VISUAL SALIENCY IN NAVIGATION:
modeling navigational behavior using saliency and depth analysis

Stamatios Psarras¹, Ava Fatah¹, Angeliki Zarkali², Sean Hanna¹

ABSTRACT

Spatial configuration is extensively used in spatial analysis methods to predict navigational behavior; several studies have shown a correlation of global and local space syntax metrics with the distribution of pedestrians in different settings. However, recent studies have also shown correlation of human behavior with the visibility of relevant objects, for example the visibility of paintings influenced navigational choices in museums. These findings suggest, that in addition to spatial configuration and depth, visually or semantically important elements are also important in wayfinding. Incorporating these additional characteristics into existing analyses could improve our ability to predict and model navigational behavior in the built environment.

The main tool of identifying which elements of a visual scene are of most importance is Saliency. Saliency is the subjective perceptual quality some stimuli possess within a visual scene, which makes them stand out from their neighbors and gain the observer’s attention; it is determined in very early stages of visual processing, implying a certain generality of saliency between different observers. Saliency detection algorithms have been extensively used in computer vision and have shown correlation with observed behavior using eye tracking. Despite the attractiveness of a global tool that can identify prominent objects within a visual scene for architectural design, Saliency detection has not been applied in spatial perception and navigation.

Here we examine the application of different saliency detection algorithms in the context of the built environment. We recorded navigational behavior of 143 pedestrians moving freely in an open space. We compared existing isovist models with observed behavior. We then tested different saliency detection models as well as a hybrid model combining isovist and saliency detection for their capacity to predict change in angular direction during navigation. Saliency had a significant negative correlation with change in angular direction; combining saliency with depth-based isovist models improved prediction performance. Our findings suggest that saliency has the potential to be a significant addition to traditional isovist models in predicting and modeling navigational behavior.

KEYWORDS

1 Stamatios Psarras, Ava Fatah and Sean Hanna are from: The Bartlett School of Graduate Studies, University College London (UCL) Gower St. WC1E
2 Angeliki Zarkali, Institute of Neurology, University College London (UCL) Queen Square WC1N
1. INTRODUCTION

In recent years, significant attempts have been made to quantify and define pedestrian behaviour within the built environment, based on the relationships between its different elements. Many of these attempts have been based on Space Syntax graph-based approaches and metrics. Indeed, several studies have established a correlation of graph measures with pedestrian movements and wayfinding; distribution of pedestrians in streets was correlated with integration values of those streets (Hillier 1987; Peponis 1989), movement rate in shopping malls with integration (Fong 2003), use of hospital corridors and integration values (Peponis 1990; Haq and Zimring 2003).

Building on these population-based approaches further attempts have been made to model and predict navigational behaviour at the level of the individual using depth-based measures such as the Isovist. First proposed by Benedikt in 1979, the Isovist from a point in space represents all areas in the environment visible from that point (Benedikt 1979). Enhanced with graph-based approaches and focused on metrics of Depth, such as the longest ray of sight, the Isovist has evolved into an important tool to model navigational behaviour since isovist based predictions have been correlated with observed natural movement (Turner and Penn 1999; Turner and Penn 2002). The theoretical model and practical applications of Visibility Graph Analysis (VGA), have been further enhanced with new metrics aiming to more richly define the environment; introducing metrics of complexity (Psarra and Grajewski 2001), or incorporating spaces with non-opaque elements, where a space may be visible but inaccessible (Varoudis and Penn 2015) have enhanced the existing isovist based approach but have not been tested against observed behaviour.

Additionally, a growing body of evidence suggests that other visibility measures other than Depth can also affect navigational behaviour: visibility of museum displays affects visitors’ movements and experience (Peponis 2004; Newhouse 2005; Zamani and Peponis 2010); visibility of common destinations influences wayfinding in airports (Churchill et al. 2008; Lam et al. 2003); visibility of intersections or landmarks affects wayfinding in buildings and urban environments (Haq and Zimring 2003; Omer and Goldblatt 2007). These studies suggest that in addition to depth, which is incorporated into isovist-based analyses, visually or semantically important elements are also important in pedestrian navigation. However, current visibility based analysis techniques do not take into account important visual qualities of space such colour, texture, or saliency despite proven to significantly affect visual perception (Peponis, Zimring, and Choi 1990; Conroy-Dalton 2001).

**Saliency**

Visual Saliency is the subjective perceptual quality some stimuli possess within a visual scene, which makes them stand out from their neighbours and gain the observer’s attention. In other words, “Saliency at a given location is determined primarily by how different this location is from its surround in colour, orientation, motion etc” (Koch and Ullman 1985, 221). This definition highlights
that an object’s saliency is not an inherent property of an object but is, partly, related to the other objects in the visual scene and the observer’s visual system and intentions.

