The Psychological Space of Common Media Impressions Held in a Media Database Retrieval System

Hideyuki, Takagi
Kyushu Institute of Design : Associate Professor

Noda, Toshihiko

Cho, Sung-Bae

The Psychological Space of Common Media Impressions Held in a Media Database Retrieval System

Hideyuki TAKAGI*   Toshihiko NODA**   Sung-Bae CHO***
* Dept. of Art & Information Design, Kyushu Institute of Design
** Graduate School, Kyushu Institute of Design
*** Dept. Computer Science, Yonsei University
* takagi@kyushu-id.ac.jp,   *** sbcho@csai.yonsei.ac.kr

Abstract— We discuss artificial computer-based *KANSEI* and its application in a media database retrieval system. In our three-step approach, we first design a psychological space in which human impressions of images, movies, or music are expressed as coordinates. Second, we determine the mapping relationship between the psychological space and physical media features. And third, we retrieve media from the database or convert media into other media through the psychological space. We then construct a universal space to express several impressions of different media using the SD method. As an application, we evaluate an interactive EC-based image retrieval method using this space.

1 INTRODUCTION

Early computers committed to numerical calculations did not deal with human factors. Computers later came to deal with speech and images as signal processing, which is still categorized as a numerical calculation. Signal processing describes the physical characteristics of sounds and images, not their contents, and deals with few human factors.

Today, in the so-called era of multimedia, it is important for computers to deal with the human factor. Sounds, images, and movies appeal to the computer user. It is important, then, how sounds and images are expressed at the *KANSEI* level to the computer user.

We define the term *KANSEI* as a psychological level of subjective preference as opposed to the psychological level of external information. We have a hierarchy of four psychological levels: sensation, perception, cognition, and *KANSEI*. While perceptual and cognitive psychologies handle the first three lower levels and do not handle the value of the external information, the *KANSEI* handles the value of the information, such as "I prefer this sound." The *KANSEI* includes preference, emotion, subjective evaluation, feeling, impression, etc. Although impression has a narrower meaning than the *KANSEI*, we will frequently use the word, *impression*, because some readers may be unfamiliar with the concept of *KANSEI*.

Computers in the multimedia era commonly interact with people at the *KANSEI* level. For example, images and music can be retrieved from a database not only with keywords such as "a picture of flowers and a puppy" or "the first movement of the Pastoral Symphony," but also with *KANSEI* phrases such as "a picture that matches my bedroom" or "music that is calm and cheerful." Furthermore, it is useful that computers have the capability to match appropriate music to movies and vice versa.

The objective of our research is to propose an approach to realize such a capability with computers. To allow computers to *be impressed* by multimedia as well as people, it is necessary that computers should use the model of impression that people get when they see or hear media. In this paper, we construct a factor space that is common to images and music as the mainstay of the *KANSEI* model and install it onto computers.

We are conducting this research according to the following steps. First, we construct a factor space of human impression that is common to images and music. Second, we obtain the mapping relationship between an image feature space and the factor space. Image feature vectors and their corresponding coordinates in the factor space subjectively selected by human subjects are used as training data to obtain the relationship. Third, we obtain the mapping relationship between a music feature space and the factor space as in the second step. Fourth, the

mutual transfer or conversion between images and music through the factor space is conducted. Finally, we expand the number of media that can be transferred or converted with other media from images and music to movies, voices, and general sounds. These new multiple mapping relationships among expanded media are installed onto computers as in the fourth step. As a result, computers can act like people feel when exposed to multimedia.

We report our progress in the first and second steps in this paper. In section 2, we propose the approach that deals with a factor space as a *KANSEI* model and discuss the applications of the model. In section 3, the construction of the core factor space for multimedia is described. Finally, in section 4, an interactive genetic algorithm (GA)-based image retrieval system is evaluated using the factor space as one of its concrete applications.

2 IMPRESSION MODEL WHOSE CORE IS A FACTOR SPACE

The *KANSEI* model-based system estimates what impression people have when an image or music is input to the system or vice versa. To realize this system, it is necessary to construct a space that handles human impressions and obtains the mapping relationship between these impressions and the physical features of images or music. Once we obtain the space and mapping relationship, we can either retrieve the media from a database according to our impression or convert it (see Figure 1.)

For this reason, we adopt a factor space as a system core. We then explain the advantages of the factor space and the applications of the model.

