars (the confidence intervals of the boostrapping procedure applied to the monocular observations) for all but the two pedestal conditions with largest blur. The confidence intervals for the conditions with largest blur, whilst not covering the black best fit line, do encompass some of the individual observers.

Overall, this reconfirms the preliminary finding that blur discrimination can be performed both monocularly and binocularly, and that there is no marked difference in discrimination between these two viewing methods – those differences that are present are similar in magnitude to the differences in results between observers.

Appendix C

New focus measures

A list of focus measures discovered in the literature is presented on page 43. Additional focus measures were conceived by the author during the course of the research described in this thesis, each of which was fully assessed using all the experiments in Sections 5 and 7. This appendix provides a brief description of each of these new measures:

randomnumber This measure does not assess the input image, but instead outputs a random number. This was used to confirm that all focus measures performed better than chance, which Table 5.2 (page 120) demonstrates to be the case – *randomnumber* scored lowest of all measures.

rawlaplace The *laplace* measure is a thresholded summation of the convolution of the input image with a 3x3 Laplace kernel [94]. As other measures were found in the literature with in both a thresholded and non-thresholded variant, a non-thresholded variant of *laplace* was created and named *rawlaplace*.

phasecongruence and phasecongruence2 Kovessi observed that many image features "give rise to points where the Fourier components of the image are maximally in phase" [22]. By searching for the congruency of Fourier components, Kovessi was able to produce a feature map. To turn this feature map into a focus score, it was summed to give an overall score for the image. As a slight variation to *phasecongruence*, the summation of the raw feature map prior to Kovessi's thresholding was used as a further focus measure, *phasecongruence2*.

alphaAdult, Red Onion and Image Ensemble Based on Billock's observation that human adults' spatial frequency contrast sensitivity was similar to that of the $1/f^{\alpha}$ power spectra observed in natural scenes [68], and the suggestion

by Tadmor and Tolhurst that α drives accommodation, a focus measure based on α was created. This family of measures outputs a score that is the ratio of the α determined from the image to that of a predetermined value, such that the peak output occurs when they are equal. For this predetermined value, *alphaAdult* uses the mean value determined by Billock's observations of humans. *alphaImageEnsemble* uses the weighted average of a number of sets of images, also as reported by Billock. Finally, *alphaRedOnion* uses the amplitude spectra slope of the modal red onion image as described in Section 4.

energylaplace5a, 5b and 5c During preliminary experiments, it was found that *energylaplace*, a measure which is the summation of the squared convolution of the original image with a 3x3 Laplacian kernel, performed reasonably. However, when used as a model observer to discriminate blur, discrimination plateaued in high blur conditions. It was found that a larger kernel more closely matched human discrimination performance in such conditions. The *5a*, *5b* and *5c* variants of the *energylaplace* measure each use a slightly different 5x5 kernel, but otherwise are identical to the original measure.

Appendix D

Hardware and software

This appendix describes some of the hardware used in the experiments described in this report, and highlights key software techniques used to control them.

D.1 Digita and the Kodak DC290

The Kodak DC290 is a 2.1 megapixel consumer digital camera that supports scripting with the Digita language. The process of installing a script onto the camera is simply to save it as a `.CSM` file in the `SYSTEM` folder of the camera's memory card. Whilst Digita does provide the ability to control all the camera's parameters, the camera itself does not fully support them. For example, whilst it is possible to set the focus distance to 53cm with the command `SetCameraState("fdst", 53)`, the camera does not support focusing at any distance other than Infinity, 20m, 10m, 5m, 3m, 2m, 1m, 70cm and 50cm.

Likewise, it is possible to specify the shutter speed, using `SetCameraState("shut", s)`, where `s` is the time for which shutter should be open (in microseconds), and aperture using `SetCameraState("aper", a)`, where `a` is the required F stop multiplied by 100. However, neither of these appeared to alter the brightness of the photos that were taken, so it is likely that they are not supported by the camera, though no documentation could be found to confirm this.

The script that was used to automatically capture a range of images was the following:

```
menu  "Add-On Scripts"
name  "Focus"
label "Focus"

mode 0  # Means the script is only available when the camera is in
        # Capture mode.
declare i:status
```

```
declare u:IPIP
declare u:uImageTaken, uImageAvail, uStorage
declare b:bSystem, bCapture, bVendor
declare i:shutter, speed, focus
declare u:min_shut, max_shut, def_shut
declare t:ShutterList, CompleteShutterList, FocusList
declare s:Head,HeadF

SetCaptureMode (still)

GetCapabilitiesRange("shut", min_shut, max_shut, def_shut)

uStorage = 1
IPIP = 0x10000000

CompleteShutterList = " 90,  0"
ShutterList = CompleteShutterList
FocusList = "050,055,060,065,070,075,080,085,090,095,100,120,150,200,  0"

status = 0
waitloop1:
GetCameraStatus (bSystem, bCapture, bVendor)
if bSystem & IPIP
   Wait (300)
   goto waitloop1
end

# SetCameraState("aper", 800) # F8.0
# SetCameraState("fdst", 65535) # Focus at infinity

SetCameraState("xmod", 1) # Programmed exposure mode.  NB. This value
                          # must be changed for different models of cameras
SetCameraState("fmod", 3) # Manual focus
SetCameraState("aper", 300) # F3.0 - wide aperture => small depth of field
SetCameraState("wmod", 11) # White balance off
SetCameraState("mcap", 1) # Still (as opposed to burst or timelapse)
SetCameraState("scpn", 7) # Photo compression, 7=lossless
SetCameraState("ssiz", 3) # Photo size, 1=High (1792x1200),
                          # 2=Medium (1440x960), 3=Standard (720x480),
                          # 4=Ultra (2240x1500)
SetCameraState("smod", 3) # Disable flash
SetCameraState("xcmp", 0) # No exposure compensation
SetCameraState("zpos", 100) # Zoom position - the DC290 can do 100-300
                          # (=1x to 3x); the default is 130
SetCameraState("irev", 0) # Instant review, units are 0.01 secs

Wait (1000)

shutloop:
   SubString(ShutterList, 0, 3, Head)
   SubString(FocusList, 0, 3, HeadF)
   StringToNumber(Head, speed)
   StringToNumber(HeadF, focus)
   if speed == 0
      if focus == 0
         goto done
      end
```

```
      if focus != 0
         SubString(FocusList, 4, 999, FocusList)
   ShutterList = CompleteShutterList
         goto shutloop
      end
   end
   shutter = 1000000 / speed
   DisplayClear()
   DisplayLine("Shutter: ",Head)
   DisplayLine("Focus:   ",HeadF,"cm")
   Wait (1000)
   SetCameraState("fdst", focus)
   SetCameraState("shut", shutter)
   StartCapture()
waitloop2:
   GetCameraStatus (bSystem, bCapture, b‘Vendor)
   if bSystem & IPIP
      Wait (250)
      goto waitloop2
   end
   SubString(ShutterList, 4, 999, ShutterList)
   goto shutloop

done:

exitscript
end
```

D.2 Sony EVI-D30 camera

The Sony EVI-D30 can be controlled using an RS232 connection at 9600 baud, 8 data bits, no parity, and 1 stop bit. Whilst the camera's control protocol, Visca, is bi-directional, it is possible to simply send commands using a fire-and-forget strategy. Full documentation of the protocol is provided by Sony [189], but for this work, the most pertinent commands are:

Command	**Description**
`0x8101043803FF`	Manual focus
`0x8101043802FF`	Auto focus
`0x810104480Z0Z0Z0ZFF`	Direct focus, where `ZZZZ` is the focus data (infinity = `1000`, close = `9FFF`)

Frequently the EVI-D30 cameras are used in conjunction with an Axis 2401 webcam. Using such a setup, the above commands can be sent directly to the camera by crafting a special URL such as:

```
http://host/axis-cgi/com/serial.cgi?port=1\&write=8101043803FF
```

On a Linux computer, the libVISCA software makes control and interrogation of the camera far easier. It can be downloaded from:

```
http://damien.douxchamps.net/libvisca
```

Appendix E

Scoring the focus measures

This appendix contains results from the experiments described in Section 5. The following tables contain full tabulated results that have been summarised in Table 5.2. They were produced using the methodology outlined in Section 5, and computed with Equation (2.8).

Measure	Accuracy	Range	False maxima	Width	Noise	Score	Rank
absolutegradient	19	95	13	16	3.74	1.11	42
absolutevariation	1	40	6	26	0.47	0.37	25
alphaAdult	0	49	5	6	0.13	0.26	15
alphaImageEnsemble	0	49	5	6	0.13	0.26	17
alphaRedOnion	0	49	5	6	0.13	0.26	15
autocorrelation	0	33	3	6	0.11	0.17	4
brennergradient	0	33	3	10	0.46	0.21	7
chernfft	0	33	4	25	0.56	0.32	24
cranepeak	1	79	6	8	0.18	0.39	30
cranesum	0	30	3	7	0.13	0.17	3
CPBD	0	56	9	5	0.15	0.37	28
energylaplace	0	53	9	3	0.35	0.37	27
energylaplace5a	0	53	9	4	0.30	0.37	26
energylaplace5b	0	49	7	4	0.21	0.30	23
energylaplace5c	0	44	6	4	0.17	0.27	19
entropy	0	84	15	13	6.09	1.58	43
groenvariance	1	42	6	13	0.39	0.29	21
histogramentropy	14	93	12	32	6.22	1.63	44
hlv	1	70	10	28	1.12	0.57	35
imagepower	21	93	14	13	2.51	0.90	39
JNBM	0	42	6	6	0.20	0.26	18
laplace	0	47	5	4	0.20	0.26	14
masgrn	0	30	6	11	0.16	0.25	12
menmay	21	93	11	26	2.31	0.85	37
normalizedgroenvariance	1	44	7	26	0.52	0.40	31
phasecongruence	22	84	7	42	3.16	1.00	41
phasecongruence2	0	77	7	3	1.05	0.46	33
randomnumber	20	93	14	42	26.53	6.45	45
range	0	63	6	8	0.73	0.37	29
rawlaplace	21	93	14	13	2.48	0.89	38
smd	0	30	3	6	0.15	0.17	2
sml	0	33	4	5	0.10	0.19	6
squaredgradient	0	33	5	7	0.18	0.22	9
standarddeviationbasedautocorrelation	1	40	6	13	0.44	0.29	20
tenengrad	0	79	12	6	0.57	0.52	34
thresholdedabsolutegradient	0	33	5	7	0.18	0.22	9
triakis11s	1	33	4	12	0.28	0.22	11
thresholdedcontent	21	93	14	13	2.51	0.90	39
thresholdedpixelcount	18	88	15	28	1.83	0.83	36
va	0	33	4	28	1.22	0.43	32
voll4	0	30	3	6	0.10	0.16	1
voll5	1	42	6	13	0.44	0.30	22
waveletw1	0	47	5	6	0.11	0.25	13
waveletw2	0	33	5	4	0.19	0.21	8
waveletw3	0	30	4	4	0.19	0.18	5

Table E.1: Scores for Chillis

Measure	Accuracy	Range	False maxima	Width	Noise	Score	Rank
absolutegradient	2	85	18	26	0.17	1.04	31
absolutevariation	2	46	10	18	0.03	0.57	12
alphaAdult	3	56	13	8	0.32	0.70	18
alphaImageEnsemble	1	43	12	6	0.10	0.57	11
alphaRedOnion	5	56	13	17	0.74	0.73	22
autocorrelation	0	44	10	5	0.09	0.53	8
brennergradient	0	55	12	17	0.02	0.68	17
chernfft	2	37	11	15	0.04	0.52	5
cranepeak	1	85	20	10	0.19	1.05	33
cranesum	0	36	8	7	0.06	0.44	4
CPBD	3	88	25	10	0.10	1.17	35
energylaplace	0	63	19	3	0.25	0.87	27
energylaplace5a	0	63	19	4	0.22	0.87	26
energylaplace5b	0	63	18	5	0.17	0.84	25
energylaplace5c	0	63	16	5	0.13	0.80	23
entropy	23	95	34	97	14.49	3.91	44
groenvariance	0	46	10	15	0.03	0.56	10
histogramentropy	72	88	9	100	3.88	1.93	42
hlv	0	69	22	32	0.99	1.04	32
imagepower	6	92	25	79	0.85	1.45	36
JNBM	0	84	19	10	0.28	1.03	30
laplace	0	46	9	5	0.16	0.53	7
masgrn	8	57	8	32	0.11	0.70	20
menmay	2	98	30	69	3.40	1.72	40
normalizedgroenvariance	0	45	13	13	0.05	0.61	16
phasecongruence	49	96	30	100	3.77	2.01	43
phasecongruence2	21	98	25	100	2.00	1.69	39
randomnumber	55	95	35	99	49.15	12.03	45
range	0	90	23	24	0.53	1.17	34
rawlaplace	52	96	25	66	1.64	1.62	38
smd	0	27	8	6	0.10	0.37	3
sml	0	84	14	10	0.08	0.94	28
squaredgradient	0	22	7	10	0.05	0.32	1
standarddeviationbasedautocorrelation	3	47	10	21	0.05	0.59	15
tenengrad	0	75	20	5	0.14	0.97	29
thresholdedabsolutegradient	0	22	7	10	0.05	0.32	1
triakis11s	0	40	13	13	0.04	0.58	13
thresholdedcontent	6	92	25	79	0.85	1.45	36
thresholdedpixelcount	28	95	25	100	2.99	1.77	41
va	0	54	12	26	0.03	0.70	19
voll4	0	46	9	6	0.09	0.53	6
voll5	3	47	10	21	0.05	0.59	14
waveletw1	0	45	10	6	0.08	0.54	9
waveletw2	0	57	19	5	0.17	0.82	24
waveletw3	0	50	17	5	0.17	0.73	21

Table E.2: Scores for Coins

Measure	Accuracy	Range	False maxima	Width	Noise	Score	Rank
absolutegradient	64	98	31	106	3.33	2.13	38
absolutevariation	12	87	19	95	0.58	1.45	11
alphaAdult	2	97	30	77	0.50	1.58	16
alphaImageEnsemble	23	98	31	90	1.06	1.72	20
alphaRedOnion	2	97	30	77	0.50	1.58	16
autocorrelation	72	96	32	77	2.83	2.02	36
brennergradient	20	85	22	106	3.12	1.74	23
chernfft	12	95	19	94	0.47	1.50	13
cranepeak	4	97	29	41	4.72	1.82	24
cranesum	71	95	24	23	2.30	1.71	19
CPBD	34	92	21	105	2.36	1.73	22
energylaplace	35	98	26	106	3.09	1.90	31
energylaplace5a	71	95	23	106	2.83	2.02	34
energylaplace5b	71	95	22	95	2.34	1.92	32
energylaplace5c	71	95	24	74	2.26	1.84	27
entropy	7	97	36	105	25.84	6.51	44
groenvariance	7	92	18	93	0.47	1.45	10
histogramentropy	29	96	15	106	3.85	1.83	26
hlv	2	28	2	43	0.11	0.51	1
imagepower	14	77	17	98	0.70	1.39	8
JNBM	72	97	28	106	3.64	2.17	39
laplace	34	93	25	106	3.86	1.94	33
masgrn	34	95	28	106	2.40	1.85	28
menmay	33	95	20	39	2.66	1.47	12
normalizedgroenvariance	7	91	16	87	0.46	1.38	7
phasecongruence	9	98	30	83	3.52	1.83	25
phasecongruence2	9	93	31	44	1.90	1.52	14
randomnumber	17	96	30	103	40.40	9.91	45
range	19	32	3	100	13.50	3.43	43
rawlaplace	34	68	13	106	3.37	1.63	18
smd	71	95	25	106	3.41	2.09	37
sml	34	95	15	106	2.71	1.72	21
squaredgradient	34	98	23	106	3.30	1.88	29
standarddeviationbasedautocorrelation	7	83	13	74	0.31	1.21	4
tenengrad	3	77	21	18	0.54	1.06	2
thresholdedabsolutegradient	34	98	23	106	3.30	1.88	29
triakis11s	9	92	27	45	0.86	1.37	6
thresholdedcontent	14	77	17	98	0.70	1.39	8
thresholdedpixelcount	34	81	19	93	2.12	1.55	15
va	18	56	10	100	0.54	1.23	5
voll4	72	96	32	77	2.83	2.02	35
voll5	7	83	13	74	0.31	1.21	3
waveletw1	71	95	31	106	3.71	2.20	40
waveletw2	71	95	33	102	4.45	2.29	42
waveletw3	71	95	32	101	4.16	2.24	41

