Abstract

Image classification is very active and promising research domain in image retrieval and management. In this paper, we propose a boosting image classification scheme with automatically selecting the discriminative features. Firstly, we present an image feature called the orientational color correlogram (OCC) and apply it to image classification. OCC is an extension of the color correlogram by adding in the orientational information which can take into account both the local color correlation and the global context structure of image. Secondly we also give a solution to feature selection for the very high dimensionality of OCC by using a boosting classification scheme which can select the most discriminative features automatically. In our experiments, only a small number of elements of OCC are selected, which can reduce the storage space of classifier models and speed up the classification process. The experimental results suggest the proposed method has preferable performances.

1. Introduction

Along with the rapid increase of the capacity of the images and videos in databases, organizing images into semantic categories can improve extremely the efficiency in CBIR. Image classification is the task of classifying images into semantic categories based on the available training data. This categorization can be helpful both in semantic organizations of digital libraries and in obtaining automatic annotations of images.

Intuitively, image classification should go along with object recognition and image content analysis. But previous works [1][2][3] have indicated that image classification can be done without object recognition procedure. Szummer and Picard [1] use the features involving Ohta color, texture and coefficients of DCT to classify the images to indoor and outdoor categories with KNN classifiers. Vailaya et al [2] describe a method to classify vacation images into classes like indoor/outdoor, city/landscape, and sunset/mountain/forest scenes. They use a Bayesian framework for separating the images in a classification hierarchy. Huang et al [3] propose using the singular value decomposition with the banded color correlograms to extract a “latent semantic” structure of images for an automatic hierarchical image classification scheme. Some useful conclusion can be drawn from the previous works: (1) relative relationships matter more than absolute properties, (2) global configuration of the relationships is importance and (3) low resolution information is sufficient.

When a large number of low-level features are extracted for representing the image, we should choose discriminative features for image classification. Some methods, such as principal component analysis (PCA) and vector quantization (VQ), have been applied to reduce the feature dimension and select the effective features. But facing to the high dimension features and a large numbers of train data, they are difficult to be applied in a practical system because for their computational complexities.

In this paper we present an effective classification scheme for image classification. Firstly, we propose the image feature called the orientational color correlogram (OCC), which is an extension of gray co-occurrence matrix and similar to the color correlogram [4]. OCC can describe both the local color spatial feature with a smaller inter-pixel distance and the global context structure with a larger inter-pixel. So another benefit of the OCC is describing the spatial information of images instead of image segmentation. So for the high dimensionality of the OCC, we apply a variant AdaBoost algorithm [14] to select discriminative features, which has great success in the real-time object detection [5].

The rest of the paper is organized as follows. Section 2 gives the detailed description for orientational color correlogram. The variant AdaBoost algorithm for automatic feature selection is presented in section 3. Then a series of experimental results are shown in the section 4. Section 5 concludes the paper.

2. Orientational color correlogram

One of the well known problems of color histogram is that it contains no spatial information which greatly limited its discriminative power. Color correlogram (CC) [4] have better descriptive and more discriminative abilities for images understanding, which has been shown to provide significant improvement over the original color histogram approach. For better description of image content, we propose the orientational color correlogram (OCC). OCC is an extension of the traditional gray-level co-occurrence matrix [6] which is widely used for texture
description, classification and segmentation. In recent years some different extensions of the co-occurrence basic ideas have been successfully employed for a number of computer vision tasks such as the estimation of the optical flow [7], image retrieval [8], and the recognition and matching of objects in numerous applications [9][10].

The co-occurrence matrix is the conditional joint probabilities of all pairwise combinations of grey level (i, j) in the spatial window of interest given two parameters: inter-pixel distance (δ) and orientation (θ) [6]. Here, we use the joint probabilities with four variables (θ, δ, i, j), the extended form of traditional gray-level co-occurrence matrix. And the (i, j) are color level instead of gray level. Compared with the color correlogram, this extended co-occurrence matrix takes the orientational information into account and we call it “orientational color correlogram”. We give the formal definition below.

