Region-based Image Classification

Qi Gu

Abstract

Image classification using low-level features is always a challenging research in computer vision. Recent years, content-based image retrieval has emerged as an important area in computer vision and multimedia computing. In this project, I'm going to introduce and implement an approach [1] that can better represent the image using low-level features and then I apply this method in image classification. To test the efficiency of this approach, I set up two experiments to compare it with a traditional image representation approach.

1. Introduction

The term image classification refers to the labeling of images into one of a number of predefined categories. Although this is always not a difficult task for humans, it has proved to be an extremely difficult problem for machines. Therefore, how to extract the key information from a bunch of low-level features and use them to represent an image is the crucial part for classification. In this project, the main procedure can be split into 3 parts: image segmentation, representation and classification. Image segmentation is trying to segment image into regions such that each region is roughly homogeneous in color and texture and Image representation concentrates on how to use low-level features to represent an image. (Low-level feature e.g.: color, texture, shape, structure, etc.).

2. Related work:

As one of the simplest representations of digital images, histograms have been widely used for various image categorization problems. [2] use k-nearest neighbor classifier on color histograms to discriminate between indoor and outdoor images. In the work of [3], Bayesian classifiers using color histograms and edge directions histograms are implemented to organize sunset/forest/mountain images and city/landscape images, respectively. [4] apply SVMs, which are built on color histogram features, to classify images containing a generic set of objects. Although histograms can usually be computed with little cost and are effective for certain classification tasks, an important drawback of a global histogram representation is that information about spatial configuration is ignored.

In this project, the representation approach I implemented is trying to find out the point in the space that maximizes the Diverse Density function. In the Diverse Density approach, an objective function, called the Diverse Density (DD) function [5], is defined over the region feature space, in which regions can be viewed as points. The DD function measures a co-occurrence of similar regions from different images with the same label. A feature point with large Diverse Density indicates that it is close to at least one region from every positive image and far away from every

negative region. The DD approach searches the region feature space for points with high Diverse Density. Once a point with the maximum DD is found, a new image is classified according to the distances between regions in the image and the maximum DD point: if the smallest distance is less than certain fixed threshold, the image is classified as positive; otherwise, the image is classified as negative.

3. Approach

The main procedure of this project can be split into 3 parts: image segmentation, representation and classification.

3.1. Image segmentation

Image segmentation procedure is based on color and texture features using a clustering algorithm.

3.1.1. Low level feature selection

To segment an image, the system first partitions the image into non-overlapping blocks of size 4*4 pixels. A feature vector is then extracted for each block. The size of block is chosen by considering the trade-off between accuracy and computation complexity. Each feature vector consists of six features. First three of them are the average L, U and V values of the pixels in the block. Here L, u and v correspond to three channels in the LUV color space, where L encodes luminance, U is saturation and V is hue angle. Both U and V encode color information (chrominance). The color space transformation, from RGB to LUV, contributes to a perceptually reasonable segmentation result. The color space conversion is first done by converting from RGB to XYZ space, then from XYZ to LUV.

The matching function from RGB to XYZ is below, based on CIE standards [6]

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{0.17697} \begin{bmatrix} 0.49 & 0.31 & 0.20 \\ 0.17697 & 0.81240 & 0.01063 \\ 0.00 & 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

The transformation from XYZ to LUV is as below:

$$L = \begin{cases} 116Y^{1/3} - 16, Y > (6/29)^3 \\ (29/3)^3 Y, \quad Y \le 6/29)^3 \end{cases}$$

$$u = 13L(u' - u'_n)$$

$$v = 13L(v' - v'_n)$$

The quantities un' and vn' are the (u', v') chromaticity coordinates of a "specified white object,"[7]
The other three represent square root of energy in the high-frequency bands of the wavelet transforms [8], that is, the square root of the second order moment of wavelet coefficients in high-frequency bands. Applying wavelet transform can average the image information and arrives at a new matrix

representing the same image in a more concise manner. It eliminates some unnecessary information. [9] shows that moments of wavelet coefficients in various frequency bands are effective for representing texture. For example, the HL band shows activities in the horizontal direction. An image with vertical strips thus has high energy in the HL band and low energy in the LH band. We use Haar wavelet transform on the L component. Haar wavelets are used as they are computationally efficient and have good performance [10].