Table E.3: Scores for Bolt

Measure	Accuracy	Range	False maxima	Width	Noise	Score	Rank
absolutegradient	6	89	17	33	1.69	1.00	36
absolutevariation	1	70	12	70	1.29	0.99	35
alphaAdult	0	29	8	12	0.04	0.35	6
alphaImageEnsemble	0	29	8	12	0.04	0.35	6
alphaRedOnion	0	29	8	12	0.04	0.35	6
autocorrelation	0	30	7	16	0.08	0.34	5
brennergradient	1	51	14	72	1.05	0.95	30
chernfft	0	67	14	70	1.06	0.99	34
cranepeak	3	80	10	16	0.10	0.69	20
cranesum	0	42	9	24	0.24	0.48	16
CPBD	0	55	13	9	0.04	0.58	18
energylaplace	0	51	9	6	0.08	0.48	13
energylaplace5a	0	51	9	7	0.06	0.48	14
energylaplace5b	0	51	9	8	0.05	0.48	15
energylaplace5c	0	41	8	9	0.04	0.40	10
entropy	11	95	22	72	16.82	4.24	44
groenvariance	0	33	8	70	1.09	0.81	24
histogramentropy	12	96	16	71	1.92	1.22	41
hlv	41	96	19	67	2.52	1.41	42
imagepower	19	96	17	5	1.22	0.98	32
JNBM	0	66	13	13	0.18	0.65	19
laplace	0	41	8	9	0.04	0.40	11
masgrn	0	49	7	26	0.49	0.50	17
menmay	1	92	15	71	0.93	1.11	39
normalizedgroenvariance	0	70	11	70	1.19	0.97	31
phasecongruence	16	82	14	19	0.71	0.82	25
phasecongruence2	0	72	15	69	0.97	1.01	38
randomnumber	27	96	24	73	34.67	8.48	45
range	2	93	5	69	0.79	1.00	37
rawlaplace	17	93	21	54	3.72	1.43	43
smd	0	42	9	16	0.16	0.45	12
sml	0	26	7	13	0.03	0.31	2
squaredgradient	0	42	8	68	0.57	0.79	22
standarddeviationbasedautocorrelation	0	33	8	70	1.18	0.82	27
tenengrad	1	79	14	22	0.47	0.77	21
thresholdedabsolutegradient	0	42	8	68	0.57	0.79	22
triakis11s	1	57	10	68	0.65	0.86	28
thresholdedcontent	19	96	17	5	1.22	0.98	32
thresholdedpixelcount	2	76	19	69	2.11	1.18	40
va	1	57	11	73	0.97	0.93	29
voll4	0	32	8	16	0.08	0.37	9
voll5	0	33	8	70	1.17	0.82	26
waveletw1	0	30	6	12	0.04	0.31	1
waveletw2	0	30	8	8	0.04	0.34	4
waveletw3	0	30	7	8	0.04	0.32	3

Table E.4: Scores for Red onion

Measure	Accuracy	Range	False maxima	Width	Noise	Score	Rank
absolutegradient	13	77	7	21	3.78	0.98	43
absolutevariation	4	0	0	16	1.59	0.41	30
alphaAdult	1	0	0	10	0.78	0.20	11
alphaImageEnsemble	1	0	0	10	0.78	0.20	12
alphaRedOnion	7	64	1	11	1.97	0.51	38
autocorrelation	2	0	0	10	0.75	0.20	10
brennergradient	1	0	0	15	1.34	0.35	25
chernfft	4	77	3	13	1.80	0.49	37
cranepeak	1	27	2	10	1.28	0.33	24
cranesum	1	0	0	11	0.82	0.21	18
CPBD	1	27	1	10	0.74	0.20	14
energylaplace	1	0	0	7	0.22	0.07	1
energylaplace5a	1	0	0	7	0.24	0.07	2
energylaplace5b	1	0	0	8	0.37	0.10	3
energylaplace5c	2	0	0	10	0.50	0.14	7
entropy	6	86	8	21	5.32	1.33	44
groenvariance	4	0	0	15	1.56	0.40	29
histogramentropy	4	91	6	19	1.88	0.56	39
hlv	4	64	2	11	1.00	0.30	22
imagepower	12	68	4	6	2.89	0.74	41
JNBM	4	18	1	9	1.04	0.27	20
laplace	1	0	0	8	0.37	0.10	4
masgrn	3	0	0	13	1.51	0.38	26
menmay	2	91	4	15	2.04	0.56	40
normalizedgroenvariance	4	64	1	14	1.44	0.39	28
phasecongruence	12	50	1	6	0.87	0.29	21
phasecongruence2	7	59	2	19	1.68	0.47	33
randomnumber	4	91	6	16	5.87	1.45	45
range	5	77	0	15	1.74	0.47	34
rawlaplace	7	59	3	21	1.70	0.48	36
smd	1	0	0	10	0.58	0.16	8
sml	1	0	0	11	0.79	0.21	15
squaredgradient	1	0	0	12	0.80	0.21	16
standarddeviationbasedautocorrelation	4	0	0	16	1.73	0.44	31
tenengrad	1	0	0	10	0.91	0.23	19
thresholdedabsolutegradient	1	0	0	12	0.80	0.21	16
triakis11s	1	0	0	14	1.18	0.31	23
thresholdedcontent	12	68	4	6	2.89	0.74	41
thresholdedpixelcount	2	73	5	16	1.19	0.39	27
va	2	23	1	18	1.89	0.48	35
voll4	2	0	0	10	0.78	0.20	13
voll5	4	0	0	16	1.73	0.44	32
waveletw1	2	0	0	10	0.66	0.18	9
waveletw2	1	0	0	9	0.41	0.12	6
waveletw3	1	0	0	8	0.38	0.11	5

Table E.5: Scores for Strawberries

Measure	Accuracy	Range	False maxima	Width	Noise	Score	Rank
absolutegradient	1	96	14	87	0.90	1.37	35
absolutevariation	2	27	7	15	0.04	0.37	1
alphaAdult	12	70	12	91	1.46	1.26	30
alphaImageEnsemble	14	72	12	84	0.76	1.19	27
alphaRedOnion	22	70	11	100	3.22	1.51	37
autocorrelation	2	33	10	10	0.04	0.46	9
brennergradient	2	33	10	36	0.12	0.57	16
chernfft	2	34	9	12	0.05	0.45	8
cranepeak	4	90	23	11	1.61	1.21	28
cranesum	2	28	10	15	0.04	0.44	7
CPBD	28	79	17	49	3.18	1.36	32
energylaplace	2	54	12	8	0.18	0.66	20
energylaplace5a	2	53	12	9	0.14	0.65	19
energylaplace5b	2	53	11	9	0.07	0.64	18
energylaplace5c	2	48	10	9	0.05	0.57	17
entropy	32	96	33	99	6.64	2.39	42
groenvariance	2	26	8	12	0.05	0.37	2
histogramentropy	32	67	8	100	2.87	1.47	36
hlv	48	98	22	100	9.64	2.87	43
imagepower	3	98	2	2	4.00	1.37	33
JNBM	31	95	21	10	2.08	1.33	31
laplace	2	95	12	12	0.11	1.02	26
masgrn	2	65	17	13	0.13	0.85	23
menmay	27	93	31	97	11.74	3.30	44
normalizedgroenvariance	2	27	8	12	0.04	0.38	3
phasecongruence	33	98	20	100	2.77	1.72	40
phasecongruence2	29	91	17	100	3.14	1.68	39
randomnumber	38	96	34	96	67.04	16.30	45
range	3	93	18	57	1.19	1.25	29
rawlaplace	53	94	19	76	1.34	1.56	38
smd	2	28	10	14	0.06	0.43	6
sml	5	72	12	17	0.05	0.83	22
squaredgradient	2	39	11	16	0.03	0.53	13
standarddeviationbasedautocorrelation	2	28	8	16	0.03	0.40	5
tenengrad	1	64	14	60	0.53	0.98	25
thresholdedabsolutegradient	2	39	11	16	0.03	0.53	13
triakis11s	1	45	16	19	0.13	0.69	21
thresholdedcontent	3	98	2	2	4.00	1.37	33
thresholdedpixelcount	17	30	0	100	7.00	1.99	41
va	1	28	10	84	0.53	0.94	24
voll4	2	42	9	10	0.05	0.51	11
voll5	1	27	8	16	0.04	0.39	4
waveletw1	2	36	11	14	0.04	0.51	12
waveletw2	2	44	11	9	0.06	0.56	15
waveletw3	2	36	11	9	0.07	0.50	10

Table E.6: Scores for Towel

Appendix F

Full ANOVA results for 'best' experiment

The ANOVA models were built and processed in R as follows:

```
library(xtable)
library(car)

bestdat<-read.table(``chillis1-withoutanomolies.txt'', header=T, sep
    =``\t'')
bestaov<-aov(Result ¬ Young30 + DecadeAge + Gender + Screen +
    Colourblind + Correction + Uncorrected + English + DecadeAge*
    Young30 + Gender*Young30 + Screen*Young30 + Colourblind*Young30 +
    Correction*Young30 + Uncorrected*Young30 + English*Young30 + Gender
    *DecadeAge + Screen*DecadeAge + Colourblind*DecadeAge + Correction*
    DecadeAge + Uncorrected*DecadeAge + English*DecadeAge + Screen*
    Gender + Colourblind*Gender + Correction*Gender + Uncorrected*
    Gender + English*Gender + Colourblind*Screen + Correction*Screen +
    Uncorrected*Screen + English*Screen + Correction*Colourblind +
    Uncorrected*Colourblind + English*Colourblind + Uncorrected*
    Correction + English*Correction + English*Uncorrected,data=bestdat)
summary(bestaov)
print(xtable(anova(bestaov),caption=``Anova for chillis1''), type=``
    latex'', file=``best-aov-chillis1.tex'')
```

Listing F.1: R code to perform analysis

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Young30	1	0.40	0.40	0.58	0.4518
DecadeAge	1	0.19	0.19	0.27	0.6042
Gender	1	0.05	0.05	0.08	0.7799
Screen	1	3.84	3.84	5.57	0.0225
Colourblind	1	3.38	3.38	4.91	0.0318
Correction	1	2.24	2.24	3.26	0.0777
Uncorrected	1	0.01	0.01	0.01	0.9231
English	1	0.00	0.00	0.00	0.9604
Young30:DecadeAge	1	0.14	0.14	0.20	0.6537
Young30:Gender	1	2.96	2.96	4.29	0.0440
Young30:Colourblind	1	2.05	2.05	2.97	0.0914
Young30:Correction	1	0.79	0.79	1.14	0.2912
Young30:Uncorrected	1	1.58	1.58	2.29	0.1372
Young30:English	1	0.12	0.12	0.17	0.6821
DecadeAge:Gender	1	0.08	0.08	0.12	0.7358
DecadeAge:Correction	1	0.62	0.62	0.89	0.3495
DecadeAge:Uncorrected	1	0.03	0.03	0.04	0.8413
DecadeAge:English	1	0.17	0.17	0.25	0.6218
Gender:Screen	1	0.04	0.04	0.05	0.8186
Gender:Correction	1	0.40	0.40	0.58	0.4509
Gender:Uncorrected	1	0.03	0.03	0.04	0.8448
Gender:English	1	0.93	0.93	1.34	0.2522
Screen:Correction	1	0.12	0.12	0.17	0.6788
Colourblind:Correction	1	0.99	0.99	1.44	0.2369
Correction:Uncorrected	1	0.00	0.00	0.00	0.9548
Correction:English	1	0.00	0.00	0.00	0.9751
Uncorrected:English	1	0.56	0.56	0.81	0.3737
Residuals	46	31.69	0.69		

Table F.1: Anova for chillis1

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Young30	1	0.83	0.83	0.32	0.5738
DecadeAge	1	1.83	1.83	0.71	0.4046
Gender	1	5.61	5.61	2.18	0.1468
Screen	1	0.42	0.42	0.16	0.6895
Colourblind	1	0.71	0.71	0.28	0.6022
Correction	1	0.18	0.18	0.07	0.7906
Uncorrected	1	1.63	1.63	0.63	0.4300
English	1	4.14	4.14	1.61	0.2113
Young30:Gender	1	1.03	1.03	0.40	0.5312
Young30:Colourblind	1	7.62	7.62	2.95	0.0923
Young30:Correction	1	13.46	13.46	5.22	0.0269
Young30:Uncorrected	1	0.31	0.31	0.12	0.7322
Young30:English	1	1.18	1.18	0.46	0.5030
DecadeAge:Gender	1	13.13	13.13	5.09	0.0288
DecadeAge:Correction	1	0.96	0.96	0.37	0.5445
DecadeAge:Uncorrected	1	0.62	0.62	0.24	0.6272
DecadeAge:English	1	1.04	1.04	0.40	0.5280
Gender:Screen	1	1.09	1.09	0.42	0.5194
Gender:Correction	1	2.10	2.10	0.82	0.3710
Gender:Uncorrected	1	4.27	4.27	1.65	0.2046
Gender:English	1	6.03	6.03	2.34	0.1332
Screen:Correction	1	1.36	1.36	0.53	0.4718
Colourblind:Correction	1	9.15	9.15	3.55	0.0659
Correction:Uncorrected	1	16.22	16.22	6.28	0.0157
Correction:English	1	4.17	4.17	1.62	0.2100
Uncorrected:English	1	0.20	0.20	0.08	0.7804
Residuals	47	121.27	2.58		

Table F.2: Anova for coin02

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Young30	1	9.84	9.84	1.72	0.1964
DecadeAge	1	13.89	13.89	2.42	0.1261
Gender	1	0.00	0.00	0.00	0.9850
Screen	1	11.80	11.80	2.06	0.1579
Colourblind	1	33.83	33.83	5.90	0.0189
Correction	1	2.14	2.14	0.37	0.5445
Uncorrected	1	38.14	38.14	6.65	0.0130
English	1	2.79	2.79	0.49	0.4890
Young30:Gender	1	1.00	1.00	0.18	0.6773
Young30:Colourblind	1	20.07	20.07	3.50	0.0674
Young30:Correction	1	9.07	9.07	1.58	0.2144
Young30:Uncorrected	1	6.18	6.18	1.08	0.3045
Young30:English	1	29.21	29.21	5.10	0.0286
DecadeAge:Gender	1	5.27	5.27	0.92	0.3423
DecadeAge:Correction	1	2.71	2.71	0.47	0.4949
DecadeAge:Uncorrected	1	24.28	24.28	4.24	0.0450
DecadeAge:English	1	4.18	4.18	0.73	0.3973
Gender:Screen	1	1.71	1.71	0.30	0.5874
Gender:Correction	1	13.53	13.53	2.36	0.1310
Gender:Uncorrected	1	1.76	1.76	0.31	0.5824
Gender:English	1	0.72	0.72	0.13	0.7238
Screen:Correction	1	21.47	21.47	3.75	0.0588
Colourblind:Correction	1	6.63	6.63	1.16	0.2874
Correction:Uncorrected	1	13.73	13.73	2.40	0.1283
Correction:English	1	3.80	3.80	0.66	0.4193
Uncorrected:English	1	15.76	15.76	2.75	0.1039
Residuals	48	275.14	5.73		

Table F.3: Anova for povraybolt

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Young30	1	0.87	0.87	0.33	0.5684
DecadeAge	1	4.95	4.95	1.88	0.1772
Gender	1	16.41	16.41	6.24	0.0162
Screen	1	1.67	1.67	0.64	0.4297
Colourblind	1	3.98	3.98	1.51	0.2254
Correction	1	2.26	2.26	0.86	0.3590
Uncorrected	1	4.97	4.97	1.89	0.1759
English	1	0.65	0.65	0.25	0.6219
Young30:DecadeAge	1	0.09	0.09	0.04	0.8522
Young30:Gender	1	3.62	3.62	1.37	0.2471
Young30:Colourblind	1	1.81	1.81	0.69	0.4108
Young30:Correction	1	2.30	2.30	0.87	0.3548
Young30:Uncorrected	1	1.34	1.34	0.51	0.4790
Young30:English	1	2.04	2.04	0.77	0.3834
DecadeAge:Gender	1	9.49	9.49	3.61	0.0640
DecadeAge:Correction	1	6.01	6.01	2.29	0.1376
DecadeAge:Uncorrected	1	6.09	6.09	2.31	0.1352
DecadeAge:English	1	0.92	0.92	0.35	0.5578
Gender:Correction	1	1.46	1.46	0.56	0.4600
Gender:Uncorrected	1	0.92	0.92	0.35	0.5568
Gender:English	1	6.12	6.12	2.32	0.1343
Screen:Correction	1	4.02	4.02	1.53	0.2226
Colourblind:Correction	1	2.05	2.05	0.78	0.3823
Correction:Uncorrected	1	0.10	0.10	0.04	0.8448
Correction:English	1	2.55	2.55	0.97	0.3303
Uncorrected:English	1	1.89	1.89	0.72	0.4017
Residuals	45	118.41	2.63		

Table F.4: Anova for redonion1

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Young30	1	1.52	1.52	1.02	0.3173
DecadeAge	1	0.01	0.01	0.01	0.9336
Gender	1	4.18	4.18	2.81	0.1002
Screen	1	0.00	0.00	0.00	0.9891
Colourblind	1	2.84	2.84	1.91	0.1733
Correction	1	4.88	4.88	3.29	0.0762
Uncorrected	1	0.00	0.00	0.00	0.9729
English	1	3.38	3.38	2.27	0.1383
Young30:DecadeAge	1	0.22	0.22	0.15	0.7016
Young30:Gender	1	1.16	1.16	0.78	0.3821
Young30:Colourblind	1	1.72	1.72	1.16	0.2872
Young30:Correction	1	2.76	2.76	1.86	0.1795
Young30:Uncorrected	1	1.92	1.92	1.30	0.2607
Young30:English	1	0.22	0.22	0.15	0.6989
DecadeAge:Gender	1	6.83	6.83	4.60	0.0373
DecadeAge:Correction	1	3.57	3.57	2.40	0.1279
DecadeAge:Uncorrected	1	8.06	8.06	5.43	0.0242
DecadeAge:English	1	0.96	0.96	0.65	0.4254
Gender:Screen	1	28.14	28.14	18.95	0.0001
Gender:Correction	1	8.59	8.59	5.78	0.0202
Gender:Uncorrected	1	4.69	4.69	3.16	0.0820
Gender:English	1	1.21	1.21	0.82	0.3706
Screen:Correction	1	2.77	2.77	1.86	0.1789
Colourblind:Correction	1	0.07	0.07	0.05	0.8265
Correction:Uncorrected	1	1.70	1.70	1.15	0.2900
Correction:English	1	2.04	2.04	1.37	0.2469
Uncorrected:English	1	1.23	1.23	0.83	0.3681
Residuals	47	69.81	1.49		

