

# COLOR PHOTO CATEGORIZATION USING COMPRESSED HISTOGRAMS AND SUPPORT VECTOR MACHINES

*Xia Feng, Jianzhong Fang and Guoping Qiu*

School of Computer Science, The University of Nottingham, United Kingdom

## ABSTRACT

In this paper, an efficient method using various histogram-based (high-dimensional) image content descriptors for automatically classifying general color photos into relevant categories is presented. Principal component analysis (PCA) is used to project the original high dimensional histograms onto their eigenspaces. Lower dimensional eigenfeatures are then used to train support vector machines (SVMs) to classify images into their categories. Experimental results show that even though different descriptors perform differently, they are all highly redundant. It is shown that the dimensionality of all these descriptors, regardless of their performances, can be significantly reduced without affecting classification accuracy. Such scheme would be useful when it is used in an interactive setting for relevant feedback in content-based image retrieval, where low dimensional content descriptors will enable fast online learning and reclassification of results.

## 1. INTRODUCTION

The fast advancement of digital imaging technology has resulted in the exponential increase of image data in both professional archives and personal leisure collections. Effectively managing large image collections and making them easily accessible poses significant technical challenges. In the past decade, there has been significant research effort in content-based image retrieval (CBIR) [1]. In a CBIR system, a user can query the database with a visual example and the system will return an ordered list of images that are similar to the query in some visual sense. Image classification, in which a collection of images are classified into categories or classes of some descriptions [6], can be used in conjunction with the CBIR technologies to boost their usability and performance. In particular, image classification techniques based on various machine learning approaches have been shown to be particularly useful in providing relevant feedback in CBIR [13].

It is well known that image data is of very high dimension. Some popular image content descriptors often used in CBIR and image classification, e.g. [6], also have very high dimensionality. In this work, we investigate and exploit the redundancy of some popular image content descriptors. In particular, we study the image classification performances and compressibility of a number of recently proposed histogram-based image content descriptors. It is found that whilst different descriptors have quite discernible classification performances, all are highly compressible. For example, reducing the dimension of some descriptors by as much as 70% causes no deterioration in performance, and reducing the dimension of some descriptors by over 90% only causes about 2% drop in classification performance. Our results indicate that many high dimensional content descriptors are highly redundant and more compact descriptors may exist. We show that principal component analysis (PCA) can be used as a

convenient and very effective tool to obtain low-dimensional content descriptors for general image classification.

Image classification scheme similar to what we are studying here is often used in an interactive setting for relevant feedback in a content-based image retrieval system [1] where realtime (or very fast) online learning and reclassification is highly desirable. Our results show that we can use much lower dimensional content descriptors than those used in the literature [6] for this purpose. Indeed, one can simply use the PCA approach in this paper to derive lower dimensional descriptors for fast relevant feedback computation.

The organization of the paper is as follows. In section 2, we briefly review related work. Section 3 gives a brief description of 5 histogram based content descriptors studied in this paper. Section 4 briefly summarizes PCA for data compression. Section 5 briefly describes support vector machine. Section 6 presents our experiment setting. Section 7 present experimental results and section 8 concludes the paper.

## 2. RELATED WORK

Organizing images into semantic categories can be extremely useful for searching and browsing through large collections of images. Recently, various methods have been proposed for this purpose. Very high dimensional color histograms and Support Vector Machines (SVM) have been used to classify color photographs into various categories [6]. SVM has also been used to perform relevant feedback in content based image retrieval [13]. AdaBoost algorithm [7] has been used for natural image retrieval [8]. The author in [9] has presented empirical results of boosted image classification.

Projecting image features onto sub spaces for classification and recognition has been investigated by a number of authors in various applications. In [12], subspace method has been used for image classification. Principal component analysis (PCA) has been used to project vectors onto eigenspace for face recognition [11]. In [4] the authors have successfully applied PCA to compress color histogram for content-based image indexing. PCA has been used in a feature selection technique to reduce feature dimensionality in a SVM based face detection system [14].

## 3. HISTOGRAM-BASED IMAGE CONTENT DESCRIPTORS

There has been many image content descriptors proposed in the content-based image indexing and retrieval literature [1]. In this paper, we examine five histogram-based descriptors. This section gives a brief description of each.

### 3.1 Opponent Color Histogram

Color histogram is one of the earliest image content descriptors used in content-based image indexing and retrieval

[1]. A color histogram can be constructed in a 3d color space. However, the dimensionality of such a histogram is often very high. Researchers have found that it is possible to construct a 2d chromaticity histogram which will often give equal performance to 3d ones [4].