Most proposed models suggest that saliency is determined in very early stages of visual processing implying a certain generality of saliency between different observers (Itti and Koch 2001; Koch and Ullman 1985; Niebur and Koch 1996). These models propose the existence of multiple “saliency maps” in the brain which determine where the observer will look next. Within each map, all visual objects of a scene are encoded in parallel for different characteristics, colour, intensity, shape, etc. Then, the next target of attention is selected via competition within each map and subsequent merging of maps. In the absence of a task, attention is drawn towards the most salient areas of the map; this then triggers a motion directing the eyes towards salient visual locations (Itti and Koch 2001; Koch and Ullman 1985).

One of the first saliency models is that of Koch which has been extensively used in computer vision to predict eye movements during free-viewing and has been validated in eye-tracking studies (Koch and Ullman 1985). Since their initial description, more than 50 different saliency models have been developed each with different strengths and weaknesses (Zhang and Sclaroff 2013). The saliency methods used in this study are:

- **Boolean Map based Saliency (BMS)** is inspired by the Gestalt principle of figure-ground segregation and attempts to create and add multiple feature maps of an image, using a random threshold. Thresholding is a post process effect often used on an image that attempts to split the color space into two categories, those above and those below the threshold. In doing so the result is a two-tone image. This particular algorithm attempts to threshold the same image multiple times using a random threshold, then it adds all the resulting images together. This allows the more distinct elements of the image to stand out in the final result. (J. Zhang and Sclaroff 2013)

- **Itti et al 1998 (ITTI)**. This method utilises a neural network that weights features of a multiscale image into a single topographical saliency map. Starting with the very basic aspects of an image like color, intensity and orientation, this algorithm extracts a series of different other feature maps (low level features of an image). The first set of feature maps is focused on contrast between light and dark, and the second set is focused on contrast between opposite colours like green-red. A final set of feature maps is focused on different types of orientation intensities. These maps are then processes and inputted to a winner-take-all neural network that determines how each feature map contributed to each region of the image and sums the result into a saliency map. (Itti, Koch, and Niebur (1998))

- **Minimum Barrier Distance Transform (MBD)** is a highly performing saliency detection method, that uses cues from the framing of the image to subtract background elements. It is applied on a pixel-by-pixel manner. MBD is using the Image Boundary Connectivity Cue which assumes that salient object tends not to touch the frame of the picture. In order, to find the salient objects MDB calculates the connectivity (distance metric) of each pixel to the
frame. The more well connected the pixel the more likely that it would be part of the background and thus not part of a salient object. (J. Zhang et al. 2015)

- **Manifold Ranking (MR)** is using a graph-based manifold ranking method on the superpixels of the image. The ranking is attempting to separate background and foreground elements. Superpixels are essential larger region of the image that have similar characteristics, therefore their pixels tend to be part of the same object but each object is comprised by multiple superpixels. The superpixels can be seen clearly in Figure 2. Each superpixel is connected to the neighbouring superpixels with the strength of that connection based on the distance of their average color, forming a graph. The graph is then queried in a two-stage process, firstly from the nodes located on the boarder (background query) to rank all regions of the image. This would form a temporary saliency map. On the second step this map is thresholded and the foreground nodes are used for a further series that forms the final saliency map. (C. Yang et al. 2013)

- **Robust Background Detection (RBD)** similarly to MBD takes advantage of the fact that salient objects tend not to touch the frame of the picture; its difference from MBD is that it applies the same set of metrics on a graph formed by superpixels of the image. This algorithm further adds extra connection to the superpixels on the boundary of the image and then computes the connectivity of the graph to find the salient object. (Zhu et al. 2014)

The purpose of saliency detection algorithms is to create a process where the obtained knowledge can be generalized and applied repeatedly, resulting to an accurate feedback map. These maps could specify the hierarchy of the visual elements in a scene with a heatmap showing their relative importance. Such processes are able to predict eye movements; Parkhurst showed significant correlation between the salient points of an image and observers’ fixation times (Parkhurst, Law, and Niebur 2002). Being able to predict eye movements based on the characteristics of a visual stimulus is a powerful tool, and has been used in marketing to identify the most attention-capturing elements and to de-emphasize salient images (Su, Durand, and Agrawala 2004).

We aimed to examine the effect of visual Saliency in navigational behaviour of pedestrians moving freely in an open space at the level of the individual and compare the performance of existing models such as the Isovist with Saliency based models in predicting change in angular direction of pedestrians.