After a one-level wavelet transform, a 4*4 block is decomposed into four frequency bands: the LL, LH, HL, and HH bands. Each band contains 2*2 coefficients. Without loss of generality, we suppose the coefficients in the HL band are

$$\{c_{k,l}, c_{k,l+1}, c_{k+1,l}, c_{k+1,l+1}\}$$

One feature is

$$f = \left(\frac{1}{4} \sum_{i=0}^{1} \sum_{j=1}^{1} c_{k+i,l+j}^{2} \right)^{\frac{1}{2}}$$

The other two features LH and HH are computed similarly to HL band.

So the complete feature vector of a block is in the form of

$$fv = [l, u, v, w_{lh}, w_{hl}, w_{hh}]$$

3.1.2. Feature vectors clustering:

I adapt Mixture Gaussian algorithm (use k-means result for initialization) to cluster the feature vectors into several classes with every class corresponding to one “region” in the segmented image.

K-means clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean [11]. In other words, k-means tries to find assignment labels $c^{(i)} \in \{1,\ldots,K\}, i = 1,\ldots,m$ and cluster centroids $\mu_1,\ldots,\mu_k$, minimizing the following objective:

$$J(c,\mu) = \sum \| x^i - \mu_{c^i} \|^2$$

The density estimation for mixture of Gaussian is trying to maximize the likelihood [12]:

$$\max_{\pi,\mu,\Sigma} l(\pi,\mu,\Sigma) = \sum_{i=1}^{m} \log p(xi;\pi,\mu,\Sigma)$$

$$p(x) = \sum_{j=1|}^{K} p(z=j)\,p(x \mid z=j) = \sum_{j=1}^{K} \pi j N(x;\mu j,\Sigma,\pi)$$

The reason I’m using mixture of Gaussians but k-means is that EM for a mixture of Gaussians does not require hand-tuning whereas in k-means, the selection of initial centroids can influence the clustering result which is not suitable for our case.

3.2. Image representation

Given a set of labeled images, finding what is in common among the positive training set and does not appear in the negative training set provides inductive clues for classifier design. In this image representation approach, such clues are captured by region prototypes computed from the DD function. An image feature space is then constructed using the region prototypes, each of which defines one dimension of the image feature space.

3.2.1. Diverse Density

DD from [5], is defined over the region feature space, in which regions can be viewed as points. The DD function measures a co-occurrence of similar regions from different images with the same label. A feature point with large Diverse Density indicates that it is close to at least one region from every positive image and far away from every negative region.

The DD function is defined as

$$DD(x) = \Pr(x \mid D) = P(x \mid B, L) = \frac{\Pr(x, B, L)}{\Pr(B, L)} = \frac{\Pr(L \mid x, B)\Pr(x, B)}{\Pr(L \mid B)\Pr(B)}$$

Here, x is a point in the region feature space, B is training set and L is labels corresponding to B. Assuming a uniform prior on the hypothesis space and independence of $< B_i, l_i >$ pairs given x, using Bayes' rule, the maximum likelihood hypothesis turns into [13]:

$$optDD(x) = \Pr(L \mid x, B) = \prod_{i=1}^{m} \Pr(l_i \mid x, B_i)$$

$$\Pr(l_i \mid x, Bi) = 1 - \mid l_i - label(B_i \mid x) \mid$$

$$label(Bi \mid x) = \max_j (e^{-\|B_{ij} - x\|^2})$$

It is not difficult to observe that values of DD are always between 0 and 1. If a point x is close to a region from a positive image Bi, then $\Pr(l_i \mid x, B_i)$ will be close to 1; if x is close to an region from a negative image Bi, then $\Pr(l_i \mid x, B_i)$ will be close to 0.

3.2.2. Learning region prototypes

Learning region prototypes can be seen as an optimization process. Since a larger value of DD at a point indicates a higher probability that the point fits better with the regions from positive images than with those from negative images, we can choose local maximizers of DD as region prototypes. Loosely speaking, a region prototype represents a class of regions that is more likely to appear in positive images than in negative images.

