Advances in Computational Research
ISSN: 0975-3273 & E-ISSN: 0975-9085, Volume 4, Issue 1, 2012, pp.-108-112.
Available online at http://www.bioinfo.in/contents.php?id=33

AUTOMATIC IMAGE CATEGORIZATION AND ANNOTATION USING K-NN FOR COREL DATASET

PATIL M.P.* AND KOLHE S.R.

Department of Computer Science,North Maharashtra University Jalgaon, MS, India.
*Corresponding Author: mpp145@gmail.com

Received: February 21, 2012; Accepted: March 06, 2012

Abstract-The search of an image in image database using keywords is made powerful due to automatic image annotation. In this paper, an automatic image annotation using K- Nearest Neighbor (K-NN) is presented. The categorization based approach is presented for annotation. Images are first segmented using k-means clustering and then processed to form feature vector. Local features are extracted from the regions of the image. The feature vectors are experimented using K-NN. Our system is validated using ten categories from the COREL images. It is observed that in multiple instance learning using K-NN with color and texture features outperforms for all type of feature vectors.
Keywords- Automatic image annotation, Color features, Texture features, K- Nearest Neighbor, Multiple Instance Learning.

Citation: Patil M.P. and Kolhe S.R. (2012) Automatic Image Categorization and Annotation using K-NN for Corel Dataset. Advances in Computational Research, ISSN: 0975-3273 & E-ISSN: 0975-9085, Volume 4, Issue 1, pp.-108-112.

Introduction

In recent years, image repositories grow exponentially with the rapid advance in information technology. A wide range of content-based image applications have become very challenging because of the rapid expansion of representations and the volume of image data, and the complex nature of the high-level semantic meanings[1]. The term image annotation refers to the labeling of images with relevant keywords. It is very important for image retrieval and object recognition. Since manual annotation is expensive and subjective, automatic image annotation becomes the research focus of a number of researchers. These automatic image annotation systems have received intensive attention in the literature of image information retrieval since this area was started years ago, and consequently a broad range of techniques have been proposed. The algorithms used in these systems perform four tasks namely feature extraction, feature selection, training annotation system, and annotation of new images. Automatic image annotation can be done using classification based systems [2].

In recent research, an image is usually characterized by a set of instances. For example, an image can be segmented into a number of homogeneous regions [3], and one feature vector is extracted from each region to capture its visual properties, such as color, texture, shape, etc. Consequently, an image can be represented as a collection of feature vectors. Images with such representation can be categorized as follows. First a set of images with known labels are used in a learning process to design a classifier. This classifier is then used to label each image in the data set. The corresponding learning problem over bags of vectors is called Multiple-Instance Learning (MIL)[4]. In this framework, an image is referred to as a bag and a feature vector corresponds to an instance.

In this paper the work carried out on COREL dataset for annotation using categorization is presented. Every image is segmented using k-means clustering to find regions. Then features of every region are extracted to form feature vector. Using K-NN multiple instance learning is carried out to classify and annotate the unlabeled image object. The results of categorization and annotation are presented.

Related Work

Visual descriptors for image categorization generally consist of either global or local features. The former ones represent global information of images and are based on global image statistics such as color histograms[5]or edge directions histograms[6]. Global feature-based methods were mostly designed to separate very general classes of images. On the contrary, local descriptors extract information at specific image locations that are relevant to characterize the visual content. Early approaches based on local features work on image blocks. In [7] color and texture features are extracted from image blocks to train a statistical model. Then several region-based approaches have been proposed, which require segmentation of images into relevant regions. E.g. in [8], an algorithm for learning region prototypes is proposed as well as a classification of regions based on Support Vector Machines (SVMs). Region based classification using DD-SVM is presented for multiple instance learning in[9].In [10], the bags-of-features representing an image are spatial pyramid aggregating statistics of local features. This approach takes into account approximate global geometric correspondences between local features. Gaussian Mixture Models (GMMs) have been used to model the distribution of bags of low-level features [11]. This approach requires both to estimate the model parameters and to compute a similarity measure to match the distributions. Measuring similarities is particularly adapted to categorization when one prototype of each category is defined in the feature space. In this context the similarity measure is used to find the prototype that best matches an unlabeled image. As pointed out in [12], this framework has provided the best results in image categorization. Human visual categories are mostly defined by similarity to prototype examples, as it results from research on cognitive psychology. In [13] the task of image categorization done using a new similarity measure on the space of Sparse Multiscale Patches (SMP).

The study on MIL was first motivated by the problem of predicting the drug molecule activity level. After that, many MIL methods have been proposed, such as learning axis-parallel concepts[16], diverse density [17], extended Citation kNN[18], etc. They have been applied to a wide spectrum of applications ranging from image concept learning and text categorization to stock market prediction.