## 2. DATASETS AND METHODS

**Recording pedestrian paths within an urban, open space**

We recorded pedestrians navigating freely within an open space in London (UCL entrance Quad) using a 1080p DSLR camera with a standard 35mm lens. The study location was chosen to allow enough freedom of movement and a mixture of seating, moving and standing locations. A high
A vantage point was used to record as much of the ground as possible; consideration was taken to include the main entry and exit points. The recording took place on a single day (September 2017), for a duration of ten minutes (from 12:05 to 12:15) recorded at 25 frames per second.

The video was converted into static images; frame extraction was performed using ffmpeg with a rate of 1 frame per 2 seconds, generating a total of 330 images. From these we manually extracted, using a custom-made python application, the position of each person in space and the direction of their movement. From these two elements, it is possible to reconstruct the direction of gaze for each subject. The aim was to capture what people were looking at while the experiment took place, their entry point and their destination. A total of 143 pedestrian paths were extracted using this method.

Reconstructing the visual space
After extracting the pedestrians’ paths, we aimed to reconstruct the visual input of each pedestrian across their individual path. To achieve this, we first needed to reconstruct an accurate representation of the three-dimensional space of the study location. To achieve this we used publicly available plans and photographs taken on the day and time of the video recording; this ensured light and weather conditions were the same. To increase the realism of the 3D model, textures were added; these were produced using perspective projections from the photographs for the large surfaces. Adding textures in this way ensured little nuances, such as windows being open, curtains drawn or shadows were included, increasing the model's fidelity. For small items and ground surfaces, approximate textures were used to match the observed materials. A visual comparison of the reconstructed model and a photograph of the study space is seen in Figure 1.
Saliency-based analysis of the pedestrians’ field of view

Next, we extracted the pedestrian's field of view for each step of the path from the 3D model; virtual cameras were placed along the paths to capture what each person was viewing at each step. Cameras were placed at 1m intervals and 1.6m height, the height of an average person; cameras were placed tangent to the direction of movement and set to cover 90 degrees of the pedestrian's field of view to include the central area of vision with the maximum visual acuity.

Subsequently, the rendered images for each pedestrian's field of view were processed using different saliency algorithms; this generated an intensity map, marking areas with high saliency and low saliency (Figure 2). Five Saliency detection algorithms used were:

- Boolean Map based Saliency (BMS) (J. Zhang and Sclaroff 2013)
- salient proto-objects, Saliency ToolBox (Itti, Koch, and Niebur (1998))
- Minimum Barrier Distance (MBD) (J. Zhang et al. 2015)
- Manifold Ranking (C. Yang et al. 2013)
- Robust Background Detection (RBD) (Zhu et al. 2014)

Figure 2. The Saliency Algorithms applied along different steps of a single pedestrian path.

BMS: Boolean Map based Saliency, ITTI: Saliency ToolBox, MBD: Minimum Barrier Distance, MR: Manifold Ranking, RBD: Robust Background Detection.
To analyse the effect the Saliency of each object has in the pedestrians’ navigational decisions, the saliency analysis had to be transformed to the object rather than the scene level (Figure 3). To achieve this, the intensity of the rendered images was then projected onto the 3D model and the values were aggregated for all pedestrians. First, the 3D model was divided into distinct surfaces and mapped the coordinates of these surfaces on each image. A surface located on the top part of the first column from the left, was calculated as located at \((x_1, y_1)\) pixel on img 1, \((x_2, y_2)\) on img 2 etc... Sequentially, saliency values from these coordinates were retrieved and aggregated, resulting on the total saliency value for that surface (Figure 3).

![Figure 3](image)

**Figure 3.** The transformation from scene to object level saliency analysis.

A: rendered images from the path and highlighted two example surfaces  
B: multiple samples on each surface were taken and their coordinates recorded  
C: the first image analysed according to its saliency and the same pixels being samples  
D: shows the average saliency of each surface for all their visible samples.

Applying this transformation to all images, an aggregated saliency value was assigned to every surface. The 3D model was then highlighted according to the aggregated saliency value (Figure 4). To avoid falsely high values for surfaces visible in multiple images, average saliency values were used instead of cumulative.
Linking Saliency and Isovist measures to pedestrian movement

We studied the effect of saliency and compared with that of the Isovist, on observed changed in direction of pedestrians in a step-by-step analysis. The tangent on each step was defined as the Current Heading and the difference with the next step, the Change of Direction (positive: left, negative: right). The last step was excluded, as there is no information on the future direction of the pedestrian and thus a change in direction can not be calculated; this resulted in 3044 steps where the change of direction was known.

