

K-DIME: An Affective Image Filtering System

Nadia Bianchi-Berthouze

University of Aizu,
Japan

Even though extracting the semantic content of information is a critical feature in information search, today's Web sites still have poor capabilities in this regard, especially when it comes to multimedia data. While most Web sites provide visual, and sometimes, auditory content, they don't provide any tools to process the resulting subjective information, and by doing so, they fail to support the user's decision process. Often, users will select a picture or a piece of music based on the impression or feeling it conveys from a pool of many others satisfying the same objective criteria, such as percentage of a given color or type of object. In fact, in fields such as design, advertising, or entertainment, affective content is the main criterion for selection.

Therefore, it's crucial that users can fully interact with the multimodal components of the information, beyond a simple search or viewing of file names, verbal descriptions, or even indexes of objective content. My group in the Database Sys-

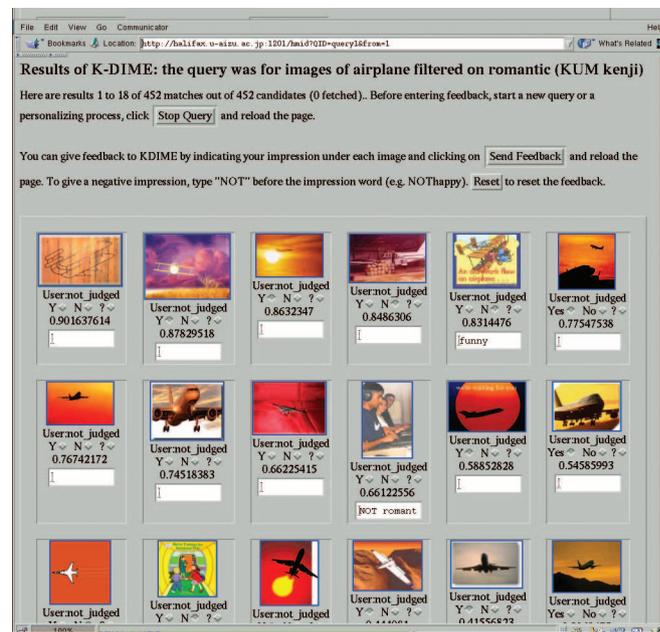
tems Lab at the University of Aizu is investigating how to endow a computer system with the ability to handle affective and subjective criteria. In this article, I present our framework for software agents to adapt to their users and perform Web-based multimedia search on the basis of objective and subjective criteria. We expect such systems to continuously tune the filtering criteria by exploiting users' feedback, thus providing flexibility against the intrinsic variability of affective experiences.

Affective Web-based filtering engine

The Kansei (Japanese for sensitivity) Distributed Information Management Environment,¹ or K-DIME, software environment lets users or software applications fetch multimedia material from the Web on the basis of textual keywords and filter it using a Kansei user model (KUM), which I'll describe later. The KUM learns a mapping between low-level features of the multimedia data and impression words—such as “romantic”—expressed in natural language.

K-DIME acts as a metasearch engine for the Web. The search process involves four main components. Given the objective criteria of a user's query—for example, “screen-saver images of airplanes”—the dispatching component selects and tunes existing search engines, analyzes the results of the selected search engines, and integrates their results to prepare for the data filtering process. The filtering component selects and configures the KUMs, which assess the data against the user's subjective criteria (such as “romantic”). The display component outputs the filtered data to the software application, which displays the data to the user (see Figure 1). The user can then provide relevance feed-

Figure 1. Result of K-DIME for the query “romantic screensaver of airplanes.” K-DIME retrieved 452 airplane images by querying different Web-based search engines and then ranked them according to their romantic impression. The user can input feedback in the form of additional impression words by entering them under each retrieved image.



