Kansei-based image Retrieval with Bayesian Decision Models and Relevance Feedback

Kansei-based Image Retrieval with Bayesian Decision Models and Relevance Feedback

Yanyu Gao*, Zhiqiang Liu (M1), Takashi Uozumi **
* SVBL **Department of Computer Science and Systems Engineering, Muroran Institute of Technology

1. Introduction

As huge amount of image collections in electronic form are available, more and more requirements have been presented for automated image retrieval. So far, four kinds of image retrieval modes have been proposed: text-based image retrieval [1], content-based image retrieval (CBIR) [2], semantic-based image retrieval (SBIR) [3], and Kansei-based image retrieval (KBIR)[4], where KBIR is mainly designed to retrieve images according to users' impression or sensitivity. It has potential applications in the fields of design, advertising, and entertainment.

In this paper, a KBIR system with scenery images as retrieval objects is presented. The whole system framework is shown in Fig.1, which consists of three parts: visual feature extraction, Kansei feature inference and relevance feedback image retrieval.

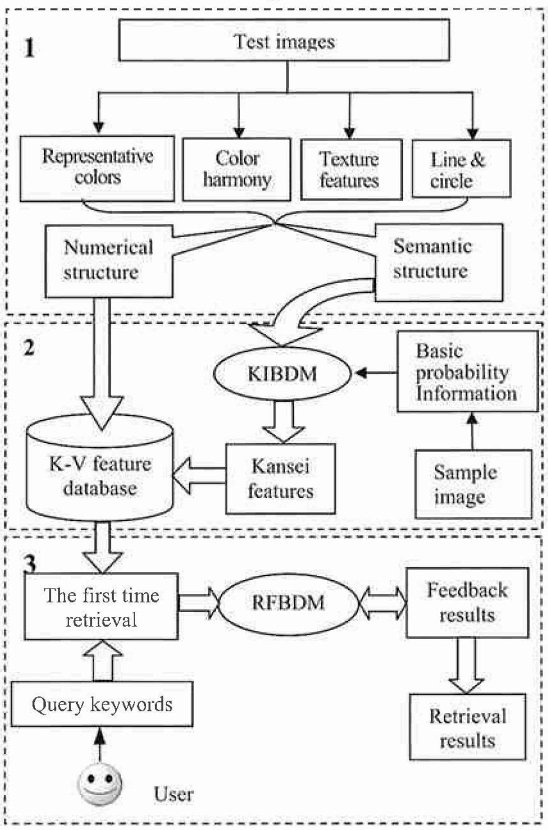

Fig.1 The framework of our KBIR system

In the first part, a set of visual feature extraction methods are proposed to extract color, texture, and shape features of an image from the perceptual viewpoints. Then, two structures are defined to describe these features from semantic and numerical angles respectively. In the second part, we construct the Kansei inference Bayesian decision model (KIBDM) to deduce Kansei feature of an image from its semantic visual content. The numerical visual features of all test images, as well as their Kansei features inferred by KIBDM, constitute the Kansei-Visual (K-V) feature database, which will be used for the first time retrieval. In the third part, the relevance feedback Bayesian decision model (RFBDM) is constructed, by which users can get their expected images by providing feedback information, such as which retrieved images meet or come closer to their goals. Such kind of relevance feedback mechanism can effectively meet personal mentality and preference.

2. Visual Feature Extraction

Our visual feature extraction scheme includes four parts: representative color extraction, color harmony evaluation, texture feature extraction, and line and circle detection.

2.1 Representative colors extraction

Based on the observation of the human visual system, we find that the color keynote of an image plays an important role to human's impression and sensitivity to the image. Therefore, the first visual feature we are interested in is the representative colors and their proportions in an image. In this paper, color histogram is adopted to evaluate color distribution of an image.