System Architecture

The system developed for classification and annotation uses standard dataset of images and follow two step process comprising feature extraction and machine learning.

a. System Framework

Classification and annotation system developed performs three basic operations one to segment the images, next to build the feature vector and the last is learning along with classification and tagging. As shown in Figure 1 the system takes input image from Corel image dataset.

The input image is segmented using clustering algorithm. The regions of image are extracted from original image with respect to clusters. These regions are used to extract color and texture features. These features form a feature vector describing the image. The feature vector is input for multiple instance learning using K-NN technique. The output generated from the classifier is tagged image.

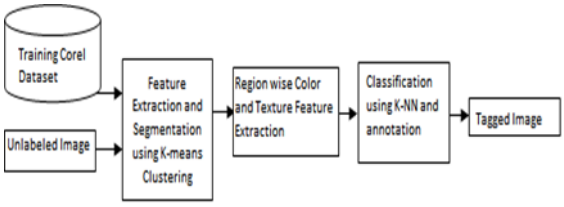

Fig. 1- System Framework

b. Data set

The Corel Database with 600 categories is used. Each category is manually labeled with few descriptive keyword. Each category consists of 100 color images of size 384 × 256. For a more controlled test on categorization, 10 distinct Corel Categories, namely Africa, beach, buildings, buses, dinosaurs, elephants, flowers, horses, mountains and food are taken. Sample images as in Figure 2 [7].

Fig. 2- Sample images of Corel Dataset

Segmentation

Image segmentation has been successfully used in content-based image analysis [7][17][18]. SIMPLIcity system [7] classifies images into textured or nontextured classes based upon how evenly a region scatters in an image.

In this section image segmentation based on color and spatial variation features using a k-means clustering algorithm[19] is described. For general-purpose images precise object segmentation is nearly as difficult as natural language semantics understanding. However, semantically precise segmentation is not crucial to the system implemented here for categorization [6]. Image segmentation is a well-studied topic [18][7]. To segment an image, the system first partitions the image into non-overlapping blocks of size 4 x 4 pixels. A feature vector is extracted for each block. Smaller block size may preserve more texture details but increase the computation time as well. Each feature vector consists of six features. Three of them are the average color components in a block. The well-known LUV color space, where L encodes luminance and U and V encode color information (chrominance) are used. The other three represent square root of energy in the high-frequency bands of the wavelet transforms [20], that is, the square root of the second order moment of wavelet coefficients in high-frequency bands. A Daubechies-4 wavelet transform is applied to the L component of the image. After a one-level wavelet transform, a 4 x 4 block is decomposed into four frequency bands, the

LL, LH, HL, and HH bands. Each contains 2x2 coefficients $\{c_{k,l}, c_{k,l+1}, c_{k+1,l}, c_{k+1,l+1}\}$. One feature is

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

The other two features are computed similarly to the LH and HH bands. The k-means algorithm is used to cluster the feature vectors into several classes with every class corresponding to one "region" in the segmented image. The algorithm does not specify the number of clusters, K , to choose. K is selected by gradually increasing until a stopping criterion met. The number of clusters in an image changes in accordance with the adjustment of the stopping criteria. A detailed description of the stopping criteria can be found in Wang et al.[7].

Segmented image with regions is taken to separate instances from the original image. The overall segmentation process is described in the figure 3.

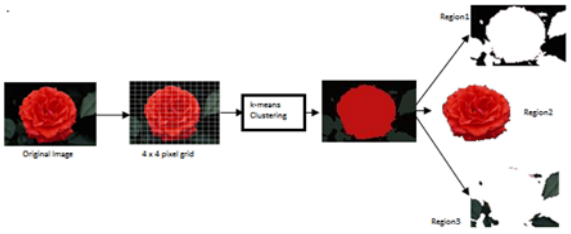

Fig. 3- Segmentation

Feature Extraction

To demonstrate the effectiveness of the feature and evaluation of classification modules, a set of simple visual features is used. The feature vector comprise of the following features:

Color Feature

Color features are widely used for image representation because of their simplicity and effectiveness. Color features are extracted at global and local both levels of an image. The color features are computed for L, U, V color channel of every region segmented in image. Three features are the average color components in a block. We use the well-known LUV color space, where L encodes luminance and U and V encode color information (chrominance). The LUV color space has good perception correlation properties. For each of the three channels average of color channel is calculated.

