Classification of Digital Photos Taken by Photographers or Home Users [*]

Hanghang Tong[1], Mingjing Li[2], Hong-Jiang Zhang[2], Jingrui He[1], and Changshui Zhang[3]

[1] Automation Department, Tsinghua University, Beijing 100084, P.R.China
{walkstar98, hejingrui98}@mails.tsinghua.edu.cn
[2] Microsoft Research Asia, 49 Zhichun Road, Beijing 100080, P.R.China
{mjli, hjzhang}@microsoft.com
[3] Automation Department, Tsinghua University, Beijing 100084, P.R.China
zcs@tsinghua.edu.cn

Abstract. In this paper, we address a specific image classification task, i.e. to group images according to whether they were taken by photographers or home users. Firstly, a set of low-level features explicitly related to such high-level semantic concept are investigated together with a set of general-purpose low-level features. Next, two different schemes are proposed to find out those most discriminative features and feed them to suitable classifiers: one resorts to boosting to perform feature selection and classifier training simultaneously; the other makes use of the information of the label by Principle Component Analysis for feature re-extraction and feature de-correlation; followed by Maximum Marginal Diversity for feature selection and Bayesian classifier or Support Vector Machine for classification. In addition, we show an application in No-Reference holistic quality assessment as a natural extension of such image classification. Experimental results demonstrate the effectiveness of our methods.

1 Introduction

With the ever-growing advance of digital technology and the advent of Internet, many home users have collected more and more digital photos. However, due to the lack of expertise, the images taken by home users are generally of poor quality compared with those taken by photographers (an example in Fig.1). Automatically grouping images into these two semantically meaningful categories is highly desirable for [11][13][19]: 1) to efficiently store and retrieve digital content; 2) to help home users better manage digital photos or assess their expertise in photographing; 3) to evaluate and compare the qualities of different images with different content.

Finding discriminative enough features and training with a suitable classifier are always the key steps in image classification [11][13]. In the past several

---

[*] This work was performed at Microsoft Research Asia.

years, there has been a lot of related work. For example, Serrano et al in [13] proposed using texture and color features and training with Support Vector Machine (SVM) for indoor/outdoor images. Oliveira et al in [11] proposed a set of features, including the prevalent color, the farthest neighbor and so on; and using Itemized Dichotomizer 3 (ID3) for photographs and graphics. Compared with these existing image classification problems, grouping images into 'by photographer' and 'by home user' is much more difficult for the following reasons: 1) it is not completely known what kinds of high level factors make the images 'by photographer' different from those 'by home-user' although it is easier for a subject to tell them apart; 2) how to express these factors (if we know them) as appropriate low-level features might be very difficult.

To address these issues, in this paper we solve our problem in a manner of 'black box' model. That is, we let the algorithm automatically find out those most discriminative features from some high-dimensional feature space in which the images belongs to these two classes might be separable; and feed them to a suitable classifier. To this end, firstly, we investigate a set of low-level features explicitly related to such high level semantic concept together with a set of general-purpose low-level features and the combination of them makes up the initial feature set. Next, to find out those most discriminative features and feed them to a suitable classifier, we propose two different schemes: one is boosting based in which situation the feature selection and classifier training are performed simultaneously, benefited from its powerful ability in combining weak learners; for the other method, we make use of the information of the label by Principle Component Analysis (PCA) [5] for feature re-extraction and decorrelation, followed by Maximum Marginal Diversity (MMD) [20] to select those most discriminative features which can be subsequently fed to Bayesian classifier or SVM [5]. While the former is very simple, the latter one is more sophisticated and leads to better performance for our problem. As a natural extension, we will show an application of such image classification in No-Reference holistic quality assessment. Experimental results on 29540 digital images and on a systematic subjective image quality assessment procedure demonstrate the effectiveness of our methods.

The rest of the paper is organized as follows: in Sect.2, we present our classification method in detail. Its application in No-Reference holistic quality assessment is shown in Sect.3. Section 4 gives the experimental results. Finally, we conclude the paper in Sect.5.

