

INFORMATION-BASED COLOR FEATURE REPRESENTATION FOR IMAGE CLASSIFICATION

S.L.Wang[†] and A. W. C. Liew[#]

[†]School of Info. Security Engg., Shanghai Jiaotong University, Shanghai, China

[#]School of Info. and Comm. Technology, Griffith University, Brisbane, Australia

ABSTRACT

For the image classification task, the color histogram is widely used as an important color feature indicating the content of the image. However, the high-resolution color histograms are usually of high dimension and contain much redundant information which does not relate to the image content, while the low-resolution histograms cannot provide adequate discriminative information for image classification. In this paper, a new color feature representation is proposed which not only takes the correlation among neighbouring components of the conventional color histogram into account but removes the redundant information as well. A high-resolution, uniform quantized color histogram is first obtained from the image. Then the redundant bins are removed and some neighbouring bins are combined together to generate a new feature component to maximize the discriminative ability. The mutual information is adopted to evaluate the discriminative power of a specific feature set and an iterative algorithm is performed to derive the histogram quantization and their corresponding feature generation. To illustrate the effectiveness of the proposed feature representation, an application of detecting adult images, i.e., image classification between erotic and benign images, is carried out. Two widely used classification techniques, SVM and Adaboost, are employed as the classifier. Experimental results show the superior performance of our color representation compared with the conventional color histogram in image classification.

Index Terms — color histogram, image classification, mutual information

1. INTRODUCTION

Color histograms are the most popular description of color distributions in an image. Due to the invariance of translation, rotation and scaling (with normalization), they usually work as color features to represent the image content. In recent years, many researchers have proposed various kinds of color histograms for image classification and the key issue related to these color histograms is the color space quantization scheme. Generally speaking, there are two major kinds of histogram generation methods: the typical rectangular-shape binning method and clustering. The typical rectangular-shape binning histogram is generated by dividing the entire color space into a number

of cubic bins and counting the number of pixels in each cubic bin. Uniform quantization is the traditional and widely-used method, however, non-uniform quantization according to the characteristics of the color space shows better performance in many image classification and retrieval applications [1-3]. However, as the histogram bins are constrained to be of rectangular shape, the color distribution may not be efficiently described. To overcome such difficulty, the clustering-based histogram generation methods have been proposed [4]. The clustering methods usually divide the color space into a large number of bins and then group them by a clustering algorithm such as the k -means.

Color histograms generated by all the generation methods mentioned above can describe the color distribution of the image so that they may serve as a discriminative feature to differentiate various kinds of images. However, for the image classification task, most of the traditional color histograms usually encounter the following main difficulties which prevent them to provide accurate classification results: i) The appropriate color resolution is difficult to determine. Low color resolution with few bins cannot provide sufficient color depth to differentiate various kinds of images; however, increasing the number of bins will cause more computation complexity for similarity calculation and more memory storage for the database. ii) The intrinsic relationship between adjacent color components is not taken into consideration, which may help indicate the image category. iii) The quantization scheme is irrelevant to the target image categories. Some histogram bins contain discriminating information for classification while others do not.

To overcome these difficulties, an information-based color histogram is proposed. The main novelties of the proposed algorithm lie in: i) mutual information (MI) is adopted linking the quantization scheme and the image category; ii) an adaptive, class-related binning technique and feature extraction method is proposed which not only helps determine the proper color resolution but also incorporates the relationship between neighbouring color components.

The paper is organized as follows. In section 2, the definition and derivation of the class-related color histogram is introduced. The evaluation of the discriminative power of a specific feature set based on the information theory is also presented in this section. Section 3 shows the experimental results and discussions on the experiments. Finally, section 4 draws the conclusion.

The work described in this paper is fully supported by the Shanghai Natural Science Fund (05ZR14080).