For each step, the saliency of the Current Heading was analysed by capturing a coloured image at 90 degrees field of view. Each object's saliency in that image was ranked from highest to lowest. Subsequently, the vector pointing from the current step to the centroid of each object was calculated and projected on the XY plane. After collecting the vectors pointing to the most salient objects, four different combinations were used: 1) the vector to the most salient object, 2) the vectors to the first and second most salient objects combined and 3) a combination of the vectors to the first, second and third most salient objects. The averaging was done both by using the unit vector and a weighted vector based on the saliency of each object. 4) Finally, an average vector was calculated from all visible objects weighted by their individual saliency. These measures were tested for each Saliency detection method to establish the best accuracy in predicting change of direction (Figure 5).
Similarly, we tested several Isovist-related measures. For each of the 3044 steps, an isovist was generated pointing towards the heading of each step. The Isovist was tested between a range of 30 to 250 degrees, for every 20 degrees. A ray was cast per degree to keep the same detail on each test. For each Isovist, the longest ray of sight was used to assess change of direction. Due to the nature of the site, there were some areas that were visible and semi-accessible (grass areas around the concrete floor). To avoid potential negative bias on the isovist results, since people preferentially walked on the concrete floor rather than these semi-accessible spaces, a second model was created that assumed a vertical wall around such areas. Each ray for each of the different parameters was compared with the current heading to assess the predicted change of direction (Figure 5).

Datasets
A total of 143 pedestrians recorded. Each of their 143 paths was split into 1m steps, resulting into 3187 steps; 3044 steps where change of direction is known (excluding the last step of each path). The observed recorded data have a range of change in direction between 0-180 degrees. However, the Saliency analysis is constrained by its 90 degree field of view to changes between 0-45 degrees. Additionally, the change of direction at small angles (<5 or <10) could be within the margin of error for the recording or attributed to a gentle sway in a path rather than a volitional change in direction. Taking into account these constrictions, our data resulted to three different datasets: a) All unfiltered data (n=3044), b) Observed change of direction >5 and <45 (n=1280) c) Observed change of direction >10 and <45 (n=604). All Saliency and Isovist measures were tested separately in each dataset.

Statistical analysis
All data analysis was performed using R (R Studio for Windows, version 1.1.423). Pearson's correlation was used to compare the paired values of observed and predicted change of direction for each step. Statistical significance was set at p value < 0.05.
Proceedings of the 12th Space Syntax Symposium

Figure 5.
A: All parameters used in the Saliency model; each parameter was tested for each of the 5 Saliency algorithms. V1: the vector pointing towards the most salient object, V2: to the second, V3: to the third. P1, P2, P3: different weightings used in the weighted saliency model. BMS: Boolean Map based Saliency, ITTY: Saliency ToolBox, MDB: Minimum Barrier Distance, MR: Manifold Ranking, RBD: Robust Background Detection.
Proceedings of the 12th Space Syntax Symposium

B. All measures used for the isovist model. Range: 30 – 250 degrees, every 20 degrees. These were tested in both an Open and Constrained for vision and movement (i.e. excluding semi-accessible, visible areas) model. The plan view on the right is showing which areas where obstructed for the constrained model in Red.

3. RESULTS

Unfiltered dataset (n=3044 steps, 0-180 degrees)
No measure of Saliency or Isovist analysis had statistically significant correlation with the observed data, in the unfiltered dataset.

Change of direction between 5 – 45 degrees (n=1280 steps)
Both Saliency and Isovist were correlated with the observed change in direction (Figure 6). Three of the five Saliency methods (MDB, RBD, ITTY) shown negative correlation in at least one parameter (first vector, combination or average); the degree of correlation differed slightly between different saliency detection methods. Overall, the best performing saliency method was MDB_1 with a negative correlation of $R=0.11$ ($p <0.001$) when using the vector of the most salient object. For the Isovist, only one parameter (Iso2_30Lon: Isovist with 30 degrees field of view) was significant, with a positive correlation $R=0.08$ ($p=0.038$). To minimise negative bias, a second more constrained for vision and movement dataset where the spaces that are visible but not so easily accessible (covered with grass) were treated as obstacles was used; this did not lead to significant improvement in performance, with only one measure again (30 degrees) reaching statistical significance.

