An Emotion-based Image Retrieval System by Using Fuzzy Integral with Relevance Feedback

Joonwhoan Lee, Lei Zhang, Eun-Jong Park
Division of Electronics & Information Engineering, Chonbuk National University, South Korea
E-mail :{chlee, zhlzhanglei, for0511}@chonbuk.ac.kr

Abstract

The emotional information processing is to simulate and recognize human sensibility, sensuality or emotion, to realize natural and harmonious human-machine interface. This paper proposes an emotion-based image retrieval method. In this method, user can choose a linguistic query among some emotional adjectives. Then the system shows some corresponding representative images that are pre-evaluated by experts. Again the user can select a representative one among the representative images to initiate traditional content-based image retrieval (CBIR). By this proposed method any CBIR can be easily expanded as emotion-based image retrieval.

In CBIR of our system, we use several color and texture visual descriptors recommended by MPEG-7. We also propose a fuzzy similarity measure based on Choquet integral in the CBIR system. For the communication between system and user, a relevance feedback mechanism is used to represent human subjectivity in image retrieval. This can improve the performance of image retrieval, and also satisfy the user's individual preference.

Keywords: CBIR, Emotion-based image retrieval, Fuzzy integral, Relevance feedback

1. Introduction

In order to overcome the drawbacks of key words based retrieval, the Content-based Image Retrieval (CBIR) was proposed in 20 century 90 years [1]. CBIR combined some technologies of image processing, pattern recognition, and database. The CBIR system should be able to automatically extract visual features (such as color, texture, shape, spatial information) that are used to describe the contents of an image and to retrieve the desired image from the multimedia database by computing the similarity between images based on the features in vector space. Therefore, a good similarity measure is essential for the effective retrieval in such system [2]. Recently, the Choquet integral is used as the similarity measure in CBIR. In this method, the fuzzy measure is defined over a power set of a given set of features and the Choquet integral is used to aggregate the similarities of features [3]. However, despite high expectations, low-level information such as color, texture and shape, does not sufficiently capture semantic information that is included in human's mind. For example, when a user want to retrieval image with blue sky, the system may retrieve blue sea image that has the same color related feature. Hence, extracting semantic information and employing it for retrieval is important to an emotion-based image retrieval system. In 1996, Eakins [4] partition the semantic information of image into three levels such scene, action and emotion. Most of the current CBIR systems that retrieve images by comparing the similarity of multi-dimensional physical features do not consider the semantic information such as emotional factor and user's subjectivity of the color patterns. In order to make the image retrieval technology is really satisfy to the user's individual requirement, we must think about the influence of emotional factor. Recently, how to express human perception has been one of the most active research topics in image retrieval [5].

We propose an emotion-based image retrieval method in this paper. In this method, user can choose a linguistic query among some emotional adjectives. Then the system shows corresponding representative images that are pre-evaluated by experts. Again the user can select a desired image among the representative images to initiate traditional content-based image retrieval (CBIR). Therefore, not only the same colors or shapes, but also users can retrieve images that provide the same emotional feeling. By this way any CBIR can be easily expanded as emotion-based image retrieval. Figure 1 shows the overall scheme of our proposed emotion-based image retrieval system.

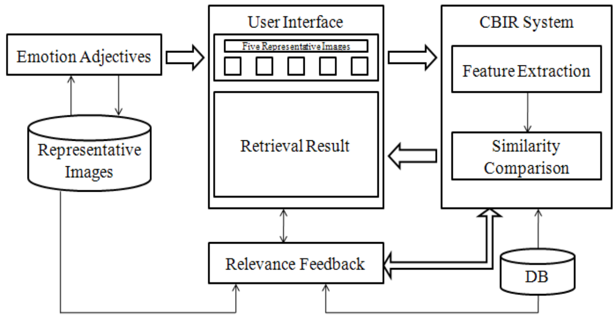

Fig. 1 Proposed emotion-based image retrieval system

In order to increase the number of images that satisfies the user's preference, relevance feedback is necessary. The relevance feedback of our system uses Choquet integral as the similarity measure [3]. In this method, the fuzzy measure is used to represent the relevance weights of features which are updated according to the positive or negative response of user's preference.