Let I be an n1×n2 image. The color in I quantized into M colors c1, …, cM. For a pixel p=(x, y) ∈ I, let I(p) denote its color. Let Ic = {I(p) = ci}. Thus, the notation p ∈ Ic is synonymous with p ∈ I. I(p) = c. We define the orientation (θ) with 8 directions 0°, 45°, 90°, 135°, 180°, 225°, 270° and 315° and calculate the inter-pixel distances (δ) in each direction. Such as in case of θ=0°, the δ = |p2 – p1|, when p1 = (x, y) and p2 = (x+δ, y), and in case of θ=45°, the δ = |p2 – p1|, when p1 = (x, y) and p2 = (x+δ, y+δ). We denote δ = |p2 – p1|b for the inter-pixel distance of p2, p1 in the orientation b.

\[
\alpha(\delta, \theta, i, j) = \Pr(p_1 \in I_c, p_2 \in I_\theta, |p_2 - p_1| = \delta)
\] (1)

In words, the OCC of an image is the probability of joint occurrence of two pixels with distance (δ) apart in direction (θ) that one pixel belongs to color cl and the other belongs to color c2. So OCC not only concludes the orientational information but describes both the local spatial correlation of colors and the global distribution of the local spatial correlation. For extraction of local features, a small value of inter-pixel is sufficient to capture the spatial content of image region.

In terms of the definition of OCC, we can get the following derivation:

\[
\alpha(\delta, \theta, i, j) = \alpha(\delta, \theta_2, i, j), \text{if} (\theta_2 - \theta_1) = 180^\circ
\] (2)

So the OCC is symmetric in the opposite direction and we can represent it just in four directions. Here we choose 0°, 45°, 90° and 135°.

The color information usually is described in 3D color space. When color of each channel is quantized to L colors c1, …, cL, the size of color space is L3. Given the size of the distance set is D and the orientation set is N. The size of such an OCC is O(D×N×L3) which is very large. In this paper, we set D=64, N=4, L=8 and the size of OCC is O(64×4×83)= 67,108,864.

Although these correlograms are quite sparse, we also try two substantive features (orientational color autocorrelogram and OCC with chromaticities) to reduce the storage space.

Similar to the color autocorrelogram [4], we calculate the orientational color autocorrelogram (OCAC) with the probability of joint occurrence of two pixels with distance (δ) apart in direction (θ) that the two pixels both belong to color cL. The size of orientational color autocorrelogram is O(D×N×L3).

\[
\alpha(\delta, \theta, i, j) = \Pr(p_1 \in I_c, p_2 \in I_\theta, |p_2 - p_1| = \delta)
\] (3)

Alternatively, we reduce the size by decreasing the dimension of color space. Researchers have found that it is possible to construct a 2D chromaticity histogram which will often give equal performance to 3D ones [11]. It is shown that red-green and blue-yellow opponent chromaticity space decorrelates color information well. The opponent chromaticities of red-green (rg) and blue-yellow (by) are defined in terms of r, g, b chromaticities.

\[
(\text{rg}, \text{by}) = \frac{r-g, r+g-b}{2}
\] (4)

When each chromaticity is quantized to L colors c1, ..., cL, the size of color space is L2. The size of OCC is O(D×N×L2), which has a considerable decrease in feature size.

3. Boosting classification scheme

Given a feature set and a training set of the positive and negative images, any machine learning approaches could be used to learn a classification function. But the size of OCC is so large that it is infeasible to directly train the classifier. Although some methods, such as PCA, VQ, have developed to effectively reduce the feature dimension, but the computation of them is expensive. If we treat the every element of OCC as a single feature, many features are not distinctive enough and we can just select the few most discriminative elements from OCC. In our work, a boosting classification scheme is used to greedily select a small number of features from the very large number of potential features.

3.1. Variant AdaBoost algorithm

In its original form, the AdaBoost learning algorithm [14] is used to combine a collection of weak classifiers to form a strong classifier. Weak classifiers are combined using a weighted majority vote through associating large weights with the good weak classifiers and small weights with the poor weak classifiers. AdaBoost is an aggressive mechanism for selecting a small set of good weak classifiers, so it can be easily interpreted as a greedy feature selection process if each weak classifier depends on a single feature. In each step of AdaBoost algorithm, a series of weak classifiers are learned, each of which depends only on a single feature and we select the
classifier with the best discriminative ability. Thus the AdaBoost process can be also viewed as a feature selection process [5].

A weak classifier \( h_j(x) \) is thus defined as the following:

\[
h_j(x) = \begin{cases} 
1 & \text{if } p_j f_j(x) < p_j \theta_j \\
-1 & \text{otherwise}
\end{cases}
\]

where \( j \) is the feature index, \( f_j(x) \) is the \( j \)-th feature, \( \theta_j \) is threshold, \( p_j \) is the parity indicating the direction of the inequality sign.

The weak classifiers are trained with single feature, their discriminating performance is low. But the final strong classifier has a good performance through the optimal combination of the weak classifiers.