Finally, a combination of Saliency and Isovist was tested combining the best performing models for each method. As Saliency had negative correlation (detractor) the following combined model was used:

Combined Model = [Isovist] – [Saliency]

The combined model surpassed all other methods; for change of direction between 5-45 degrees, $R=0.15$ ($p<0.001$).
Figure 6. Absolute correlation values between the different Saliency and Isovist measures and observed data for Change of direction between 5-45 degrees. Note the Saliency results had negative correlation, but their values were flipped to be comparable with the Isovist. Red highlights the best performing Saliency method and Orange the best performing Isovist method; Black highlights the combination of best performing Saliency and best performing Isovist method.

**Change of direction between 10 – 45 degrees (n=604 steps)**

For both Isovist and Saliency, results improved when constraining the dataset to change of direction between 10-45 degrees, with similar inner dynamics; saliency models exhibiting negative correlation thus saliency acting as a detractor and a single Isovist model with a positive correlation. The best Saliency method was again MDB_1 with negative R=0.15 (p<0.001) and the best Isovist measure Iso2_30Lon with positive R=0.10 (p=0.018). The combined model again outperformed single measures with R=0.19 (p<0.001). The results for all measures used is seen in Figure 7.
Figure 7. Absolute correlation values between the different Saliency and Isovist measures and observed data for Change of direction between 10-45 degrees. Note the Saliency results had negative correlation, but their values were flipped to be comparable with the Isovist.

Red highlights the best performing Saliency method and Orange the best performing Isovist method; Black highlights the combination of best performing Saliency and best performing Isovist method.

4. CONCLUSIONS

Our study showed that visual Saliency is correlated with observed navigational behavior. Many studies have linked saliency detection algorithms with observed eye movements and fixation (Parkhurst, Law, and Niebur 2002); this is the first study to our knowledge where Saliency detection has been applied to architecture.

In our cohort of 143 pedestrians navigating freely in an open space, we found that Saliency improved the fidelity of visibility-based analysis in predicting change in direction. Specifically, we have shown a negative correlation between Saliency and navigational behaviour in a step-by-step analysis. It is yet unclear why Saliency has a negative correlation with change in direction. It could represent an avoidance behaviour by pedestrians towards salient objects, for example to avoid collision. However, there could be instances where certain salient objects act as attractors whilst others act as detractors therefore saliency models’ performance could have been underestimated and could improve by determining which salient objects act as attractors and which as detractors. We have only examined cumulative effects of saliency and further research is required to assess this.

In addition, we have confirmed the importance of depth in navigation; Isovist metrics were significantly correlated with observed behavior, albeit in a constricted visual field of 30 degrees. The hallmark isovist study in Tate Gallery by Turner et al. (A. Turner and Penn 1999; A. Turner and Penn 2002) established correlation between isovist-based predictions and observed natural movement. Our
study has confirmed that isovist measures are correlated with step-by-step navigational decisions in a larger sample of 143 pedestrians and in an open space. However the performance of isovist metrics was poorer in our cohort. This poorer performance could be due to the open-space study area, less constrained than interior spaces where isovists have previously been tested. As isovists point towards the longest ray, if there is a far destination, as long as it is within the field of view the isovist will prefer it; in contrast saliency models are more relational.

Importantly, we have shown a cumulative effect of Saliency and isovist measures; combining depth with saliency analysis led to a better performing model than each individual metric in our cohort. Saliency could represent a novel and powerful tool of spatial analysis that takes into account the visual characteristics of space and does not solely rely on spatial configuration, thus enriching existing Space Syntax based approaches.

Our study has of course certain limitations, that need to be taken into account when generalising these results. Firstly, only one building was used; this result to bias towards the dominant features of the specific architecture. Secondly, the mixture of users of the space was not recorded; their familiarity with the space or their professional background could have had affected their navigational capabilities (Hölscher, Brösamle, and Vrachliotis 2012 ; Maguire, Woollett, and Spiers 2006). Finally, their individual tasks (attending event, leaving, looking for seating) were similarly not recorded; this could have an effect to the pedestrians’ navigational behaviour (Emo, 2014).

Despite these limitations, our study has demonstrated in a cohort of 143 pedestrians, that visual Saliency plays an important role with pedestrian navigation. We have recorded pedestrians in an unobtrusive, natural way, moving freely in an open space thus limiting researcher-influence in the results. Incorporating Saliency detection analysis in existing Isovist models, led to a better prediction of the observed behaviour. Saliency could be a powerful, emerging metric that could enhance the fidelity of depth-based approaches in modelling pedestrian behaviour.
REFERENCES


Omer, Itzhak, and Ran Goldblatt. 2007. “The Implications of Inter-Visibility Between Landmarks on Wayfinding Performance: An Investigation Using a Virtual Urban Environment.” Computers,