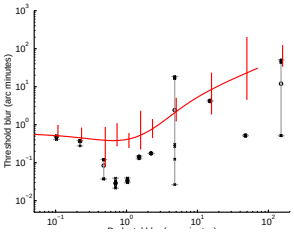
Table F.5: Anova for strawberries1

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Young30	1	1.36	1.36	0.25	0.6199
DecadeAge	1	2.63	2.63	0.48	0.4910
Gender	1	15.80	15.80	2.90	0.0956
Screen	1	1.89	1.89	0.35	0.5586
Colourblind	1	13.72	13.72	2.52	0.1197
Correction	1	0.45	0.45	0.08	0.7752
Uncorrected	1	1.70	1.70	0.31	0.5793
English	1	2.42	2.42	0.44	0.5089
Young30:Gender	1	0.74	0.74	0.14	0.7140
Young30:Colourblind	1	5.88	5.88	1.08	0.3044
Young30:Correction	1	8.10	8.10	1.49	0.2292
Young30:Uncorrected	1	8.64	8.64	1.59	0.2145
Young30:English	1	13.20	13.20	2.43	0.1267
DecadeAge:Gender	1	3.80	3.80	0.70	0.4082
DecadeAge:Correction	1	19.49	19.49	3.58	0.0652
DecadeAge:Uncorrected	1	0.44	0.44	0.08	0.7787
Gender:Correction	1	8.43	8.43	1.55	0.2200
Gender:Uncorrected	1	0.08	0.08	0.01	0.9031
Gender:English	1	5.04	5.04	0.93	0.3413
Screen:Correction	1	0.20	0.20	0.04	0.8499
Colourblind:Correction	1	0.19	0.19	0.04	0.8520
Correction:Uncorrected	1	6.33	6.33	1.16	0.2870
Correction:English	1	0.47	0.47	0.09	0.7710
Uncorrected:English	1	2.76	2.76	0.51	0.4799
Residuals	43	234.06	5.44		

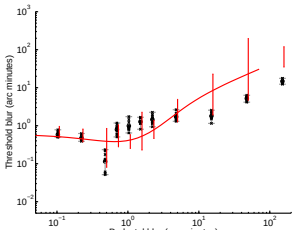
Table F.6: Anova for towel01

Appendix G

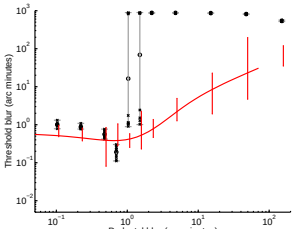
Full results for model observers



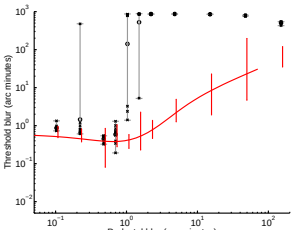
(a) absolutegradient $nf = 0.02\%$, $similarity = 0.456$



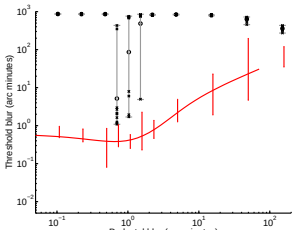
(b) absolutevariation $nf = 2.78\%$, $similarity = 0.348$



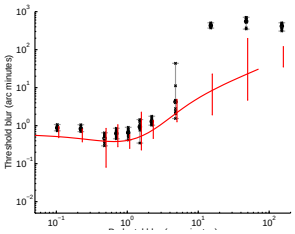
(c) alphaAdult $nf = 4.64\%$, $similarity = 1.13$



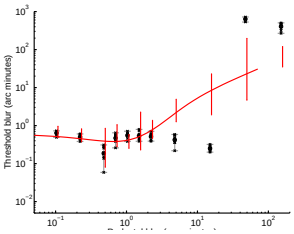
(d) alphaImageEnsemble $nf = 7.74\%$, $similarity = 0.755$



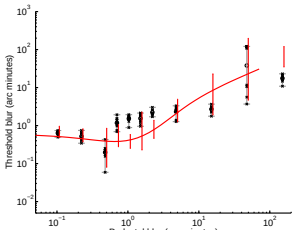
(e) alphaRedOnion $nf = 7.74\%$, $similarity = 859$



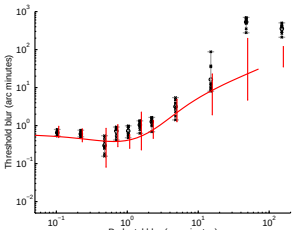
(f) autocorrelation $nf = 35.9\%$, $similarity = 0.688$



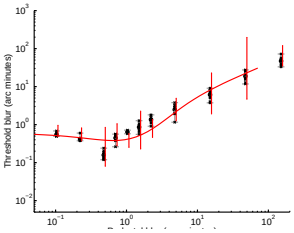
(g) brennergradient $nf = 4.64\%$, $similarity = 0.84$



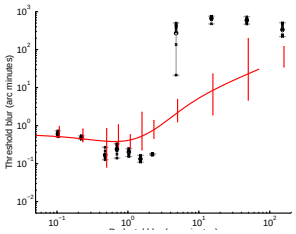
(h) chernfft $nf = 7.74\%$, $similarity = 0.317$



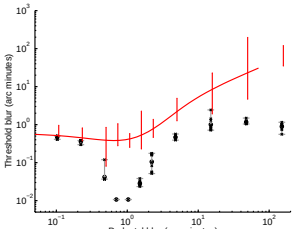
(i) cranepeak $nf = 21.5\%$, $similarity = 0.688$



(j) cranesum $nf = 12.9\%$, $similarity = 0.281$

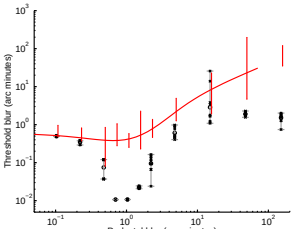


(k) CPBD $nf = 12.9\%$, $similarity = 1.7$

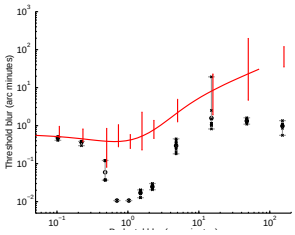


(l) energylaplace $nf = 0.22\%$, $similarity = 0.389$

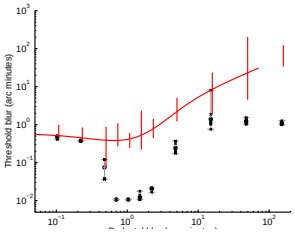
Figure G.1: Discrimination plotted against pedestal blur for each focus measure, where the noise factor is that which resulted in the greatest similarity between discrimination shape and results from human observers when considering Van Hateren 01342.



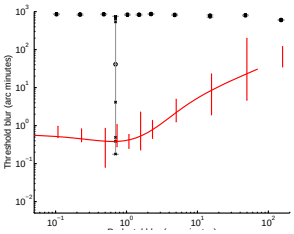
(a) energylaplace5a $nf = 0.36\%$, $similarity = 0.44$



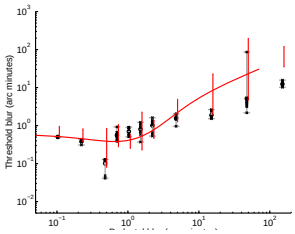
(b) energylaplace5b $nf = 0.22\%$, $similarity = 0.4$



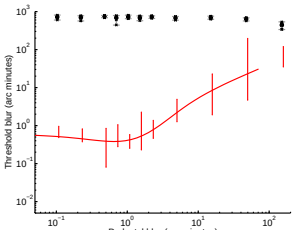
(c) energylaplace5c $nf = 0.22\%$, $similarity = 0.392$



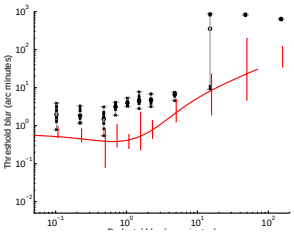
(d) entropy $nf = 0.05\%$, $similarity = 843$



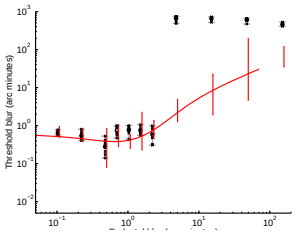
(e) groenvariance $nf = 4.64\%$, $similarity = 0.342$



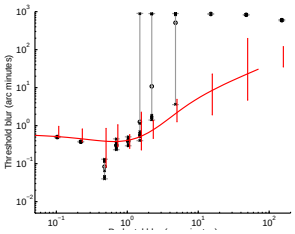
(f) histogramentropy $nf = 100\%$, $similarity = 949$



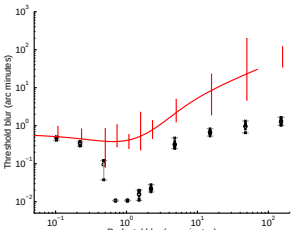
(g) hlv $nf = 0.01\%$, $similarity = 1.83$



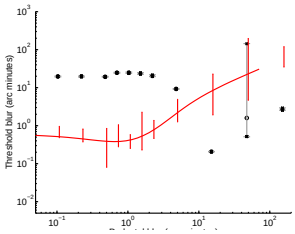
(h) imagepower $nf = 35.9\%$, $similarity = 0.732$



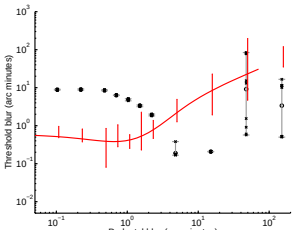
(i) JNBM $nf = 7.74\%$, $similarity = 1.08$



(j) laplace $nf = 0.22\%$, $similarity = 0.375$

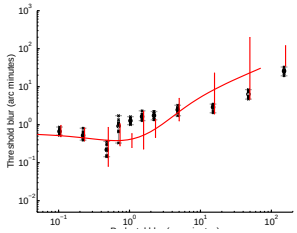


(k) masgrn $nf = 0.05\%$, $similarity = 23.3$

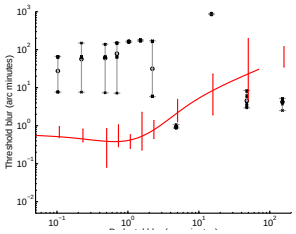


(l) menmay $nf = 0.01\%$, $similarity = 9.05$

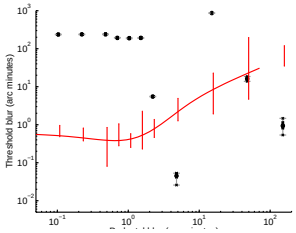
Figure G.2: Discrimination plotted against pedestal blur for each focus measure, where the noise factor is that which resulted in the greatest similarity between discrimination shape and results from human observers when considering Van Hateren 01342.



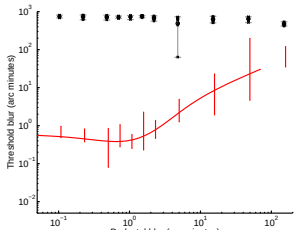
(a) normalizedgroenvariance $nf = 7.74\%$, $similarity = 0.367$



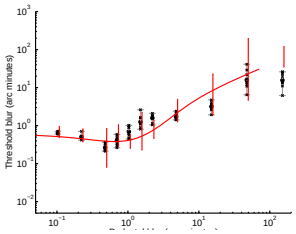
(b) phasecongruence $nf = 0.60\%$, $similarity = 26.6$



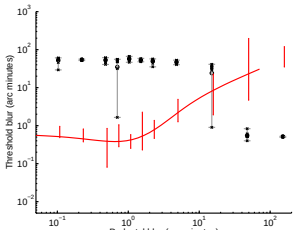
(c) phasecongruence2 $nf = 0.02\%$, $similarity = 229$



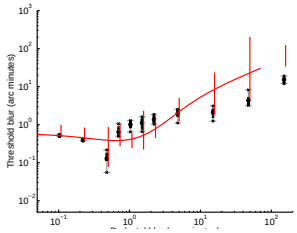
(d) randomnumber $nf = 0.05\%$, $similarity = 881$



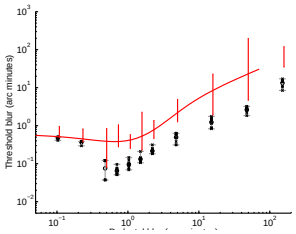
(e) range $nf = 4.64\%$, $similarity = 0.254$



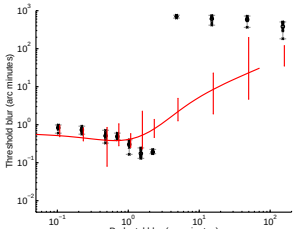
(f) rawlaplace $nf = 0.01\%$, $similarity = 67.6$



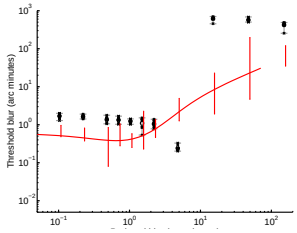
(g) rmscontrast $nf = 2.78\%$, $similarity = 0.364$



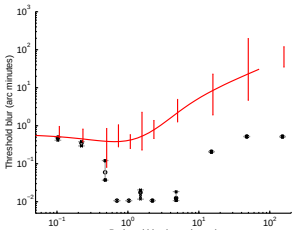
(h) smd $nf = 2.78\%$, $similarity = 0.329$



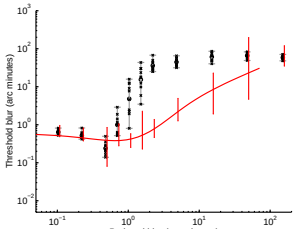
(i) sml $nf = 21.5\%$, $similarity = 1.65$



(j) squaredgradient $nf = 59.9\%$, $similarity = 4.82$

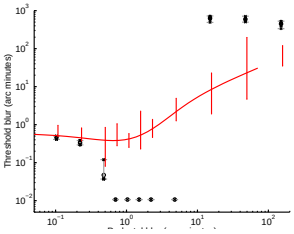


(k) standarddeviationbased-autocorrelation $nf = 0.03\%$, $similarity = 0.369$

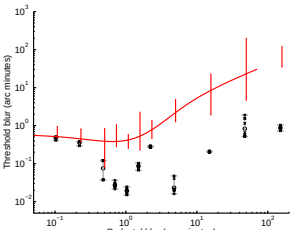


(l) tenengrad $nf = 35.9\%$, $similarity = 0.345$

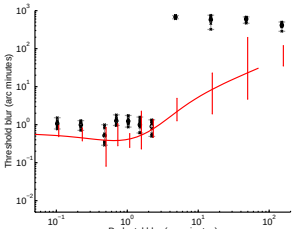
Figure G.3: Discrimination plotted against pedestal blur for each focus measure, where the noise factor is that which resulted in the greatest similarity between discrimination shape and results from human observers when considering Van Hateren 01342.



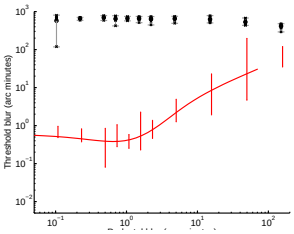
(a) thresholdedabsolutegradient $nf = 0.13\%$, $similarity = 2.37$



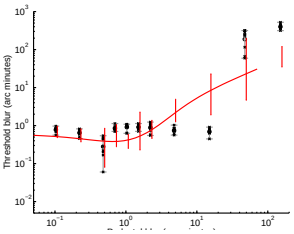
(b) triakis11s $nf = 0.03\%$, $similarity = 0.511$



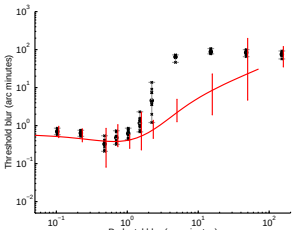
(c) thresholdedcontent $nf = 59.9\%$, $similarity = 0.802$



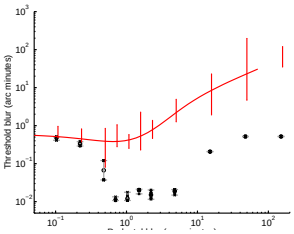
(d) thresholdedpixelcount $nf = 35.9\%$, $similarity = 767$



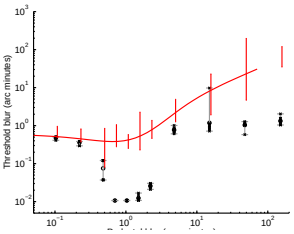
(e) va $nf = 1.67\%$, $similarity = 0.54$



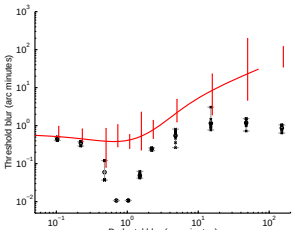
(f) voll4 $nf = 21.5\%$, $similarity = 0.336$



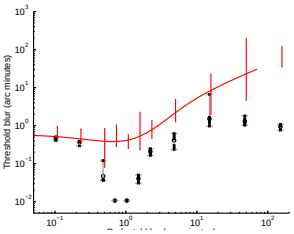
(g) voll5 $nf = 0.05\%$, $similarity = 0.371$



(h) waveletw1 $nf = 0.22\%$, $similarity = 0.379$

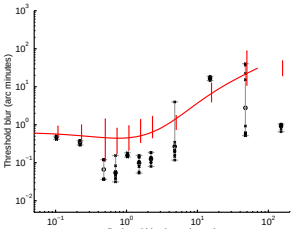


(i) waveletw2 $nf = 0.22\%$, $similarity = 0.394$

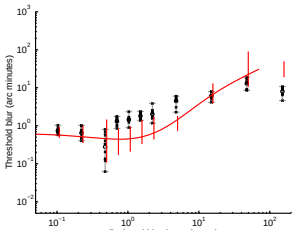


(j) waveletw3 $nf = 0.22\%$, $similarity = 0.398$

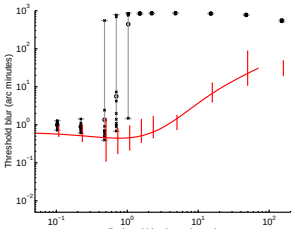
Figure G.4: Discrimination plotted against pedestal blur for each focus measure, where the noise factor is that which resulted in the greatest similarity between discrimination shape and results from human observers when considering Van Hateren 01342.