The rest of this paper is organized as follows: Section 2 discusses the MPEG-7 standard descriptors for CBIR and how to use the Choquet integral as similarity measure. Section 3 describes how to model the user's feedback using fuzzy sets and how to update the feature relevance values. In Section 4, we show the experimental results with discussion. Conclusion is given in Section 5.

2. Backgrounds

2.1 MPEG-7 visual descriptors for CBIR

In order to construct CBIR system, proper features such as color and texture are needed to be defined and extracted. There are so many features that have been used for CBIR. But successful features should well represent the image contents. Also, the features or descriptors need to be standardized to improve interoperability. This is the reason why we take the MPEG-7 standard in our work. MPEG-7 standard has low level and high level these two type descriptors. Low level descriptors represent the visual construction of the scene including color, shape, texture and motion. High-level descriptors are the special descriptors which can be used for some special domain like human face recognition. In our work we are focused on some low-level descriptors. In this case, each descriptor comes in the form of vectors involving a number of bins and has a recommended similarity measure.[6] In our system we use seven MPEG-7 low-level visual descriptors as follows:

- Color Layout Descriptor (CLD)
- Scalable Color Descriptor (SCD)
- Color Structure Descriptor (CSD)
- Dominant Color Descriptor (DCD)
- Edge Histogram Descriptor (EHD)
- Homogeneous Texture Descriptor(HTD)
- Texture Browser Descriptor (TBD)

2.2 Choquet Integral as similarity measure

The fuzzy integral can be interpreted as a fuzzy expectation, the maximal grade of agreement between two opposite tendencies, or the maximal grade of agreement between the objective evidence and the expectation. In this work, we develop a retrieval system that uses the Choquet integral as a aggregated similarity measure.

Let $X=\{f_1,\dots f_n\}$ be a set of the n descriptors used in the CBIR system, where f_i can represent a color, texture, or shape descriptor in the previous section. Let x^k and x^l be two n-dimensional descriptor vectors that represent the two images to be compared. To use the Choqoet integral as an aggregated similarity measure, we treat the difference in each dimension between x^k and x^l as an information source. The Choquet integral as the aggregated similarity measure is defined as:

$$C_g(x^k,x^l)=\sum_{i=1}^{n}\left[\left|x^k_{(i)}-x^l_{(i)}\right|-\left|x^k_{(i-1)}-x^l_{(i-1)}\right|\right]\times g(\{f_{(i)},\dots f_{(n)}\}) \quad (1)$$

where the notation (i) means that the feature indices have been permuted so that

$$\left|x_{(1)}^k - x_{(1)}^l\right| \le \left|x_{(2)}^k - x_{(2)}^l\right| \le \ldots \le \left|x_{(n)}^k - x_{(n)}^l\right|.$$

Note that each difference in Eq. (1) such as $\left|x_{(i)}^k - x_{(i)}^l\right|$ actually denote the dissimilarity measure according to the descriptors defined in MPEG-7 standard and $g(\{f_{(i)},\ldots f_{(n)}\})$ represent the sugeno's fuzzy measures defined on the power set of descriptors.

2.3 Emotional adjectives and initial fuzzy density

In [7] [8] [9], the author describes 13 emotion pairs of a human subject feels, these emotion pairs are given based on evaluating emotion features of image according to the average brightness and average hue of an image. In our previous research these 13 pairs are shown to be redundant and 3 pairs of adjectives are sufficient via factor analysis. We used the three pairs of those adjectives such as: "dynamic-static", "warm-cool", "heavy-light" in our experiment.

Also our previous research, we showed the MPEG-7 visual descriptors for images are enough to represent the emotions that can be described by the pairs of adjectives. In the research at first we evaluated predefined set of images by human subjects with respect to the adjective image scales. Also we did the similarity-based clustering of the set of images with respect to each descriptor [10]. Then we calculated inclusion degrees of the clusters to the human evaluated partitions. Note that the descriptor with larger inclusion degree means the more effective descriptor for the emotion. Table 1 sows the result of inclusion degrees.