2. AN INFORMATION-BASED COLOR HISTOGRAM

Since they are easy to compute and tolerate against the small changes of the view points, the traditional color histograms are widely used in image retrieval and classification. For an RGB-formatted color image I , the entire color space can be uniformly divided into $N \times N \times N$ histogram bins. Then the color histogram can be generated by calculating the number of pixels contained in each bin and the histogram vector can be represented as $\langle h_{111}, h_{112}, \dots, h_{NNN} \rangle$. For high-resolution color histogram, there are inevitably lots of empty bins since some particular color seldom appears in natural images. These bins are apparently useless to describe the image and considering them will increase the feature dimension and computational complexity in the subsequent classification process. Hence, the empty bins are first excluded from the histogram vector. In addition, some non-empty bins also do not contain discriminative information for image classification. Furthermore, as there are definitely much relationship between neighbouring bins, taking it into account may improve the performance. In the proposed algorithm, the neighbouring relationship is described by a linear combination of data contained in several bins. It should be noted that only the relationship between adjacent bins (their color distance is below a preset threshold) is approximately represented by linear correlation because the relationship among all the histogram bins may not be linear and using high-order statistics to describe such non-linear relationship is very complex and computational expensive.

2.1 Information-based Discriminative Power Evaluation

How to describe and compute the discriminative power of various bins and their linear combinations is the key issue in the proposed algorithm. Classification accuracy can reflect the discriminative power of a certain feature to some extent; however, they may vary drastically when selecting different classifiers. Recent research shows that the mutual information (MI) between the class label and the transformed feature (a single bin or a linear combination between adjacent bins) can be a more general criterion to represent discriminative power of the feature [5]. The major advantages of MI are i) it is invariant to the classifier selection; ii) it accounts for high-order statistics so that non-linear relationship between the feature and the class label can be well-represented by MI. However, high computational complexity of Shannon's MI prevents it from being widely used currently. Recently, Principe et al. [6] present how using Renyi's formulation rather than Shannon's leads to non-parametric entropy estimators when coupled with Parzen density estimation. Then Torkkola [7] gives a formulation of MI between the continuous variables and discrete class labels. With continuous-value Y and discrete class label C , the mutual information is as follows [7]:

$$I(C, Y) = V_{IN} + V_{ALL} - 2V_{BTW} \quad (1)$$

$$= \sum_c \int_y p(c, y)^2 dy + \sum_c \int_y P(c) p(y)^2 dy - 2 \sum_c \int_y p(c, y)^2 P(c) p(y) dy$$

With the assumption of Parzen distribution for $P(y)$ and $p(c, y)$, three terms in (1) can be rewritten as follows.

$$V_{IN} = \sum_c \int_y p(c, y)^2 dy = \frac{1}{N^2} \sum_{p=1}^C \sum_{j=1}^{N_p} \sum_{k=1}^N G(\mathbf{y}_{pj} - \mathbf{y}_{pk}, \sigma^2 \mathbf{I}) \quad (2)$$

$$V_{ALL} = \sum_c \int_y P(c) p(y)^2 dy = \frac{1}{N^2} \left(\sum_{p=1}^C \left(\frac{N_p}{N} \right)^2 \right) \sum_{j=1}^N \sum_{k=1}^N G(\mathbf{y}_j - \mathbf{y}_k, \sigma^2 \mathbf{I}) \quad (3)$$

$$V_{BTW} = \sum_c \int_y p(c, y)^2 P(c) p(y) dy = \frac{1}{N^2} \sum_{p=1}^C \frac{N_p}{N} \sum_{j=1}^{N_p} \sum_{k=1}^N G(\mathbf{y}_{pj} - \mathbf{y}_k, \sigma^2 \mathbf{I}) \quad (4)$$

where C is the total number of classes, N_p and N are the number of training samples in class p and the total number of training samples respectively. G is the Gaussian kernel and σ is the kernel width.

