Affect-Based Retrieval of Landscape Images Using Probabilistic Affective Model

Yunhee Shin[1], Eun Yi Kim[2], and Tae-Eung Sung[3]

[1] Department of Patent Valuation & Licensing, Korea Invention Promotion Association, Seoul, Korea
[2] Visual Information Processing Lab., Konkuk University, Seoul, Korea
[3] Department of Technology Commercialization Information, Korea Institute of Science and Technology Information, Seoul, Korea
yhshin827@gmail.com, eykim@konkuk.ac.kr, ts322@kisti.re.kr

Abstract. We consider the problem of ranking the web image search using human affects. For this, a Probabilistic Affective Model (PAM) is presented for predicting the affects from color compositions (CCs) of images, then the retrieval system is developed using them. The PAM first segments an image into seed regions, then extracts CCs among seed regions and their neighbors, finally infer the numerical ratings of certain affects by comparing the extracted CCs with pre-defined human-devised color triplets. The performance of the proposed system has been studied at an online demonstration site where 52 users search 16,276 landscape images using affects, then the results demonstrated its effectiveness in affect-based image annotation and retrieval.

Keywords: Affect-based image retrieval, probabilistic affective model, mean-shift clustering, color image scale.

1 Introduction

With increasing the importance of affective computing, it becomes necessary to retrieve and process images according to human affects or preference, as even same categorized image can be differently interpreted depending on mood and affects. In particular, among several image domains such as photographic, medical, and artistic, it is very important to use affective meanings in landscape images.

However, judging such affective qualities of images is not easy task as affective meanings are always hidden, i.e. there is no direct mapping from the image to the meanings. Thus, to discover the hidden affective meanings from observable visual image features is a key step towards an affect-based image retrieval.

Generally, images provide color, texture, shape, and pattern information. Thus, various studies have been proceeded to investigate the relationships between these visual features and human affects and to identify certain visual features that predict human affects [1-4], and some retrieval systems have been developed [1,3,4].

M. Kurosu (Ed.): Human-Computer Interaction, Part V, HCII 2013, LNCS 8008, pp. 362–371, 2013.

Datta et al. investigated the relationships between several visual features and the aesthetic quality in photographic images [1], where color, texture and color compositions were extracted. Based on those relationships, they have developed a classification model to assign aesthetic quality of a photographic image, thereafter have applied this model to Photo.net, where is photo sharing community site on Web. In [3], to predict 8 affective classes such as anger, despair, interest, joy, pleasure, pride and sadness from landscape images, they used the dominant colors of an image, and developed multiple linear regressions to establish the mapping between such colors and 8 affective classes. From them, the affective classification system was implemented on Web. In [4], they investigated relationship between visual features, such as color and pattern, and human affects. With these features, they automatically predict human affects associated with a textile image using machine learning algorithms. And, they have implemented textile retrieval system on Web.

Among various visual features above-mentioned, the most widely used feature is color including dominant colors and color compositions [1-4]. Then, it is hard to represent various affects, based solely on color, so this study focuses on using color compositions that constitute an image.

In this paper, a novel affect-based retrieval of landscape images using a Probability Affective Model (PAM) is presented. The PAM is developed for automatically predicting the affects from color compositions of images, and it is used for annotating Web images in our retrieval system. Our system was tested with 16,276 online landscape images the affective judgments of 52 users on an online demonstration site. Then, it produced the performance of 85.22% in annotation and that of 62.5% in retrieval. Consequently, these results proved the potential of the proposed system in affect-based annotation and retrieval of web images.

2 Image Annotation

For affect-based image retrieval, it should be first performed to annotate a given image using human affects. In this section, we define affective classes and then introduce PAM to automatically predict human affects from visual image features.

2.1 Affective Classes

For affective classes to annotate photographic images, this study defines 15 affective classes based Kobayashi's vocabularies, where they selected 180 affective adjectives that widely used in applications such as visual media, design and art [2].

To find representative words among such 180 adjectives, a survey on Yahoo! search results was performed. First, we collected 8,300,000 landscape images from Yahoo!, and investigated affective adjectives to be used in annotating images. Then, the affective adjectives were ranked according to the frequencies of occurrence in image tags. Based on these results, we re-defined 15 representative affects: {pretty, colorful, dynamic, gorgeous, wild, romantic, natural, graceful, quiet, classic, dandy, majestic, pure, cool and modern}.