	Cool-Warm	Heavy-Light	Dynamic-Static
CLD	1	1	0.313
SCD	0.911	0.664	0.464
CSD	1	0.862	0.497
DCD	0.639	0.713	0.455
EHD	0.481	0.462	1
HTD	0.395	0.721	0.829
TBD	0.525	0.531	0.42

Table.1 Initial values of descriptors of each pair

For simplicity of the system, we choose 3 descriptors for a pair of emotional adjectives based on Table 1.For example, we use CSD, EHD and HTD to evaluate the similarity when user chooses "dynamic" or "static".

3. Relevance feedback

3.1 Initialization

Let $g^i = g\{f_i\}$, g^i is called fuzzy density, where i means the ith descriptor. All the fuzzy densities g^i are considered have a special value for each emotional adjective pair at the first time retrieval, and are initialized to a value for all 3 descriptors. For example, we set

$$g^1 = 0.497, g^2 = 0.99, g^3 = 0.829 \qquad (2)$$

where, g^1, g^2, g^3 means the value of descriptors CLD, EHD,HTD if the emotional adjective is "dynamic" or "Static". The other values which have marked with different colors are also the same use for the rest of emotional adjective pairs. The system will retrieve similar images that the feature values of these descriptors are closely to the value of query image.

Then we can compute the fuzzy measures recursively as:

$$g\{f_i\} = g^i \qquad (3)$$

$$g\{f_{(i)},\ldots,f_{(n)}\} = g\{f_{(i)},\ldots,f_{(n-1)}\} + \lambda g\{f_{(n)}\} g\{f_{(i)},\ldots,f_{(n-1)}\} \qquad (4)$$

where, λ is computed by:

$$(\lambda + 1) = \prod_{i=1}^{n} (1 + \lambda g^i) \qquad (5)$$

These fuzzy measures are used to compute the Choquet integral in (1).

3.2 Relevance feedback

For the relevance feedback processing, we update the fuzzy density weight for each round of retrieval. Based on that similarity values, the system retrieves the K most similar images. The user then classifies each retrieved images as relevant or irrelevant. We denotethe set of relevant images by X^+, and the set of irrelevant images by X^-. The feature relevance weight update is as follows:

$$g^i = g^i + \eta \Delta g^{i+} - \eta \Delta g^{i-} \tag{6}$$

where

$$\Delta g^{i+} = \sum_{x^k \in X^+} \mu_{SF}(x_i^k) \cdot \mu_{v_1^+}(x^k) \tag{7}$$

$$\Delta g^{i-} = \sum_{x^k \in X^-} \mu_{SF}(x_i^k) \cdot \mu_{v_1^-}(x^k) \tag{8}$$

The value of η reflect the importance of relevance feedback in the retrieval. Assumed that the descriptor vector of the k^{th} retrieved image (x^k) has been labeled v_1^+ if it was relevant, and v_1^- if it was irrelevant. m means the number of relevant or irrelevant images. The $\mu_{v_1^+}(x^k)$ and $\mu_{v_1^-}(x^k)$ are the membership functions of similarity and dissimilarity respectively. We define these two functions as:

$$\mu_{v_1^+}(x_i^m) = \frac{1}{1 + \max(0, C_g(x^Q, x^m) - \alpha)} \tag{9}$$

$$\mu_{v_1^-}(x_i^m) = \frac{1}{1 + \max(0, \beta - C_g(x^Q, x^m))} \tag{10}$$

where x^Q is the query image and x^m is the m^{th} retrieved image. The $\mu_{SF}(x_i^k)$ is the membership function of similarity for each descriptor of each retrieved image and query image. It is defined as:

$$\mu_{SF}(x_i^k) = \frac{1}{1 + 10 * \max(0, \left| x_i^Q - x_i^k \right| - \delta_i)} \tag{11}$$

In our experiment, we set δ =0.1, $\alpha = C_g(x^Q, x^1)$, and $\beta = C_g(x^Q, x^{30})$.

These membership functions can be learned offline by analyzing the images in the database, or specified by expert. After updating the feature relevance weights, the system restarts a new iteration of retrieval based on the new fuzzy densities.