For my project, the dimension of the optimization problem is 6 because the dimension of the region features is 6. Since the DD functions are smooth, we can apply gradient based methods to find local maximizers. Region prototypes are selected from those maximizers with two additional constraints: (a) they need to be distinct from each other; and (b) they need to have large DD values.

Each image feature is defined by one region prototype and the region that is closest to the region prototype. Here, the distance is measured by the Euclidean distance. Hence, it can also be viewed as a measure of the degree that a region prototype shows up in the image.

3.3. Image classification

After we get the image feature vectors, the final step is to use a classifier to do the classification and the classifier I chose in the project is Logistic Regression classifier (we implemented during the homework).

4. Experiment

The image data set employed in our empirical study consists of 400 images, 4 categories taken from CDROM published by COREL Corporation. Each COREL CD-ROM of 100 images represents one distinct topic of interest. All the images are in JPEG format with size 384 * 256 or 256*384. Some randomly selected sample images from each category are shown in Figure 1

To test the performance of my proposed image representation approach, I compare it with a traditional image representation method, which simply aligns the regions from each image according to its size. This approach is based on the assumption that the larger the region is, the more important it is to the image. However, for most of the real cases, this does not make sense, for example, suppose the largest area of image1 is a building and the largest region of image2 is a car, so these two regions maybe set to the same dimension in the feature vector, but totally irrelevant.

The classification experiment is conducted on 4 categories (shown in figure1). The two representation approach will share the same region features generated by image segmentation and also use the same classifier to do the test. For each category, I use 10-fold cross validation to estimate the result.