2 Grouping Image into 'by Photographer' and 'by Home User'

2.1 Initial Feature Extraction

Despite of the difficulties mentioned above, it is still possible to represent some high level concepts explicitly related with whether a given image is taken 'by

photographer' or 'by home user' as suitable low-level features. We have performed extensive experiments and have come up with the following low-level features:

* **Blurness:** we use a two-dimensional feature $blur_i = [ib, be]^T$ proposed in our previous work [19] to indicate whether image i is blurred (ib) and to what extent it is blurred (be).
* **Contrast:** at current stage, we use a two-dimensional feature $contrast_i = [p_u, p_l]^T$ to indicate whether image i is over-bright (p_u) or over-dark (p_l).
* **Colorfulness:** the colorfulness of image i is measured by a one-dimensional feature $colorful_i$ [4].
* **Saliency:** we use a three-dimensional feature $saliency_i = [s_1, s_2, s_3]^T$ to indicate the saliency of image i, where s_1, s_2 and s_3 are the mean, variance and third-order moment of its saliency map (SM) [8].

To compensate for the limited understanding of the relationship between the high level concepts and its low-level features, a set of general-purpose low-level features are also used as Table 1:

Table 1. General-purpose low-level features

Category	Name	Dim.	Category	Name	Dim.
	Band Difference[1]	1		MRSAR[10]	15
	Color Moment [15]	9		Tamura[17]	18
	Color Histogram[16]	64	**Texture**	Wavelet[21]	18
Color	Lab Coherence[12]	128		WaveletPwt[9]	24
	Luv Coherence[12]	128		WaveletTwt[2]	104
	HSV Coherence[12]	128		Canny Histogram[6]	15
	Correlogram[7]	144	**Shape**	Sobel Histogram	15
Energy	DFT moment	6		Laplace Histogram	15
	DCT moment	6			

Note that 1) "Sobel Histogram" and "Laplace Histogram" are the modified versions of "Canny Histogram" which use Sobel and Laplace operators to detect edges instead of Canny operator, respectively; 2) "DFT moment" and "DCT moment" contains the mean and variance of the coefficients of Discrete Fourier Transformation and Discrete Cosine Transformation for red, green and blue channels, respectively.

The combination of all above features makes up the initial feature set for classification which contains 21 different kinds of low-level features and is 846-dimensional.

2.2   Finding Discriminative Features and Feeding to Classifier

It is always a challenge to select a good feature set for image classification [10][13]. We propose two different schemes for our task in this paper.

Boosting Based Scheme Recent developments in machine learning field have demonstrated that boosting based methods may have a satisfactory combined performance by combing weak learners [3][5]. Furthermore, the boosting procedure can also be viewed as a feature selection process if the weak learner uses a single feature in each stage. Benefiting from such cherished properties, our first scheme is very simple. That is, we just use some boosting based method to train on the initial low-level feature set and in this context, boosting performs both feature selection and classifier training simultaneously. To be specific, we will examine both Ada-Boost and Real-AdaBoost for our classification task.

Feature Re-extraction Based Scheme There are two other kinds of effective classifiers: one is Bayesian classifier which theoretically produces the minimum classification error; the other is SVM which has not only strong theoretical foundations but also excellent empirical successes. However, we can not directly apply these classifiers to our task since the dimensions of the initial feature set is very high. In such high-dimensional feature space, the following two issues become very difficult: 1) the high accuracy of probability estimation that is necessary for Bayesian classifier; and 2) the optimization of quadratic problem in SVM. To take the advantage of Bayesian classifier or SVM, we have to select a small subset from the initial feature set, whose elements are most discriminative.

On the other hand, we find out by experiments that the discriminative power for most features in the initial feature set is too weak, which means a small subset of it might not be adequate for a satisfactory classification performance.

Based on the above observations, we propose the following algorithm to re-extract some more discriminative features from the initial feature set, select those most discriminative ones by MMD and feed them to Bayesian classifier or SVM, hoping to further improve the classification performance compared with the first scheme. For denotation simplicity, we use S^+ and S^- denote the subset of images taken 'by photographer' and 'by home user'; N^+ and N^- denote the number of images in S^+ and S^-; and Σ^+ and Σ^- are the covariance matrices for S^+ and S^-, respectively.