4. Experiment and Discussion

In our experiment, we used about 1000 images in the database. For each emotion adjective, we carried out 5 representative images that are pre-evaluated by experts. These images are used as query images for CBIR. For each query image, the system retrieves 30 images that are most similar to the query image by the emotion adjective which is chosen by user. The retrieval result is showed as follows:

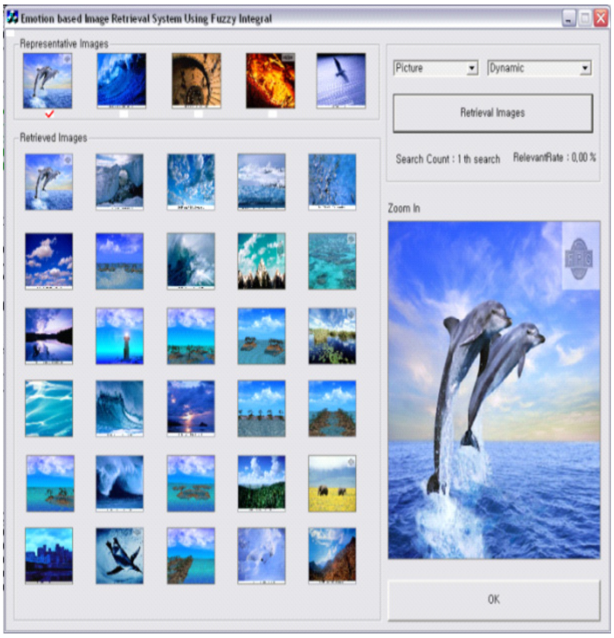

Fig.1 Retrieval result without relevance feedback

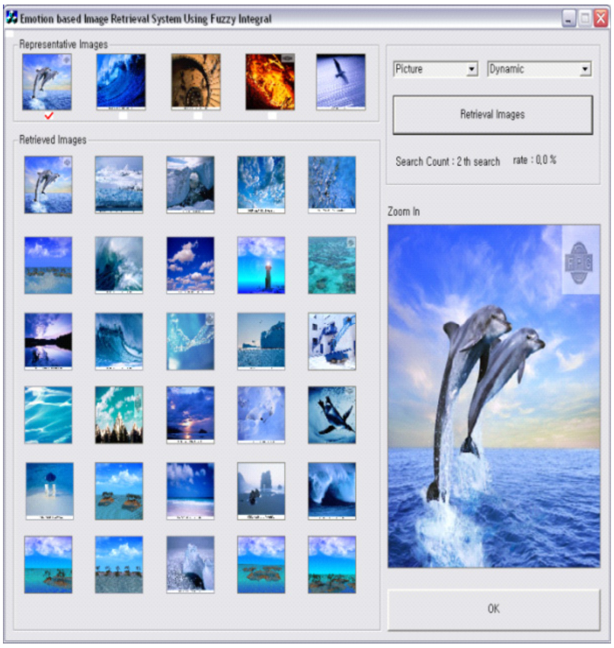

Fig.2 Retrieval result after one relevance feedback

Figure 1 is the first retrieval result without relevance feedback. In figure 1, the emotion adjective is "dynamic", we choose representative image 1 which marked with "√"

as the query image. Figure 2 showed the retrieval result after one time relevance feedback. From the result, we can see the result is better than that without feedback. After the 4 times of relevance feedback, the rate of relevant image can reach to 83%. We do the experiment of all the cases and choose out the best result to do the experiment analysis. Figure 3 shows the change of the number of relevant images with the times of feedback.

From Figure.3 we can see that, the number of relevant images from query image 1, 2 and 5 is more than that from query image 3 and 4. There can be several reasons why the system does not provide uniform results.

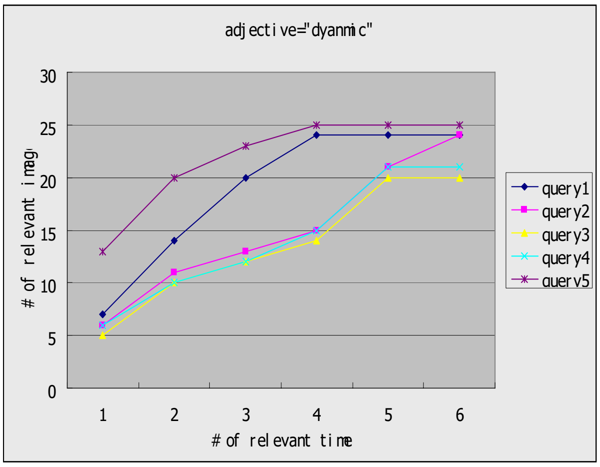

Fig. 3 Relevant degree of all query images

At first there may be not sufficient number of images similar to query 3 and 4 in our database. The lack of similar images produces the early saturation of relevance feedback. The other reason is that query 3 and 4 may not be representative for adjective "dynamic". Even though the query 3 and 4 are pretty dynamic, we can remove them for further retrievals.