2.1 Why a factor space?

Adjective pair or adjective scale space has been frequently used for conventional *KANSEI*-based image retrieval systems [11, 2]. Adjective pair space is a space consisting of several adjective pairs as in Table 1 to express human impressions. The advantage of this space is that it is easy to make. The disadvantages include the possibility of biased coverage by adopted adjective pairs for the whole impression; sometimes the same adjective has a slightly different nuance for images and music; and the adjective space has redundancy to express impressions and a problem with orthotropism.
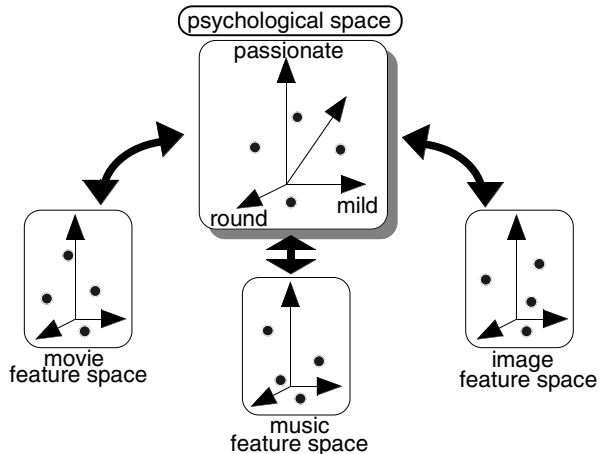

Figure 1: Media retrieval and conversion through a psychological space.

A factor space solves these disadvantages. We can reduce redundancy of the adjective space and obtain universal variables by analyzing subjective data that is distributed on the adjective pair space using the principal component analysis. Another advantage is that it may be easier to express several impressions for media database retrieval or media conversion because the number of dimensions of the factor space is fewer than that of the adjective pair space.

The number of dimensions of the factor space depends on how precisely a source adjective pair space is approximated (see section 3 for concrete numerical data). The balance is adjustable between the approximation precision of the media database retrieval and media conversion and the usability of their systems.

2.2 Application #1: media database retrieval based on impression

Research on database retrieval without keywords has increased. Auto-indexing and other trials have been attempted, especially for image database retrieval [1]. Auto-indexing is a method used to extract image features that characterize the image and use the features as keys.

Another content-based database retrieval method is to use the corresponding relationship between a psychological space expressing impressions and the physical features of the media. The first approach uses an interactive evolutionary computation (EC) that has became popular recently [8, 9]. Lee and Cho's image database retrieval system taken up in section 4 is based on this approach [3, 4]. The

interactive EC-based database retrieval does not construct a psychological space explicitly but uses subjective evaluation or decisions as coordinates of a psychological space that defines impression. The second approach is to retrieve databases by explicitly constructing the psychological space and obtaining the mapping relationship between media features and the psychological space. The method introduced in this section is the second type of media database retrieval.

The first step of our approach is to construct the factor space introduced in section 3 using the SD (semantic differential) method. The flow of the SD method to construct the space is: (1) adjective pairs that express the impression of media are collected, (2) human subjects rate their impression to the given image, music, or other media on the scale for each adjective pair, and (3) coordinates of the adjective pair space, the ratings from (2), are analyzed by a principal component analysis, and an orthogonal space, i.e. a factor space, is constructed. This core factor space is the impression model that expresses how people feel when they see or hear media.

The second step is to obtain physical features of media. There are many such features, such as brightness or frequency space distribution for images, or the fluctuation of pitch or duration of each musical tone, etc.

The third step is to obtain the mapping relationship between the factor space and the physical feature space. Usually, the number of dimensions of physical feature space is bigger than that of the factor space. The mapping relationship can be obtained simply by multiple regression analysis [5]. Generally, the relation is nonlinear and may be expressed using a fuzzy system (FS) or a neural network (NN). Once we obtain the mapping relationship, any new image or music can be mapped to one coordinate on the factor space. Specifically, if an image or music is provided to a computer, the computer can output its impression as if it could *be impressed*.

The fourth step is to make a system that obtains the reverse of the mapping relationship obtained in the third step. In the *KANSEI-based* media database retrieval, we give our impression of an image or music to a computer, and make the computer retrieve the image or music. In the interactive EC-based retrieval systems, users evaluate how close retrieved images or music are to the impression in their minds and evaluate the system using

fitness values. Conversely, the coordinate of a factor space is given to the system mentioned in this section. When one point on the factor space is inversely mapped to the physical feature space, it can be mapped to multiple points on the physical feature space because the number of dimensions of the feature space is usually larger than that of the factor space. This means that there are several images or pieces of music that give similar impressions, which are appropriate. Retrieval systems display these multiple candidates with similar impression to users and let them choose the best one.