2.2 Probabilistic Affective Model

A Probabilistic Affective Model (PAM) is used to annotate a given image using affective features. Let denote $W = \{w_1, w_2,...,w_L\}$ is affective classes, where L is 15. As such, the affective features belonging to an image i are denoted by a 15-D vector $e(i) = (e_{i,1},..., e_{i,L})$, where $e_{i,m}$ means the probability that the image i belongs to the corresponding m_{th} affective class.

Given an image i, each probability of affective feature $e_{i,m}$ is computed as follows:

$$e(i) = \left(e_{i,1}, e_{i,2}, \dots, e_{i,L}\right) = (p(w_1|i), p(w_2|i), \dots, p(w_L|i)) \tag{1}$$

Generally, an image i is composed of N regions, that is, $i = \{R_1, R_{2,...} R_N\}$, each of which can be described by the visual features extracted from that region. In this work, the region is described by color compositions, λ_j. The sense of "color composition" includes both the characteristics of separate colors that occur in images and the organization for combining these parts into a whole.

Then, we assume that the color composition of a whole image is obtained by combining the respective color compositions of the segmented regions, and that the color compositions among regions are independent. Accordingly, m_{th} feature of an affective feature vector $e(i)$, $e_{i,m}$ in Eq.(1) can be rewritten as follows:

$$\begin{aligned} e_{i,m} &= p(w_m|i) = p\big(w_m\big|(R_1, R_2, \dots, R_N)\big) \\ &= p\big(w_m\big|(\lambda_1, \lambda_2, \dots, \lambda_P)\big) = \prod_{j=1}^{J} p(w_m|\lambda_j) \end{aligned} \tag{2}$$

Consequently, $p(w_m|\lambda_j)$ indicates the probability for a color composition λ_j to be mapped onto an affective word w_m.

The process to compute the PAM consists of following three steps:

- Image segmentation using mean-shift clustering
- Color composition extraction through RAG analysis
- Affect mapping onto human-devised color compositions

2.2.1 Image Segmentation

One of the classical image segmentation approaches is using clustering algorithm. After performing color quantization to transpose from RGB color space into 130 basic colors, we segment the mean-shift clustering algorithm [5].

2.2.2 Image Segmentation

In this module, the color compositions between segmented regions are extracted. Since considering those all combinations is very computationally intensive, we firstly select more influential regions having the larger importance than other regions, which is defined as *seed regions*.

We assumed that regions that have larger areas and are closer to the center of the image are more important than others. Thus, the importance, $\phi(R_j)$, is assigned using a criterion based on its area $A(R_j)$ and its Gaussian distance to the image center, $G(R_j)$, such as $\phi(R_j) = A(R_j) \times G(R_j)$, where $A(R_j) = \frac{area\ of\ R_j}{width \times heigh}$ and $G(R_j) = \mathcal{G}_d(\mu, \Sigma) = \frac{1}{2\pi|\Sigma|} e^{-\frac{1}{2}(d-\mu)^T \Sigma^{-1}(d-\mu)}$.

Regions on an image are ranked in decreasing order of their importance values, and then the top M regions among all regions are selected as seed regions, $S(i)$, using threshold T_1.

Thereafter, the RAG is draw from the $S(i)$, and the analysis between seed regions and their adjacent regions is performed. Let denote $\mathcal{C}(\cdot)$ is function to compute color composition between seed region S_j and $\eta(S_j)$, where $\eta(S_j)$ is the set of adjacent regions of S_j. Then, a color composition λ_j is described by a triplet of colors of S_j and $\eta(S_j)$,

$$\Lambda(i) = \{(\lambda_1, \lambda_2, \dots, \lambda_P) \mid \lambda_j = \mathcal{C}(S_j, \eta(S_j))\} \tag{3}$$

Accordingly, in case of a seed region with four neighbors, there are ${}_4C_2$ cases when selecting secondary colors. By summing each color compositions for all seed regions, the $\Lambda(i)$ is finally obtained.

2.2.3 Affect Mapping

We predict the affective features from color compositions obtained from an image. For this, all of color compositions in $\Lambda(i)$ is mapped onto 1170 human-devised color compositions, Ξ, which is developed by Kobayashi [2].

$$\Xi = \{(\xi_1, w_1), (\xi_2, w_2), \dots, (\xi_K, w_K)\} \tag{4}$$

A pair (ξ_k, w_k) is composited of triplet $\xi_k = \{C'_{k,0}, C'_{k,1}, C'_{k,2}\}$, and affective class w_k $(w_k \in W)$ which is manually assigned to the corresponding triplet, ξ_k.