Even though any CBIR can be expanded to emotion-based image retrievals and the CBIR by the proposed method, there are several disadvantages in the system. Inherently our system indirectly accesses the images for emotion-based retrieval via pre-evaluated representative images by human expert. Therefore human expert should have quite general emotions for the evaluation. If the expert is abnormal then the user can have difficulties to get the desired retrievals. This problem can be reduced when the possible users can be involved in the pre-evaluation process.

In addition, the value of fuzzy densities become to equal at the last round of relevance feedback. Once all the fuzzy densities are equal, the images of relevance feedback do not change any more and saturate. This can be solved by so called consistency principle of relevance feedback that we are developing now.

5. Conclusion

We propose an emotion-based image retrieval method in this paper. In this method, user can choose a linguistic query among some emotional adjectives. Then the system shows corresponding representative images that are pre-evaluated by experts. Again the user can select a desired image among the representative images to initiate traditional content-based image retrieval (CBIR). Therefore, not only the same colors or shapes, but also users can retrieve images that provide the same emotional feeling. By this way any CBIR can be easily expanded as emotion-based image retrieval. The proposed retrieval system uses an aggregated similarity measure based on Choquet integral. The fuzzy integral can have an excellent performance on image retrieval than other methods which have been discussed in [3]. MPEG-7 descriptors are used in the system because it can provide interoperability.

At present, research on emotion-based image retrieval is still an open topic. It has the affinity of human visual information processing. So that, the integration of psychics, pattern recognition, image processing and information retrieval will be necessary direction for emotion-based image retrieval.

References

[1] S.K.Chang, A.Hsu, "Image information systems: where do we go from here?", IEEE Trans. on Knowledge and Data Engineering , Vol.5, No.5, pp.431-442, 1992.

[2] Hichem Frigui, "Interactive image retrieval using fuzzy sets", Pattern Recognition Letters, Vol.22, Issue 9, pp.1021-1031, July, 2001.

[3] Junli Li, Zherr Chi, Gang Chen, "Image Retrieval Based on Sugeno Fuzzy Integral", IEEE International Conference on Image and

Graphics, 2004.
[4] Eakins J P. "Automatic image content retrieval- Are we getting anywhere?", In Proc. of Third International Conference on Electronic Library and Visual Information Research, pp.123-135, 1996.
[5] Yu Wang, MingYue Ding, Chenping Zhou, Ying Hu, "Interactive relevance feedback mechanism for image retrieval using rough set", Knowledge-Based Systems, Vol.19, pp.696-703, July, 2006.
[6]Tomasz Kaczmarzyk, Witold Pedrycz, "Content-based image retrieval: an application of MPEG-7 Standard and Fuzzy C-Means", IEEE,2006.
[7] Jinsub Um, Kyoungbae Eum, Joonwhoan Lee, "A Study of the Emotional Evaluation Models of Color Patterns Based on the Adaptive Fuzzy System and the Neural Network", Vol.27, No.3, 2002.
[8] Joonwhoan Lee, Young-Min Cheon, Soon-Young Kim, Eun-Jong Park, "Emotional Evaluation of Color Patterns Based on Rough Sets", IEEE International Conference on Natural Computation, 2007.
[9] Hun-Woo Yoo, "Visual-Based Emotional Descriptor and Feedback Mechanism for Image Retrieval",
Journal of Information Science and Engineering, Vol.22, pp.1205-1227, 2006.
[10] Mark Girolami, "Mercer Kernel-Based Clustering in Feature Space", IEEE TRANSACIONS ON NEURAL NETWORKS, Vol.13, No.3, pp.780-784, May, 2002.