This reverse mapping is conducted by combining GA to the FS or NN obtained in the third step (see Figure 2.) The searching data of the GA are images or music from a database, distributed on the physical feature space. The GA fitness value is the distance between the coordinate on the factor space that a user specifies as his or her impression for media retrieval and the coordinate of the image or music mapped from the feature space. Thus, the closest image or music to his or her impression is determined from a media database by the mapping relationship and GA.
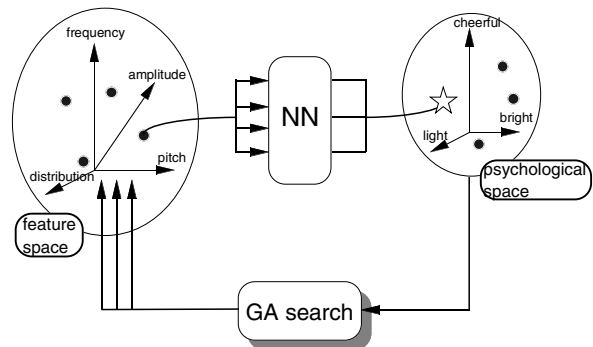

Figure 2: Mapping between a physical space and a psychological space using NN and GA.

2.3 Application #2: media conversion based on impressions

The application introduced in section 2.2 is single media database retrieval. When the mapping and reverse mapping relationships in the third and fourth steps in section 1 for multiple media databases are obtained, we can realize a media converter through the factor space. When we input music into a computer, the computer can display images whose impression is similar to the original music; when we input an image into a computer, the

computer can play music whose impression is similar to the original image. Figure 1 shows this media conversion system.

The key point to realize this converter is to construct a common psychological space from the impressions of several media. Unlike the adjective pair space, the factor space, i.e. a psychological space, is made by reducing the redundancy and by condensing the dimensions of the adjective pair. Therefore, the dependency of the factor space on the media is less than that of the adjective pair space, and can be commonly used for several media. In section 3, we describe how we constructed the common factor space by eliminating adjective pairs that have a different nuance between images and music when adjective pairs are collected to construct the factor space.

3 Construction of a Factor Space

A factor space was constructed using the **SD** method; images were displayed to our subjects, they rated their impressions of each image on a seven-level scale for each adjective pair in Table 1, and the rating data were analyzed by a principal component analysis.

The adjective pairs used to construct the factor space were chosen as follows. First, in a preliminary experiment, we showed **50** images to **3** subjects and let them write down as many adjectives describing their impression as they could. They finally obtained a list of **151** different adjectives. Second, we eliminated adjectives that had similar meanings, that depended on the semantic contents of the images, and that may have depended only on one person's evaluation. We made adjective pairs from the remaining adjectives. Third, we eliminated adjective pairs whose nuances are slightly different for images and music. Since one of our research objectives is to realize a media converter, such adjectives may be useless in constructing a universal factor space for media. Finally, we obtained the **14** adjective pairs listed in Table 1.

Images displayed to our **10** subjects were selected as follows. Given that we already had **3,000** images and their adjective pair ratings from our preliminary experiment, we eliminated similar images using the similarity on the adjective space. These adjectives were just for preliminary use and are different from those in Table 1. Since similar images in the preliminary adjective space also become neighbors in the factor space, we measured the distances among images in the prelimi-

Table 1: 14 pairs of adjectives used to construct a psychological scale space for image. Original words used in the experiment in section 4 are in Japanese.

bright	—	*dim*
vivid	—	*subdued*
clear	—	*fainted*
gaudy	—	*plain*
passionate	—	*dispassionate*
hard	—	*soft*
jaunty	—	*placid*
pure	—	*impure*
warm	—	*cool*
simple	—	*complex*
comical	—	*serious*
emotionally attractive	–	emotionally unattractive
perspectively wide	—	*perspectively narrow*
dry	—	*wet*

nary adjective space and chose **610** representatives from the original **3,000** images.

Our **10** subjects watched and rated each of these **610** images at **7** levels for the **14** adjective pairs in Table 1. Figure 3 shows the interface for this evaluation. We analyzed the obtained 610×10 fourteen dimensional data using the principal component analysis and formed factor spaces. When we approximated the **14** dimensional adjective space with a 3- to 6-dimensional factor space, their accumulated cover rates were **59%, 67%, 73%,** and **77%,** respectively. Figure 4 shows the case of a three-dimensional factor space. Adjectives at the coordinate axes are temporal factor names representative of the original adjectives belonging to each factor.

4 Evaluation of *KANSEI*-Based Image DB Retrieval System

4.1 Image database retrieval system

As an application example of the factor space, we introduce how to choose target image tasks for evaluating an interactive EC-based image database retrieval system [10].