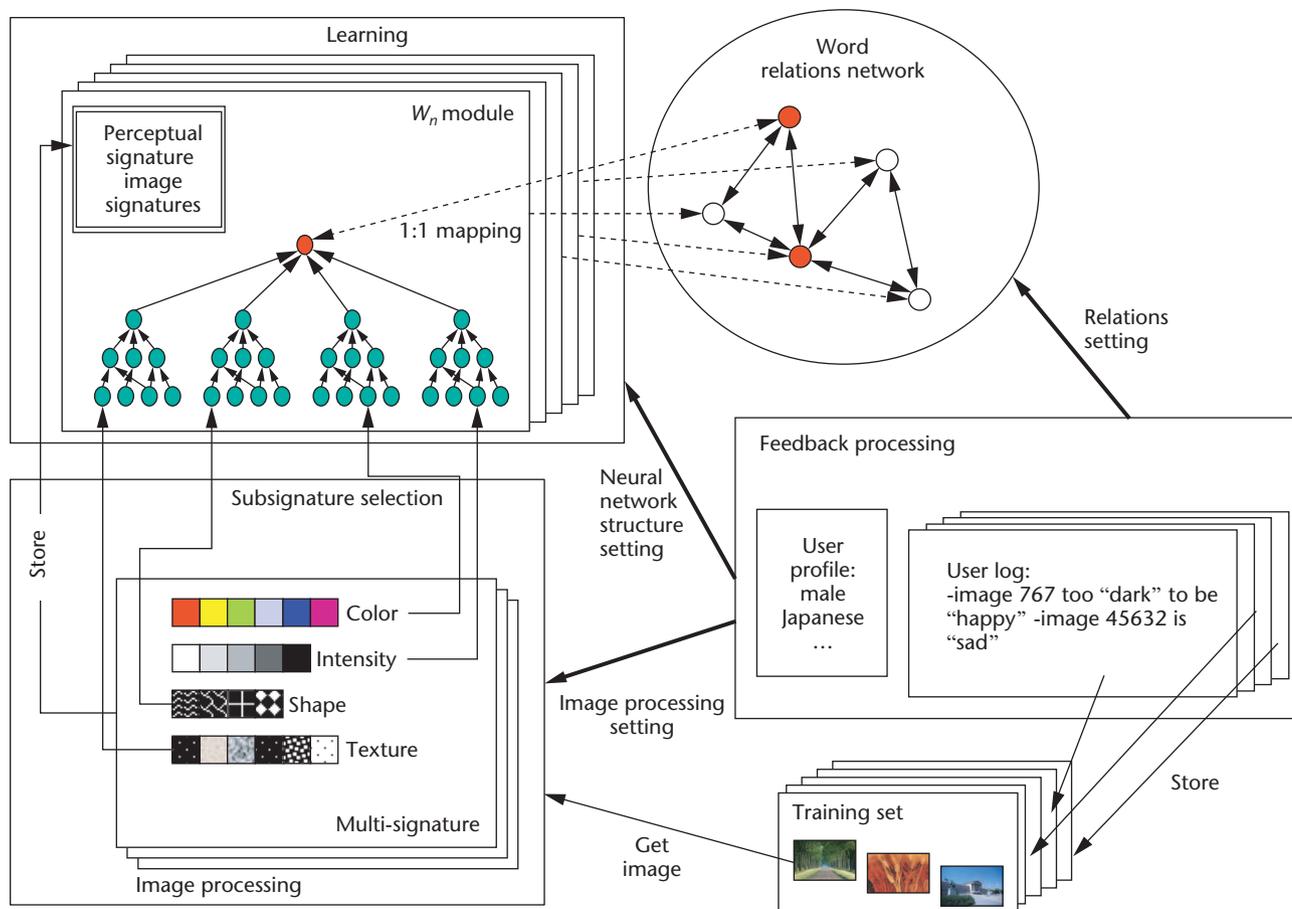


Figure 2. Kansei user model architecture. Each impression-word module consists of a set of neural networks fed by subsets of the images' perceptual signature. A feedback-processing component is in charge of analyzing the feedback and triggering the refinement of word modules.

back.² K-DIME's feedback-processing component analyzes this feedback. It aims to adapt existing KUMs or to create new ones for newly entered impression words.

Kansei user model

The user model assumes that the meaning of an impression word is partially grounded in the low-level characteristics of the multimedia data from which it originates. We can therefore set the KUMs to learn associations between perceptual states and impression words under user supervision. The perceptual state is the signature, or set of low-level descriptive variables, extracted by a set of image processing algorithms that describes the image in terms of color, texture, and shape features.

Figure 2 depicts the modeling process' information flow. A word module consists of a set of neural networks that learn the relationships between subsets of the perceptual state and impression words. A saliency factor, or weight, is associated with each network and updated

through user feedback. With this mechanism, a word module can deal with different subjective impression words by relating to different categories of perceptual characteristics. For example, "romantic" or "fresh" might relate to color features, whereas "imposing" or "brave" might relate to shape characteristics. Since the lexicon evolves as interaction unfolds—that is, as new words are introduced—each word is independently dealt with by a word module in charge of learning its particular associations with perceptual states.