Texture Feature

Texture features are extracted using discrete Gabor wavelet transform [21]. For a given image I(x, y) with size P X Q, its discrete Gabor wavelet transform is given by a convolution [22]:

$$G_{mn}(x,y) = \sum_{s}\sum_{t} I(x-s, y-t)\Psi^{*}_{mn}(s,t) \quad (2)$$

where, s and t are the filter mask size variables, and Ψ^{*}_{mn} is the complex conjugate of Ψ_{mn} which is a class of self-similar functions generated from dilation and rotation of the following mother wavelet:

$$\Psi(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)\right]\cdot\exp(j2\pi Wx) \quad (3)$$

where W is called the modulation frequency. The self-similar Gabor wavelets are obtained through the generating function:

$$\Psi_{mn}(x,y) = a^{-m}\Psi(\tilde{x},\tilde{y})$$

where m and n specify the scale and orientation of the wavelet respectively, with

m = 0, 1, ...M-1, n = 0, 1, ..., N-1, and

$$\tilde{x} = a^{-m}(x\cos\theta + y\sin\theta)$$

$$\tilde{y} = a^{-m}(-x\sin\theta + y\cos\theta)$$

where a >1 and θ = n π /N.

The variables in the above equations are defined as follows:

$$a = (U_h/U_l)^{\frac{1}{M-1}}$$

$$\sigma_{x,m,n} = \frac{(a+1)\sqrt{2\ln 2}}{2\pi a^m(a-1)U_l}$$

$$\sigma_{y,m,n} = \frac{1}{2\pi\tan\left(\frac{\pi}{2N}\right)\sqrt{\frac{U_h^2}{2\ln 2}-\left(\frac{1}{2\pi\sigma_{x,m,n}}\right)^2}}$$

In our implementation, we used the following constants as commonly used in the literature:

Ul= 0.05, Uh=0.4, s and t range from 0 to 60, i.e, filter mask size is 60 x 60.

After applying Gabor filters on the image with different orientation at different scale, we obtain an array of magnitudes:

$$E(m,n) = \sum_{x}\sum_{y}|G_{mn}(x,y)| \quad (4)$$

m = 0, 1, ..., M-1; n = 0, 1, ..., N-1

These magnitudes represent the energy content at different scale and orientation of the image. The main purpose of texture-based retrieval is to find images or regions with similar texture. It is assumed that we are interested in images or regions that have homogenous texture, therefore the following mean μ_{mn} and standard deviation σ_{mn} of the magnitude of the transformed coefficients are used to represent the homogenous texture feature of the region:

$$\mu_{mn} = \frac{E(m,n)}{P\ X\ Q} \quad (5)$$

$$\sigma_{mn} = \frac{\sqrt{\sum_x\sum_y(|G_{mn}(x,y)|-\mu_{mn})^2}}{P\ X\ Q} \quad (6)$$

A feature vector is created using μ_{mn} and σ_{mn} as the feature components. Five scales and six orientations are used in implementation and the feature vector is given by:

$$\text{feature vector} = (\mu_{00}, \sigma_{00}, \mu_{01}, \sigma_{01}, \ldots \mu_{45}, \sigma_{45})$$

Automatic Image Categorization and Annotation

System becomes capable of the autonomous acquisition and integration of knowledge is machine learning. This learning results in a system that can improve its own speed or performance of the process. The overall objective of machine learning is to improve efficiency and/or effectiveness of the system.

Multiple Instance Learning using K- Nearest Neighbour (K-NN)

Multiple Instance Learning (MIL) is proposed as a variation of supervised learning for problems with incomplete knowledge about labels of training examples. In supervised learning, every training instance is assigned with a discrete or real-valued label. In comparison, in MIL the labels are only assigned to bags of instances [23].

The K-nearest neighbor rule, also called the majority voting K-nearest neighbor, is one of the oldest and simplest non-parametric techniques in the pattern classification literature. The intuition underlying Nearest Neighbor Classification is quite straight forward, examples are classified based on the class of their nearest neighbors. It is often useful to take more than one neighbor into account so the technique is more commonly referred to as K-NN Classification where k nearest neighbors are used in determining the class. Because classification is based directly on the training examples it is also called Example-Based Classification or Case-Based Classification.

K–NN classification has two stages; the first is the determination of the nearest neighbors and the second is the determination of the class using those neighbors by
simple majority voting or by distance weighted voting.

Following the k-nearest neighbor rule, we give a high level summary of the nearest neighbor classifier[16]. Let $T = \{x_1, x_2, \dots, x_n\}$ be the training set. Given a query object (x' , c') randomly, its unknown class c' is determined as follow:

1) Select the set $T' = \{x_1^{NN}, x_2^{NN}, \dots, x_k^{NN}\}$, the set of k nearest training objects to the selected query object x' , arranged in an ascending order in terms of the distance (or dissimilarity) measure between x_i^{NN} and x' . The distance between two bags is computed by using Hausdorff distance.The Hausdorff distance provides such a metric function between subsets of a metric space. By definition, two sets A and B are within Hausdorff distance d of each other iff every point of A is within distance d of at least one point of B, and every point of B is within distance d of at least one point of A[24].