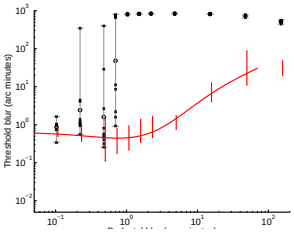
(a) absolutegradient $nf = 0.22\%$, $similarity = 0.394$



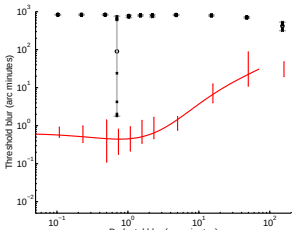
(b) absolutevariation $nf = 2.78\%$, $similarity = 0.385$



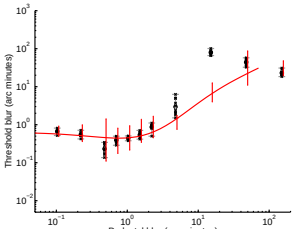
(c) alphaAdult $nf = 7.74\%$, $similarity = 0.861$



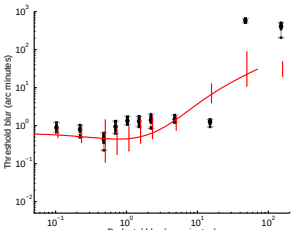
(d) alphaImageEnsemble $nf = 12.9\%$, $similarity = 0.833$



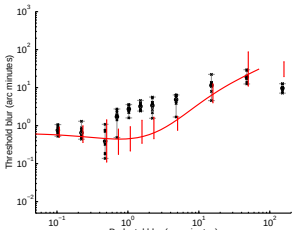
(e) alphaRedOnion $nf = 12.9\%$, $similarity = 829$



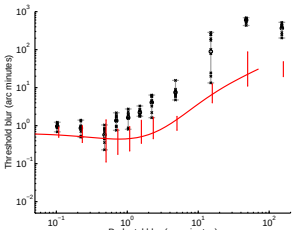
(f) autocorrelation $nf = 12.9\%$, $similarity = 0.412$



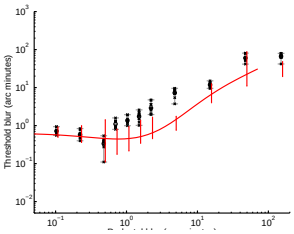
(g) brennergradient $nf = 7.74\%$, $similarity = 0.749$



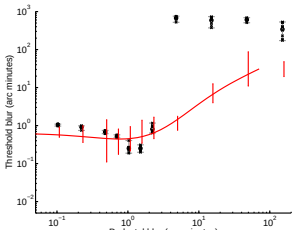
(h) chernfft $nf = 7.74\%$, $similarity = 0.361$



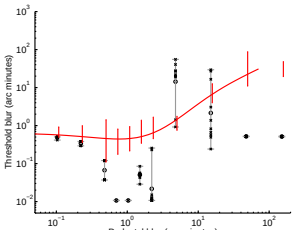
(i) cranepeak $nf = 35.9\%$, $similarity = 0.75$



(j) cranesum $nf = 21.5\%$, $similarity = 0.381$

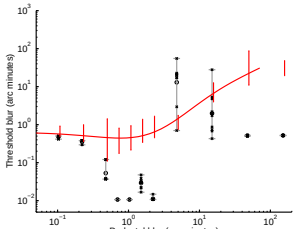


(k) CPBD $nf = 21.5\%$, $similarity = 1.22$

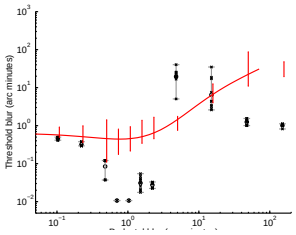


(l) energylaplace $nf = 0.03\%$, $similarity = 0.498$

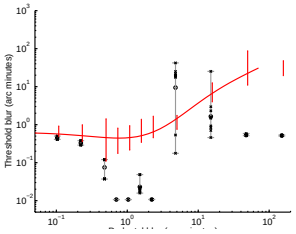
Figure G.5: Discrimination plotted against pedestal blur for each focus measure, where the noise factor is that which resulted in the greatest similarity between discrimination shape and results from human observers when considering Van Hateren 00005.



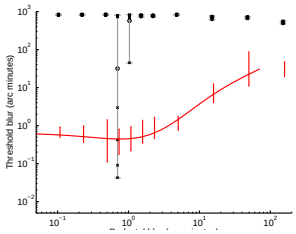
(a) energylaplace5a $nf = 0.05\%$, $similarity = 0.5$



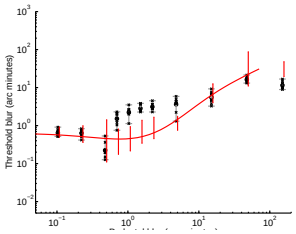
(b) energylaplace5b $nf = 0.22\%$, $similarity = 0.472$



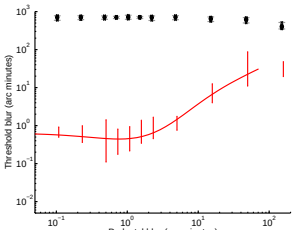
(c) energylaplace5c $nf = 0.08\%$, $similarity = 0.505$



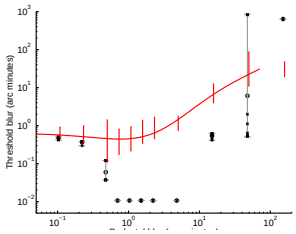
(d) entropy $nf = 2.78\%$, $similarity = 818$



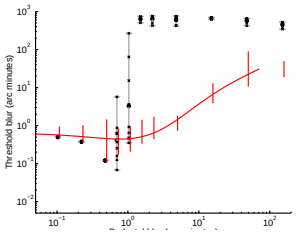
(e) groenvariance $nf = 7.74\%$, $similarity = 0.365$



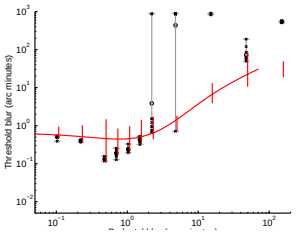
(f) histogramentropy $nf = 100\%$, $similarity = 923$



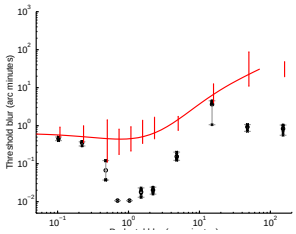
(g) hlv $nf = 0.00\%$, $similarity = 4.31$



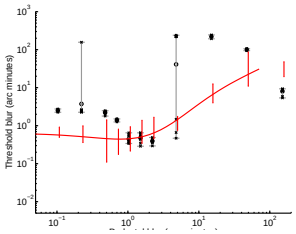
(h) imagepower $nf = 59.9\%$, $similarity = 0.993$



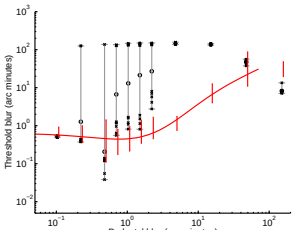
(i) JNBM $nf = 4.64\%$, $similarity = 0.596$



(j) laplace $nf = 0.13\%$, $similarity = 0.451$

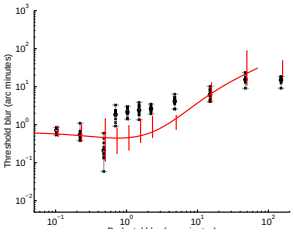


(k) masgrn $nf = 1.67\%$, $similarity = 2.51$

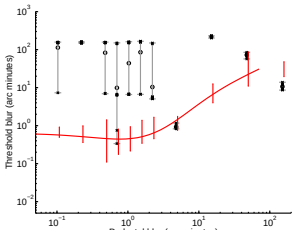


(l) menmay $nf = 0.60\%$, $similarity = 0.438$

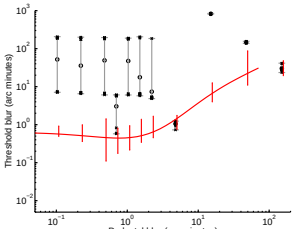
Figure G.6: Discrimination plotted against pedestal blur for each focus measure, where the noise factor is that which resulted in the greatest similarity between discrimination shape and results from human observers when considering Van Hateren 00005.



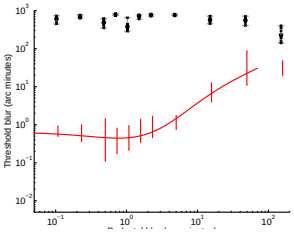
(a) normalizedgroenvariance $nf = 7.74\%$, $similarity = 0.39$



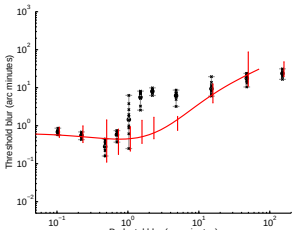
(b) phasecongruence $nf = 4.64\%$, $similarity = 109$



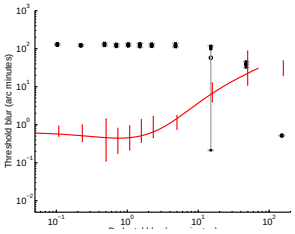
(c) phasecongruence2 $nf = 4.64\%$, $similarity = 50.1$



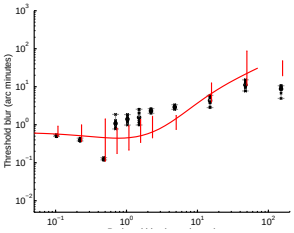
(d) randomnumber $nf = 0.00\%$, $similarity = 703$



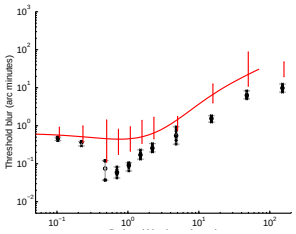
(e) range $nf = 7.74\%$, $similarity = 0.35$



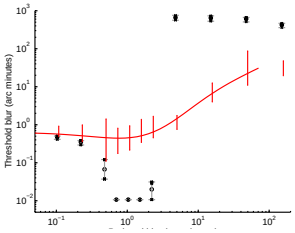
(f) rawlaplace $nf = 0.02\%$, $similarity = 137$



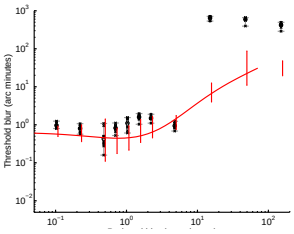
(g) rmscontrast $nf = 2.78\%$, $similarity = 0.426$



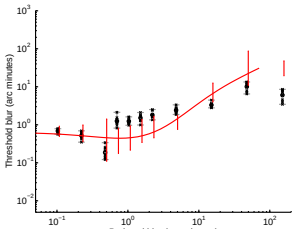
(h) smd $nf = 2.78\%$, $similarity = 0.421$



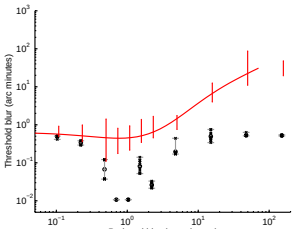
(i) sml $nf = 0.22\%$, $similarity = 1.71$



(j) squaredgradient $nf = 21.5\%$, $similarity = 0.789$

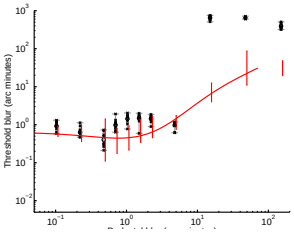


(k) standarddeviationbased-autocorrelation $nf = 4.64\%$, $similarity = 0.406$

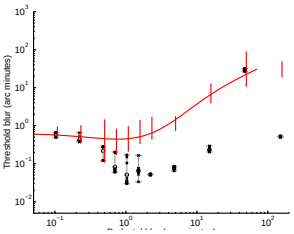


(l) tenengrad $nf = 0.22\%$, $similarity = 0.433$

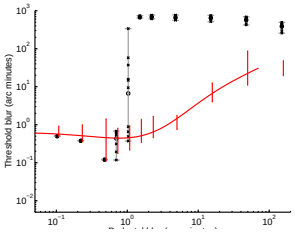
Figure G.7: Discrimination plotted against pedestal blur for each focus measure, where the noise factor is that which resulted in the greatest similarity between discrimination shape and results from human observers when considering Van Hateren 00005.



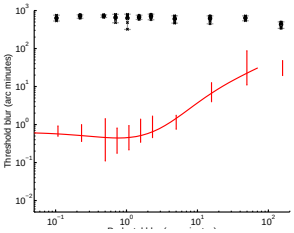
(a) thresholdedabsolutegradient $nf = 21.5\%$, $similarity = 0.798$



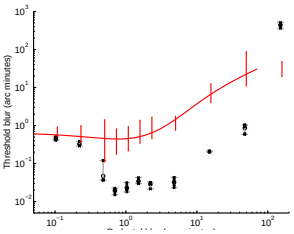
(b) triakis11s $nf = 0.02\%$, $similarity = 0.844$



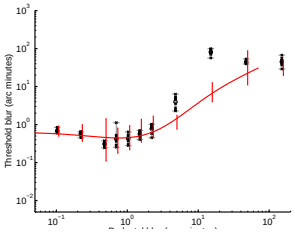
(c) thresholdedcontent $nf = 59.9\%$, $similarity = 0.991$



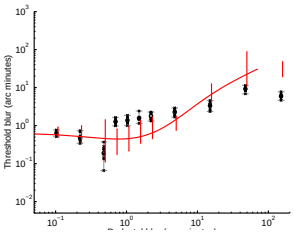
(d) thresholdedpixelcount $nf = 7.74\%$, $similarity = 859$



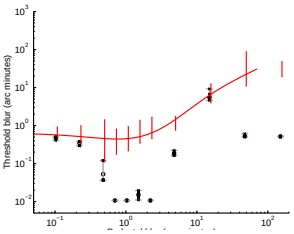
(e) va $nf = 0.08\%$, $similarity = 0.434$



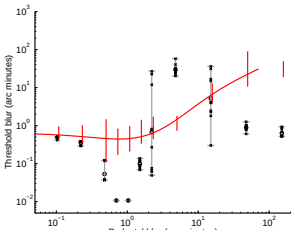
(f) voll4 $nf = 12.9\%$, $similarity = 0.359$



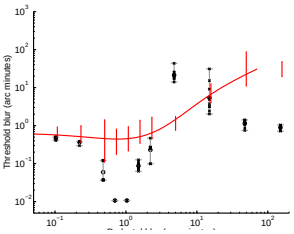
(g) voll5 $nf = 4.64\%$, $similarity = 0.399$



(h) waveletw1 $nf = 0.05\%$, $similarity = 0.5$



(i) waveletw2 $nf = 0.13\%$, $similarity = 0.443$



(j) waveletw3 $nf = 0.22\%$, $similarity = 0.463$

Figure G.8: Discrimination plotted against pedestal blur for each focus measure, where the noise factor is that which resulted in the greatest similarity between discrimination shape and results from human observerss when considering Van Hateren 000005.

Appendix H

Software implementation

This appendix provides implementation details for all the focus measures used throughout this work, together with their dependencies. The mathematical theory and justification behind each focus measure is not considered, except where this has a direct impact on the implementation. Further details about the respective measures can be found in the references from Table 2.3 (page 43).

Note: It was noted that two algorithms (*autocorrelation* and *voll*4) are essential identical. They differ solely in boundary conditions, and thus produce a different (though very similar) output when assessing the same input.