Since the HSV color space (hue, saturation, value) corresponds better to how people experience color than other color spaces do, we adopt it as our work space and discretize it into 20 bins corresponding to 20 different colors—red, orange, yellow, green, cyan, blue, purple, magenta, bright red, bright orange, bright yellow, bright green, bright cyan, bright blue, bright purple, and bright magenta. Then we count the pixel number of each bin for each image. Four colors with the biggest area percentages are regarded as the representative colors of the image. An example of color quantization is shown in Fig.2.

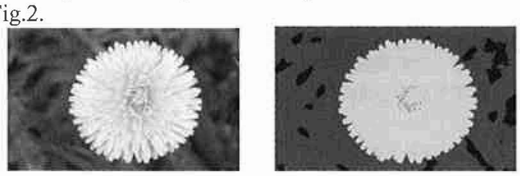

Fig.2 An image and its color quantization result

2.2 Color harmony

Color harmony has long been of interest to researchers in various fields. An objective definition of color harmony was given by Judd and Wyszecki [5] as "when two or more colors seen in neighboring areas produce a pleasing effect, they are said to produce a color harmony". In this paper, we evaluate the color harmony feature based on the color schemes wheel of [6], which defined six color schemes based on the representative colors of an image. We summarized these schemes and gave their harmony evaluation in Table 1.

Table 1. The harmony description and evaluation

	Harmony scheme name	Hue disparity	Harmony description	Ev
1	Monochromatic color scheme	0~5°	Engenders clean, elegant, and soothing effect	1
2	Analogous color scheme	6°~60°	Similar to scheme 1, but offers more nuances.	0.8
3	Complementary color scheme	180°±10°	Intrinsically high-contrast and looks best use a warm color against a cool color	0.7
4	Split complementary color scheme (3 major colors)	165°±5°, 165°±5°, 30°±5°	Provides high contrast without the strong tension	0.8
5	Triadic color scheme	120°±10°	Offers strong visual contrast while retaining harmony and color richness	0.8
6	Tetradic color scheme (4 major colors)	180°±10° 180°±10°	Varied and unbalanced	0.6
7	Otherwise			0.5

2.3 Perceptual texture feature

Texture representation based on human visual perception was first proposed by Tamura et al. [7]. In this paper, we adopt three highly essential Tamura texture—coarseness, contrast and directionality as our texture features.

Coarseness refers to the size and number of the primitives. A coarse texture contains a small number of larger primitives, whereas a fine texture contains a large number of small primitives. Contrast measures vividness of the texture. An image has a high "contrast" if the differences in intensity between neighborhood pixels are large, whereas has a low "contrast" if the differences are small. Directionality refers the shape of texture primitives and their placement rule. A directional texture has one or more recognizable orientation of primitives, whereas an isotropic texture has no recognizable orientation of primitives. Semantically, if the angle of texture direction is within $[0, \pi/4]$ or $[3\pi/4, \pi]$, the image is regarded as having a horizontal or slight-gradient texture direction; if the angle is within $(\pi/4, 3\pi/4)$, the image is regarded as having a steep-gradient or vertical texture direction.

2.4 Line and circular detection

Line slope and circular shape can induce different emotions. For example, an oblique slope evokes dynamism and action, while a flat slope arouses calmness and relaxation. The image with circular shape at center may induce the feeling of romantic easily, because circular shape always associates with romantic objects, such as moon, sun, or flower. Therefore, detecting line slope and circle shape for an image would be helpful to deduce its Kansei features.

Hough transform is a powerful tool for graphic elements detection due to its global vision and robustness in noisy or degraded environment. Generally, four steps are needed to detect lines (circles) using Hough transform: 1) Obtain edge of the detected image using some edge detector (e.g. Canny edge detector); 2) Perform Hough transform on each edge pixel in the image space and accumulate hits for each parameter in the Hough space; 3) Detect peaks in the Hough parameter space; 4) Verify the lines (circles) indicated by the peak parameters.

