A Statistic Approach for Photo Quality Assessment

Li-Yun Lo
Department of Computer Science and Information Engineering
National Kaohsiung University of Applied Sciences
Kaohsiung, Taiwan, R.O.C.
liyun.lo0914@gmail.com

Ju-Chin Chen*
Department of Computer Science and Information Engineering
National Kaohsiung University of Applied Sciences
Kaohsiung, Taiwan, R.O.C.
jc.chen@cc.kuas.edu.tw

***Abstract*—This study proposes a photo quality assessment based on the spatial relations of image patches. In order to investigate the components of high-quality photos, the image is decomposed into patches based on the color information. Then the color moment and histogram of oriented gradients (HOG) are extracted for the feature representation. Because the diverse types of photos, the photo with the segmented patches is assigned to a subtopic before further modeling. Different from the prior researches which model the spatial relations of image patches obtained from high quality photo, in our work the negative models are learned from the low quality photos as well to provide more discriminate assessment results. Note that the spatial information of location and size of image patch is modeled by Gaussian mixture model (GMM), and the likelihood probabilities in accordance with the positive and negative context models are integrated as the assessment score. The experimental results demonstrates that the usage of the low-quality photos can provide the significant improvement and the proposed system have the promising potential for the photo quality assessment.**

Keywords-Color moment; HOG; Context modeling

I. INTRODUCTION

With the progress of digital cameras, photography has become more popular. Some issues about the capability improvement of digital camera have been attracted, including automatic focus, white balance, face detection, etc. Recently, more and more photographers focus on how to shoot high-quality or pleasing photos. However, the aesthetic feeling is subjective and it is hard to establish standards for the quality assessment. Although the feeling is subject, there is still large difference between photos captured by the professional and amateur photographers, as shown in Fig. 1.

Several systems [11, 12, 17, 21] are proposed based on the visual attention and the composition rules to assess the photo quality. Thus the region of interest (ROI) or the visual attention is a key issue for the photo quality [1, 2, 9, 14, 20, 21]. On the other hand, the works [3, 13, 15, 22] analyzed the spatial context of individual objects (or patches) in the corresponding photo; however, the relations between objects are discarded. In 2010, Cheng *et al.* [4] proposed the context model to learn the spatial distribution of features and automatically search the professional view in the photo.

In this study an automatic photo quality assessment system is proposed and it can evaluate the quality of photos taken ordinarily. Motivated by the work [4], not only the spatial context of individual object but also the spatial relations of pair of objects are modeled. Moreover, the high- and low-quality photos are collected and modeled simultaneously to assess the photo quality more discriminatively.

The reminder of this paper is organized as follows. In Section 2, we introduce the proposed system and the context modeling. Then the system performance is evaluated and analyzed in Section 3 and we make a conclusion in Section 4.

II. STATISTICAL CONTEXT MODELING FOR PHOTO QUALITY ASSESSMENT

As shown in Fig. 2, the flowchart of the training and the test process of the proposed photo quality assessment system, the training process commences by image segmentation. The color and texture features for each patch are subsequently extracted. In order to reduce the variances of photo contexts, *k*-means is used to cluster the extracted features into several subtopics. Then for each subtopic, the contexts of spatial relations are modeled separately. Note that not only positive photos, the negative photos are also used in the proposed system. In the test process, as shown in Fig. 2(b), the patches of the test photo are obtained using the same processes as the used in the training process. Then the likelihood probabilities of the individual patches are evaluated and integrated for the photo quality assessment.

Fig. 1. Photos with higher scores (left) and lower scores (right) on the website, http://www.dpchallenge.com/

