COLOR-FREQUENCY-ORIENTATION HISTOGRAM BASED IMAGE RETRIEVAL

Zhuo Zhang[1], *Xiaodong Gu*[1,2], *Sunyuan Kung*[1]

Princeton University, Electrical Engineering Department, Princeton NJ, US[1]
Fudan University, Electronic Engineering Department, Shanghai, China[2]

ABSTRACT

This paper proposes a multiclass image retrieval method using combined color-frequency-orientation histogram. Shape information, obtained via edge detector and Hough Transform, is also incorporated into the new feature. The feature has shown advantage in both unsupervised and supervised learning on Corel image dataset containing 10 categories of 1000 complex scenes. In unsupervised learning, comparing with histogram-based method [1], SIMPLIcity [2], FIRM [3], edge-based method [4], multi-resolution-based method [5], our approach respectively shows 25%, 14%, 10%, 7% and 2% improvement in accuracy. In supervised learning, we implement both one-against-one SVM and one-against-all SVM for multiclass classification. One-against-all SVM beats one-against-one SVM, achieving 95% accuracy with sufficient training.

***Index Terms*— Image classification, image retrieval, color-frequency-orientation histogram, support vector machines (SVMs)**

1. INTRODUCTION

Image retrieval and pattern recognition become widely used in our daily life since too much information comes to us at the same time. By the nature of the task, the key point is how to represent an image with a mathematical way and how to measure their distance based on an efficient model [6]. For object recognition and feature representation, Lowe proposed a very popular method, called SIFT, to capture some key points from image as features. However, it performs not good when recognizing different objects of same category. Therefore, local region detection is proposed, focusing on the main part of image and deducing large dimension of data [3]. They both use local information as image features but ignore the location information of original images. For image matching and distance measurement part, BoW model was proposed [7] and became widely used for image retrieval, especially for content-based image clustering while combining with Dirichlet models. Other methods were also proposed such as Integrated Region Matching (IRM) [8] for image retrieval, reducing the influence of inaccurate segmentation. But these models are all parameter-sensitive and specific for applications.

The main contributions of this paper are as follows: firstly, we combine multiple histogram methods together for feature selection. We employ not only HSV color channel histogram and frequency histogram, but also Canny edge detector and Hough transform for shape histogram. Secondly, in unsupervised learning, we cluster 1000 images, a subset of 10 categories of Corel dataset, and our method achieves 61%, higher than 47% (SIMPLIcity [2]), 51% (FIRM [3]), 54% (edge-based method [4]), 59%(multi-resolution-based method [5]). Finally, in supervised learning, we apply multiclass SVM to test its performance on the same dataset. The accuracy improves with increasing the training set until 80% of data is used for training. Also when more that 50% of data is trained, one-against-all SVM beat one-against-one SVM, showing 95% accuracy.

The paper is organized as follows. Section II gives a system overview and mainly focuses on feature selection. Section III demonstrates K-means clustering performance comparing with other current methods. Section IV further applies multiclass SVM, i.e. the one-against-all SVM and one-against-one SVM for multiclass classification, giving comparison results. Section V draws the conclusions.

2. FEATURE SELECTION

In order to better extract features from images on different conditions, we apply the following preprocess procedure: 1) cut off the high frequency component using mean-filter smoothing and 2) reduces the effect of different exploring time of images using brightness normalization. After preprocessing, color histogram, frequency histogram and orientation information are extracted as global features, while edge information is used as the local feature. After feature selection, we first use K-means clustering in unsupervised learning image retrieval. Then, we apply one-against-one and one-against-all SVMs for multiclass classification in supervised learning image retrieval. Such a system structure is shown in Fig.1.

2.1. Color Histogram

Originally image is represented by RGB channels. However here we use (R-B,2B-R-G,R+G+B) to isolate brightness variation. Moreover we apply a color projection, changing the

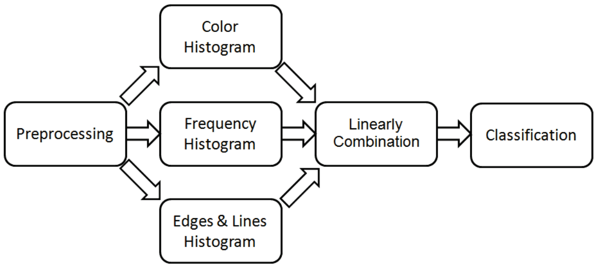

Fig. 1. System Overview

image representation to HSV channels, yielding better edge preserving and high color contrast for image retrieval. Here we use 16 histograms for each channel.