Algorithm 1. Feature re-extraction based scheme

---
1. Normalize the feature $F(i)(i = 1, 2, \ldots, (N^+ + N^-))$ on each dimension to $[0, 1]$;
2. Calculate covariance matrix Σ [5]:

$$\Sigma = (N^- \cdot \Sigma^- + N^+ \cdot \Sigma^+)/(N^- + N^+) \qquad (1)$$

3. Perform PCA on Σ. Let $u_j(j = 1, 2, \ldots, 846)$ denote the jth principle axis;
4. The new feature set is denoted as $F^{'}(i) = [x_1, x_2, \ldots, x_{846}]^T$, where $x_j(j = 1, 2, \ldots, 846)$ denote the projection of $F(i)$ on u_j;
5. Use MMD to select the most N discriminative feature $F^{'}_s(i)$;
6. Feed $F^{'}_s(i)$ to Bayesian classifier or SVM.

Note that by taking the covariance matrix as (1), we can make use of the information of the label in PCA to re-extract some more discriminative features from the initial feature set. Moreover, de-correlation on different dimensions by PCA also makes the subsequent feature selection step more reliable.

2.3 Application in No-Reference Holistic Quality Assessment

No-Reference (NR) quality assessment is a relatively new topic. Compared with the traditional assessment methods, it dose not require any kind of reference information and can be applied when the original un-distorted image might not exist or be very difficult to obtain. In recently years, it has been attracting more and more research attention.

However, due to the limited understanding of HVS (Human Vision System), most, if not all, of the existing NR assessment algorithms are based on the following philosophy [14][18]: "*all images are perfect, regardless of content, until distorted*". While this philosophy simplifies NR into measuring the introduced distortion, it can not evaluate the holistic quality for different images with different content since cognitive and aesthetic information within images is ignored in these methods and all undistorted images are treated as equally perfect.

As a natural extension of our image classification problem, we might solve NR holistic quality assessment from another point of view. Generally speaking, the images taken 'by photographer' are of relatively higher quality than those taken 'by home user'. Thus we have actually got a classifier which separates the images of 'high quality' and those of 'low quality' in Sect.2. By converting the output of the classifier to a continuous value, we get a confident coefficient indicating a given image i being of 'high quality' or being of 'low quality', which can be used as its holistic quality metric.

$$Qm(i) = \sum_{t=1}^{T} h_t(F(i)) \tag{2}$$

where $h_t(t = 1, 2, \ldots, T)$ denote the tth weak learner of Real-AdaBoost; T is the total number of weak learners; and $F(i)$ is the initial feature vector for image i.

Finally, the quality score of the given image $Ps(i)$ can be predicted as (3) so that it will be consistent with the result given by human observers [18]:

$$Ps(i) = \alpha + \beta \cdot Qm(i)^{\gamma} \tag{3}$$

where α, β and γ are unknown parameters and can be determined by minimizing the MSE (mean-square-error) between prediction scores and mean human scores.

3 Experimental Results

3.1 Image Classification

We examine our classification methods on a large image database: 16643 images from both COREL and Microsoft Office Online compose the subset of the images

'by photographer', and 12897 images taken by the staff in Microsoft Research Asia compose the subset of the images 'by home user'.

A set of parameters and operations need to be set:

- For both Ada-Boost and Real-AdaBoost, the bin number $bin = 20$; and the weak learner number $T = 100$;
- The number N of features selected in Algorithms 1 is determined by the elbow point on the plot of MMD of the feature $F_s'(i)$ in descending order;
- The adopted kernel function in SVM is the RBF kernel; the scale factor $\sigma = 0.05$ in the kernel and penalty factor $C = 10$;
- The probabilities for Bayesian classifier are obtained by Parzen Window Density Estimation [5]; and $P(S^+)/P(S^-) = N^+/N^-$.

We have performed 5-fold cross-validation on all 29540 images. The testing error is given in Table 2. It can be shown that 1) both schemes are effective; 2) SVM and Bayesian classifier produce better performance than Ada-Boost and Real-AdaBoost.

Table 2. The cross-validation results for image classification

	Ada-Boost	**Real-AdaBoost**	**SVM**	**Bayesian**
testing error	8.9%	6.6%	6.1%	4.9%

3.2 NR Holistic Quality Assessment

A systematic subjective experiment is performed on 379 images which possess different content. The subjective experiment is conducted in a similar way as [14] did: 16 human observers (8 men and 8 women) are asked to rate each image as 'Bad', 'Poor', 'Fair', 'Good' or 'Excellent' on the same computer. The images are displayed on the gray-level background one by one in a random order. Mean human scores are acquired after normalizing the original raw scores and removing outliers. All these 379 images are divided randomly into two sets: one as "training set" to determine the parameters in (3); and the other as "testing set" to examine the performance of our method for NR holistic quality assessment.