H.1 absolutegradient

```
function score = absolutegradient(I)

    % "This function was proposed by Jarvis [2] and is obtained by
       setting
    % n = 1, m = 1, theta = 0 in F1" [1]
    %
    % References:
    % [1] "A Comparison of Different Focus Functions for Use in
    % Autofocus Algorithms" by Groen et al in Cytometry 6:81-91 (1985)
    % [2] "Focus optimisation criteria for computer image processing."
       by
    % Jarvis in Microscope 24:163, 1976
    %

    % Ensure the image is 8-bit (0-255) grey intensity.
    if max(I(:))<1, I = I * 255; end;

    z = gradient(I);
    E = find(z>0);
    score = sum(sum(E));

end
```

H.2 absolutevariation

```
function score = absolutevariation(I)

    % "Absolute variation. As the computation of the variance is rather
    % complicated, a comparable result could be expected from the much
    % easier to calculate absolute difference. Thus m = 1 and c = A (
       the
    % image area) [using equation F3]." [1]
    %
    % References:
    % [1] "A Comparison of Different Focus Functions for Use in
    % Autofocus Algorithms" by Groen et al in Cytometry 6:81-91 (1985)
    %

    % Ensure the image is 8-bit (0-255) grey intensity.
    if max(I(:))<1, I = I * 255; end;

    % Calculate image area:
    imagearea = prod(size(I));

    % Calculate difference from mean:
    d = I - mean(I(:));
    d = abs(d);

    % Finally, the score:
    score = 1/imagearea * sum(sum(d));

end
```

H.3 alphaadult

```
function score = alphaAdult(I)

    % Use Alpha from Billock's results showing adults tuned to alpha=-
       1.15

    score = alphaEquals(I,-1.15);

end
```

H.4 alphaEquals

```
function score = alphaEquals(I, desiredalpha)

    % Score is based on difference between the image's alpha, and the
    % desired alpha.  This function is used by several wrappers which
       set
    % pre-determined desired alpha values.

    % Ensure the image is 8-bit (0-255) grey intensity.
    if max(I(:))<1, I = I * 255; end;

    % DoPowerPlot, which calculates alpha, requires a greyscale
    % square image.
    [d1 d2]=size(I);
    if d2>d1,
        offset = floor((d2-d1)/2);
        Igc = I(1:d1, (1+offset):(offset+d1));
    else
        offset = floor((d1-d2)/2);
        Igc = I((1+offset):(offset+d2), 1:d2);
    end;

    % First: Determine current image alpha
    [currentfreq,currentscale_hist,currentorient,currentorient_hist,
       ...
        currentalpha,currentalphaintercept,currentdummy] = ...
            DoPowerPlot(Igc,32);

    % Second: Return the score, such that the peak value (1) is
       returned
    % when image alpha = desired alpha (and a negative value is
       returned
    % should it be necessary).

    if currentalpha ≤ desiredalpha,
        score = desiredalpha / currentalpha;
    else
        score = currentalpha / desiredalpha;
    end;
end
```

H.5 alphaimageensemble

```
function score = alphaImageEnsemble(I)

    % Use Alpha from Billock's results showing the average alpha over
       an
    % ensemble of natural scenes is 1.08

    score = alphaEquals(I,-1.08);


end
```

H.6 alpharedonion

```
function score = alphaRedOnion(I)

    % Red Onion modal ground truth was image 19 (64134421-640x480.jpg).
    % Its alpha is 1.392 calculated as follows:
    %     I = imread('64134421-640x480-modal-best.jpg');
    %     Ig = rgb2gray(I);
    %     size(Ig) % 640x480
    %     Igc = Ig(1:480,80:80+479); % Crop out 480x480 from the centre
    %     doPowerPlot(Igc,32)

    score = alphaEquals(I,-1.392);

end
```

H.7 autocorrelation

```
function score = autocorrelation(I)

    % "AutoCorrelation (Vollath, 1987, 1988)." [1]
    %
    % References:
    % [1] "Autofocusing in Computer Microscopy: Selecting the Optimal
       Focus
    % Algorithm" by Sun et al in MICROSCOPY RESEARCH AND TECHNIQUE
    % 65:139149 (2004)
    %
    % score = sum over all image of i(x, y) . i(x + 1, y)
    %          - sum over all image of i(x, y) . i(x + 2, y)
    %

    [w h] = size(I);
    score = 0;
    for i=1:w-2,
        for j=1:h,
            score = score + (I(i,j)*I(i+1,j) - I(i,j)*I(i+2,j));
        end;
    end;
end
```

H.8 brennergradient

```
function score = brennergradient(I, T)

    % "Evaluation of autofocus functions in molecular cytogenetic
       analysis"
    % by Santos et al in Journal of Microscopy Vol.188 Issue 03, pp 264
       -272
    % http://dx.doi.org/10.1046/j.1365-2818.1997.2630819.x
    %
    % score = sum over height
    %             sum over width
    %                abs(i(x+2,y)-i(x,y))^2
    % where abs(i(x+2,y)-i(x,y))^2 > T
    %
    % Table 1 shows that a threshold of 5 is used.

    step = 2;
    T1 = 5;

    [width height] = size(I);
    ML = zeros(width,height);
    for x=(1+step):(width-step),
        for y=(1+step):(height-step),
            ML(x,y) = abs(I(x+step,y)-I(x,y))^2;
        end;
    end;

    score = sum(sum(find(ML≥T1)));
end
```

H.9 chernfft

```
function score = chernfft(theimage)

    % Chern et al (2001) "Practical issues in pixel-based auto-
    % focusing for machine vision", Proceedings of the 2001 IEEE
    % International Conference of Robotics and Automation, Seoul 2001
    %
    % "The image's grey levels are placed row by row into a 1D
    % array, and its FFT evaluated."
    %

    theimage = double(theimage);
    [w h] = size(theimage);

    % Matlab's reshape takes elements columnwise, so rotate
    % the image before reshaping:
    theimage = theimage';
    oned = reshape(theimage,w*h,1);

    % Perform an FFT.  Chern stated that zero padding was used
    % but does not comment on the length of the FFT.  As such
    % no padding is performed in this implementation.
    f = fft(oned);

    re = real(f);
    im = imag(f);

    score = 1/(w*h) * sum(abs((re.^2+im.^2).*atan(im./re)));

end
```

H.10 cpbd

```
%=========================================================================
% File: CPBD_compute.m
% Original code written by Niranjan D. Narvekar
% IVU Lab (http://ivulab.asu.edu)
% Last Revised: October 2009 by Niranjan D. Narvekar
%=========================================================================
% Copyright Notice:
% Copyright (c) 2009-2010 Arizona Board of Regents.
% All Rights Reserved.
% Contact: Lina Karam (karam@asu.edu) and Niranjan D. Narvekar (
   nnarveka@asu.edu)
% Image, Video, and Usabilty (IVU) Lab, ivulab.asu.edu
% Arizona State University
% This copyright statement may not be removed from this file or from
% modifications to this file.
% This copyright notice must also be included in any file or product
% that is derived from this source file.
%
% Redistribution and use of this code in source and binary forms,
% with or without modification, are permitted provided that the
% following conditions are met:
% - Redistribution's of source code must retain the above copyright
% notice, this list of conditions and the following disclaimer.
% - Redistribution's in binary form must reproduce the above copyright
% notice, this list of conditions and the following disclaimer in the
% documentation and/or other materials provided with the distribution.
% - The Image, Video, and Usability Laboratory (IVU Lab,
% http://ivulab.asu.edu) is acknowledged in any publication that
% reports research results using this code, copies of this code, or
% modifications of this code.
% The code and our papers are to be cited in the bibliography as:
%
% The code and our papers are to be cited in the bibliography as:
%
% N.D. Narvekar and L. J. Karam, "CPBD Sharpness Metric Software",
% http://ivulab.asu.edu/Quality/CPBD
%
% N.D. Narvekar and L. J. Karam, "A No-Reference Perceptual Image
   Sharpness
% Metric Based on a Cumulative Probability of Blur Detection,"
% International Workshop on Quality of Multimedia Experience (QoMEX
```

```
   2009),
% pp. 87-91, July 2009.
%
% N. D. Narvekar and L. J. Karam, "An Improved No-Reference Sharpness
   Metric Based on the
% Probability of Blur Detection," International Workshop on Video
   Processing and Quality Metrics
% for Consumer Electronics (VPQM), http://www.vpqm.org, January 2010.
%
% DISCLAIMER:
% This software is provided by the copyright holders and contributors
% "as is" and any express or implied warranties, including, but not
% limited to, the implied warranties of merchantability and fitness for
% a particular purpose are disclaimed. In no event shall the Arizona
% Board of Regents, Arizona State University, IVU Lab members, or
% contributors be liable for any direct, indirect, incidental, special,
% exemplary, or consequential damages (including, but not limited to,
% procurement of substitute goods or services; loss of use, data, or
% profits; or business interruption) however caused and on any theory
% of liability, whether in contract, strict liability, or tort
% (including negligence or otherwise) arising in any way out of the use
% of this software, even if advised of the possibility of such damage.
%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
   %%%%%
% function      : CPBD_compute
% description   : This function computes the CPBD metric which
   determines
%                 the amount of sharpness of the image. Larger the
   metric
%                 value, sharper the image.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
   %%%%%

function [sharpness_metric] = CPBD_compute(input_image)

%%%%%%%%%%%%% pre-processing %%%%%%%%%%%%%
% convert to gray scale if color image
[x y z] = size(input_image);
if z > 1
    input_image = rgb2gray(input_image);
end

% Robert Shilston's fix - ensure 8bit images are integers:
```

```
input_image = round(input_image);

% convert the image to double for further processing
input_image = double(input_image);
% get the size of image
[m,n] = size(input_image);

%%%%%%%%%%%%% parameters %%%%%%%%%%%%%
% threshold to characterize blocks as edge/non-edge blocks
threshold = 0.002;
% fitting parameter
beta = 3.6;
% block size
rb = 64;
rc = 64;
% maximum block indices
max_blk_row_idx = floor(m/rb);
max_blk_col_idx = floor(n/rc);
% just noticeable widths based on the perceptual experiments
%widthjnb = [5*ones(1,57) 3*ones(1,200)];
widthjnb = [5*ones(1,51) 3*ones(1,205)];

%%%%%%%%%%%%% initialization %%%%%%%%%%%%%
% arrays and variables used during the calculations
total_num_edges = 0;
hist_pblur = zeros(1,101);
cum_hist = zeros(1,101);

%%%%%%%%%%%%% edge detection %%%%%%%%%%%%%
% edge detection using canny and sobel canny edge detection is done to
% classify the blocks as edge or non-edge blocks and sobel edge
% detection is done for the purpose of edge width measurement.

input_image_canny_edge = edge(input_image,'canny');
input_image_sobel_edge = edge(input_image,'Sobel',[2],'vertical');

%%%%%%%%%%%%% edge width calculation %%%%%%%%%%%%%
[width] = marziliano_method(input_image_sobel_edge, input_image);

%%%%%%%%%%%%% sharpness metric calculation %%%%%%%%%%%%%
% loop over the blocks
for i=1:max_blk_row_idx
    for j=1:max_blk_col_idx

        % get the row and col indices for the block pixel positions
```

```
        rows = (rb*(i-1)+1):(rb*i);
        cols = (rc*(j-1)+1):(rc*j);
        % decide whether the block is an edge block or not
        decision = get_edge_blk_decision(input_image_canny_edge(rows,
            cols), threshold);

        % process the edge blocks
        if (decision==1)

            % get the edge widths of the detected edges for the block
            local_width = width(rows,cols);
            local_width = local_width(local_width ~= 0);

            % find the contrast for the block
            blk_contrast = blkproc(double(input_image(rows,cols)),[rb
                rc],@get_contrast_block)+1;
            % get the block Wjnb based on block contrast
            blk_jnb = widthjnb(blk_contrast);
            % calculate the probability of blur detection at the edges
            % detected in the block
            prob_blur_detection = 1 - exp(-abs(local_width./blk_jnb).^
                beta);

            % update the statistics using the block information
            for k = 1:numel(local_width)
                % update the histogram
                temp_index = round(prob_blur_detection(k)* 100) + 1;
                hist_pblur(temp_index) = hist_pblur(temp_index) + 1;
                % update the total number of edges detected
                total_num_edges = total_num_edges + 1;
            end
        end
    end
end

% normalize the pdf
if(total_num_edges ~=0)
    hist_pblur = hist_pblur / total_num_edges;
else
    hist_pblur = zeros(size(hist_pblur));
end

% calculate the sharpness metric
sharpness_metric = sum(hist_pblur(1:64));
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    %%%%%
% function      : marziliano_method
% description   : This function calculates the edge-widths of the
    detected
%                 edges and returns an matrix as big as the image with
    0's
%                 at non-edge locations and edge-widths at the edge
%                 locations.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    %%%%%

function [edge_width_map] = marziliano_method(E, A)

% edge_width_map consists of zero and non-zero values. A zero value
% indicates that there is no edge at that position and a non-zero value
% indicates that there is an edge at that position and the value itself
% gives the edge width
edge_width_map = zeros(size(A));

% converting the image to type double
A = double(A);

% find the gradient for the image
[Gx Gy] = gradient(A);

% dimensions of the image
[M N] = size(A);

% initializing the matrix to empty which holds the angle information of
     the
% edges
angle_A = [];

% calculate the angle of the edges
 for m=1:M
     for n=1:N
            if (Gx(m,n)¬=0)
                angle_A(m,n) = atan2(Gy(m,n),Gx(m,n))*(180/pi); % in
                    degrees
            end
```

```
            if (Gx(m,n)==0 && Gy(m,n)==0)
                angle_A(m,n) = 0;
            end
            if (Gx(m,n)==0 && Gy(m,n)==pi/2)
              angle_A(m,n) = 90;
            end
     end
 end


if(numel(angle_A) ~= 0)

    % quantize the angle
    angle_Arnd = 45*round(angle_A./45);

    count = 0;
    for m=2:M-1
        for n=2:N-1
            if (E(m,n)==1)

                %%%%%%%%%%%%%%%%%%%% If gradient angle = 180 or -180 %%
                    %%%%%%%%%%%%%%%%%%%%
                if (angle_Arnd(m,n) ==180 || angle_Arnd(m,n) ==-180)
                    count = count + 1;
                    for k=0:100
                        posy1 = n-1 -k;
                        posy2 = n-2 -k;
                        if ( posy2≤0)
                            break;
                        end

                        if ((A(m,posy2) - A(m,posy1))≤0)
                            break;
                        end
                    end
                    width_count_side1 = k + 1 ;
                    for k=0:100
                        negy1 = n+1 + k;
                        negy2 = n+2 + k;

                        if (negy2>N)
                            break;
                        end

                        if ((A(m,negy2) - A(m,negy1))≥0)
```

```
            break;
        end

    end
    width_count_side2 = k + 1 ;

    edge_width_map(m,n) = width_count_side1+width_count
        _side2;

end


%%%%%%%%%%%%%%%%%%%% If gradient angle = 0 %%%%%%%%%%%%
    %%%%%%%%%%
if (angle_Arnd(m,n) ==0)
    count = count + 1;
    for k=0:100
        posy1 = n+1 +k;
        posy2 = n+2 +k;
        if ( posy2>N)
            break;
        end

        if ((A(m,posy2) - A(m,posy1))≤0)
            break;
        end
    end
    width_count_side1 = k + 1 ;
    for k=0:100
        negy1 = n-1 - k;
        negy2 = n-2 - k;

        if (negy2≤0)
            break;
        end

        if ((A(m,negy2) - A(m,negy1))≥0)
            break;
        end

    end
    width_count_side2 = k + 1 ;
    edge_width_map(m,n) = width_count_side1+width_count
        _side2;
end
```

```
            end
        end
    end
end




%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    %%%%%
% function       : get_edge_blk_decision
% description    : Gives a decision whether the block is edge block or
    not.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    %%%%%

function [im_out] = get_edge_blk_decision(im_in,T)

[m,n] = size(im_in);
L = m*n;
im_edge_pixels = sum(sum(im_in));
im_out = im_edge_pixels > (L*T) ;




%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    %%%%%
% function       : get_contrast_block
% description    : Returns the contrast of the block.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    %%%%%

function contrast = get_contrast_block(A)

A = double(A);
[m,n] =size(A);
% get constrast locally
contrast = max(max(A)) - min(min(A));
```

H.11 cranepeak

```
function score = cranepeak(theimage)

    % H. D. Crane, "A theoretical analysis of the visual accommodation
    % system in humans" Tech. Rep. NASA CR-606, NASA - Ames Research,
    % Sep 1966.  http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/
    % 19660027855_1966027855.pdf on PDF page 41 or paper page 29:
    % To be definite, we assume a first spatial derivative over the
    % receptive field. The relevant measure is the integral of the
       absolute
    % magnitude of the derivative over the entire field. Suppose, for
    % example, that a defocused edge falls across the field. Then the
    % derivative will everywhere have the same polarity, and the
       integral
    % of the derivative over the field will (except for an arbitrary
    % constant) simply be equal to the total light intensity. If
       instead
    % of a single edge, however, we focused a bar across the field,
       then
    % the derivative would have opposite polarity at the two edges, and
    % integrating the absolute magnitude would result in a "measure of
    % derivative over the field" that would be twice the value obtained
    % with only a single edge.
    %
    % This implies an equation:
    % score = sum(sum(abs(diff(diff(theimage)'))));
    %
    % However, this does not work for the image:
    %   1   1   1   1
    %   1   1   1   1
    %   0   0   0   0
    %   0   0   0   0
    %
    % >> theimage = [1 1 1 1; 1 1 1 1; 0 0 0 0; 0 0 0 0];
    % >> score = sum(sum(abs(diff(diff(theimage)'))))
    %
    % score =
    %
    %      0
    %
    % >>
    %
```

```
    % As the first diff() function determines that there's no
       horizontal
    % difference, and so outputs an array of zeros.  The second diff()
       then
    % sees that the input is all zeros, so outputs zero.  ie score = 0.
    % When, in actual fact, we'd expect the score to be non-zero  This
       can
    % be found by instead computing the two-dimensional gradient, then
    % evaluating the euclidean distance at each point:
    %
    % >> theimage = [1 1 1 1; 1 1 1 1; 0 0 0 0; 0 0 0 0];
    % >> [px,py] = gradient(double(theimage))
    %
    % px =
    %
    %      0     0     0     0
    %      0     0     0     0
    %      0     0     0     0
    %      0     0     0     0
    %
    %
    % py =
    %
    %           0          0          0          0
    %    -0.5000   -0.5000   -0.5000   -0.5000
    %    -0.5000   -0.5000   -0.5000   -0.5000
    %           0          0          0          0
    %
    % >> score = sum(sum(sqrt(px.^2+py.^2)))
    %
    % score =
    %
    %      4
    %
    % >>
    %
    [px,py] = gradient(double(theimage));
    score = max(max(sqrt(px.^2+py.^2)));
end
```