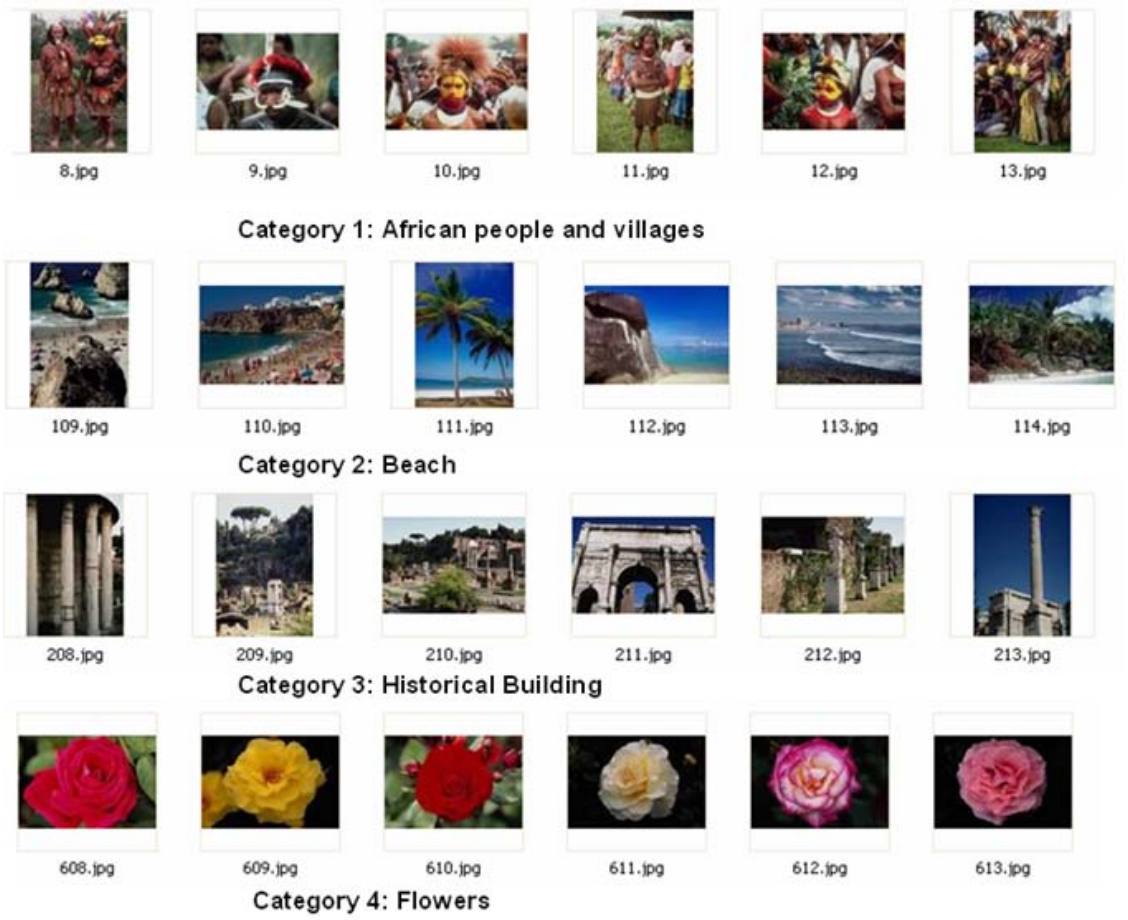


Figure 1: sample images taken from 4 categories

5. Result

5.1. Classification Results

The classification results provided in Figure 2 are based on images in Category 1 to Category 4. Compared with traditional representation, the average accuracy of DD is about 5% higher. When we take a close look at the precision in each category, it shows that the precision varies from each category, e.g. the accuracy for Flowers can be over 90% but for His-building, it's only about 60%.

This result is because of two reasons. One is that the layout in His-building is more complex than Flowers, so the information held in His-building is too much that cause noise and disturbance to image representation. The other one is due to the fact that the cluster number for different category should be adjusted according to their complexity, e.g. the result in Figure 2 is using a default K equals to 16, but for Flowers, 5 regions maybe already enough.

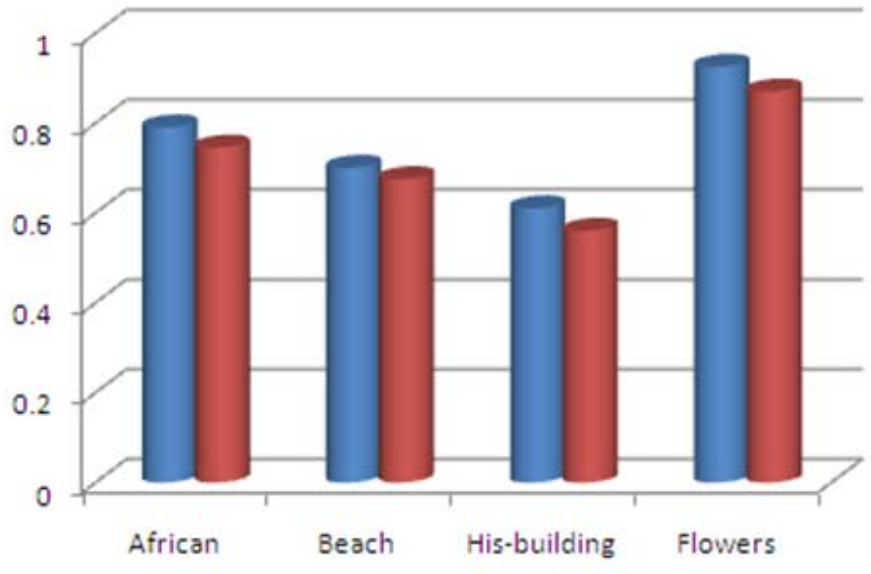


Figure 2: Classification Precision
Blue: proposed image representation approach
Red: comparison approach

5.2. Sensitivity to Image Segmentation

Because image segmentation cannot be perfect, being robust to segmentation-related uncertainties becomes a critical performance index for a region-based image classification method. Table 1 shows two images, “African” and “Flowers,” and the segmentation results with different numbers of regions. We can see from Table 1 that, when K is small, objects totally different in semantics may be clustered into the same region (under-segmented). While under some other stopping criteria, one object may be divided into several regions (over-segmented).

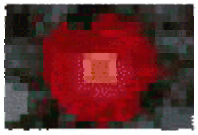
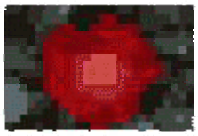
	Original image	16 regions	8 regions	4 regions
African				
Flowers				

Table 1: Segmentation results given by the clustering algorithm with 3 different K

In this section, I compare the performance of DD approach with traditional representation when the region number of image varies. I pick 5 K from 3 to 16 and record down the average accuracy of 4 categories corresponding to different K.