Using Mallow distance [6], each color composition $\lambda_j = \{C_{j,0}, C_{j,1}, C_{j,2}\}$, is compared with triplets, ξ_k in Ξ.

$$Sim(\lambda_j, \xi_k) = -min \sum_{x=0}^{2} \sum_{y=0}^{2} \phi(R_{j,x}) \cdot \left|\left|C_{j,x} - C'_{k,y}\right|\right|^2 \tag{5}$$

Here, $C_{j,x}$ and $C'_{k,y} (x, y = 0,1,2)$ are colors of λ_j and ξ_k, respectively. The importance value $\phi(R_{j,x})$ of the region with the color $C_{j,x}$ is used as weight for this distance.

A color composition, λ_j, is projected to its nearest neighbor with the highest similarity in all of ξ_{k*}, and then k^*_{th} affective class is assigned to λ_j and the reflection, finally $p(w_{k^*}|\lambda_j)$, is calculated based on its importance and its similarity.

$$k^* = arg\max_k Sim(\lambda_j, \xi_k), (1 \le \xi_k \le 1170)$$
$$p(w_{k^*}|\lambda_j) = \frac{\phi(S_j)}{|Sim(\lambda_j, \xi_{k^*})|}, \; (w_{k^*} \in W) \quad (6)$$

While scanning a whole of $\Lambda(i)$, $p(w_{k^*}|\lambda_j)$ are accumulated for affective classes, w_{k^*}, thus the $e(i)$ is computed.

3 Image Retrieval

We developed retrieval system of landscape images using human affects. When given a query, q, our system retrieves the results ranked according to similarities between query q and images in database. Then, for the comparison with query, the images in DB were annotated by PAM. Thus, these images were represented as $E = \{e(i) = (e_{i,1}, e_{i,2}, \ldots, e_{i,L}) |\; i = 1, \ldots, n\}$, where n and L are the total number of images in DB and that of affective classes, respectively.

Accordingly, when searching the images, the query q, given by a user, is first transformed to an affective feature vector $e(q)$, and then, the $e(q)$ is compared with all of affective vectors in the E.

In this work, the similarities between images and query are calculated using the cosine similarity. Thus, the similarity between query q and image i is

$$s(q, i) = Cos(e(q), e(i)) = \frac{\sum_{m=1}^{L} e_{q,m} \times e_{i,m}}{\sum_{m=1}^{L} e_{q,m}^2 \times e_{i,m}^2}, \quad (e(i) \in E) \quad (7)$$

The $e_{q,m}$ and the $e_{i,m}$ are the m_{th} affective feature values of query q and image i from DB, respectively. The images are ranked based on these similarities $s(q,i)$, where $i = 1,\ldots,n$. Thus, the higher the similarity is, the more relevant the image is.

Figure 1 shows the scene of our retrieval system on Web. In our system, a user can submit the query using either affective class or example image. In case of text query to denote m_{th} affective class, w_m $(w_m \in W)$, only the affective feature corresponding to m_{th} affective class has non-zero value, and others have zero values. For example, when selected a ‘romantic’ class as query q, the $e(q)$ is represented as $e(q) = (0,0,0,0,0,1,0,0,0,0,0,0,0,0,0)$. On the other hand, for the query by example, the query image q is transformed to an affective vector $e(q)$ by the PAM. Thereafter, for the respective query type, the cosine similarity is performed with images in DB and the relevant images are retrieved. In addition, our system recommends some images that are mostly selected by users within a certain period of time, as shown in below of Figure 1.

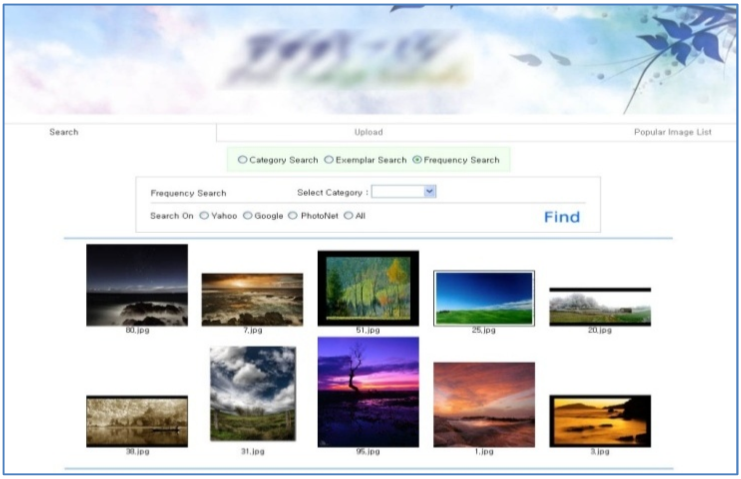

Fig. 1. The scene of our retrieval system on Web

Consequently, through such query schemes, a user can easily search more preferable images.