Personalized holiday planner

We have integrated K-DIME into the Personalizable Hotel Retrieval System (PHRS). PHRS aims to combine the ability to satisfy objective criteria, such as price range or type of hotel amenities, with visual impression criteria. PHRS forwards users' queries (see Figure 3) to K-DIME, which will gather data from various Web sites. When such data include pictures, K-DIME indexes the accommodation on the basis of its atmos-

phere—that is, the visual impression conveyed by its pictures—as evaluated by the user’s KUM (see Figure 4).

Retrieved and indexed hotels are stored in a database for further manipulation by the user. Such manipulation includes defining and selecting different points of view, or levels of visualization, and refining the selection criteria. Indeed, users rarely fully specify initial queries, and it’s only after the first return of information that they might add specifications, which they weren’t initially aware of. Any relevance feedback is returned to K-DIME for further refinement of the user’s KUM, including the creation of new impression words (see Figure 5, next page).

Aspects of the visual subjective experience

Early experimental results obtained with the prototype have demonstrated the feasibility of the framework for constructing real-world applications.³ More importantly, they showed that the prototype was a good testbed to explore the requirements for computational models of subjective experiences, and as a result, we identified four issues that should be addressed in future research.

Variability of users’ subjective visual experience

When we asked subjects to classify, over a month interval, the same set of landscape images by the visual impression they convey, assessments show a variability of about 80 percent. We attribute this significant result to intrinsically variable factors that affect the perception of images, through a different organization of brain mechanisms of selection and attention.⁴ Typically, these factors are state of mind, the task the user wants to accomplish, and so on.

Context dependence

Experiments using different classes of images show that context—that is, high-level content and the user’s task—affects the mapping between low-level features of the image and impression. For example, “quietness” in a landscape might involve different features than “quietness” in the picture of a home interior.

Language generalization

When we asked subjects to browse through a set of images that they had previously classified as conveying a similar impression and arrange the images in subgroups according to nuances of that impression, they showed great difficulty doing this. However, when we provided subjects

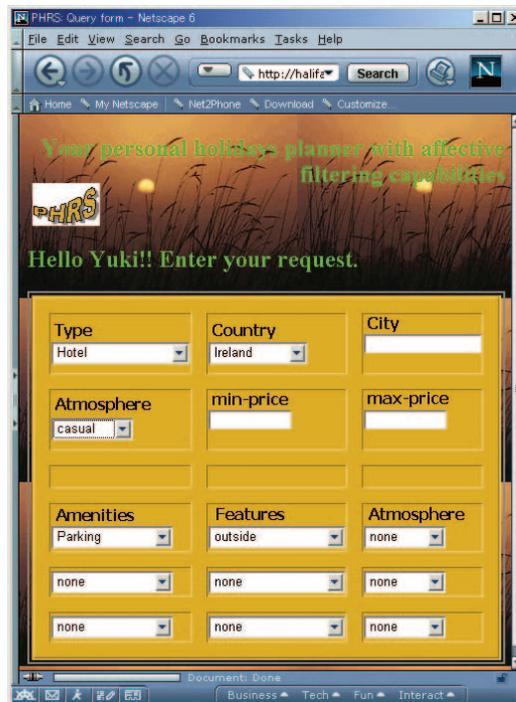


Figure 3. Web form for a bed and breakfast search. The user can enter complex queries that include objective as well as affective criteria.

See details	Name	Type	Atmosphere	Address	Description	Pool	Restaurant	Parking	Rank
Show	Glen House	1	dark, casual, country	93a Process Rd, Ballymoney, Ireland	Situated on A26 suitable for touring North Antrim Coast & Giant's Causeway a modern family bungalow. Fully central heated, all rooms ensuite, and suitable for wheel chairs. A complimentary cup of tea on arrival and complimentary snacks on arrival.	yes	no	yes	Rank 4
Show	Mountview	1	casual, country	23 Carrowcroey Rd Arroy, Ballymoney, Ireland	This family run business is located 5 miles from the Causeway coast. A warm welcome is guaranteed.	no	no	yes	Rank 2
Show	Hollow	1	dark, casual	107 Knocknacarry Rd, Cushendun, Ireland		yes	no	yes	Rank 4
Show	Country House	1	casual, natural	33 Whitepark Road, Ballycastle, Ireland	Comfortable family run house set in beautiful surroundings. Ideally situated for touring Causeway Coast and Glens of Antrim. Convenient to shops beach, tennis courts etc.	no	yes	yes	Rank 4

Figure 4. Result of the query “B&B in Ireland with parking and a casual atmosphere.” The user can customize the presentation’s layout.

with a list of words, they were able to perform the task. Therefore, we can explain previous inconsistencies because users tend to use a same impression word to identify a large range of impressions—for example, using the word “quietness” to indicate either a peaceful feeling or a sense of loneliness. Providing users with a more refined lexicon helps solve inconsistencies between low-level features within images of a same impression.