Based on the mechanism of boosting, a variant of AdaBoost algorithm is applied here.

1. For the \( N \) training samples \((x_1, y_1), \ldots, (x_n, y_n)\), where \( y_i = 1, -1 \) respective represent the positives and the negatives.
2. Initialize weights \( w_i = 1/N, i=1,\ldots,N \).
3. For \( t = 1 \ldots T \),
   1) Normalize weights so that \( w_t \) is a distribution.
   2) For each feature \( j \), train a classifier \( h_j \) which is restricted to using a single feature using weight \( w_t \) on the training samples. The error \( \varepsilon_j \) is evaluated,
   \[
   \varepsilon_j = \sum_{y_i \neq h_j(x_i)} w_t.
   \]
   3) Chose the classifier \( h_j \) with lowest error \( \varepsilon_j \) as the classifier \( h_t \).
   4) Update weights according to:
   \[
   w_{t+1} = w_t \beta_t^{1-e_t},
   \]
   where, \( e_t = 0 \) if sample \( x_i \) is classified correctly, \( e_t = 1 \) otherwise, and \( \beta_t = \frac{1}{1-\varepsilon_t} \).
4. The final strong classifier is:

\[
h(x) = \text{sign} \left[ \sum_{t=1}^{T} \alpha_t \cdot h_t(x) \right], \text{ where } \alpha_t = \log \frac{1}{\beta_t}
\]

Figure 1. A variant AdaBoost algorithm

In this way, the final model of the classifier will be extremely small because we just store an index of the selected features, the respective weights, thresholds and parity signs for the weak classifiers.

3.2. Image classification

In most of the previous work, the images are segmented into sub-blocks for better describing the local spatial information and the global context structure. In this paper, we avoid the segmentation by setting the inter-pixel distance of OCC parameter in \([1, \min(n_1,n_2)]\), where \( n_1, n_2 \) are the height and width of image. When the inter-pixel distance is small, the correlogram reflect well the local information of images, and when the inter-pixel distance is large, the correlogram can also describe the global spatial context structure.

For a given image category, we choose the images in this category as the positive samples and the images in the other categories as the negative samples. By applying the variant AdaBoost algorithm, our scheme can automatically select the discriminative features which can involve the local information with the smaller inter-pixel distance and global information with the larger inter-pixel distance.

4. Experiments

Figure 2. Examples of the 10 image categories

We test our method on 10 fairly representative categories which are basketball, beach, building, classroom, land, meeting room, store setting, studio setting, vegetation and weather report. Figure 2 shows some examples of the 10 image categories. These image categories are defined in the IBM VideoAnnEx MPEG-7 Collaborative Annotation System [13]. The experimental images were obtained from keyframes of the common annotation results of TRECVID2003 [12] development data, which contains ABC World News Tonight, CNN Headline News and C-SPAN news. Totally 2,200 images are collected. In each category, 50% images are randomly chosen for training and the rest 50% for test. All images normalized to 64×64 by bilinear interpolation and quantize the RGB color space into 8×8×8 =512 colors, which are good enough for representing the images semantics. We set size of distance set \( D=64 \), direction number \( N=4 \) and color level \( L=8 \). To greatly reduce the
feature number, we also represent the orientational color correlogram by two alternative approaches, calculating the autocorrelogram (OCAC) and reducing the color space into 2D with opponent chromaticities (OCCOP).

We separately test the 10 classifiers on the whole test set with the OCC, OCAC, OCCOP and color correlogram (CC), the color histogram (CH) also is tested as a benchmark. The results are shown in Figure 3 and the numbers of selected features is 500.

![Figure 3](image)

Figure 3. Classification accuracies of different image features for the 10 image categories. OCC: orientational color correlogram, OC: color correlogram, OCAC: orientational color autocorrelogram, OCCOP: OCC using opponent chromaticities color space, CH: color histogram.

In our image classification scheme, although the training process is time-consuming, but the test process is very fast because we just use a small feature set, which is possible to be used in practical system.

5. Conclusion

The paper has described a scheme for image classification. Firstly, we propose the orientational color correlogram. Secondly, we gave a solution to automatically select the discriminative features from the high dimension features. Compared with the original features, only a small number of discriminative features are extracted, which can reduce the storage space of classifier models and speed up the classification process. The experimental results show the proposed scheme has preferable performances than the benchmark. Because of involving the orientational information in OCC, the directions of our future work will concentrate on the image classification with automatic image orientation adjustment.

6. Acknowledgement

The research described in this paper has been supported by NSF of China (69935010).

7. References