978-1-4673-2588-2/12/$31.00 ©2012 IEEE

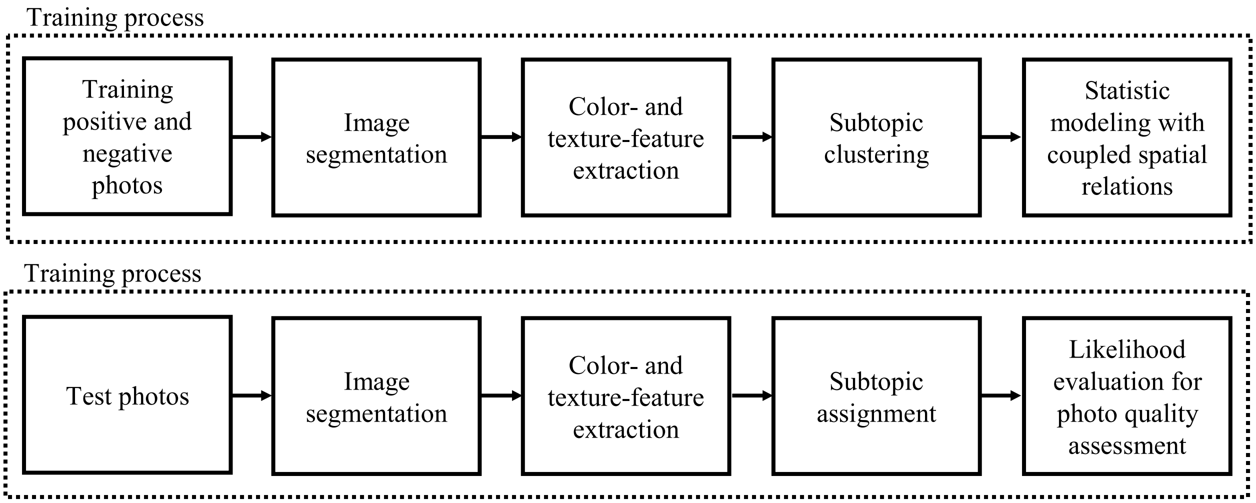

Fig. 2. Flowchart of the proposed photo quality model: (a) Training process (b) Test process.

A. *Image segmentation and Feature extraction*

In order to analyze the ingredients of high-quality photos for quality assessment, each training positive and negative photo is firstly segmented [8] into atomic patches. Note that the image is not resized to avoid distortion. Additionally, to avoid the over-segmented results, the minimum size of patch is constrained, and each photo can be decomposed into 30 to 60 atomic patches on average. Fig. 3 illustrated the segmentation results.

Based on the segmentation results, the color and the texture features are extracted to represent these atomic patches. For the color features, the color moment [16, 18, 19], including mean, standard deviation and skewness, are used to represent the patches. Note that to satisfy the human visual perception, the YIQ color space is used for color feature extraction. Then three moments can be obtained from each color channel. In short, the color feature vector for each patch is 9-dimension, i.e. 3 moments for each of 3 channels, and it is denoted as f_{ij}^c for the *j*-th patch of the *i*-th image. However, the color statistics provides limited information, and thus the complementary features, i.e. texture features, are introduced. HOG [6] feature is widely used in object recognition because of its rotation invariance property, which is obtained by evaluating the histograms of patch gradient orientations in a dense grid. In our study, each image patch is divided into 4×4 regions, and a histogram of 8 gradient orientations is accumulated over the pixels within the corresponding regions. Then the texture feature is combined into a 128-dimension vector and denoted as f_{ij}^H for the texture representation of *j*-th patch of the *i*-th image. Finally, by concatenating the color and the texture features, each atomic patch can be represented by a 137-dimension vector.

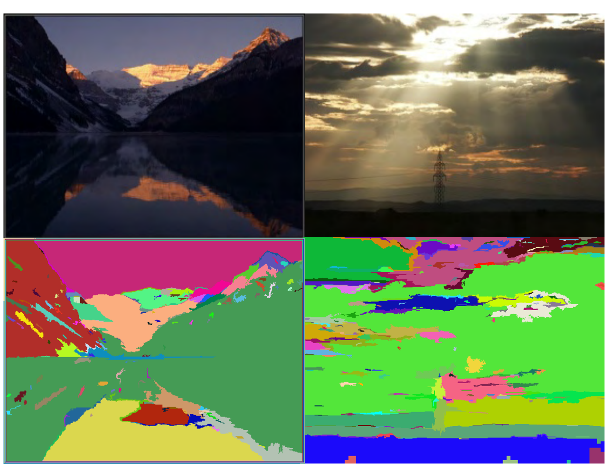

Fig 3. The original images (top row) and the color segmentation results (bottom row). Each image is segmented into 30 to 60 image patches.