H.12 cranesum

```
function score = cranesum(theimage)


    % H. D. Crane, "A theoretical analysis of the visual accommodation
    % system in humans" Tech. Rep. NASA CR-606, NASA - Ames Research,
    % Sep 1966.  http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/
    % 19660027855_1966027855.pdf on PDF page 41 or paper page 29:
    % To be definite, we assume a first spatial derivative over the
    % receptive field. The relevant measure is the integral of the
       absolute
    % magnitude of the derivative over the entire field. Suppose, for
    % example, that a defocused edge falls across the field. Then the
    % derivative will everywhere have the same polarity, and the
       integral
    % of the derivative over the field will (except for an arbitrary
    % constant) simply be equal to the total light intensity. If
       instead
    % of a single edge, however, we focused a bar across the field,
       then
    % the derivative would have opposite polarity at the two edges, and
    % integrating the absolute magnitude would result in a "measure of
    % derivative over the field" that would be twice the value obtained
    % with only a single edge.
    %
    % This implies an equation:
    % score = sum(sum(abs(diff(diff(theimage)'))));
    %
    % However, this does not work for the image:
    %   1   1   1   1
    %   1   1   1   1
    %   0   0   0   0
    %   0   0   0   0
    %
    % >> theimage = [1 1 1 1; 1 1 1 1; 0 0 0 0; 0 0 0 0];
    % >> score = sum(sum(abs(diff(diff(theimage)'))))
    %
    % score =
    %
    %      0
    %
    % >>
```

```
    %
    % As the first diff() function determines that there's no
       horizontal
    % difference, and so outputs an array of zeros.  The second diff()
       then
    % sees that the input is all zeros, so outputs zero.  ie score = 0.
    % When, in actual fact, we'd expect the score to be non-zero  This
       can
    % be found by instead computing the two-dimensional gradient, then
    % evaluating the euclidean distance at each point:
    %
    % >> theimage = [1 1 1 1; 1 1 1 1; 0 0 0 0; 0 0 0 0];
    % >> [px,py] = gradient(double(theimage))
    %
    % px =
    %
    %      0     0     0     0
    %      0     0     0     0
    %      0     0     0     0
    %      0     0     0     0
    %
    %
    % py =
    %
    %           0         0         0         0
    %    -0.5000   -0.5000   -0.5000   -0.5000
    %    -0.5000   -0.5000   -0.5000   -0.5000
    %           0         0         0         0
    %
    % >> score = sum(sum(sqrt(px.^2+py.^2)))
    %
    % score =
    %
    %      4
    %
    % >>
    %
    [px,py] = gradient(double(theimage));
    score = sum(sum(sqrt(px.^2+py.^2)));
end
```

H.13 dopowerplot

```
function [freq,scale_hist,orient,orient_hist,alpha,alphaintercept,dummy
    ]...
    =DoPowerPlot(image,nbins)

    % Source code provided by Steven Dakin.

    dummy = 0;

    a           = size(image);
    scale_hist  = zeros(1,nbins);
    freq        = zeros(1,nbins);
    orient      = zeros(1,nbins);
    orient_hist = zeros(1,nbins);

    f1=fftshift(abs(fft2(double(image))));
    [X,Y]=meshgrid(-a(1)/2:a(1)/2-1,-a(2)/2:a(2)/2-1);

    A=angle(X+Y.*i);
    B=abs(X+Y.*i);
    crit_angle_range=(2*pi)/nbins;
    crit_scale_range=(sqrt(2)*a(1)/2.0)/nbins;
    noctaves=log2(sqrt(2)*(a(1)/2));

    for j=1:nbins
        low_scale=2^(((j-1.0)/nbins) *noctaves);
        high_scale=2^(((j)/nbins) *noctaves);
        low_angle=(j-0.5)*crit_angle_range-pi;
        high_angle=low_angle+crit_angle_range;
        s1=((B <= high_scale)&(B>low_scale));
        s2=((A <= high_angle)&(A>low_angle));
        scale_sum=sum(sum(s1));
        scale_energy=sum(sum(s1.*f1));
        if (scale_sum>0)
            scale_hist(j)=scale_energy/scale_sum;
        end
        freq(j)=2^(((j)/nbins) *noctaves   );
        orient_sum=sum(sum(s2));
        orient_energy=sum(sum(s2.*f1));
        if (orient_sum>0)
            orient_hist(j)=orient_energy/orient_sum;
        end
```

```
    orient(j)=pi/2+(low_angle+high_angle)/2;
end

iScaleHist=interp1(scale_hist,8);
iFreq=2.^(([1:(8*nbins)]./(8*nbins)) .*noctaves   );

[maxVal maxLoc]=max(iScaleHist);
maxFreq=iFreq(maxLoc);

aboveT=find((iScaleHist./max(iScaleHist))>0.5);
fprintf('peak sf. %3.3f FWHH %3.3f\n',...
    maxFreq,iFreq(aboveT(end))./iFreq(aboveT(1)));

goodVals=find(scale_hist>0);

% Find alpha:
p1=polyfit(log(freq(goodVals)),log(scale_hist(goodVals)),1);
alpha = p1(1);
alphaintercept = p1(2);
fprintf('Slope parameter, alpha = %3.3f\n',alpha);
pred=polyval(p1,log(freq));

fweight=scale_hist(2)/(1/(freq(2).^1.5));

if nargout < 7,
    subplot(3,1,1);
    ishow(image);

    subplot(3,1,2);
    loglog(freq,scale_hist,'o');
    hold on;
    loglog(freq,exp(pred),'b-');
    hold off;

    subplot(3,1,3);
    plot(orient,orient_hist,'o-');
end;
```

H.14 energylaplace

```
function score = energylaplace(theimage)

    % Subbarao et al (1993) "Focusing Techniques"
    % Journal of Optical Engineering

    theimage = double(theimage);
    L = [-1 -4 -1; -4 20 -4; -1 -4 -1];
    S = conv2(theimage,L).^2;
    score = sum(sum(S));
end
```

H.15 energylaplace5a

```
function score = energylaplace5a(theimage)
    theimage = double(theimage);

    % http://www.ph.tn.tudelft.nl/¬imap/library/wouter/laplace5.html
    L =[ 0   0  -1   0   0
         0  -1  -2  -1   0
        -1  -2 +16  -2  -1
         0  -1  -2  -1   0
         0   0  -1   0   0];

    S = conv2(theimage,L).^2;
    score = sum(sum(S));
end
```

H.16 energylaplace5b

```
function score = energylaplace5b(theimage)

    % Energy laplace, plus a new 5x5 kernel.

    theimage = double(theimage);

    % http://rsb.info.nih.gov/nih-image/more-docs/macros.html
    L =[-1  -1  -1  -1  -1
        -1  -1  -1  -1  -1
        -1  -1  24  -1  -1
        -1  -1  -1  -1  -1
        -1  -1  -1  -1  -1];

    S = conv2(theimage,L).^2;
    score = sum(sum(S));
end
```

H.17 energylaplace5c

```
function score = energylaplace5c(theimage)

    % Energy laplace, plus a new 5x5 kernel.

    theimage = double(theimage);

    % http://rsb.info.nih.gov/nih-image/more-docs/macros.html
    L =[-1  -3  -4  -3  -1
        -3   0   6   0  -3
        -4   6  20   6  -4
        -3   0   6   0  -3
        -1  -3  -4  -3  -1];

    S = conv2(theimage,L).^2;
    score = sum(sum(S));
end
```

H.18 entropy

```
function score = entropy(I)

    % "Entropy Algorithm (Firestone et al., 1991). This algorithm
       assumes
    % that focused images contain more information than defocused
       images."
    % [1]
    %
    % References:
    % [1] "Autofocusing in Computer Microscopy: Selecting the Optimal
       Focus
    % Algorithm" by Sun et al in MICROSCOPY RESEARCH AND TECHNIQUE
    % 65:139149 (2004)
    %
    % score = - sum over all i [p(i) log2(p(i))]
    %
    % where p(i) = h(i) / (width*height)
    % and h(i) = probability of a pixel with intensity i

    score = 0;
    [width height] = size(I);
    for i=min(min(I)):max(max(I)),
        h = length(find(I==i));
        p = h / (width*height);
        if (p~=0) score = score + (p * log2(p));
    end;

    score = - score;

end
```

H.19 groenvariance

```
function score = groenvariance(I)

    % "Variance (Groen et al., 1985; Yeo et al., 1993). This algorithm
    % computes variations in gray level among image pixels. It uses the
    % power function to amplify larger differences from the mean
       intensity
    % mu instead of simply amplifying high-intensity values." [1]
    %
    % References:
    % [1] "Autofocusing in Computer Microscopy: Selecting the Optimal
       Focus
    % Algorithm" by Sun et al in MICROSCOPY RESEARCH AND TECHNIQUE
    % 65:139149 (2004)
    %
    % score = 1/(HW) * 2d sum of [I(x,y) - mean(I)]^2
    %

    [w h] = size(I);
    score = 1/(h*w) * sum(sum( (I - mean(mean(I))).^2 ));
end
```

H.20 histogramentropy

```
function score = histogramentropy(I)

    % Chern et al (2001) "Practical issues in pixel-based auto-
    % focusing for machine vision", Proceedings of the 2001 IEEE
    % International Conference of Robotics and Automation, Seoul 2001

    % "If the intensity histogram is, h(i), where h(i) is the frequency
    % of pixels of intensity i, then the histogram entropy is defined
    % as -sum(h(i)*ln(h(i))) if h(i)!=0"

    % Ensure the image is 8-bit (0-255) grey intensity.
    if max(I(:))<1, I = I * 255; end;

    % First: Determine number of pixels with each intensity, and store
    % in a histogram H

    h = hist(I(:),0:255);

    nonzerobins = find(h¬=0);

    score = sum(h(nonzerobins).*log(h(nonzerobins)));

end
```

H.21 hlv

```
function score = hlv(I)

    % Chern et al (2001) "Practical issues in pixel-based auto-
    % focusing for machine vision", Proceedings of the 2001 IEEE
    % International Conference of Robotics and Automation, Seoul 2001

    % "In the histogram of local variations, the intensity
    % histogram is evaluated with pixel intensities compressed
    % logarithmically and the gradient of the line of best fit through
    % the points, m, is evaluated. The quantity, m, is at a minimum
    % for the sharpest image. Since there are 256 gray levels, i = 0
    % to 255, sum[ln(i+1)] and sum[ln(i+1)]^2 is known and m may be
    % evaluated.

    % Ensure the image is 8-bit (0-255) grey intensity.
    if max(I(:))<1, I = I * 255; end;

    % First: Determine number of pixels with each intensity, and store
    % in a histogram H
    h = hist(I(:),0:255);

    % Construct the first term
    term1 = 0;
    for i = 1:256,
        term1 = term1 + log(i+1)*h(i);
    end;

    % Compute m
    m = (256 * term1 - 1167.26*sum(h)) / 60354.1;

    % Then, as m is a minimum for the sharpest image, invert it:
    score = 1/m;

    % Finally, as some scores were negative, shift the score up by 1.
    % All scores are normalised before comparison, so this has low
    % impact.
    score = score + 1;
end
```

H.22 imagepower

```
function score = imagepower(I)

    % "Image Power (Santos et al., 1997). This algorithm sums the
       square of
    % image intensities above a given threshold."
    % [1]
    %
    % "A threshold of 150 was set for focus algorithms (F-15 F-18)
       because
    % these algorithms exhibit satisfactory behavior with this
       threshold
    % value." [1]  I presume this is with 8 bit pixel values (ie 0-255)
       .
    %
    % References:
    % [1] "Autofocusing in Computer Microscopy: Selecting the Optimal
       Focus
    % Algorithm" by Sun et al in MICROSCOPY RESEARCH AND TECHNIQUE
    % 65:139149 (2004)
    %
    % score = sum over entire image where I(x,y)>threshold
    %

    % First, make the image into a 1-D array:
    [w h] = size(I);
    I = reshape(I,w*h,1);

    % Determine the thesold:
    if (max(max(I))<=1),
        % Assume this image is between 0-1, and so use a threshold of
        % 150/255:
        threshold = 150/255;
    else
        % Assume this is an 8 bit image
        threshold = 150;
    end;

    % Find the matching values:
    [r c v] = find(I>threshold);

    % Square the intensities:
```

```
    v = v.^2;

    % And sum the values:
    score = sum(v);
end
```

H.23 jnbm

```
%=========================================================================
% File: JNBM_compute.m
% Original code written by Rony Ferzli, IVU Lab (http://ivulab.asu.edu)
% Code modified by Lina Karam
% Last Revised: September 2009 by Lina Karam
%=========================================================================
% Copyright Notice:
% Copyright (c) 2007-2009 Arizona Board of Regents.
% All Rights Reserved.
% Contact: Lina Karam (karam@asu.edu) and Rony Ferzli (rony.ferzli@asu.
   edu)
% Image, Video, and Usabilty (IVU) Lab, http://ivulab.asu.edu
% Arizona State University
% This copyright statement may not be removed from this file or from
% modifications to this file.
% This copyright notice must also be included in any file or product
% that is derived from this source file.
%
% Redistribution and use of this code in source and binary forms,
% with or without modification, are permitted provided that the
% following conditions are met:
% - Redistribution's of source code must retain the above copyright
% notice, this list of conditions and the following disclaimer.
% - Redistribution's in binary form must reproduce the above copyright
% notice, this list of conditions and the following disclaimer in the
% documentation and/or other materials provided with the distribution.
% - The Image, Video, and Usability Laboratory (IVU Lab,
% http://ivulab.asu.edu) is acknowledged in any publication that
% reports research results using this code, copies of this code, or
% modifications of this code.
% The code and our papers are to be cited in the bibliography as:
% R. Ferzli and L. J. Karam, "JNB Sharpness Metric Software",
% http://ivulab.asu.edu
%
% R. Ferzli and L. J. Karam, "A No-Reference Objective Image Sharpness
% Metric Based on the Notion of Just Noticeable Blur (JNB)," IEEE
% Transactions on Image Processing, vol. 18, no. 4, pp. 717-728, April
% 2009.
%
% DISCLAIMER:
% This software is provided by the copyright holders and contributors
```

```
% "as is" and any express or implied warranties, including, but not
% limited to, the implied warranties of merchantability and fitness for
% a particular purpose are disclaimed. In no event shall the Arizona
% Board of Regents, Arizona State University, IVU Lab members, or
% contributors be liable for any direct, indirect, incidental, special,
% exemplary, or consequential damages (including, but not limited to,
% procurement of substitute goods or services; loss of use, data, or
% profits; or business interruption) however caused and on any theory
% of liability, whether in contract, strict liability, or tort
% (including negligence or otherwise) arising in any way out of the use
% of this software, even if advised of the possibility of such damage.
%
function [metric] = JNBM_compute(A)

% Robert Shilston's fix - ensure 8bit images are integers:
A = round(A);

%A = imfilter(A,1/9*ones(3,3));
beta = 3.6;
T = 0.002;
A = double(A);
[m,n] = size(A);
rb = 64;
rc = 64;
count = 1;

C=1:255;

widthjnb = [5*ones(1,60) 3*ones(1,30) 3*ones(1,180)];

for i=1:floor(m/rb)

    for j=1:floor(n/rc)
        row = rb*(i-1)+1:rb*i;
        col = rc*(j-1)+1:rc*j;
        A_temp = A(row,col);
        % check if block to be processed
        decision = get_edgeblocks_mod(A_temp,T);
        if (decision==1)
           local_width = edge_width(A_temp);
           Ac_meas = blkproc(A_temp,[rb rc],@get_contrast_block);
           Ajnb = widthjnb(Ac_meas+1);
           temp(count) = sum(abs(local_width./Ajnb).^beta).^(1/beta);
           count = count + 1;
         end
```

```
    end
end

blockrow = floor(m/rb);
blockcol = floor(n/rc);
L = blockrow*blockcol;
metric = (L/(sum(temp.^beta).^(1/beta)));

%
function [local] = edge_width(A)
% Compute edge width based on following paper:
% P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi,
% Perceptual blur and ringing metrics: Applications to JPEG2000,
% Signal Proc.: Image Comm., vol. 19, pp. 163172, 2004.


A = double(A);
E = edge(A,'Sobel',[],'vertical');
%E = edge(A,'Sobel');
[Gx Gy] = gradient(A);
% Magnitude
graA = abs(Gx) + abs(Gy);
[M N] = size(A);
 for m=1:M
     for n=1:N
            if (Gx(m,n)~=0)
                angle_A(m,n) = atan2(Gy(m,n),Gx(m,n))*(180/pi); % in
                   degrees
            end
            if (Gx(m,n)==0 && Gy(m,n)==0)
                angle_A(m,n) = 0;
            end
            if (Gx(m,n)==0 && Gy(m,n)==pi/2)
              angle_A(m,n) = 90;
            end
     end
 end

% quantize the angle
angle_Arnd = 45*round(angle_A./45);
width_loc = [];
count = 0;
for m=2:M-1
    for n=2:N-1
        if (E(m,n)==1)
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    %%%%%%
%%%%%%%%%%%%%%%%%%% If gradient = 0 %%%%%%%%%%%%%%%%%%%%%
    %%%%
%%%%%%%%%%%%%%%%%%% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    %%%
if (angle_Arnd(m,n) ==180 || angle_A(m,n) ==-180)
    count = count + 1;
        for k=0:100
        posy1 = n-1 -k;
        posy2 = n-2 -k;
        if ( posy2<=0)
            break;
        end

        if ((A(m,posy2) - A(m,posy1))<=0)
            break;
        end
    end
    width_count_side1 = k + 1 ;
    for k=0:100
        negy1 = n+1 + k;
        negy2 = n+2 + k;

        if (negy2>N)
            break;
        end

        if ((A(m,negy2) - A(m,negy1))>=0)
            break;
        end

    end
    width_count_side2 = k + 1 ;
    width_loc = [width_loc width_count_side1+width_count_
       side2];
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    %%%%%%
%%%%%%%%%%%%%%%%%%% If gradient = 0 %%%%%%%%%%%%%%%%%%%%%
    %%%%
%%%%%%%%%%%%%%%%%%% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    %%%
if (angle_Arnd(m,n) ==0)
```

```
                count = count + 1;
                    for k=0:100
                    posy1 = n+1 +k;
                    posy2 = n+2 +k;
                    if ( posy2>N)
                        break;
                    end

                    if ((A(m,posy2) - A(m,posy1))≤0)
                        break;
                    end
                end
                width_count_side1 = k + 1 ;
                for k=0:100
                    negy1 = n-1 - k;
                    negy2 = n-2 - k;

                    if (negy2≤0)
                        break;
                    end

                    if ((A(m,negy2) - A(m,negy1))≥0)
                        break;
                    end

                end
                width_count_side2 = k + 1 ;
                width_loc = [width_loc width_count_side1+width_count_
                    side2];
            end


        end
    end
end


local = width_loc;

%
function im_out = get_edgeblocks_mod(im_in,T)

im_in = double(im_in);

[im_in_edge,th] = edge(im_in,'canny');
```

```
[m,n] = size(im_in_edge);
L = m*n;
im_edge_pixels = sum(sum(im_in_edge));
im_out = im_edge_pixels > (L*T) ;

%
function contrast = get_contrast_block(A)

A = double(A);
[m,n] =size(A);
contrast = max(max(A)) - min(min(A));
```