The parameters of line detection are the normal length r from the origin to the line, and the angle θ between the line and the X-axis. We quantize the angle scope of detected lines into three classes. If $0\leq\theta<\pi/6$ or $5\pi/6<\theta\leq\pi$, the detected line has a horizontal direction; if $\pi/6\leq \theta \leq\pi/3$ or $2\pi/3< \theta \leq5\pi/6$, the detected line has a slight gradient; if $\pi/3\leq \theta \leq2\pi/3$, the detected line has a steep gradient or vertical direction. The parameters of circle detection are the coordinates of the circle center (a, b) and the radius ρ. Examples of line and circle detection are shown in Fig.3.

Fig.3 Examples of line and circle detection

2.5 Visual feature structures

In order to organize visual features obtained above, we build two structures—semantic visual feature structure (SVFS) and numerical visual feature structure (NVFS). SVFS is used to deduce Kansei features and defined as {names of representative colors, area percentage of representative colors, color harmony scheme, coarseness, contrast, directionality, line detection result, circle detection result}. NVFS is used for constructing K-V database and calculating feedback results. It is composed of "HSV values of representative colors, color harmony evaluation, coarseness degree, contrast degree, angle of texture direction, angle of detected line, center and radius of detected circle" and can be represented by a 18-dimension vector: $S_{nvf}=\{f_{h1}, f_{s1}, f_{v1}, \ldots, f_{h4}, f_{s4}, f_{v4}, f_{har}, f_{crs}, f_{con}, f_{dir}, f_{line}, f_{circle}\}$.

Using Fig.2 as an example, its SVFS description is {green, 55.66%, bright yellow, 36.18%, black, 6.62%, yellow, 1.15%, analogous color scheme, no texture, middle contrast, nondirectional, nonline, circle}, while its NVFS description is {[0.2567, 0.5346, 0.2913], [0.1373, 0.9807, 0.9496], [0.2367, 0.8861, 0.1199], [0.1584, 0.8446, 0.3696], 0.8, 0.634979, 0.443357, 0, 0, 0.8156}.

3. Deduce Kansei Features

In this section, a Kansei inference Bayesian decision model (KIBDM) is constructed to deduce general Kansei reflection from visual features. The reason to use Bayesian decision model is that it allows all features to be integrated in common terms of probabilities, although these features are measured by different metrics [8]. In this paper, 5 pairs of Kansei keywords—cold vs. warm; depressed vs. cheerful; placid vs. active; austere vs. romantic, and plain vs. gorgeous are selected to represent human's Kansei.

Before construct the KIBDM, a subjective experiment was performed to help us obtain its basic probability information, in which 40 volunteers (35 male and 5 female) with normal color vision and normal or corrected-to-normal vision attended. 20 color patches and 25 texture images in the size of 100×72 pixels, besides, 35 scenery images in the size of 900×600 (or 600×900) pixels were selected as sample images. Participants watched each image and gave 5 Kansei scores in 7 SD scales. According to semantic visual feature of the 80 images and questionnaire results, we can easily calculate the basic probabilistic information $p(Kfa_i|svf_m)$, where Kfa_i represents the value of the *i-th* pair of Kansei factors, Kfa_i =-3,…,3, with i=1,…,5; svf_m represents the *m-th* related semantic visual features (RSVF) [9] with m=1,…,3 (as shown in Table 2).