2.2. Frequency Histogram

Even though color histogram is a commonly used classic method for image retrieval, especially for clustering different background images, the texture and shape information is discarded and the method is only sensitive to the images with intensive variations or color distortions.

In order to represent the color gradient and the extent of image's variation in the global scale, the original image is transformed into frequency domain using FFT. Then calculate 48-histogram features as frequency domain information. Notice that here DC component is discarded due to the high variation with little useful information.

2.3. Shape Features(Edge and Line Histograms)

2.3.1. Edge Histogram using Canny Detector

Although we get the color information and color gradient from spatial and frequency domain, it is far more enough for feature extractions. Using histogram for feature selection usually loses object's size, shape and location. Therefore it is necessary to add some shape information as extra features. Edges is detected by Canny Edge Detector and histogram is used then since it is not necessary to know an object's exact location but its presence.

A binary image is derived by Canny Edge Detector. We segment the image into 8*5 parts and calculate the histogram for edges in each image. We will get another 40-dimension feature for edge information.

Note that suitable parameters are used for the simulation study. For the size of Gaussian filter, large size causes more blurring, detecting larger, smoother edges while smaller filters cause less blurring, and allow detection of small, sharp lines. In the following simulation, we use 5*5 order Gaussian filter for 384*256 pixel image. For the threshold settings, a high threshold can miss important information while the low threshold will falsely identify irrelevant information (such as noise) as important information. Here we set 0.5*mean and

Algorithm 1 Canny Edge Detector

Given a color image I_{color}
Turn into gray image $I_{gray} = rgb2gray(I_{color})$
Denoise $I_{denoise} = filter(I_{gray}, 'Gaussian')$
Calculate gradient $G = \frac{d(I_{denoise})}{dp}$
Here $G = \sqrt{G_x^2 + G_y^2}, \theta = arctan\frac{G_y}{G_x}$
$\theta = round(0, \pi/4, \pi/2, \pi * 3/4)$
if $G(x, y) > \gamma_{high}$ **then**
  $E(x, y) = 1$
else
  if $G(x, y) > \gamma_{low}$ and $G(x+\Delta x, y+\Delta y) > \gamma_{high}$ **then**
    $E(x, y) = 1$
  else
    $E(x, y) = 0$
  end if
end if
return E

2.5*mean respectively as low threshold and high threshold for simulation.

2.3.2. Line Histogram using Hough Transform

Besides edge's location information, the distribution of orientation of edges still represents much for content of images. For example, buses or buildings usually contain many parallel lines while for beaches or mountains there are only one boundary to distinguish sea and beach or mountain and sky.

Here Hough Transform is implemented for deriving lines and orientation information, which is to find imperfect instances of objects within a certain class of shapes by voting. This voting procedure is carried out in a parameter space, from which object candidates are obtained as local maxima in a so-called accumulator space that is explicitly constructed by the algorithm for computing the Hough transform.

Algorithm 2 Hough Transform

Get edges $E = Canny(I_{original})$
Transform $r(\theta) = x_0 * cos(\theta) + x_0 * sin(\theta)$
Reconstruct along $\theta = round(0, \pi/4, \pi/2, \pi * 3/4)$
Histogram h based on (r, θ)
return h

3. K-MEANS CLUSTERING

In this section, we will study the performance of new combined features applying to unsupervised clustering applications. Linear combination of features is used and L1-norm distance measure is applied in this K-means simulation. We run simulation on a subset of Corel dataset images, which was used in [1–5]. It has 1000 images within 10 categories, namely Africa, Beach, Buildings, Buses, Dinosaurs, Flowers, Elephants, Horses, Mountains and Food. The overview pictures are shown on Fig. 2.