B. *Subtopics clustering*

Since the diverse contexts of photos would results in difficulty in context modeling, all positive photos are grouped into several subtopics before further modeling. The *k*-means method [10] is firstly applied to all patches of positive photos. Hence, each clustering center can be taken as the patch prototype [8] and the number of patch prototypes is set to be 200 in our work. Note that because the dimensions of color and texture feature vector is not consistent, each patch is assigned to its corresponding nearest cluster by

$$l_{ij} = \arg\min_{k}(\|f_{ij}^{c} - c_{k}^{c}\|^{2} + \alpha\|f_{ij}^{H} - c_{k}^{H}\|^{2}) \qquad (1)$$

where α is a constant value to balance the importance of color and texture features, c_k^c and c_k^H are the color and the texture components of the *k-th* cluster, respectively.

After obtaining the patch prototype, each photo is then represented as a histogram by counting the occurrence of each patch as in [7]. Finally, based on the bag-of-feature

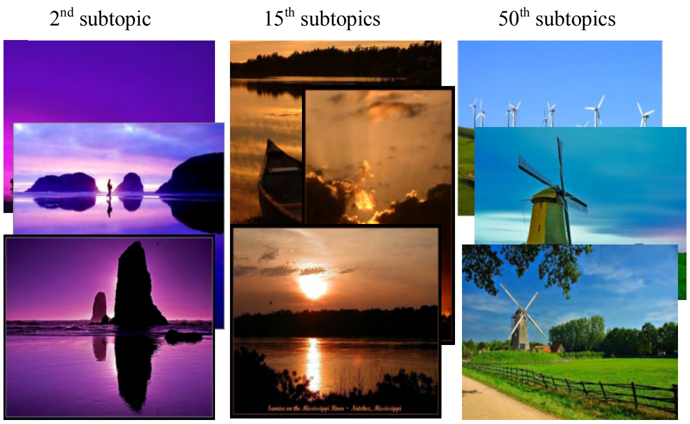

Fig 4. The illustration of clustering results (from left to right): The 2nd, 15th and 50th subtopics.

representation [7] the k-means method is applied to the all positive photos and they are categorized into R subtopics (R=20 in our work). Fig. 4 illustrates the photos for some subtopics. Note that the subtopic clustering is only applied to positive photos; while the negative photos are assigned to the nearest subtopic based on the similarity between its corresponding bag-of-feature representation and each subtopic center.

C. Statistical Context Modeling via Copuled Spatial Realtions

To analyze the ingredients of high- and low-quality photos in each subtopic, the spatial information [4] of each patch in the corresponding photo are obtained and devoted as $x_i^s = (u_i^s v_i^s, s_i^s)$ where (u_i^s, v_i^s) is the coordinate of center point of the i-th patch within the s-*th* subtopic and s_i^s is the ratio of patch size to the photo size.

For each subtopic, Gaussian mixture model (GMM) is applied to model the spatial information of patches for high-quality (positive) and low-quality (negative) photos. The likelihood probability of each patch is given by

$$p(x_i^s \middle| \Omega^s) = \sum_{k=1}^{c} \pi_k^s P(x_i^s \middle| \mu_k^s, \Sigma_k^s) \tag{2}$$

where Ω is either the positive or negative class in the s-th subtopic; c is the number of Gaussian models (c=5 in our case); (μ_k, Σ_k) is the parameters of the k-th Gaussian models; and π_k is the weight of the likelihood probability $P(x_i^s \middle| \mu_k^s, \Sigma_k^s)$ and $\sum_{k=c}^{c} \pi_k = 1$. The parameter set of GMM, i.e. $(\pi_k^s, \mu_k^s, \Sigma_k^s)$ is estimated using the Expectation-Maximization (EM) algorithm [5].

Besides, the relative spatial information of pair of patches is useful to analyze the ingredients of photos as well [4]. The concatenated vector $\hat{x}_{ij} = [x_i^T \; x_j^T]^T$ is used to couple the spatial information of the pair of patches x_i^s and x_j^s. Here the index for the subtopic is omitted. The likelihood probability is given by

$$p(\hat{x}_{ij} \mid \Phi) = \sum_{k'=1}^{c'} \hat{\pi}_{k'} P(\hat{x}_{ij} \mid \hat{\mu}_{k'}, \hat{\Sigma}_{k'}) \tag{3}$$

where Φ is either the positive or negative class; c' is the number of Gaussian models (c'=5 in our case); The parameter set of GMM, i.e. $(\hat{\pi}_{k'}, \hat{\mu}_{k'}, \hat{\Sigma}_{k'})$ is estimated by EM algorithm [5] as well. Note that different from [4], the spatial relations for single or pair of patches are modeled for positive and negative photos, respectively, and the experiment results will show that the usage of spatial relations obtained from negative photos can provide significant improvement for system.