Table 2. Visual elements of the five pairs of Kansei keywords

	Kansei	RSVF	Visual elements	RNVF
1	Cool(-) Warm(+)	svf_1	The hue of representative colors	$f_{h1}, f_{h2}, f_{h3}, f_{h4}$
		svf_2	The saturation of representative colors	$f_{s1}, f_{s2}, f_{s3}, f_{s4}$
		svf_3	The luminance of representative colors	$f_{v1}, f_{v2}, f_{v3}, f_{v4}$
2	Depressed(-) Cheerful(+)	svf_1	The saturation of representative colors	$f_{s1}, f_{s2}, f_{s3}, f_{s4}$
		svf_2	The luminance of representative colors	$f_{v1}, f_{v2}, f_{v3}, f_{v4}$
		svf_3	Color harmony	f_{har}
3	Placid (-) Active(+)	svf_1	The saturation of representative colors	$f_{s1}, f_{s2}, f_{s3}, f_{s4}$
		svf_2	Directionality	f_{dir}
		svf_3	Color harmony	f_{har}
4	Austere(-) Romantic(+)	svf_1	Line and circle detection	f_{line}, f_{circle}
		svf_2	The saturation of representative colors	$f_{s1}, f_{s2}, f_{s3}, f_{s4}$
		svf_3	Overall coarseness	f_{crs}
5	Plain(-) Gorgeous(+)	svf_1	The saturation of representative colors	$f_{s1}, f_{s2}, f_{s3}, f_{s4}$
		svf_2	Color harmony	f_{har}
		svf_3	Overall contrast	f_{con}

Based on the basic probability information and principles of Bayesian decision model, we set up our KIBDM. Assuming all visual features we proposed above are conditionally independent, Kansei features of an image Kfa_i can be deduced by maximum a posteriori (MAP) criterion:

$$Kf_i = \arg\max_{j=-3,\cdots,3} \sum_{m=1}^{3} p_j(Kfa_i|svf_m)p(svf_m) \quad (1)$$

where $p_j(Kfa_i|svf_m)$ represents the probability of $p(Kfa_i|svf_m)$ when Kfa_i=j, and i=1,…,5; $p(svf_m)$ represents the weight of RSVF, and is set as $p(svf_1)$=0.5, $p(svf_2)$=0.3, $p(svf_3)$=0.2 respectively. If RSVF is one of HSV color features, its basic probabilistic information is defined as:

$$p_j(Kfa_i|svf_m) = \sum_{n=1}^{N} \lambda_n p_j(Kfa_i|svf_m^n) \quad (2)$$

where λ_n is the area percentage of region n; svf_m^n represents the m-th RSVF in region n; N is the number of segmented regions and defined as N=4 in this paper.

After deducing Kansei features for all test images, we combine these Kansei features Kf_i and numerical visual features S_{nvf} together to constitute the K-V feature database, in which each image is described as a 23-dimension feature vector.

4. Image Retrieval

The K-V feature database reflects the general relationship between visual feature and Kansei semantics. Therefore, we use it only for the first time retrieval. For adjusting the retrieval results to meet user's personal sensitivity and sensibility, relevance feedback mechanism is adopted in our system, which can be considered as a Bayesian classification problem [10].

Assuming the initial retrieval results as S_0, according to user's feedback information, S_0 can be classified into two classes (Ω^+ and Ω^-) by the relevance feedback Bayesian decision model (RFBDM), where Ω^+ includes the images regarded as "relevant", while Ω^- includes the images regarded as "non-relevant". The posterior probabilities of an image X belonging to the Ω^+ can be calculated as:

$$P(\Omega^+|X) = \frac{p(X|\Omega^+)p(\Omega^+)}{p(X|\Omega^+)p(\Omega^+) + p(X|\Omega^-)p(\Omega^-)} \quad (3)$$

where $p(\Omega^+)$ and $p(\Omega^-)$ are the priori probabilities of class Ω^+ and Ω^-, and defined as 0.5 initially. Conditional probability density function $p(X|\Omega^+)$ and $p(X|\Omega^-)$ can be computed as the normalized distances between image X and the centroids of Ω^+ and Ω^-, namely C^+ and C^-:

$$p(X|\Omega^+) = \sqrt{\sum_{q=1}^{3}\left(\frac{x_q - C_q^+}{D(nvf_q)}\right)^2}, \quad p(X|\Omega^-) = \sqrt{\sum_{q=1}^{3}\left(\frac{x_q - C_q^-}{D(nvf_q)}\right)^2} \quad (4)$$