H.24 laplace

```
function score = laplace(theimage)

    % Chern et al (2001) "Practical issues in pixel-based auto-
    % focusing for machine vision", Proceedings of the 2001 IEEE
    % International Conference of Robotics and Automation, Seoul 2001
    %
    % Whilst the paper doesn't explicitly state that this is
    % measure is a convolution, its context implies that this
    % is the case.
    %

    theimage = double(theimage);
    [w h] = size(theimage);
    T = 0;

    L = 1/6 * [1 4 1; 4 -20 4; 1 4 1];
    S = conv2(theimage,L);
    score = 1/(w*h) * sum(sum(S(find(S>T))));
end
```

H.25 masgrn

```
function score = masgrn(I)

    % "Mason and Green's histogram method differs from the Mendelsohn
       and
    % Mayall histogram method in the way the threshold is selected.
       They
    % weighed the importance of picturepoints by estimates of the
       gradient
    % at that point." [1]
    %
    % References:
    % [1] "Comparison of autofocus methods for automated microscopy" by
    % Firestone et al in CYTOMETRY 12:195-206 (1991)
    % [2] "Automatic Focusing of a Computer-Controlled Microscope" by
    % Mason and Green IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING,
    % VOL. BME-22, NO. 4, JULY 1975.
    %
    % This implementation uses equations 10-12 from [1].

    % Ensure the image is 8-bit (0-255) grey intensity.
    if max(I(:))<1, I = I * 255; end;

    % First: Compute the 'importance' of each pixel, except for the
       edge
    % pixels where importance is not defined.
    [width height]=size(I);
    importance = zeros(size(I));
    for i=1+1:width-1,
        for j=1+1:height-1,
            term1 = 2 * (I(i,j-1)-I(i,j+1))^2;
            term2 = 2 * (I(i-1,j)-I(i+1,j))^2;
            term3 = (I(i-1,j-1)-I(i+1,j+1))^2;
            term4 = (I(i-1,j+1)-I(i+1,j-1))^2;
            importance(i,j) = term1 + term2 + term3 + term4;
        end;
    end;

    % Then determine the threshold:
    T = sum(sum(importance.*I))/sum(sum(importance));

    % Build the histogram:
```

```
    H = hist(I(:),0:255);

    % Finally, the score:
    score = 0;
    for k=ceil(T):255,
        score = score + H(k) * (k-T);
    end;
end
```

H.26 menmay

```
function score = menmay(I)

    % "Mendelsohn and Mayall's histogram method.  The focus function is
    % found by computing the weighted sum of picturepoints in histogram
    % bins that are above a given threshold." [1]
    %
    % References:
    % [1] "Comparison of autofocus methods for automated microscopy" by
    % Firestone et al in CYTOMETRY 12:195-206 (1991)
    % [2] "Computer-oriented analysis of human chromosomes - III Focus"
    % by Mendelsohn and Mayall in Comput. Biol. Med. 2:137-150.
    %
    % This implementation uses equation 9 from [1].

    % Ensure the image is 8-bit (0-255) grey intensity.
    if max(I(:))<1, I = I * 255; end;

    % First: Determine number of pixels with each intensity, and store
    % in a histogram H
    H = hist(I(:),0:255);

    % Second: Compute mean intensity of the image, rounding up to the
    % nearest integer, and store this as T.
    T = ceil(mean(I(:)));

    % score = sum of multiplying the bin index with the number of
    % pixels in that bin, for all bins > T.
    score = 0;
    for k=(T+1):255,
        score = score + k * H(k);
    end;

end
```

H.27 normalizedgroenvariance

```
function score = normalizedgroenvariance(I)

    % "Normalized Variance (Groen et al., 1985; Yeo et al., 1993). By
    % normalizing the final output with the mean intensity mu, this
    % algorithm compensates for the differences in average image
       intensity
    % among different images." [1]
    %
    % References:
    % [1] "Autofocusing in Computer Microscopy: Selecting the Optimal
       Focus
    % Algorithm" by Sun et al in MICROSCOPY RESEARCH AND TECHNIQUE
    % 65:139149 (2004)
    %
    % score = 1/(H*W*mean(I)) * 2d sum of [I(x,y) - mean(I)]^2
    %

    [w h] = size(I);
    score = 1/(h*w*mean(mean(I))) * sum(sum( (I - mean(mean(I))).^2 ));
end
```

H.28 phasecongruence

```
function score = phasecongruence(I)

    % 2d summation of phasecongruency metric proposed by Kovesi, using
    % original parameters [1, section 9.1].

    % Depends on functions from
    % http://mitpress.mit.edu/e-journals/Videre/001/articles/Kovesi/

    [pc orient ft] = phasecong(I);
    nonmax = nonmaxsup(pc, orient, 1.5);
    features = hysthresh(nonmax, 0.3, 0.15);
    score = sum(sum(features));

end
```

H.29 phasecongruence2

```
function score = phasecongruence2(I)

    % 2d summation of phasecongruency metric proposed by Kovesi,
       ommitting
    % the original thresholding parameters parameters [1, section 9.1].

    % Depends on functions from
    % http://mitpress.mit.edu/e-journals/Videre/001/articles/Kovesi/

    [pc orient ft] = phasecong(I);
    nonmax = nonmaxsup(pc, orient, 1.5);
    score = sum(sum(nonmax));

end
```

H.30 randomnumber

```
function score = randomnumber(theimage)

    % This simply returns a random number.

    score = rand;
end
```

H.31 range

```
function score = range(I)

    % "Range Algorithm (Firestone et al., 1991). This algorithm
       computes
    % the difference between the highest and the lowest intensity
       levels."
    % [1]
    %
    % References:
    % [1] "Autofocusing in Computer Microscopy: Selecting the Optimal
       Focus
    % Algorithm" by Sun et al in MICROSCOPY RESEARCH AND TECHNIQUE
    % 65:139149 (2004)
    %
    % [2] "Comparison of autofocus methods for automated microscopy",
       by
    % Firestone et al in Cytometry, vol 12, 3:195-206, 1991
    %
    % score = max(I) - min(I)
    %

    score = max(max(I)) - min(min(I));
end
```

H.32 rawlaplace

```
function score = rawlaplace(theimage)
    theimage = double(theimage);
    L = [-1 -4 -1; -4 21 -4; -1 -4 -1];
    S = conv2(theimage,L);
    score = sum(S(:));
end
```

H.33 rmscontrast

```
function score = rmscontrast(theimage)

    % RMS Contrast
    %
    % Equations 4a and 4b from  E. Peli (Oct. 1990). "Contrast in
       Complex
    % Images". Journal of the Optical Society of America A 7 (10):
    % pp 2032 2040. doi:10.1364/JOSAA.7.002032.

    [w h] = size(theimage);

    % Mean (eq 4b)
    xmean = mean(theimage(:));

    % RMS contrast (eq 4a: rms = ((1/(n-1)) * sum((x - xmean)^2) )^2)
    % First, compute the centre of the summation:
    sumvals = (theimage - xmean).^2;
    sumresult = sum(sumvals(:));
    score = ((1 / w*h) * sumresult)^0.5;
end
```

H.34 scorefocusmeasure

```
function [A R Rmax Rscoring F W N] = ...
    scorefocusmeasure(score, groundtruthindex),

    % Prune the input data, in case we've a dual peak.
    groundtruthindex = floor(mean(groundtruthindex));
    peakindex = floor(mean(find(score==max(score(:)))));

    % Ensure that the score is scaled 0-1
    score(find(score==0))=nan;
    activerange = length(score) - length(find(isnan(score)));
    scalefactor = (max(score)-min(score));
    if scalefactor==0, scalefactor=1; end;
    score = (score-min(score))/scalefactor;
    score(find(isnan(score)))=0;

    % * Accuracy (A): Distance between maxima of the focus curve and
       the
    % ground truth of `best' image, measured in number of image frames
       of
    % distance.
    A = abs(peakindex-groundtruthindex);

    % * Range (R): The distance (in number of images) between the first
    % minima on either side of the global maxima. This should be large,
        as
    % there should not be any local maxima on the focus curve.

    % First, find the near side range:
    nearrange=inf;
    for i = peakindex:-1:2,
        if score(i-1) > score(i),
            nearrange = i;
            break;
        end;
        if score(i) == 0,
            nearrange = i;
            break;
        end;
    end;
    if nearrange==inf,
        nearrange = peakindex;
```

```
else,
    nearrange = peakindex - nearrange;
end;

% Then the far side range:
farrange=inf;
for i = (peakindex+1):(length(score)-1),
    if score(i+1) > score(i),
        farrange = i;
        break;
    end;
    if score(i) == 0,
        farrange = i;
        break;
    end;
end;
if farrange==inf,
    farrange = length(score) - peakindex;
else,
    farrange = farrange - peakindex;
end;

R = farrange + nearrange;
Rmax = activerange;
Rscoring = Rmax - R;

% * Number of false maxima (F): The number of maxima appearing in a
% focus cuve, excluding the global maximum.
F = 0;
for i = 2:length(score)-1,
    if i¬=peakindex,
        if score(i)>score(i-1) & score(i)>score(i+1),
            F = F + 1;
        end;
    end
end

% * Width (W): The width of the curve (in number of images) at 50%
   of
% the maxima's height. Ideally this should be small.
maximaheight = score(peakindex);
I = find(score > (maximaheight/2));
if length(I),
    W = max(I) - min(I);
else
```

```
        W = 1;
    end;


    % * Noise level (N): This describes the speed of the direction of
    % change between two false maxima of a focus curve. It is computed
       by
    % taking the sum of squares of the second derivatives obtained by
    % convolving the curve (ommitting the peak value) with the kernel
    % (-1; 2;-1).
    trimmedscore = [score(1:peakindex-1) score(peakindex+1:end)];
    derivatives = conv(trimmedscore, [-1 2 -1]);
    N = sum(derivatives.^2);

    % NB. The score cannot be computed here, as we need to normalise
       across
    % all the images being assessed within this group.
    % S = sqrt(A^2 + (length(score) - R)^2 + W^2 + N^2 + F^2);

end
```

H.35 setalpha

```
function [outputim,newalpha]=SetAlpha(im,alpha)

    % This was the script that resulted after writing ExploreAlpha. The
    % matrix Z was discovered by trial and error, with a little bit of
    % thought.
    %
    % If this starts failing to set alpha to a particular desired value
       ,
    % then further analysis with ExploreAlpha will be required.
    % Specifically it is like that the equation y=mx+c might not hold
       for
    % the parameter variations, or m might be different and so rrr is
    % being erroneously calculated.

    if (size(im)~=[512 512]),
        error('This function is only implemented for 512x512 images');
    end;

    % Get the image in, into the frequency domain
    A = im;
    [dummy, dummy, dummy, dummy, oldalpha, dummy, dummy] ...
        = DoPowerPlot(A,32);

    Afft = fft2(double(A));
    Affts = fftshift(Afft);
    Affts_abs = abs(Affts);
    Affts_angle = angle(Affts);

    % Work out the parameter by which to adjust the image.  This
       equation
    % was determined by measuring how alpha varied with rrr for two
       images
    % ('farm' and Van Hateren's 'imk00005').
    % Farm was approx 0.995, and VH5 was around 0.987, but these were
    % determined by reading from graphs, so it's likely that they're
       the
    % same.
    rrr = (alpha - oldalpha)/0.995;

    % Build a matrix to manipulate it:
    [X,Y]=meshgrid(linspace(-1,1,512));
```

```
    Z = sqrt(X.^2+Y.^2);
    Z = Z.^rrr;

    % Ensure that the DC component doesn't change, and normalise the
       matrix
    % magnitude:
    Z(257,257) = 0;
    Z = Z./max(Z(:));
    Z(257,257) = 1;

    % Now apply the scaling:
    NA = Z.*Affts_abs;
    Nffts = NA .* (cos(Affts_angle)+i*sin(Affts_angle));
    Nfft = ifftshift(Nffts);
    N = ifft2(Nfft);
    N = real(N); % remove any residuals.
    N = uint8(N);

    [dummy, dummy, dummy, dummy, newalpha, dummy, dummy] ...
        = DoPowerPlot(N,32);
    outputim = N;

end
```

H.36 smd

```
function score = smd(theimage)

    % Chern et al (2001) "Practical issues in pixel-based auto-
    % focusing for machine vision", Proceedings of the 2001 IEEE
    % International Conference of Robotics and Automation, Seoul 2001

    ydiff = diff(theimage);
    xdiff = diff(theimage');

    score = sum(abs(ydiff(:))) + sum(abs(xdiff(:)));

end
```

H.37 sml

```
function score = sml(I)

    % [1] eq F-5, defines this as:
    % score = sum over height ( sum over width ( abs(Lx(x,y))+abs(Ly(x,
       y) )
    %
    % but the original paper [2, equations 7-11] define SML as:
    %
    % ML(x,y) = abs(2I(x,y) - I(x-step,y) - I(x+step,y)) +
    %             abs(2I(x,y) - I(x,y-step) - I(x,y+step))
    %
    % then the focus at a point is defined by the sum of the modified
    % laplacian around some small window:
    %
    % F(i,j) = Sum from x=i-N to x=i+N of
    %           Sum from y=j-N to y=j+N of
    %             ML(x,y) where ML(x,y)≥T1
    % where N = 1 or 2.
    %
    % Nayar [2, p14] shows that SML better than Tenengrad in textured
    % images, and that Tenengrad computed with T=1, SML with T1=7, step
       =1.
    % However, the paper doesn't explicitly say which N to use.  This
       is
    % not a significant problem for the use in this application, as we'
       re
    % looking for a global score, not a score in a particular location.
    % As such, the equation that's implemented here is:
    %
    %  score = sum over all points of
    %            ML(x,y) where ML(x,y)≥T1
    %
    % References:
    % [1] "Autofocusing in Computer Microscopy: Selecting the Optimal
       Focus
    % Algorithm" by Sun et al in MICROSCOPY RESEARCH AND TECHNIQUE
    % 65:139149 (2004)
    %
    % [2] "Shape from Focus", Shree Nayar, 1989,
    % http://www1.cs.columbia.edu/CAVE/publications/pdfs/Nayar_TR89.pdf
    %
```

```
    step = 1;
    T1 = 7;

    [width height] = size(I);
    ML = zeros(width,height);
    for x=(1+step):(width-step),
        for y=(1+step):(height-step),
            term1 = abs(2*I(x,y)-I(x-step,y)-I(x+step,y));
            term2 = abs(2*I(x,y)-I(x,y-step)-I(x,y+step));

            ML(x,y) = term1 + term2;
        end;
    end;

    score = sum(sum(find(ML≥T1)));
end
```

H.38 squaredgradient

```
function score = squaredgradient(I, T)

    % "Evaluation of autofocus functions in molecular cytogenetic
       analysis"
    % by Santos et al in Journal of Microscopy Vol.188 Issue 03, pp 264
       -272
    % http://dx.doi.org/10.1046/j.1365-2818.1997.2630819.x
    %
    % score = sum over height
    %             sum over width
    %                abs(i(x+1,y)-i(x,y))^2
    % where abs(i(x+1,y)-i(x,y))^2 > T
    %
    % Table 1 shows that a threshold of 25 is used.

    step = 1;
    T1 = 25;

    [width height] = size(I);
    ML = zeros(width,height);
    for x=(1+step):(width-step),
        for y=(1+step):(height-step),
            ML(x,y) = abs(I(x+step,y)-I(x,y))^2;
        end;
    end;

    score = sum(sum(find(ML≥T1)));
end
```

H.39 standarddeviationbasedautocorrelation

```
function score = standarddeviationbasedautocorrelation(I)

    % "Standard Deviation-Based Correlation (Vollath, 1987, 1988)." [1]
    %
    % References:
    % [1] "Autofocusing in Computer Microscopy: Selecting the Optimal
       Focus
    % Algorithm" by Sun et al in MICROSCOPY RESEARCH AND TECHNIQUE
    % 65:139149 (2004)
    %
    % score = sum over all image of i(x, y) . i(x + 1, y)
    %                                    - H * W * mean(i)^2
    %

    [w h] = size(I);
    score = 0;
    for i=1:w-2,
        for j=1:h,
            score = score + (I(i,j)*I(i+1,j));
        end;
    end;
    score = score - h*w*mean(mean(I))^2;
end
```