D. Likelihood evaluation for photo qualiy assessment

In the test process, as shown in Fig. 2(b), the test image is segmented into image patches, and the color and the texture features are extracted from each patch. To reduce the context variances of photos, the image is subsequently assigned to a subtopic with minimum distance between feature vector and cluster centers of subtopics. Then the likelihood probability of each image patch and of every image pair can be estimated by Eq. (2) and (3), respectively. Note that the likelihood probabilities for single/pair patches are estimated for positive and negative photos. Hence, the assessment score based on the likelihood ratio is given by

$$score = \sum_{i=1}^{s} \frac{\log(p(x_i \mid \Omega_i^P))}{\log(p(x_i \mid \Omega_i^N))} + \sum_{i=1}^{s} \sum_{\substack{j=1 \\ i \neq j}}^{s} \frac{\log(p([x_i^T \; x_j^T]^T \mid \Phi_{ij}^P))}{\log(p([x_i^T \; x_j^T]^T \mid \Phi_{ij}^N))} \tag{4}$$

where x_i is i-th patch for the test image, s is the total number of patches, Ω_i^P and Ω_i^N is the statistical context models of positive and negative photos for image patch x_i, and Φ_{ij}^P and Φ_{ij}^N is the statistical context models of positive and negative photos for pair patches of x_i and x_j.

III. Experimental Results

We collect photos from the website, DPChallenge.com, which can provide a platform for photographers to share their photos. In addition, all the website viewers can provide the score for each photo according to their subjective preference. All photos in the website are categorized into several topics, including landscapes, seascapes, animals, fashion, etc. The category of landscapes is selected in our experimental setting, and 20,251 photos with preference scores are automatically crawled. In our system, those photos with top 10% score are assumed to have higher qualities and favored by more users, and hence collected as the positive data. While those photos with bottom 10% scores are assumed to have lower qualities and taken as negative data. Then the contexts of spatial relations for positive and negative photos are statistically modeled in the training process.

To analyze the effects of negative data for the confidence of the photo assessment, the experimental results of the system which takes the likelihood to positive class as the assessment score is compared. Fig. 5 shows the Receiver Operator Characteristic (ROC) curves of the proposed system and the comparison one. It can be seen that the system can automatically assess photo quality with the promising results. Additionally, the usage of the negative data can provide significant improvement for the photo quality assessment.

IV. Conclusion

We propose a photo quality assessment system by statistically modeling the context from the higher and lower quality photos. The experimental results show that the system can automatically assess photo quality with the promising results and the usage of negative data can significantly improve the accuracy rate for the system.

In the future work, we would like to take more photography rules as photo features, such as white balance, exposure and against photo contents, to raise the accuracy of quality assessment. In addition, we could apply our model to automatically find the professional view in the photos [4].

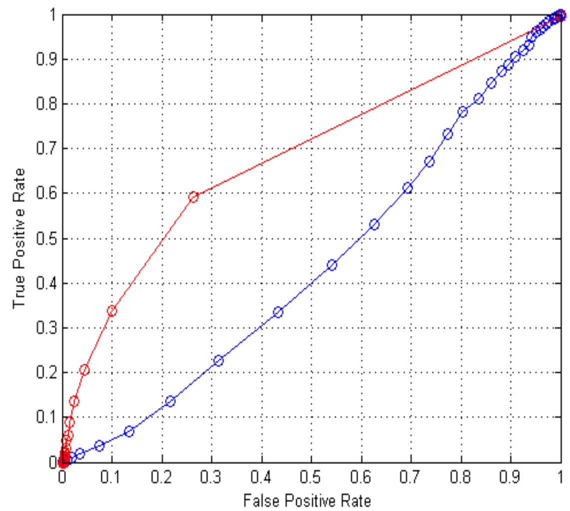

Fig 5. Comparison of ROC curves obtained by the proposed system (red line) and Cheng *et al.* [4] (blue line).

V. References

[1] S. Avidan and A. Shamir, “Seam carving for content-aware image resizing,” ACM Trans. Graphics, vol. 26, issue 3, pp. 1-9, Jul. 2007.