where x_q is the related numerical visual feature (RNVF) of image X, $D(nvf_q)$ is used for normalization, q=1, …, 3. C^+ and C^- are defined as:

$$C^+ = \frac{(\delta_1^+ f_1^+ + \cdots + \delta_i^+ f_i^+)}{(\delta_1^+ + \cdots + \delta_i^+)}, \quad C^- = \frac{(\delta_1^- f_1^- + \cdots + \delta_j^- f_j^-)}{(\delta_1^- + \cdots + \delta_j^-)} \quad (5)$$

where δ_i^+ and δ_j^- represent the weights of "relevant" and "non-relevant" respectively. $f_i^+ = (nvf_{i1}^+, \cdots, nvf_{i3}^+)$ and $f_j^- = (nvf_{j1}^-, \cdots, nvf_{j3}^-)$ are RNVF vectors of those "relevant" and "non-relevant" images; i and j represent the numbers of "relevant" and "non-relevant" images respectively. If RNVF is a kind of HSV color features, e.g. hue feature, then $nvf = \sum_{n=1}^{4} \lambda_n f_{hn}$, where λ_n is the area percentage of region n. The posterior probabilities $p(\Omega^-|X)$ can be computed by $p(\Omega^-|X) = 1 - p(\Omega^+|X)$.

Under the optimal condition, for each image in S_0, if its $p(\Omega^+|X) > p(\Omega^-|X)$, we regard it as "relevant" to user's preference. All "relevant" images constitute the first feedback results Q_1. However, it should be noted that Q_1 are limited to the initial retrieval results S_0, which potentially restricts the retrieval results of this feedback iteration and retrieval scope of next feedback iteration. Therefore, we propose to enlarge Q_1 by drawing its near neighbors in the feature space into its scope.

After t times feedback, the prior probabilities, class conditional PDFs, and posterior probabilities of "relevant" evaluation are described as $p(\Omega^+(t))$, $p(X|\Omega^+(t))$ and $p(\Omega^+(t)|X)$ respectively, where $\Omega^+(t)$ represents the cumulated "relevant" image set from the iteration 1 to t. Since $p(\Omega^+(t))$ and $p(\Omega^-(t))$ are difficult to be estimated, we approximate them by $p(\Omega^{+}(t\text{-}1) \mid X)$ and $p(\Omega^{-}(t\text{-}1) \mid X)$ similar to the idea of [10].

5. Experiments

A prototype system has been developed in Matlab 6.5. Totally 4,104 scenery images are selected as our test images, which include 6 categories: seaside, mountain, sky, flower, city, and firework.

For evaluating our image retrieval system, two commonly used evaluation parameters? precision and recall are calculated and compared. They are defined as *Precision*(K)=C_K/K and *Recall*(K)=C_K/M, where K is the number of retrieval results, C_K is the number of matched images among all retrieval results, M is the total number of matched images in the database. 12 undergraduate/ graduate students with normal color vision were asked to test and evaluate our system by answering how many retrieved images match with their Kansei in the first time process and how many feedback circles are needed to find their satisfying images. The average precision and recall rate to 452 "sky" images are shown in Figs.5 and 6. The feedback times for obtaining the satisfactory images are shown in Fig.7.