For a single feature, \mathbf{y} is one dimensional and the MI value can be calculated directly. For a feature vector, a dimension-reducing linear transformation is performed to project the feature vector \mathbf{x} onto a single value with the maximum MI value. Thus,

$$\mathbf{w} = \arg \max_{\mathbf{w}} (I(C, Y)); \quad \mathbf{y} = \mathbf{w}^T \mathbf{x} \quad \text{subject to } \mathbf{w}^T \mathbf{w} = 1 \quad (5)$$

As directly calculate \mathbf{w} in (5) is difficult, an iterative optimization method is adopted. The linear discriminant analysis (LDA) is performed on the training set to derive the initial estimate of the weighting parameter vector, $\mathbf{w}_{\text{initial}}$. According to the partial derivatives of the mutual information function I with respect to the weighting parameter vector \mathbf{w} , the conjugate gradient (CG) algorithm is employed to derive the weighting parameter vector with maximum mutual information iteratively. Detailed derivation of the derivatives is shown in Appendix and interested readers may refer to [7] for more information.

2.2 Bin-grouping and feature extraction method

Based on maximization of the mutual information criterion, the bin-grouping and feature extraction method is carried out as follows:

- I. Detect and mask out the empty bins (the average value of a specific bin below a preset threshold, T_{Emp} , is defined as an empty bin) and set the feature index $i=1$.
- II. Calculate the discriminative power of each single feature (the MI value between the feature and the class label). Select the most discriminative feature with highest MI value (denoted by $I_{previous}$).
- III. Form the candidate set by combining the bin(s) with most discriminative feature and each of its (their) neighbouring bin(s). Remove the groups from the candidate set whose color compactness (denoted by the maximum color distance) is greater than a preset threshold, Max_{dist} .
- IV. From the candidate set, select the group of bins with highest MI (denoted by $I_{current}$). If $I_{current} > I_{previous} + T_{MI}$ (T_{MI} is a preset threshold), combine the bins in the group and update the most discriminative feature set

and its MI value ($I_{previous}$) with $I_{current}$. Return to step III.

- V. If $I_{current} \leq I_{previous} + T_{MI}$, extract the i th feature and record its group components and the weighting parameter w_i which leads to the maximum MI value. Mask all the bins in the group set from further processing. If there is any bin not being processed yet, increase i by 1 and go to step II. Otherwise stop.

Finally, the class-related color histogram can be generated by i) partitioning the high-resolution color histogram into groups according to the grouping scheme mentioned above; ii) from each group, extracting the corresponding histogram feature vector and multiplying by w_i to generate the i th feature. Note that from the generation method of the new histogram feature, the former components have higher discriminative power. Hence, the new histogram feature can be further dimension-reduced by extract first several components.

3. EXPERIMENTAL RESULTS

In order to evaluate the proposed color image representation, we have applied it for nude image detection. It is a specific application of image classification, i.e., differentiating all the nude images from the benign ones. Four thousand images have been collected from internet and half of them are nude images. The training data set for determining the group partition and the corresponding weighing parameters is composed of one-tenth of the entire nude and benign image set (200 nude and 200 benign images). After feature extraction, two kinds of classifiers, i.e., the Support Vector Machine (SVM) [8] and the Adaboost algorithm [9] are adopted for classification.

3.1 Bin-grouping and histogram generation

The uniform quantized color histogram vector is described by a $10 \times 10 \times 10$ -dimension vector. Then normalization is performed to reduce the variations caused by different image size. With T_{Emp} set as 0.0001, the bins whose average value for both nude and benign image classes are below the threshold is masked out for the subsequent processing. In our experiments, almost half of the histogram bins are masked out (547 bins remained) which demonstrates that for natural images, their color elements do not evenly distributed in RGB color space and non-uniform quantization is more suitable for such color distribution.

After empty-bin masking, the grouping procedure is carried out with T_{MI} as 0.01 empirically. The maximum color distance will affect the grouping result. Table 1 lists the grouping result and the recognition performance with various settings of Max_{dist} , where EER_{SVM} and EER_{ADA} are the equal error rate by SVM and Adaboost classifier, respectively.

3.2 Recognition test

In order to evaluate the recognition performance of

the proposed color representation, two sophisticated classifiers, the support vector machine (SVM) and Adaboost algorithms are employed. SVM projects the data onto the high-dimension feature space and then performs the maximal-margin linear classification [8]. The incorporated kernel function helps SVM to handle many non-linear data classification. In our experiments, the Radial Basis function (RBF) kernel is adopted which produces accurate results in many image retrieval and classification algorithm in the literature. Adaboost is a recently developed iterative learning algorithm which boosts the classification performance by combining a set of weak classifiers to form a strong one [9]. In each iteration, the classifier will focus on the examples which are difficult to classify in the previous iterations.