4 Experiments

To assess the effectiveness of our system, we first assess performance of PAM and then evaluate the retrieved results.

4.1 Image Collection

In this study, two different DBs were selected. The first is image DB from Photo.net, which is one of large online photo sharing community. The second is image DB from Yahoo! search. To collect landscape images from those sites, ten queries to describe the landscape were used such as *coast, desert, field, forest, lake side, mountain, snow scene, sky, sunset and waterfall.*

From Photo.net, 2,000 images were downloaded, 32,280 images were collected using Yahoo! openAPI. Consequently, 34,280 images were used in experiments.

For a complete evaluation of the proposed system, users' affective judgments are needed. For the user study, 52 participants (16 female and 36 male) were asked to participate. It was conducted through a survey system, which was implemented on the Web.

However, as there is wide variability in people's individual feelings, it is very difficult to establish the ground-truth data for affects. As such, we need to consider the varieties between people's sensitivities. Moreover, even the same person can make different judgments according to the mood, context, motivation, etc. Thus, we considered issues of variability within the same person and between persons. To deal with these varieties within a person and between persons, the repeated subject evaluations are adapted during user study and the fuzzy system is used as illustrated in [4].

Through fuzzy system, only the 16,272 images among 32,380 images were used as experimental data, which was categorized into positive and negative groups for the respective affects.

4.2 Experimental Results

The annotation of the PAM was compared with the users' affective judgments, that is, the ground-truth through fuzzy rules. They were categorized into positive and negative ones for the respective affects. Meanwhile, the PAM has a real value ranged 0 to 1, so a threshold value was applied on the output of PAM, for classifying the image into positive and negative one each affective class. This threshold value is set to 0.4 by experiments. Thus, if the output of PAM of certain affective class was larger than 0.4, the image was classified as relevant to the corresponding affective class, otherwise it was classified as irrelevant one.

Table 1. Performance of our method for two DBs

Affective classes	# of images		Photo.net		Yahoo!	
	Photo.net	Yahoo!	Recall	Precision	Recall	Precision
Pretty	81	925	82.72	97.50	93.25	75.00
Colorful	83	995	91.57	95.34	89.31	68.00
Dynamic	81	1,031	88.75	80.00	82.06	67.88
Gorgeous	85	1,029	90.59	89.47	87.30	66.39
Wild	93	1,130	84.62	80.00	89.62	73.33
Romantic	73	1,076	91.78	85.71	89.82	68.72
Natural	88	1,131	81.71	91.89	70.26	80.29
Graceful	81	930	93.83	88.89	71.27	66.67
Quiet	92	909	91.30	83.33	70.16	69.13
Classic	77	879	94.67	71.43	70.46	62.36
Dandy	83	893	86.75	76.67	70.36	65.40
Majestic	89	995	92.13	90.00	80.14	79.11
Pure	91	1,099	90.11	84.62	85.69	80.85
Cool	91	1,122	90.11	88.18	88.00	84.93
Modern	88	856	95.51	76.67	72.68	56.92
Average	1,276	15,000	89.74	85.31	80.69	71.00

Table 1 summarizes the performance of the proposed method in annotating the Photo.net and Yahoo! image DBs, respectively. When comparing the results for DBs, the proposed method showed a better performance of Photo.net DB than that of Yahoo! DB, which was likely caused by a difference in the quality of the images between the two DBs.

When comparing the images in the respective DBs as regards their beauty or artistic merit, the images from the former were rated higher than those from the latter. This difference clearly influenced the results of the human affective judgments, which the variances for the former were much smaller than those for the latter; on average, the values were 1.23 and 3.33, respectively. This difference resulted in greater human ambiguity in the results, thereby decreasing the accuracy of the PAM. Nonetheless, despite the difference in performance for the DBs, the PAM produced on average a recall of 85.22% and precision of 78.16%.