Fig. 2. Dataset Overview

3.1. Simulation Results for K-means

For k-means clustering, we simulate 20 times repeatedly and calculate the accuracy r for class k shown below:

$r(k) = \frac{1}{100} \sum_{i:T(i)=k} I_{C(i)=k}$,

where $T(i)$ denotes the true label for image i and $I_{C(i)=k}$ is the indicator function whether image i is classified as class k. The overall results are compared with other methods [1–5]. Histogram-based [1] method mainly focuses on color space histogram as its main features. SIMPLIcity [2] and FIRM [3] are region based detection. They segment images for catching up the local information, achieving higher accuracy than simple histogram based method. [4] has proposed an image retrieval scheme, where the relative weights of the features can be updated adaptively to specify its importance, achieving better results than the previous two. Multi-resolution method [5] extracts color, shape information from different resolution image. It uses most significant highest priority principle for matching scheme as adapted in [5]. However, our method uses both global information such as color, frequency, orientation and local details like edge. As evidenced by the results in Table 1, the new feature yields higher accuracy over the previously proposed features.

Table 1. K-means Results comparing with Current Methods

	Accuracy					
Class	**Purposed Method**	Histogram Based	SIMPLIcity	FIRM	Edge Based	Multi resolution
Africa	**.50**	.30	.48	.47	.45	.54
Beach	**.51**	.30	.32	.35	.35	.38
Building	**.42**	.25	.35	.35	.35	.40
Bus	**.83**	.26	.36	.60	.60	.64
Dinosaur	**.99**	.90	.95	.95	.95	.96
Elephant	**.39**	.36	.38	.25	.60	.62
Flower	**.86**	.40	.42	.65	.65	.68
Horse	**.64**	.38	.72	.65	.70	.75
Mountain	**.47**	.25	.35	.30	.40	.45
Food	**.47**	.20	.38	.48	.40	.53
Overall	**.61**	.36	.47	.51	.54	.59

Table 1 illustrates that shape and orientation information extracted using edge detector and Hough Transform significantly improves the recognition accuracy for classes of Beach, Bus and Mountain, which contains simple straight contour. However for elephant and horse cases, since the edges in these images a little mess up, extra information for orientations misleads its clustering to some extent, showing an even poorer result comparing with Edge Based method. But for overall average result, our proposed method beats all of other methods.

4. SVM CLASSIFICATION

In this section, we will study the performance of the new features when applied to supervised classification applications.

4.1. Introduction to Multiclass SVM

The standard Support Vector Machine (SVM) takes a set of input data and predicts for each given input, which makes the SVM a non-probabilistic binary linear classifier. In order to extend it into a multiclass classifier, several methods have been proposed where typically we construct it by combining several binary classifiers. Specifically, we will compare two multiclass classifiers based on binary classifications: "one-against-one"SVM and "one-against-all"SVM [9].

4.1.1. One-against-all SVM

One popular implementation for multiclass SVM classification is one-against-all SVM. It constructs k SVM models where k is the number of classes. The ith SVM is trained with all of the data in the ith class with positive labels, and all other data with negative labels. Thus given l training data $(x_1, y_1), ..., (x_l, y_l)$, where $x_i \in R^n$, $i = 1, ..., l$ and $y_i \in \{1, ..., k\}$ is the class label for x_i, the ith SVM solves the following problem:

$$\min_{\omega^i, b^i, \xi^i} \frac{1}{2}(\omega^i)^T \omega^i + C \sum_{j=1}^{l} \xi_j^i (\omega^i)^T$$

$$\begin{aligned} \text{subject to } & (\omega^i)^T \phi(x_j) + b^i \geq 1 - \xi_j^i, \text{if } y_j = i, \\ & (\omega^i)^T \phi(x_j) + b^i \leq 1 - \xi_j^i, \text{if } y_j \neq i, \\ & \xi_j^i \geq 0, j = 1, ..., l, \end{aligned}$$

where the training data x_i is mapped to a higher dimensional space by the function ϕ and C is the penalty parameter. For testing data, we apply k SVM respectively and predict it as the class that yields the largest value.