In [4], it was shown that the red-green and blue-yellow opponent chromaticity space decorrelates color information well and the 2d color histograms constructed in this space are also uniformly distributed hence render them more compressible. The opponent chromaticities of red-green (rg) and blue-yellow (by) are defined in terms of  $r$ ,  $g$ ,  $b$  chromaticities:

$$(rg, by) = (r - g, \frac{r + g}{2} - b) \quad (1)$$

The opponent color histogram investigated in this paper is constructed in the  $rg$  and  $by$  space according to (1).

### 3.2. Color Correlogram

One of the well know problems of color histogram is that it contains no spatial information which has greatly limited its discriminative power. Recently researchers have developed a better and more discriminative technique known as the Color Correlogram (CC) [5], which has been shown to provide significant improvement over the original color histogram approach. Color Correlogram is very similar to co-occurrence matrix developed some 20 year ago for grey scale texture classification. Formally, the CC of image  $\{Z(x,y), x = 1, 2 \dots M, y = 1, 2, \dots, N\}$  is defined as

$$CC(i, j, k) = \Pr\{Z(x_1, y_1) \in C_i \mid Z(x_2, y_2) \in C_j\} \\ k = \max\{|x_1 - x_2|, |y_1 - y_2|\} \quad (2)$$

where the original image  $Z(x,y)$  is quantized to a fixed number of colors  $C_1, C_2, \dots, C_L$  and the distance between the two pixel  $k = 1, 2, \dots, K$  is fixed *a priori*. In words, the CC of an image is the probability of joint occurrence of two pixels some  $k$  distance apart that one pixel belongs to color  $C_i$  and the other belongs to color  $C_j$ . The size of CC is  $O(L^2K)$ . To reduce storage requirement, [5] concentrated on auto-correlogram whereby  $i = j$  and its size is of  $O(LK)$ . We also only investigate the auto-correlogram in this paper.

### 3.3 MPEG-7 Color Structure Descriptor

The MPEG-7 color structure descriptor is a color feature descriptor that captures both color content (similar to a color histogram) and information about the structure of this content [16]. The extraction method embeds color structure information into the descriptor by taking into account all colors in a structuring element of  $8 \times 8$  pixels that slides over the image, instead of considering each pixel separately. This descriptor is more discriminative than color histogram because it can distinguish between two images in which a given color is present in identical amounts but where the structure of the groups of pixels having that color is different in the two images. MPEG-7 standard specifies that color values are represented in the double-coned HMMD color space, which is quantized non-uniformly. Detail description of this descriptor can be found in [16].

### 3.4 Colored Pattern Appearance Histogram

The colored pattern appearance model (CPAM) which has two channels capturing the characteristics of the chromatic and achromatic spatial patterns of small image regions has been

used to compile content descriptors for image retrieval [2]. In this method, the visual appearance of a small image block is modeled by three components: the stimulus strength, the spatial pattern and the color pattern. In [2], the YCbCr space is used, and the stimulus strength  $S$  is approximated by the local mean of the Y component. The pixels in Y normalized by  $S$  form the achromatic spatial pattern (ASP) vector. Because  $C_b$  and  $C_r$  have lower bandwidth, they are sub-sampled. The sub-sampled pixels of  $C_b$  and  $C_r$  are normalized by  $S$ , and then concatenated together to form the chromatic spatial pattern (CSP) vector. Vector quantization (VQ) is used to encode the ASP and CSP vectors. Let  $Q_c$  with a codebook  $C_c = \{C_c(j); j \in \hat{I} \ M\}$  be the VQ quantizer for the CSP vectors and  $Q_A$  with a codebook  $C_A = \{C_A(i); i \in \hat{I} \ N\}$  be the VQ quantizer for the ASP vectors. The probabilities that the CSP vectors are coded by the  $i^{\text{th}}$  codeword of  $C_C$  and the probabilities of the ASP vectors are coded by the  $j^{\text{th}}$  codeword of  $C_A$ , for all possible  $i$  and  $j$  form the colored pattern appearance histogram (CPAH):

$$CPAH = \{\{\Pr(Q_A(ASP) = i), \forall i\}, \{\Pr(Q_C(CSP) = j), \forall j\}\} \quad (3)$$

### 3.5 Layered Color Indexing

From human perception's point of view; different frequency components of the visual stimulus may be treated differently by the visual system. Physically, different frequency components of the signal may correspond to different objects or object parts. Consequently, different frequency components of the visual stimulus may not only have different perceptual significance, they may also correspond to different physical objects. Incorporating such concept is helpful to the development of more effective image retrieval techniques. In [3] a layered color indexing method was presented. An image is first passed through a filter bank (each filter of the filter bank covers a specified spectral bandwidth). The output of the filter bank is used to classify the pixels. Pixels in an area with similar spatial frequencies are then retained on the same layer. Each individual layer, which contains only those pixels in areas with similar frequency distributions, is used to form its own color index. The aggregation of the feature indices from all the layers then forms the overall index of the image.