The results in Figure 3 indicate that DD outperforms the comparison approach on all 5 cluster numbers. In addition, for DD, there are no significant changes in the average classification accuracy for different K. While the performance for comparison vibrates more severely. This appears to support the claim that DD has low sensitivity to image segmentation.

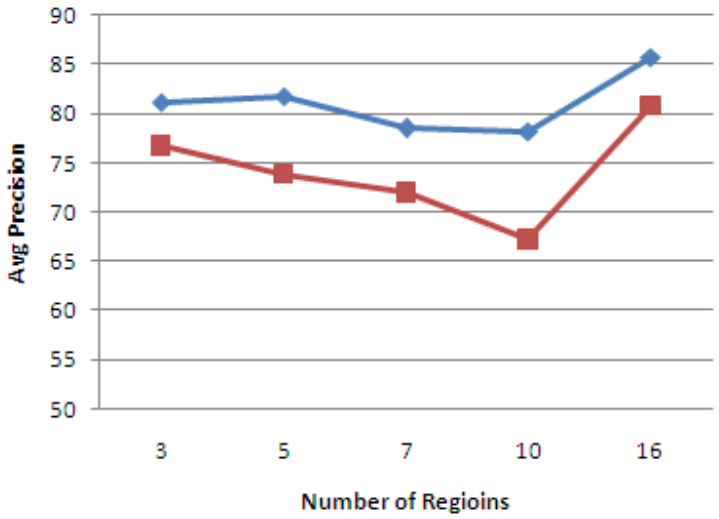


Figure 3: Comparing two image representation approaches on the robustness to image segmentation, performed on 4 categories, using average precision.
Blue: proposed image representation approach;
Red: comparison approach.

6. Conclusion

The approach of mapping the original image feature vectors into the region prototype space which is determined by maximizing the DD likelihood can help to better represent the image, so to increase the precision of classification.

7. Reference

[1] Chen, Wang, 2004, Image Categorization by learning and Reasoning with regions, Journal of Machine Learning Research 5 (2004) 913–939

[2] M. Szummer and R. W. Picard. Indoor-outdoor image classification. In Proc. IEEE Int'l Workshop on Content-Based Access of Image and Video Databases, pages 42–51, 1998.

[3] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H.-J. Zhang. Image classification for content-based indexing. IEEE Transactions on Image Processing, 10(1):117–130, 2001.
[4] O. Chapelle, P. Haffner, and V. N. Vapnik. Support vector machines for histogram-based image classification. IEEE Transactions on Neural Networks, 10(5):1055–1064, 1999.
[5] O. Maron and T. Lozano-P´erez. A framework for multiple-instance learning. In Advances in Neural Information Processing Systems 10, pages 570–576. Cambridge, MA: MIT Press, 1998.
[6] Fairman H.S., Brill M.H., Hemmendinger H. (February 1997). "How the CIE 1931 Color-Matching Functions Were Derived from the Wright–Guild Data". Color Research and Application 22 (1): 11–23.
[7] Mark D Fairchild, Color Appearance Models. Reading, MA: Addison-Wesley, 1998
[8] D. A. Forsyth and J. Ponce. Computer Vision: A Modern Approach. Prentice Hall, 2002.
[9] M. Unser. Texture classification and segmentation using wavelet frames. IEEE Transactions on Image Processing, 4(11):1549–1560, 1995.
[10] Natsev, A., Rastogi, R., & Shim, K. (2004). WALRUS: A Similarity Retrieval Algorithm for Image Databases IEEE Transactions on Knowledge and Data Engineering, 16(3), 301-316.
[11] http://en.wikipedia.org/wiki/K-means_clustering
[12] lecture nodes of CS134 machine learning, 2009 spring term
[13] Zhang and S. A. Goldman. EM-DD: An improved multiple-instance learning technique. In Advances in Neural Information Processing Systems 14, pages 1073–1080. Cambridge, MA: MIT Press, 2002.