H.40 tenengrad

```
function score = tenengrad(theimage)

    % The method is to estimate the gradient VI(x,y)
    % at each image point (x,y), and simply to sum all
    % the magnitudes greater than a threshold value.
    % The gradient magnitude is
    %
    % | ∆ I(x,y) | = sqrt(Ix^2+Iy^2).
    %
    % The partials can be estimated by many discrete
    % operators. We employ the Sobel operator with the
    % convolution kernels
    %
    %     ix = 0.25 * [-1 0 1; -2 0 2; -1 0 1];
    %     iy = 0.25 * [1 2 1; 0 0 0; -1 -2 -1];
    %
    % We compute the gradient magnitude as
    %
    % S(x,y) = sqrt([ix   I(x,y)]^2 + [iy   I(x,y)]^2)
    %
    % and state the criterion function as
    %
    % max of 2d summation of S(x,y)^2, for S(x,y)>T
    %
    % where T is a threshold [1]
    %
    %
    % It is noted [2, p265] that "Although this function made use of a
    % threshold in its initial form, following Krotkov (1987) no
       threshold
    % is proposed".
    %
    % [1] "Focusing", Krotkov in Journal of Computer Vision, Vol 1,
    % pp 223-237, 1987
    %
    % [2] "Evaluation of autofocus functions in molecular cytogenetic
    % analysis" by Santos et al in Journal of Microscopy Vol.188 Issue
       03,
    % pp 264-272 http://dx.doi.org/10.1046/j.1365-2818.1997.2630819.x
    %
```

```
    thethreshold=0; % several papers don't use a threshold

    theimage = double(theimage);
    ix = 0.25 * [-1 0 1; -2 0 2; -1 0 1];
    iy = 0.25 * [1 2 1; 0 0 0; -1 -2 -1];

    % We should do:
    % S = sqrt(conv2(theimage(:,:),ix).^2+conv2(theimage(:,:),iy).^2);
    % score = sum(sum(find(S>thethreshold).^2));

    % But, it's more efficient not to sqrt, but instead compare against
        a
    % squared threshold:
    thethreshold = thethreshold^2;
    S = conv2(theimage(:,:),ix).^2+conv2(theimage(:,:),iy).^2;

    if thethreshold==0,
        score = sum(sum(S.^2));
    else
        error('score = sum(sum(find(S>thethreshold).^2));');
    end;
end
```

H.41 thresholdedabsolutegradient

```
function score = thresholdedabsolutegradient(I, T)

    % "Evaluation of autofocus functions in molecular cytogenetic
       analysis"
    % by Santos et al in Journal of Microscopy Vol.188 Issue 03, pp 264
       -272
    % http://dx.doi.org/10.1046/j.1365-2818.1997.2630819.x
    %
    % score = sum over height
    %             sum over width
    %                abs(i(x+1,y)-i(x,y))
    % where abs(i(x+1,y)-i(x,y)) > T
    %
    % Table 1 shows that a threshold of 5 is used.

    step = 1;
    T1 = 5;

    [width height] = size(I);
    ML = zeros(width,height);
    for x=(1+step):(width-step),
        for y=(1+step):(height-step),
            ML(x,y) = abs(I(x+step,y)-I(x,y));
        end;
    end;

    score = sum(sum(find(ML≥T1)));
end
```

H.42 thresholdedcontent

```
function score = thresholdedcontent(I)

    % "Thresholded Content (Groen et al., 1985; Mendelsohn and Mayall,
    % 1972). This algorithm sums the pixel intensities above a
       threshold."
    % [1]
    %
    % "A threshold of 150 was set for focus algorithms (F-15 F-18)
       because
    % these algorithms exhibit satisfactory behavior with this
       threshold
    % value." [1]  I presume this is with 8 bit pixel values (ie 0-255)
       .
    %
    % References:
    % [1] "Autofocusing in Computer Microscopy: Selecting the Optimal
       Focus
    % Algorithm" by Sun et al in MICROSCOPY RESEARCH AND TECHNIQUE
    % 65:139149 (2004)
    %
    % score = sum over entire image where I(x,y)>threshold
    %

    % First, make the image into a 1-D array:
    [w h] = size(I);
    I = reshape(I,w*h,1);

    % Determine the thesold:
    if (max(max(I))<=1),
        % Assume this image is between 0-1, and so use a threshold of
        % 150/255:
        threshold = 150/255;
    else
        % Assume this is an 8 bit image
        threshold = 150;
    end;

    % Find the matching values:
    [r c v] = find(I>threshold);

    % And sum the values:
```

```
    score = sum(v);
end
```

H.43 thresholdedpixelcount

```
function score = thresholdedpixelcount(I)

    % "Thresholded Pixel Count (Groen et al., 1985). This algorithm
       counts
    % the number of pixels having intensity below a given threshold."
    % [1]
    %
    % "A threshold of 150 was set for focus algorithms (F-15 F-18)
       because
    % these algorithms exhibit satisfactory behavior with this
       threshold
    % value." [1]  I presume this is with 8 bit pixel values (ie 0-255)
       .
    %
    % References:
    % [1] "Autofocusing in Computer Microscopy: Selecting the Optimal
       Focus
    % Algorithm" by Sun et al in MICROSCOPY RESEARCH AND TECHNIQUE
    % 65:139149 (2004)
    %
    % score = sum over entire image of s(x,y)
    %
    % where s(x,y) = 1 when i(x,y)<threshold; 0 otherwise

    % First, make the image into a 1-D array:
    [w h] = size(I);
    I = reshape(I,w*h,1);

    % Determine the thesold:
    if (max(max(I))≤1),
        % Assume this image is between 0-1, and so use a threshold of
        % 150/255:
        threshold = 150/255;
    else
        % Assume this is an 8 bit image
        threshold = 150;
    end;

    % Find the matching values:
    matchingindexes = find(I<threshold);
```

```
    % And count the matching pixels
    score = length(matchingindexes);
end
```

H.44 triakis11s

```
function score = triakis11s(I)

    % References:
    % [1] "Comparison of autofocus methods for automated microscopy" by
    % Firestone et al in CYTOMETRY 12:195-206 (1991)

    % Ensure the image is 8-bit (0-255) grey intensity.
    if max(I(:))<1, I = I * 255; end;

    % Resize to 64x64 (as used by [1]).  First, square the image:
    [d1 d2]=size(I);
    if d2>d1,
        offset = floor((d2-d1)/2);
        I = I(1:d1, (1+offset):(offset+d1));
    else
        offset = floor((d1-d2)/2);
        I = I((1+offset):(offset+d2), 1:d2);
    end;

    % Then sample the image down to 64x64
    I = imresize(I,[64 64]);

    % Determine boundaries:
    [width height] = size(I);

    % Create and populate a 3D binary space:
    v = zeros(width,height,255,'uint8');
    filterop = logical(v);
    v = logical(v);

    % Fill the space:
    for i = 1:width,
        for j = 1:height,
            if I(i,j),
                v(i,j,1:ceil(I(i,j))) = true;
            end;
        end;
    end;


    % For triakis11, we're looking to pass neighbourhoods which have
```

```
% 11 voxels set.


% However, [1] used a face-centred cubic close pack,
% meaning that each voxel had 12 neighbouring pixels (a
% ring of six around it in one plane, then three sitting
% above and beneath).  No explanation was made as to how
% the cartesian-image was transformed into a hexagonal
% packed arrangement for searching.
%
% This implementation pushes the cartesian voxel
% arrangement slightly, so that it becomes hexagonal.
% Then, for an arbitrary point (i,j,k), the neighbours are:
%
% Beneath (k-1)
% NB. o denotes the site of the original pixel in the adjacent
   plane.
%
%                (i,j,k-1)
%                    o
%       (i,j-1,k-1)      (i+1,j-1,k-1)
%
%
%
% Same level (k):
% NB. x denotes site of neighbours in adjacent planes
%
%      (i-1,j+1,k)    (i,j+1,k)
%                  x2
% (i-1,j,k)    ORIGINAL   (i+1,j,k)
%          x1           x3
%      (i,j-1,k)     (i+1,j-1,k)
%
%
% Above (k+1)
% NB. o denotes the site of the original pixel in the adjacent
   plane.
%
%                (i,j,k+1)
%                    o
%       (i,j-1,k+1)      (i+1,j-1,k+1)
%

% Create a mask space:
neighbourmask = zeros(3,3,3);
```

```
    % Define centre of mask:
    i=2; j=2; k=2;

    % Set mask based on the above voxel selection:
    neighbourmask(i,j,k-1)=1;
    neighbourmask(i,j-1,k-1)=1;
    neighbourmask(i+1,j-1,k-1)=1;

    neighbourmask(i-1,j+1,k)=1;
    neighbourmask(i,j+1,k)=1;
    neighbourmask(i-1,j,k)=1;
    neighbourmask(i+1,j,k)=1;
    neighbourmask(i,j-1,k)=1;
    neighbourmask(i+1,j-1,k)=1;

    neighbourmask(i,j,k+1)=1;
    neighbourmask(i,j-1,k+1)=1;
    neighbourmask(i+1,j-1,k+1)=1;

    % Search the space for matching criteria:
    for i=2:width-1,
        for j=2:height-1,
            for k=2:254,
                % Extract region:
                extract = v(i-1:i+1,j-1:j+1,k-1:k+1);

                % Reduce from 26 neighbour voxels to 12:
                extract = extract .* neighbourmask;

                % See if this is a matching neighbourhood:
                if sum(extract(:))==11,

                    % Merge this neighbour into the filter output:
                    term1 = filterop(i-1:i+1,j-1:j+1,k-1:k+1);
                    term2 = v(i-1:i+1,j-1:j+1,k-1:k+1);
                    filterop(i-1:i+1,j-1:j+1,k-1:k+1) = term1 | term2 ;
                end;
            end;
        end;
    end;

    % Finally, determine the score:
    score = sum(filterop(:));
end
```

H.45 va

```
function score = va(theimage)

    % Convert the input into 24 bit colour, but still greyscale:
    if max(theimage(:))>1, theimage = theimage ./ 256; end;
    theimage24(:,:,1) = theimage;
    theimage24(:,:,2) = theimage;
    theimage24(:,:,3) = theimage;

    % First: write the image to disk as a BMP:
    vafn = ['va' num2str(floor(rand*100000)) '.bmp'];

    imwrite(theimage24, vafn, 'BMP');

    % And compute the score:
    theexe = 'C:\Docume¬1\RobShi¬1\Desktop\Phd\TransferReport\
       focusmeasures\va.exe ';
    theoptions = ' --tests=2 --colourcomparison=eucrgb --showprogress=
       no';
    thecmd = [ theexe theoptions ' ' vafn];
    [stat result]=system(thecmd);
    score = str2num(result);
    delete(vafn)
end
```

H.46 voll4

```
function score = voll4(I)

    % References:
    % "Evaluation of autofocus functions in molecular cytogenetic
       analysis"
    % by Santos et al in Journal of Microscopy, Vol 188(3), pp264-272,
    % December 1997.

    % Ensure the image is 8-bit (0-255) grey intensity.
    if max(I(:))<1, I = I * 255; end;

    % Determine boundaries:
    [width height] = size(I);

    % Initialise variables:
    score = 0;

    % Part one of the equation:
    for i = 1:(width-1),
        for j = 1:height,
            score = score + I(i,j)*I(i+1,j);
        end;
    end;

    % Part two of the equation:
    for i = 1:(width-2),
        for j = 1:height,
            score = score - I(i,j)*I(i+2,j);
        end;
    end;

end
```

H.47 voll5

```
function score = voll5(I)

    % References:
    % "Evaluation of autofocus functions in molecular cytogenetic
       analysis"
    % by Santos et al in Journal of Microscopy, Vol 188(3), pp264-272,
    % December 1997.

    % Ensure the image is 8-bit (0-255) grey intensity.
    if max(I(:))<1, I = I * 255; end;

    % Determine boundaries:
    [width height] = size(I);

    % Initialise variables:
    score = 0;

    for i = 1:(width-1),
        for j = 1:height,
            score = score + I(i,j)*I(i+1,j);
        end;
    end;

    score = score - width*height*(mean(I(:)).^2);

end
```

H.48 waveletw1

```
function score = waveletw1(I)

    % "This algorithm uses the Daubechies D6 wavelet filter, applying
       both
    % high-pass (H) and lowpass (L) filtering to an image. The
       resultant
    % image is divided into four subimages: LL, HL, LH, and HH. The
    % algorithm sums the absolute values in the HL, LH, and HH regions.
       "
    %
    % "Autofocusing in Computer Microscopy: Selecting the Optimal Focus
    % Algorithm" by Sun et al in MICROSCOPY RESEARCH AND TECHNIQUE
    % 65:139149 (2004)

    % First, install wavelab

    I = double(I);
    I = makedyadic(I);
    d6 = MakeONFilter('Daubechies', 6);
    wlI = FWT2_PO(I,5,d6);
    [w h] = size(wlI);
    ll = wlI(1:w/2, 1:h/2);
    lh = wlI(1:w/2, h/2:h);
    hl = wlI(w/2:w, 1:h/2);
    hh = wlI(w/2:w, h/2:h);

    score = sum(sum(abs(hl))) + sum(sum(abs(lh))) + sum(sum(abs(hh)));
end
```

H.49 waveletw2

```
function score = waveletw2(I)

    % "This algorithm sums the variances in the HL, LH, and HH regions.
        The
    % mean values, mu, in each region are computed from absolute values
       ."
    %
    % "Autofocusing in Computer Microscopy: Selecting the Optimal Focus
    % Algorithm" by Sun et al in MICROSCOPY RESEARCH AND TECHNIQUE
    % 65:139149 (2004)
    %
    % "Wavelet-Based Autofocusing and Unsupervised Segmentation of
    % Microscopic Images" by Yang and Nelson in Proceedings of the 2003
    % IEEE/RSJ lntl Conference on Intelligent Robots and Systems, Las
       Vegas
    % Nevada October 2003, p2144 -- 2148.
    %
    % NB. Equation F8 of Sun et al appears to contain a typographical
       error
    % when compared with Equation 5 of Yang and Nelson.
    %
    % score = 1/(HW) * 2d sum of
    %                      [abs(HL) - mean(abs(HL))]^2 +
    %                      [abs(LH) - mean(abs(LH))]^2 +
    %                      [abs(HH) - mean(abs(HH))]^2
    %
    % Furthermore, Sun indicates the summation is over the height and
       width
    % of the image, whilst Yang sums over the set of i,j present in the
    % arrays HL, LH and HH respectively.  It's this latter approach
       which
    % is implemented.  It should also be noted that [1]'s equation F8
       is
    % simply not present in [2] - it looks to be a later version of Eq
       5.
    %
    % This requires Wavelab (http://www-stat.stanford.edu/~wavelab/)

    I = double(I);
    I = makedyadic(I);
    d6 = MakeONFilter('Daubechies', 6);
```

```
    wlI = FWT2_PO(I,5,d6);
    [w h] = size(wlI);
    ll = wlI(1:w/2, 1:h/2);
    lh = wlI(1:w/2, h/2:h);
    hl = wlI(w/2:w, 1:h/2);
    hh = wlI(w/2:w, h/2:h);

    hlsum = sum(sum((abs(hl) - mean(mean(abs(hl)))).^2));
    lhsum = sum(sum((abs(lh) - mean(mean(abs(lh)))).^2));
    hhsum = sum(sum((abs(hh) - mean(mean(abs(hh)))).^2));

    score = 1/(h*w) * (hlsum + lhsum + hhsum);
end
```

H.50 waveletw3

```
function score = waveletw3(I)

    % "This algorithm differs from [waveletw2] in that the mean values,
        mu,
    % in the HL, LH, and HH regions are computed without using absolute
    % values." [1]
    %
    % References:
    % [1] "Autofocusing in Computer Microscopy: Selecting the Optimal
       Focus
    % Algorithm" by Sun et al in MICROSCOPY RESEARCH AND TECHNIQUE
    % 65:139149 (2004)
    %
    % [2] "Wavelet-Based Autofocusing and Unsupervised Segmentation of
    % Microscopic Images" by Yang and Nelson in Proceedings of the 2003
    % IEEE/RSJ lntl Conference on Intelligent Robots and Systems, Las
       Vegas
    % Nevada October 2003, p2144 -- 2148.
    %
    % NB. Equation F8 of Sun et al appears to contain a typographical
       error
    % when compared with Equation 5 of Yang and Nelson.  This uses the
       Yang
    % formula.
    %
    % score = 1/(HW) * 2d sum of
    %                      [HL - mean(HL)]^2 +
    %                      [LH - mean(LH)]^2 +
    %                      [HH - mean(HH)]^2
    %
    % Furthermore, Sun indicates the summation is over the height and
       width
    % of the image, whilst Yang sums over the set of i,j present in the
    % arrays HL, LH and HH respectively.  There's also discrepancy as
       to
    % the definition of the mean that's used, as well as whether or not
        to
    % take absolutes.  For the avoidance of doubt, the original paper
       [2]'s
    % equations are used.
    %
```

```
    % This requires Wavelab (http://www-stat.stanford.edu/~wavelab/)

    I = double(I);
    I = makedyadic(I);
    d6 = MakeONFilter('Daubechies', 6);
    wlI = FWT2_PO(I,5,d6);
    [w h] = size(wlI);
    ll = wlI(1:w/2, 1:h/2);
    lh = wlI(1:w/2, h/2:h);
    hl = wlI(w/2:w, 1:h/2);
    hh = wlI(w/2:w, h/2:h);

    hlsum = sum(sum((hl - mean(mean(hl))).^2));
    lhsum = sum(sum((lh - mean(mean(lh))).^2));
    hhsum = sum(sum((hh - mean(mean(hh))).^2));

    score = 1/(h*w) * (hlsum + lhsum + hhsum);
end
```