Max_{dist}	No. groups	EER_{SVM} (%)	EER_{ADA} (%)
2	124	16.48	12.46
3	109	15.77	12.10
4	108	14.68	12.09
5	104	15.10	12.11
6	99	15.00	12.40
7	97	15.14	12.65
8	91	15.06	12.98
9	91	15.15	12.98
10	89	16.90	13.11

Table. 1 The grouping result and recognition performance with various Max_{dist}

To illustrate the effectiveness of the proposed color feature representation, Table.2 shows the recognition performance using the class-related color histogram ($Max_{dist}=4$) by SVM and Adaboost compared with the traditional color histogram feature.

EER(%)	$f_{\text{histogram}}$	$f_{\text{class-related}}$
SVM	16.52	14.68
Adaboost	17.27	12.09

Table. 2 Recognition performance using traditional color histogram feature and the proposed class-related color histogram feature by SVM and Adaboost.

From Table.2, it is shown that the proposed color feature representation always outperforms the traditional color histogram feature using SVM or Adaboost. And the performance improvement using Adaboost is more significant than that of using SVM. It is because the Adaboost algorithm is a combination of weak classifiers, and each weak classifier considers one single feature independently. The intrinsic connection between each neighbouring color histogram components is not much exploited in the Adaboost classifier using the traditional color histogram. By bin-grouping, class-related information in such correlation among neighbouring histogram components is extracted and greatly improves the recognition performance. For SVM, the EER by the proposed feature is also smaller than that by the traditional histogram, which demonstrates that the proposed feature

representation is of more discriminative power than the traditional histogram.

By the optimal parameter set, the proposal algorithm reaches 87.91% classification accuracy in detecting adult images. Compared with the original color histogram, over 5% accuracy improvement is achieved and the computational complexity in the recognition stage is reduced due to the lower dimensional feature vector. Fig.3 shows some false acceptance samples, i.e. benign images false-classified as erotic ones. It can be observed from the figures that the detection error may be caused by the color distributions of these images are quite similar to those of erotic images. Similar results can be observed from the false-rejection results, i.e., nude images classified as benign ones. (To avoid causing upset feeling for the readers, the false-rejection images are not shown in the paper.) Some techniques considering the spatial information (the color coherence vector [10]) and intensity texture information [11] may help improve the performance. However, such techniques inevitably require additional computation and are more time-consuming.

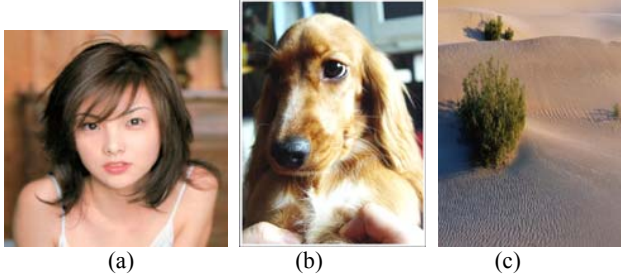


Fig. 3 Some false-acceptance samples

4. CONCLUSIONS

The paper proposes a new color feature for image classification. Mutual information between the class label and the specific feature set is adopted to evaluate its discriminative power. Based on the MI criterion, a new histogram quantization scheme and feature extraction method is proposed, which considers the intrinsic relationship between adjacent color components and thus can provide low-dimension, discriminative feature to differentiate various kinds of images. To evaluate the performance of the proposed color feature, experiments of nude image detection using SVM and Adaboost are carried out. From the experimental results, it is observed that the information-based color histogram provides superior performance compared with the conventional, high-resolution color histogram.