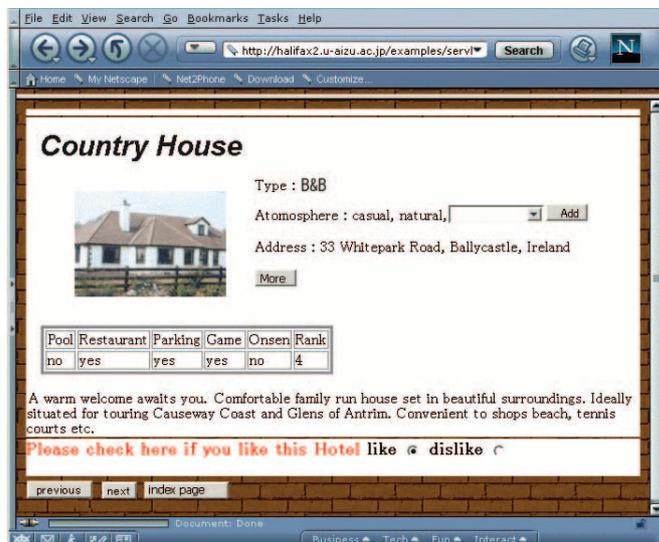


Figure 5. Detailed view of one of the retrieved bed and breakfast locations. The user can enter feedback on the retrieved bed and breakfast's saliency and atmosphere.

Externalization process

Relevance feedback has only relative value and is often inconsistent over time. It's even more so when information allows for multiple interpretations. To overcome this limit, we've explored the feasibility of the user actively participating in the user model adaptation process. Our experiments showed that users can be led, through externalization processes, to be more aware of the cognitive process involved in their subjective interpretations of an image. Externalization processes,⁵ whereby the user provides a physical representation of his naive models (high-level explanation of a phenomenon), can be triggered

by detecting inconsistencies in judgment, detecting similarities, and providing users with a visual representation of externalized cognitive pathways. This issue, as well as the previous three issues, certainly poses an interesting challenge to the human-machine interaction community. **MM**

References

1. N. Bianchi-Berthouze, "Mining Multimedia Subjective Feedback," *J. Intelligent Information Systems*, vol. 19, no.1, July 2002, pp. 43-59.
2. Y. Rui et al., "Relevance Feedback: A Power Tool in Interactive Content-Based Image Retrieval," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 8, no. 5, Sept. 1998, pp. 644-655.
3. N. Bianchi-Berthouze et al., "Supporting the Interaction between User and Web-Based Multimedia Information," to appear in *Proc. IEEE Int'l Conf. Web Intelligence*, IEEE CS Press, 2003.
4. H. Pashler, "Attention and Visual Perception: Analysing Divided Attention," *Visual Cognition*, vol. 2, S.M Kosslyn and D.N. Osherson, eds., MIT Press, 1996, pp. 71-100.
5. N. Bianchi-Berthouze and C. Lisetti, "Modeling Multimodal Expression of Users' Affective Subjective Experience," *Int'l J. User Modeling and User-Adapted Interaction*, vol. 12, no.1, Feb. 2002, pp. 49-84.

Readers may contact Nadia Bianchi-Berthouze at the Database Systems Lab, University of Aizu, Aizu Wakamatsu, 965-8580, Japan; nadia@u-aizu.ac.jp.

Readers may contact editor Tiziana Catarci at the Dept. of Information Systems, Univ. of Rome "La Sapienza," Via Salaria 113, 00198 Rome, Italy; catarci@dis.uniroma1.it.

**SET
INDUSTRY
STANDARDS**

wireless networks
gigabit Ethernet
enhanced parallel ports
802.11 FireWire
token rings

IEEE Computer Society members work together to define standards like IEEE 802, 1003, 1394, 1284, and many more.

HELP SHAPE FUTURE TECHNOLOGIES • JOIN AN IEEE COMPUTER SOCIETY STANDARDS WORKING GROUP AT

computer.org/standards/