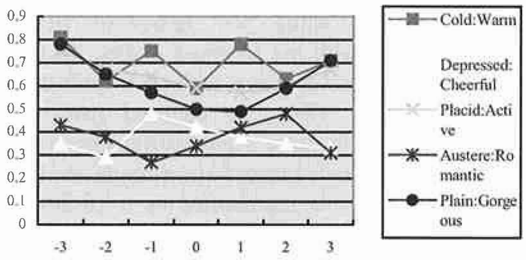

Fig.5 Precision evaluation for Kansei keywords

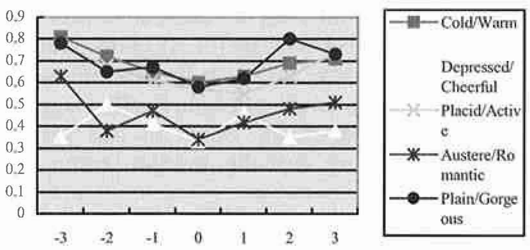

Fig.6 Recall evaluation for Kansei keywords

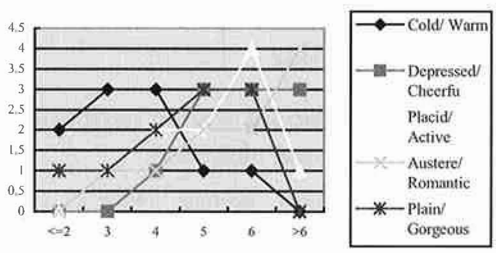

Fig.7 Feedback times for obtaining the satisfactory images

From Figs.5 and 6, we find that for the query keywords cold/warm, placid/active, and plain/gorgeous, our system can obtain preferable precision degrees. However, to the abstract query keywords depressed/cheerful and austere/romantic, its precision degrees are relatively low. Therefore, the relationship between abstract Kansei features and visual features are more complex and need to be investigated further. From Fig.7, we can calculate the average feedback times needed to get the satisfactory images. They are 3.6, 6.1, 5.2, 7.2, and 4.6 respectively. To the query keywords placid/active and austere/romantic, it needs more feedback circles to get satisfactory results.

6. Conclusions and Discussion

In this paper, we develop a KBIR system with scenery images as retrieval objects. Our study mainly focuses on applying Bayesian decision model to infer subjective Kansei from visual feature. Experimental results show that the proposed methods are feasible.

However, what we have done is just the first step in KBIR. The potential research orientation includes: 1) Semantic contents of an image are very important to its Kansei perception to people. For example, rose always represents romance, while defoliated tree always make us depressed. Constructing the relationship between semantic contents and Kansei feature is helpful to improve retrieval results of KBIR. 2) Human's reaction to an image is led not only by biological, physiological and psychological, but also by social and cultural factors. Therefore, further studies on the relationship between human's Kansei and image's visual feature in the fields of psychometrics and sociology are useful for KBIR.

References

[1] Y. Rui, T. S. Huang, and S. Chang, "Image retrieval: current techniques, promising directions and open issues," Journal of VCIR, vol.10, pp.39-62, 1999.

[2] W. M. Smeulders, M. Worring, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," IEEE Trans. on PAMI, vol.22, no.12, pp.1349-1380, 2000.

[3] S. C. Cheng, T.C. Chou, C. L. Yang, and H. Y. Chang, "A semantic learning for content-based image retrieval using analytical hierarchy process," Expert System with Applications, vol.28, pp.495-505, 2005.

[4] Chia-Cheih Cheng, A Research of Image Retrieval System to Assist Color Matching—A case study of fabric images, Master Thesis, National Cheng Kung University, Taiwan, 2002.

[5] Judd DB, Wyszecki G. Color in Business, Science and Industry, Third Edition, Join Wiley & Sons, New York, 1975.

[6] http://www.color-wheel-pro.com/color-theory-basics.html

[7] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception", IEEE Trans. on SMC, vol.8, pp. 460-473, 1978.

[8] J. Luo, A. E. Savakis, and A. Singhal, "A Bayesian network-based framework for semantic image understanding", Pattern Recognition, vol.38, no.6, pp.919-934, 2005.

[9] Yanyu Gao, Zhiqiang Liu, Takashi Uozumi, An Investigation of the Relationship between Visual Features and Kansei Perception for Scenery Images.

[10] L. Duan, W. Gao, W. Zeng, and D. Zhao, "Adaptive relevance feedback based on Bayesian inference for image retrieval", Signal Processing, vol.85, pp.395-399, 2005.