Let  $H_k$  be the color histogram of layer  $L_k$ , then layered color indexing (LCI).

$$H(x) = \{H_k\}, \forall k \quad (4)$$

That is, multiple color histograms for an image, each taking colors from pixels in areas with similar sharpness. This approach not only indexes the color, but also associates colors with their surface roughness. It has been argued that such an association not only has perceptual significance, it also makes sense physically. It has been shown that the layered color indexing method significantly enhances the power of color indexing but at the same time retains its simplicity and elegance.

## 4. DATA COMPRESSION USING PCA

Principal component analysis (PCA) [17] is a classical statistical method which has been widely used in data analysis and compression. The basic idea in PCA is to find the  $n$  linearly transformed components  $s_1, s_2, \dots, s_n$  so that they explain the maximum amount of variance. In practice, PCA is done by calculating the  $n$  largest eigenvectors of the (sample) covariance matrix  $\Sigma = E(x^t x)$ , where  $x$  is the  $m$  dimensional sample vectors.

The basic goal in PCA is to reduce the dimension of the data. Thus one usually chooses  $n < m$ . Indeed, it can be proven that the representation given by PCA is an optimal linear dimension reduction technique in the mean-square sense. Such a reduction in dimension has important benefits. First, the computational overhead of the subsequent processing stages is reduced. Second, noise may be reduced, as the data not contained in the  $n$  first components may be mostly due to noise. Third, a projection into a subspace of a very low dimension is useful for visualizing the data. Note that often it is not necessary to use the  $n$  principal components themselves, since any other orthonormal basis of the subspace spanned by the principal components (called the PCA subspace) has the same data compression or denoising capabilities.

In the literature, PCA has been successfully applied to dimensionality reduction in many fields of scientific and engineering pursuits. Since the dimensionality of typical histogram descriptors are quite high which will result in high computational cost in distance calculation, we use PCA to reduce the dimensionality of histograms described in the last section before they are presented to the classifier. Our goal is to investigate the amount of redundancy exist in each of these descriptors so that we can develop computationally efficient methods using lower dimensional features without sacrificing classification accuracy performances.

## 5. SUPPORT VECTOR MACHINES

Support vector machine [15] is an emerging machine learning technology that has been successfully used in a variety of pattern recognition tasks. We shall consider SVMs in the binary classification setting. We are given training data  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  that are vectors in some space  $\mathbf{x} \in \mathbb{R}^d$ . We are also given their labels  $\{y_1, \dots, y_n\}$  where  $y_i \in \{-1, 1\}$ . SVMs are simply hyperplanes that separate the training data by a maximal margin. All vectors lying on one side of the hyperplane are labeled as -1, those on the other side are labeled as 1. Training instances that lie closest to the hyperplane are called *support vectors*. More generally, SVMs allow us to project the training data in space  $\mathbf{x}$  to a higher dimensional feature space  $\mathbf{F}$  via a Mercer kernel operator  $K$ . In other words, we consider the set of classifiers of the form:

$$f(x) = \sum_{i=1}^n \mathbf{a}_i k(x_i, x) \quad (5)$$

When  $f(x) \geq 0$  we classify  $\mathbf{x}$  as +1 (class 1); otherwise we classify  $\mathbf{x}$  as -1 (class 2).

SVMs are designed for binary classification [15]. When dealing with several classes, such as in image classification, one needs an appropriate multiclass method. Our approach is One per Class (OPC), also known as ‘‘One against the others’’, which compares a given class with all the others put together [6]. In the ‘‘one against the others’’ algorithm,  $N$  hyper planes are constructed, where  $N$  is the number of classes. Each hyperplane separates one class from the other classes. In this way, we get  $N$  decision functions,  $f_k(x)$ ,  $k = 1, \dots, N$ . The class label of a new sample  $\mathbf{x}$ ,  $C_x$ , is given by

$$C_x = \arg \left( \max_{v_k} (f_k(x)) \right) \quad (6)$$

i.e., the class with the largest decision function.