[2] S. Banerjee and B. L. Evans, “In-camera automation of photographic composition rules,” Image Processing, vol. 16, issue 7, pp. 1807-1820, Jul. 2007.

[3] S. Birchfield and S. Rangarajan, “Spatiograms versus histograms for region-based tracking,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, IEEE Press, San Diego, CA, USA, vol. 2, pp. 1158-1163, Jun. 2005.

[4] B. Cheng, B. Ni, S. Yan and Q. Tian, “Learning to photograph,” Proc. ACM International Conference on Multimedia, Firenze, Italy, pp. 291-300, Oct. 2010.

[5] A. Dempster, N. Laird, and D. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” Journal of the Royal Statistical Society, Series B, vol. 39, No. 1, pp. 1-38, 1977.

[6] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, IEEE Press, San Diego, CA, USA, pp. 886-893, Jun. 2005.

[7] C. Dance, J. Willamowski, L. Fan, c. Bray, and G. Csurka, “Visual categorization with bags of keypoints”, *ECCV International Workshop on Statistical Learning in Computer Vision*, 2004.

[8] P. Felzenszwalb,and D. Huttenlocher, “Efficient graph-based image segmentation,” International Journal of Computer Vision, vol. 59, issue 2, pp. 167-181, Sep. 2004.

[9] X. Hou and L. Zhang, “Saliency detection: A spectral residual approach,” Proc. European Conference on Computer Vision, Minneapolis, Minnesota, USA, pp. 386-399, Oct. 2007.

[10] S. C. Johnson, “Hierarchical clustering schemes,” Psychometrika, vol. 32, No. 3, pp. 241–254, 1967.

[11] Y. Ke, X. Tang, and F. Jing, “The design of high-level features for photo quality assessment,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, New York, NY, USA, pp. 419-426, Jun. 2006.

[12] Y. Luo and X. Tang, “Photo and video quality evaluation: Focusing on the subject,” Proc. European Conference on Computer Vision, Marseille, France, pp. 386-399, Oct. 2008

[13] J. Li, W. Wu, T. Wang, and Y. Zhang, “One step beyond histogram: Image representation using markov stationary features,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, IEEE Press, Anchorage, Alaska, USA, pp. 1-8, Jun. 2008.

[14] T. Mei, X.-S. Hua, H.-Q. Zhou, and S. Li, “Modeling and mining of users’ capture intention for home videos,” IEEE Trans. Multimedia, vol. 9, No. 1, pp. 66-77, Jan. 2007.

[15] B. Ni, S. Yan, and A. Kassim, “Contextualizing histogram,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, IEEE Press, Miami, Florida, USA, pp. 1682-1689, Jun. 2009.

[16] Y. Rui and T. Huang, “A novel relevance feedback technique in image retrieval,” Proc. ACM International Conference on Multimedia, New York, NY, USA, pp. 67-70, Oct. 1999.

[17] H. Sheikh, A.Bovik, and G. Veciana, “An information fidelity criterion for image quality assessment using natural scene statistics,” Image Processing, vol. 14, issue 12, pp. 2117-2128, Dec. 2005.

[18] K. Sande, T. Gevers, and C. Snoek, “Evaluating color descriptors for object and scene recognition,” IEEE Trans. on Pattern Analysis and Machine Intelligence (in press), 2009.

[19] M. Stricker and M. Orengo, “Similarity of color images,” Storage and Retrieval of Image and Video Databases III, vol. 2, pp. 381-392, Feb. 1995.

[20] V. Setlur, S. Takagi, R. Raskar, M. Gleicher, and B. Gooch, “Automatic image retargeting,” Proc. ACM International conference on Mobile and ubiquitous multimedia, Christchurch, New Zealand, pp. 59-68, Dec. 2005.

[21] X. Sun, H. Yao, R. Ji, and S. Liu, “Photo assessment based on computational visual attention model,” Proc. ACM International Conference on Multimedia, Beijing, China, pp. 541-544, Oct. 2009.

[22] Y. Zhang, M. Zhao, S. -Y. Neo, T. -S. Chua, and Q. Tian, “Visual synset: towards a higher-level visual representation,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, IEEE Press, Anchorage, Alaska, USA, pp. 1-8, Jun. 2008.