Fig. 2. Top-12 ranked results from our system: (a) for text query “cool” and (b) for query by example image

Top-12 ranked results of our retrieval system for respective query type are shown in Figure 2.

In practice, the important issue in image retrieval is to filter images according to the highest relevant images for a given query. Therefore, two experiments are used to evaluate the effectiveness of the proposed system.

The first experiment was designed to examine a conservative version of relevance for the ranking results. For this experiment, the top 20 images from the proposed system were compared with the top 20 images from the ground-truth. Figure 3 shows the number of intersections between them for 15 affective classes for each dataset.

As shown in Figure 3, the proposed system produced on average a relevance accuracy of 63% and 62% for the two DBs, respectively. In the case of annotation, a significant difference was observed between the results for the two DBs, whereas for retrieval only a slight difference was observed between the two DBs. This was because the goal of retrieval, unlike annotation, is not to order the full set of images, but only to select the best ones to show. That is, the proposed system provided generally satisfactory search results for the user.

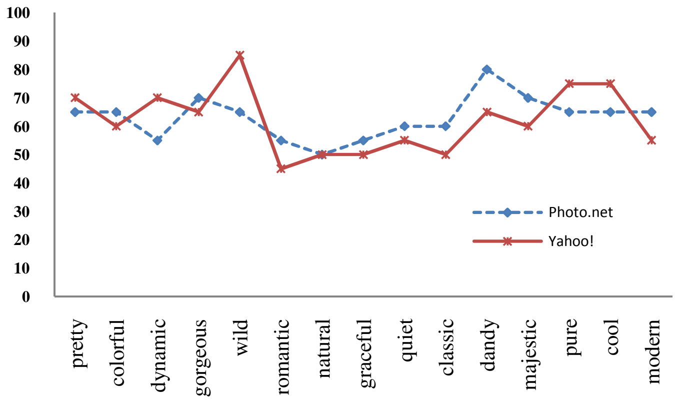

Fig. 3. Relevance graphs at top twenty ranked images

Table 2. Irrelevant images for two DBs

	Photo.net DB	**Yahoo! DB**
Top-3	**0.00**	**0.13**
Top-5	**0.20**	**0.93**
Top-10	**0.93**	**2.73**
Top-20	**3.13**	**5.93**

Meanwhile, minimizing the number of irrelevant images is also important for image retrieval [7]. For this, the second experiment to evaluate the performance of the proposed system is performed by counting the number of irrelevant images in the top 20, top 10, top 5, and top 3 ranked images for each DB. Table 2 shows the results. Among the top 20 images, 3.13 and 5.93 irrelevant results were produced on average for each DB, respectively. However, when looking at the top 3 images, the number of irrelevant images dropped to 0.0 and 0.13, respectively. Although there were some variances in the number of images shown, the proposed retrieval system achieved a good performance.

Consequently, the experimental results demonstrated that the proposed system can be successfully used in affect-based annotation and retrieval of photographic images.

5 Conclusion

In this paper, we suggested retrieving landscape images using human affects. For this, we presented the PAM to automatically predict various human affects from image and used it as annotation process in our retrieval system. Our system was tested on 16,276 online landscape images with the affective judgments of 52 users. Then, it produced the performance of 85.22% in annotation and that of 62.5% in retrieval. To improve the accuracy of the system, we intend to explore about the incorporation of other features such as textures or shapes in future work.

References

1. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Studying Aesthetics in Photographic Images Using a Computational Approach. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 288–301. Springer, Heidelberg (2006)
2. Kobayashi, S.: Color Image Scale. Publishing of Kodansha (1991)
3. Wei, K., He, B., Zhang, T., He, W.: Image emotional classification based on color semantic description. In: Tang, C., Ling, C.X., Zhou, X., Cercone, N.J., Li, X. (eds.) ADMA 2008. LNCS (LNAI), vol. 5139, pp. 485–491. Springer, Heidelberg (2008)
4. Shin, Y., Kim, E.Y., Kim, Y.: Automatic textile image annotation by prediction emotional concepts from visual features. Image and Vision Computing 28, 526–537 (2010)
5. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. 24(5), 603–619 (2002)
6. Li, J., Wang, J.Z.: Real-time computerized annotation of pictures. IEEE Trans. Pattern Anal. Mach. Intell. 30(6), 985–1002 (2008)
7. Jing, Y., Baluja, S.: Pagerank for product image search. In: ACM WWW 2008, pp. 307–316 (2008)