Lee & Cho's image retrieval system evaluated in this section uses Wavelet coefficients as physical image features and retrieves images based on an interactive GA [3, 4]. A retriever evaluates how well each of 12 displayed images matches his or her search criteria and inputs the subjective evaluation values to their system. The image retrieval system conducts GA operation based on the evaluation value as fitness value and retrieves
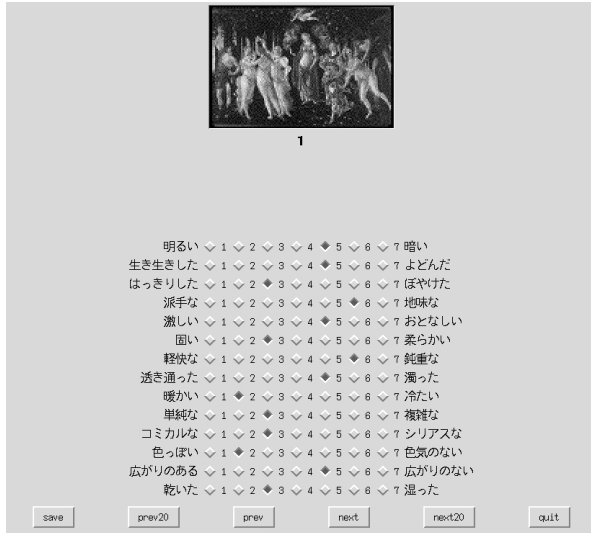
Figure 3: User interface used to obtain a coordinate in the 14-dimensional adjective pair space for 610 images.

Figure 4: The case of a three-dimensional factor space. The names at the axes of coordinates are used in the experiment in section 4.

another 12 images from their image database. They are displayed to the retriever for the next generation search. This process is iterated until a satisfactory image is obtained.

In our subjective test, the same fixed target image tasks and number of searching generations are given to human subjects to control experimental conditions. To compare the interactive GA-based image retrieval system, a random-based image retrieval system whose user interface is identical to that of the GA-based system is prepared.

4.2 Target image tasks from a factor space

When we evaluate *KANSEI*-based image retrieval systems, it is important to determine the target image tasks that widely cover the impression areas without bias, otherwise reliable evaluation is not obtained. Generally, it is difficult to quantitatively show impressions without bias. The factor space in section 3 can offer a solution to this problem. We can measure the unique distribution of target image tasks using the coordinates of the factor space, avoiding biased target image tasks, which provides reliable impartial evaluation.

We chose a three-dimensional factor space from the experimental time point of view. Since there are eight quadrants in a three-dimensional space, we chose four quadrants that are not adjacent to each other (see Figure 4.) Four target image tasks determined by these four quadrants of the factor space are:
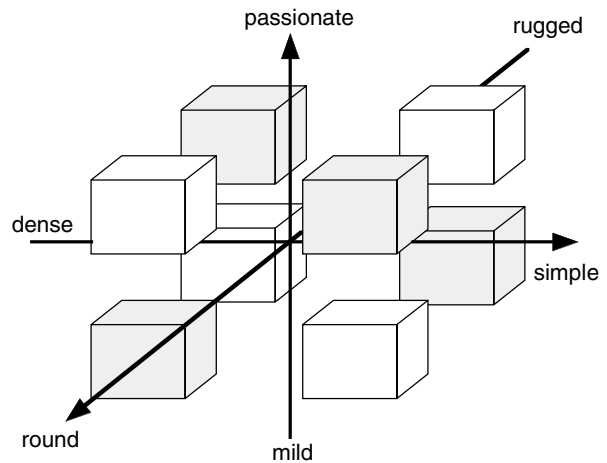
(1) passionate, simple, and round images,
(2) mild, simple, and rugged images,
(3) passionate, dense, and rugged images, and
(4) mild, dense, and round images.

4.3 Experimental condition of evaluation

Twelve subjects were asked to operate two image database retrieval systems and subjectively evaluate the retrieved images. Each subject was given two of four target image tasks and asked to retrieve an image whose impression most closely matched the given target image task. Four target image tasks were not given to avoid subject fatigue and maintain similar experimental conditions.

Two image database retrieval systems were evaluated; one was the Lee & Cho's image retrieval system, and other was a system in which the GA search of Lee & Cho's system was replaced by a random search. The user interfaces of both systems were identical, and subjects were not informed as to which system was which.

In the retrieval experiment, each subject operated the both systems with two target image tasks per system. The number of retrieving generations was 13 per task per system. The image whose subjective evaluation, or fitness value, was the highest for each generation was saved and used in the following subjective test.