5. APPENDIX

Let \mathbf{x} be the original feature vector, y be the one-dimensional transformed feature and $y = \mathbf{w}^T \mathbf{x}$. The partial derivative of $\frac{\partial I}{\partial \mathbf{w}}$ can be rewritten as,

$$\frac{\partial I}{\partial \mathbf{w}} = \sum_{i=1}^N \frac{\partial I}{\partial y_i} \cdot \frac{\partial y_i}{\partial \mathbf{w}} = \sum_{i=1}^N \frac{\partial I}{\partial y_i} \cdot \mathbf{x}_i \quad (\text{A1})$$

If y_i is in the p th class, we denote y_i as y_{pi} .

$$\frac{\partial I}{\partial y_i} = \frac{\partial I}{\partial y_{pi}} = \frac{\partial V_{IN}}{\partial y_{pi}} + \frac{\partial V_{ALL}}{\partial y_{pi}} - 2 \frac{\partial V_{BTW}}{\partial y_{pi}} \quad (\text{A2})$$

$$\frac{\partial V_{IN}}{\partial y_{pi}} = \frac{1}{N^2 \sigma^2} \sum_{k=1}^{N_p} G(y_{pk} - y_{pi}, 2\sigma^2)(y_{pk} - y_{pi}) \quad (\text{A3})$$

$$\frac{\partial V_{ALL}}{\partial y_{pi}} = \frac{1}{N^2 \sigma^2} \left(\sum_{p=1}^C \left(\frac{N_p}{N} \right)^2 \right) \sum_{k=1}^N G(y_k - y_{pi}, 2\sigma^2)(y_k - y_{pi}) \quad (\text{A4})$$

$$\frac{\partial V_{BTW}}{\partial y_{pi}} = \frac{1}{N^2 \sigma^2} \sum_{j=1}^C \frac{N_j + N_i}{2N} \sum_{k=1}^{N_j} G(y_{pk} - y_{pi}, 2\sigma^2)(y_{pk} - y_{pi}) \quad (\text{A5})$$

6. REFERENCES

- [1] Yining Deng, B.S. Manjunath, C. Kenney, M.S. Moore and H. Shin, "An efficient color representation for image retrieval", *IEEE Trans. on Image Processing*, vol.10, issue 1, pp. 140-147, Jan 2001.
- [2] HongJiang Zhang, Yihong Gong, Chien Y. Low and Stephen W. Smoliar, "Image Retrieval Based on Color Features: An Evaluation Study", *Proc. of SPIE Conf. on Digital Storage and Archival*, Pennsylvania, Oct. 25-27, 1995.
- [3] Wan Xia and C. -C. J. Kuo, "A new approach to image retrieval with hierarchical color clustering", *IEEE Trans. on Circuits and Systems for Video Technology*, vol.8, issue 5, pp. 628-643, Sept. 1998..
- [4] Ju Han and Kai-Kuang Ma, "Fuzzy color histogram and its use in color image retrieval", *IEEE Trans. on Image Processing*, vol. 11, issue 8, pp. 944-952, Aug. 2002.
- [5] C. E. Shannon, "A mathematical theory of communication", *The Bell System Technical Journal*, vol. 27, pp. 379-423, 1948.
- [6] Jose C. Principe, Dongxin Xu, John W. Fisher III, "Information theoretic Learning", in *Unsupervised Adaptive Filtering*, Wiley, New York, NY, 2000.
- [7] Kari Torkkola, "Feature Extraction by Non-Parametric Mutual Information Maximization", *Journal of Machine Learning Research* 3, pp. 1415-1438, March, 2003.
- [8] F. Melgani, L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines", *IEEE Trans. on Geoscience and Remote Sensing*, vol. 42, issue 8, pp. 1778-1790, Aug. 2004.
- [9] R. Schapire, "A Brief Introduction to Boosting", *Proc. of the Sixteen International Joint Conference on Artificial Intelligence (IJCAI99)*, pp. 1401-1406, 1999.
- [10] Greg Pass, Ramin Zabih and Justin Miller, "Comparing Images Using Color Coherence Vectors", *ACM Conference on Multimedia*, pp. 65-74, Nov. 1996.
- [11] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 18, pp. 837-842, Aug. 1996.