The result is encouraging: the linear correlation value between the prediction result and mean human score on "testing set" is 84.7%. The MSE between the prediction result and mean human score on "testing set" is 11.1. An example of applying our algorithm to evaluate holistic quality of different images is shown in Fig.1.

4 Conclusion

In this paper, we have dealt with a specific image classification problem: i.e. to group images according to the person who takes them: 'by photographer'

or 'by home user'. A set of low-level features which are explicitly related to such specific high level semantic concept are investigated together with a set of general-purpose low-level features. To find out those most discriminative features and feed them to suitable classifiers, we propose two different schemes: one is boosting based, in which situation we make use of the cherished properties of boosting methods to perform feature selection and classifier training simultaneously; the other is feature re-extraction based, in which context we resort to PCA in a supervised manner to re-extract some more discriminative features from the initial weak features; then we use MMD to select those most discriminative ones and feed them to SVM or Bayesian classifier. Moreover, de-correlation on different dimensions of features by PCA also makes the subsequent feature selection step more reliable. While the first scheme is very simple, the latter one is more sophisticated and produces higher performance for our problem. As a natural extension, we show an application of such image classification in No-Reference holistic quality assessment. Experimental results on 29540 digital images and on a systematic subjective image quality assessment procedure demonstrate the effectiveness of our method.

5 Acknowledgements

References

[1] Athitsos, V., et al: Distinguishing photographs and graphics on the World Wide Web. IEEE Workshop on CBAIVL (1997)

[2] Chang, T., et al: Texture analysis and classification with tree-structured wavelet transform. IEEE Trans. on Image Processing **2** (1993) 429-441

[3] Friedman, J., et al: Additive logistic regression: a statistical view of boosting. The Annual of Statistics **28(2)** (2000) 337-374

[4] Hasler, D., et al: Measuring colorfulness in real images. SPIE **5007** (2003) 87-95

[5] Hastie, T., et al: The Elements of Statistical Learning. Springer Verlag (2001)

[6] He, J.R., et al: W-Boost and its application to web image classification. Proc. ICPR (2004)

[7] Huang, J., et al: Image indexing using color correlogram. Proc. CVPR (1997) 762-768

[8] Ma, Y.F., et al: A user attention model for video summarization. ACM Multimedia (2002) 533-542

[9] Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. on PAMI **11** (1989) 674-693

[10] Mao, J. et al: Textureclassification and segmentation using multiresolution simultaneous autoregressive models. Pattern Recognition **25** (1992) 173-188

[11] Oliveira, C.J.S., et al: Classifying images collected on the World Wide Web. SIGGRAPH (2002) 327-334

[12] Pass, G.: Comparing images using color coherence vectors. ACM Multimedia (1997) 65-73

[13] Serrano, N., et al: A computational efficient approach to indoor/outdoor scene classification. Proc. ICPR (2002) 146-149

[14] Sheikh, H.R., et al: Blind quality assessment for JPEG2000 compressed images. ICSSC (2002)

[15] Stricker, M., et al: Similarity of color images. SPIE **2420** (1995) 381-392

[16] Swain, M., et al: Color indexing. Int. Journal of Computer Vision **7(1)** (1991) 11-32

[17] Tamura, H., et al: Texture features corresponding to visual perception. IEEE Trans. on SMC **8** (1978) 460-473

[18] Tong, H.H., et al: No-reference quality assessment for JPEG2000 compressed images. Proc. ICIP (2004)

[19] Tong, H.H., et al: Blur detection for digital images using wavelet transform. Proc. ICME (2004)

[20] Vasconcelos, N., et al: Feature selection by maximum marginal diversity. Proc. CVPR (2003) 762-769

[21] Wang, J.Z., et al: Content-based image indexing and searching using Daubechies' wavelets. IJDL **1** (1998) 311-328

Fig. 1. An example of evaluating the holistic quality for different images. Ps: the prediction result; Mhs: the mean human score. Note that (a) and (b) are taken by 'home user'; while (c) and (d) are taken by 'photographer'.

(a) $Ps = 9.5\ Mhs = 11.7$

(b) $Ps = 28.6\ Mhs = 36.7$

(c) $Ps = 65.3\ Mhs = 70.0$

(d) $Ps = 82.6\ Mhs = 78.3$