## 6. LEARNING IMAGE CLASSIFICATION USING SVM IN HISTOGRAM SUBSPACES

We investigate the classification performance and compressibility of the various histogram-based image content

descriptors using a system set up as shown in Fig. 1. From an input image, we derive its histogram descriptors, which are then passed to the PCA module to reduce their dimensionality. The compressed histogram features are then fed to the SVM classifiers which then output the identity of the input image. Let  $\mathbf{y}_i$  be the  $i$ th eigenvector of the co-variance matrix of the  $m$  dimensional image descriptor  $\mathbf{x}$ ,  $s = \{s_i, i = 1, \dots, m\}$  the projections of  $\mathbf{x}$  onto the eigenspace, and  $s_i = \mathbf{x} \cdot \mathbf{y}_i$ . It is well known that the vector formed from the first few  $s_i$ ,  $\mathbf{s}' = \{s_i, i = 1, \dots, n\}$ ,  $n < m$ , can be used to approximate  $\mathbf{x}$ .  $\mathbf{s}'$  is then presented to the SVMs for classifying the input images. We investigate how the selection of  $n$  affects the classification performances and report our findings in the next section.

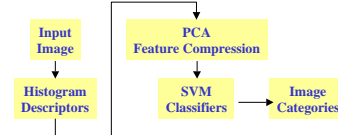


Fig. 1. Image categorization experiment set up

## 7. EXPERIMENTAL RESULTS

### 7.1 Data Sets

We use 10 categories of color photographs from the Corel color photo collection. These categories are labeled by human. The categories are: Lions, Elephants, Tigers, Horses, Sky scenes, Cheetahs, Eagles, Night Scene, Sunset and Roses. Each category consists of 100 color photos (It should be pointed out that some images in one category can have visual features that are very similar to those found in images belong to other categories, hence making automatic categorization difficult). For each class, we randomly select 50 images as the training samples and use the other 50 for testing.

For each type of histogram, one co-variance matrix is constructed using the 500 training samples to calculate the eigenvectors for feature reduction. For each category, we train a SVM for each type of histogram descriptor.

### 7.2 Results

To evaluate the classification performance, we measure the overall correct classification rate (CCR). In all experiments, we used a 256d opponent color histogram (OCH), a 256d color auto-correlogram (CAC), a 256d MPEG-7 color structure histogram (M7CS), a 384d colored pattern appearance histogram (CPAH) and a 256d layered color indexing (LCI). In a first experiment, we experiment on the five raw histograms without compressing them. In order to investigate how sensitive the performances are to the training parameters, we train the SVMs with different margins ( $\epsilon$ ). For details of the adjustable margins, please refer to [15].

Table 1 shows the number of support vectors for different training margins and Table 2 shows the correct classification rates for different descriptors for SVMs trained using different margins. It is seen that the training is very stable, both the number of support vectors and the correct classification rates do not vary significantly for different SVM training parameters. From Table 2, we also see that M7CS gives the best performance. It is also interesting to note that the number of supporting vectors for this descriptors is three times as many as those of other four descriptors. This may explain the good performance of the M7CS descriptor. It would be very

useful and interesting to understand the significance of this correlation.

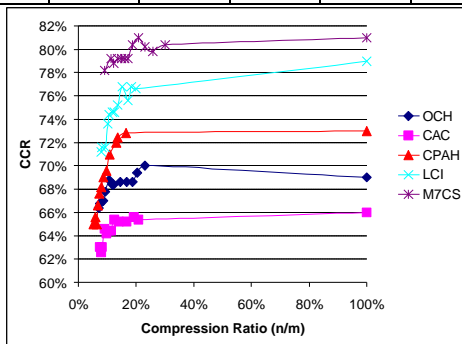
Now that we have an idea how large the training margin should be, we choose the one that performs well ( $\epsilon = 1.8$ ) for the rest of our experiments. We investigate how the amount of compression affects the classification performances. Various factors of compression are tried, and results are shown in Fig. 2. These results clearly demonstrated that these descriptors are highly compressible. We found that reducing the dimension of the descriptors by 70% does not affect the performance accuracy. By reducing the dimension by 90% we only notice 2% drop in performance. Fig. 3 shows some typical confused images.