In the subjective test, each subject compared images from the two image retrieval systems. The best image of the same generation saved by the retrieval experiment for each system was displayed to the subject si-

Table 2: Significance test of comparison in each generation. * and ** means that the interactive EC-based system retrieved significantly closer images to the given target image task than that of the random-based system with ($p < 0.05$) and ($p < 0.01$), respectively.

generation #	2	3	4	5	6	7
sign test	–	**	–	–	**	*
Wilcoxon test	–	*	–	–	–	**
generation #	8	9	10	11	12	13
sign test	–	–	**	*	–	–
Wilcoxon test	*	–	*	–	*	–

Table 3: Significance test of comparison on target image tasks. See the caption of Table 2 for * and **.

given target image task #	1	2	3	4	total
sign test	**	**	–	*	**
Wilcoxon test	**	**	**	**	**

multaneously. Each subject then evaluated the pair images that he or she retrieved from the retrieval experiment. A five-level subjective rating was given, and the evaluation data were statistically tested using a sign test and the Wilcoxon test [6, 7]. These tests were applied to both comparisons for each generation and each target image task. These tests clarified how the performance of the two systems is different in the searching convergence and depends on the given target image task.

4.4 Subjective test results

The convergence comparison in each generation is shown in Table 2. Since the two systems were initialized with random values, the first generation was not compared. For all cases, it was statistically significant that the interactive EC-based image retrieval system was superior to the random-based system.

Table 3 shows the comparison of target image tasks. For all cases, it was statistically significant that the interactive EC-based image retrieval system was superior to the random-based system, too.

5 CONCLUSION

To allow computers to understand the impressions of images, music, movies, or other media, we introduced an approach that uses a factor space as a model of KANSEI for artistic media. We showed our current research and the perspectives gained from our research. As an application of a factor space essential to our KANSEI model, we evaluated an interactive GA-based image database retrieval system using the factor space.

REFERENCES

[1] Chan, C.-Y. and Pau, L. F., "A survey of access methods for image data," Int'l J. of Software Engineering and Knowledge Engineering, vol.7, no.3, pp.305–319 (1997)

[2] Hayashi, T. and Hagiwara, M., "An image retrieval system to estimate impression words from images using neural network", IEEE Int'l Conf. on Systems, Man and Cybernetics (SMC'97), Orlando, Florida, USA, pp.150–155 (Oct., 1997).

[3] Lee, J.-Y. and Cho, S.-B., "Interactive genetic algorithm for content-based image retrieval," Asian Fuzzy Systems Symposium (AFSS'98), Masan, Korea, pp.470–484, (June, 1998).

[4] Lee, J.-Y. and Cho, S.-B., "Interactive genetic algorithm with wavelet coefficients for emotional image retrieval," 5th Int'l Conf. on Soft Computing and Intelligent/Information Systems (IIZUKA'98), Iizuka, Fukuoka, Japan, pp. 829–832, World Scientific (Oct., 1998).

[5] Moriyama, T., Saito, H., and Ozawa, S., "Evaluation of the relation between emotional concepts and emotional parameters in speech," Trans. of IEICE D-II, vol.J82-D-II, no.4, pp.703–711 (1999) (in Japanese).

[6] Siegal, S., "Non-parametric Statistics for the behavioral sciences," McGraw-Hill, New York, pp.68–75 (1956).

[7] Sprent, P., "Quick Statistics, – An Introduction to Non-parametric Methods –," Penguin Books, Harmondworth, Middlesex, England, pp.85–99 (1981).

[8] Takagi, H., "Interactive Evolutionary Computation: System Optimization Based on Human Subjective Evaluation," IEEE Int'l Conf. on Intelligent Engineering Systems (INES'98), Vienna, Austria, pp.1–6 (Sept., 1998)

[9] Takagi, H., "Interactive Evolutionary Computation – Cooperation of computational intelligence and human KANSEI —," 5th Int'l Conf. on Soft Computing and Intelligent/Information Systems (IIZUKA'98), Iizuka, Fukuoka, Japan, pp.41–50, World Scientific (Oct., 1998).

[10] Takagi, H., Cho, S.-B., and Noda, T., "Evaluation of an IGA-based Image Retrieval System Using Wavelet Coefficients," IEEE Int'l Conf. on Fuzzy Systems (FUZZ-IEEE'99), Seoul, Korea (Aug., 1999)

[11] Yoshida, K., Kato, T., and Yanaru, T., "Image retrieval system based on subjective interpretation," 5th Int'l Conf. on Soft Computing and Intelligent/Information Systems (IIZUKA'98), Iizuka, Fukuoka, Japan, pp. 247–250, World Scientific (Oct., 1998).