4.1.2. One-against-one SVM

Another major method is called one-against-one SVM. This method constructs $k(k-1)/2$ classifiers for any pair-wise classes. For training data from the ith and the jth class, we solve the following binary classification problem:

$$\min_{\omega^{ij},b^{ij},\xi^{ij}} \frac{1}{2}(\omega^{ij})^T\omega^{ij} + C\sum_t \xi_t^{ij}(\omega^{ij})^T$$
$$\text{subject to } (\omega^{ij})^T\phi(x_t) + b^{ij} \geq 1 - \xi_t^{ij}, \text{if } y_t = i$$
$$(\omega^{ij})^T\phi(x_t) + b^{ij} \leq 1 - \xi_t^{ij}, \text{if } y_t = j$$
$$\xi_j^i \geq 0, j = 1, ..., l.$$

If the classifier says the data belongs to the ith class, then the vote for the ith class is added by one. Otherwise, the jth is increased by one. Finally we predict the test sample as the class with the largest vote.

4.2. Simulation Results for SVM

We apply multiclass SVM classification to the same dataset above. First randomly choose a fixed number of images (from 5 to 95, step size 5, and finally a maximal training 99) out of 100 in each category as training data and test the rest. The performance of one-against-one SVM and one-against-all SVM is shown in Fig. 3.

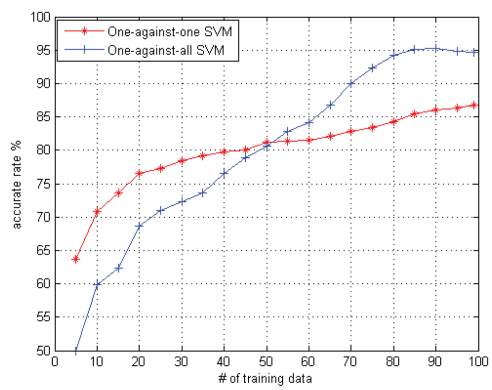

Fig. 3. Comparison between two SVMs

As expected, the performance for supervised learning is much better, comparing to K-means unsupervised learning. Fig. 3 illustrates that originally when training set is smaller than 50, one-against-one SVM performs better. Intuitively, pair-wise comparisons for the whole training dataset make better use of limited data, yielding better results. However, when training is sufficient, one-against-all SVM performs better because in this dataset, Beaches and Mountains share a very similar structure in both color and shape histogram. Therefore when we perform a pair-wise comparison, it will choose these two classes with high probability when they are compared with other eight classes. Finally when voting, the voting number of two classes will get almost the same number, producing error classification. Another interesting point in Fig.3 is that in one-against-all SVM method, when training data are more than 90%, the classification accuracy slightly decreases due to overfitting, or limitation of dataset.

5. CONCLUSION

This paper has proposed a color-frequency-orientation histogram based multiclass image retrieval method. It is used in unsupervised learning image retrieval using K-means and supervised learning image retrieval using SVM. Simulation results on a subset of Corel dataset (1000 images, 10 categories) show it is an advantageous method in accuracy.

6. REFERENCES

[1] Y. Rubner, L.J. Guibas, and C. Tomasi, "The earth movers distance, multi-dimensional scaling, and color-based image retrieval," in *Proceedings of the ARPA Image Understanding Workshop*. Citeseer, 1997, pp. 661–668.

[2] J.Z. Wang, J. Li, and G. Wiederhold, "Simplicity: Semantics-sensitive integrated matching for picture libraries," *IEEE Transactions on pattern analysis and machine intelligence*, pp. 947–963, 2001.

[3] Y. Chen and J.Z. Wang, "A region-based fuzzy feature matching approach to content-based image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1252–1267, 2002.

[4] M. Banerjee, K. M K, and PK Das, "Image retrieval with visually prominent features using fuzzy set theoretic evaluation," in *IEEE International Conference on Visual Information Engineering VIE*. Citeseer, 2006, pp. 298–303.

[5] PS Hiremath and J. Pujari, "Content based image retrieval based on color, texture and shape features using image and its complement," *International Journal of Computer Science and Security*, vol. 1, no. 4, pp. 25–35, 2007.

[6] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 12, pp. 1349–1380, 2000.

[7] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 2, pp. 524–531.

[8] J. Li, J.Z. Wang, and G. Wiederhold, "Irm: integrated region matching for image retrieval," in *Proceedings of the eighth ACM international conference on Multimedia*. ACM, 2000, pp. 147–156.

[9] C.W. Hsu and C.J. Lin, "A comparison of methods for multiclass support vector machines," *Neural Networks, IEEE Transactions on*, vol. 13, no. 2, pp. 415–425, 2002.