**Table 1 The number of support vectors for different descriptors trained using different margins ( $\epsilon$ ).**

$\epsilon$	OCH	CAC	M7CS	CPAH	LCI
1	123	115	350	111	116
1.4	121	113	353	110	113
1.8	117	110	354	109	115
2.2	113	110	353	107	113

**Table 2 Correct classification performances for different descriptors trained using different margins ( $\epsilon$ ).**

$\epsilon$	OCH	CAC	M7CS	CPAH	LCI
1	68%	65%	80%	70%	76%
1.4	68%	66%	80%	72%	77%
1.8	69%	66%	81%	73%	79%
2.2	70%	66%	81%	73%	78%



**Fig. 2** Correct classification rates for different descriptors for different compression factors.



**Fig. 3** Typical images that are miss classified. This is not a surprising result. For example, although two images may contain different animals (hence classified as different category by human), the large backgrounds are very similar. Histogram based methods will be unable to distinguish such cases. Other more sophisticated methods are needed, however, such methods are still being researched.

## 8. CONCLUDING REMARKS

In this paper, we have investigated five histogram-based based image descriptors for general color photo categorization. We show different descriptors do give different performances. We also found that all these histograms are highly redundant and can be significantly compressed without affecting the classification performance. Histograms are a descriptor used extensively in image content management and they normally have very high dimension. From the results of this study, we

conjecture that there may exist lower dimensional alternative to the existing high dimensional histogram based image content descriptors, it may well worth pursuing these lower complexity alternatives. One possibility is to use the technique used in this paper, i.e., PCA, to reduce the histograms' dimensionality. Other better methods may also exist and warrant further investigation. This would be useful when such scheme is used in an interactive setting for relevant feedback, where a low dimensional feature will allow fast training and reclassification of results.

## References

1. A. W. M. Smeulders et al, "Content-based image retrieval at the end of the early years", IEEE Trans PAMI, vol. 22, pp. 1349 - 1380, 2000
2. G..Qiu, "Indexing chromatic and achromatic patterns for content-based color image retrieval", Pattern Recognition, Vol 35, Issue 8, pp. 1675-1686, Aug. 2002
3. G. Qiu and K-M Lam, "Spectrally layered color indexing", in Image and Video Retrieval, CIVR2002, M.S. Lew, N. Sebe, J.P. Eakins (Eds.), Lecture Notes in Computer Science, vol. 2383, pp. 100-107, Springer, July 2002
4. J. Berens, G. D. Finlayson and G. Qiu, "Image indexing using compressed color histogram", IEE Proceedings, Vision, Image and Signal Processing, vol. 147, no. 4, pp. 349 - 355, 2000
5. J. Huang, S.R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image Indexing Using Color Correlograms", International Journal of Computer Vision, pp. 245 - 268, 1999
6. O. Chapelle, P. Haffner, and V. Vapnik. "SVMs for histogram-based image classification". IEEE Transaction on Neural Networks, 1999.
7. Y. Freund and R. Schapire. "A decision-theoretic generalization of online learning and an application to boosting". J. Comp. & Sys. Sci. , 55 (1):119-139, 1997.
8. K. Tieu and P. Viola. "Boosting image retrieval". Proc. of CVPR , 1:228-235, 2000.
9. N. Howe, "Boosted Image Classification:An Empirical Study",
10. K-S Goh, E. Chang and K-T Cheng, "Support Vector Machine Pairwise Classifiers with Error Reduction for Image Classification", 3rd Intl Workshop on Multimedia information retrieval (MIR2001), Ottawa, Canada, October 5,2001.
11. M. Turk. and A. Pentland., "Eigenfaces for Recognition", J. Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86,1991.
12. T. Balachander and R. Kothari, "Introducing Locality and Softness in Subspace Classification" ,Pattern Analysis & Applications Volume 2 Issue 1 (1999) pp 53-58 Springer-Verlag London Limited 1999
13. Q.Tian, P.Hong and T.S.Huang, "Update Relevant Image Weights for Content-Based Image Retrieval Using Support Vector Machines", IEEE ICME'2000
14. T. Serre, et al, "Feature Selection for Face Detection", MIT A.I. Memo No. 1697, September, 2000
15. M. Hearst, "Support vector machines", IEEE Intelligence Systems, pp. 18 - 28, July/August, 1998
16. B. S. Manjunath, et al, " Color and Texture Descriptors", IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 6, JUNE 2001
17. C. M. Bishop, Neural Networks for Pattern Recognition, Clarendon Press, Oxford (1995).