Formally speaking, given two sets of points $A = \{a_1, \dots, a_n\}$ and $B = \{b_1, \dots, b_n\}$,

The Hausdorff distance is defined as:

$$H(A, B) = \max\{h(A, B), h(A, B)\} \qquad (7)$$

Where maximal and minimal distance is calculated for $h(A, B)$ as

$$h(A, B) = \max_{a \in A} \min_{b \in B} \| a - b \|$$

$$h(A, B) = \min_{a \in A} \min_{b \in B} \| a - b \|$$

2) Assign the class label to the query object x', based on the majority voting class of its nearest neighbors:

$$c' = \arg\max \sum_{x_i^{NN} \epsilon T'} I[c = c_i^{NN}] \qquad (8)$$

where c is a class label , c_i^{NN} is the class label for the i-th neighbor among k nearest neighbors of the query object. $I[c = c_i^{NN}]$, an indicator function, takes a value of one if the class c_i^{NN} of the neighbor x_i^{NN} is the same as the class c and zero otherwise.

Experimental Results

Corel dataset described in section 3 is used for automatic image categorization and annotation experimentations. Out of 1000 images, 500 images are used for training purpose whereas remaining 500 images are used for testing.

The image is first taken to extract features for segmentation. For each block of 4 x 4 pixels six features as described in segmentation section are computed. The feature vector comprising 3 color and texture features is used for clustering using k-means clustering. Every in a image represents a region. The images are representing color and texture features. Color features are extracted for L,u,v color components for every region. This forms feature vector viz. Color Feature (CF) of size k by 3 where k is number of regions in image. Gabor filter is used to extract texture given Texture Feature (TXF) vector of size k by 60. In combination of color with texture gives a vector of Color Texture Feature (CTXF) vector of size k by 63.

K-NN experiments are carried on all feature vectors with varying the K parameter and distance formula. All three combinations of features are experimented with maximal and minimal Hausdorff distance. The performance along with value of nearest neighbors is given in Table 1. For k-NN classifier at k=13 and features vector is CTXF the accuracy obtained is 60.8%. The classification performance using confusion matrix is given in Table 3. Category wise accuracy comparison with [2,7,8] is done in Table 2. Category dinosaur is containing single object images so easy to classify. While food contains many mix objects misleading classification. Africa is also a category containing mix types of textures and colors and due to it objects of other categories are misclassified as Africa.

Here the improvement is achieved in categories dinosaur and mountain because the texture features extracted using gabor wavelet gives magnitude of energy of the regions which is similar for these categories.

The classified image is tagged with the category label. Although these image categories do not share annotation words, they may be semantically related. For example, both the “beach” and the “mountains” categories contain images with rocks, sky, and trees. Therefore, the evaluation method we use here only provides a lower bound for the annotation accuracy of the system.

Table 1- Analysis of Categorization and Tagging Performance

Feature Extraction	K-NN			
	MaxHD		MinHD	
	Nearest Neighbors	Output (%)	Nearest Neighbors	Output (%)
Color Features (CF)	1	57.8	10	40.4
Texture Features (TXF)	21	39.2	18	27.8
Color and Texture Features(CTXF)	13	60.8	19	31.4

Table 2- Categorization Performance on Corel

Class	Categorization Performance (%)			
	Chen and Wang [7]	Setia and Burkhardt [8]	P. Piro et al. [2]	Proposed Approach
Africa	67.7	66	74.8	60
Beaches	68.4	32	56.4	54
Building	74.3	76	74.0	54
Bus	90.3	64	73.4	80
Dinosaur	99.7	94	88.8	100
Elephant	76.0	50	87.8	44
Flower	88.3	78	89.0	56
Horses	93.4	72	91.4	40
Mountain	70.3	70	62.0	92
Food	87.0	76	55.6	28

Conclusion

In this paper the authors have put forward analysis of multiple instance learning using k-NN machine learning technique for color and texture features based on classification performance. The image splitting in regions using k-means clustering is demonstrated. It is observed that L, u, v color strength in the image regions along with texture out performs other feature combinations with K-NN classifier. It is experimented that with K-NN classification accuracy is improved. The results show that single object images with features strongly discriminating e.g. dinosaur classify up to 100%. While in case of beach images features are similar to building misclassify beach to buildings.

References

[1] Vu K., Hua K.A., Cheng H., Lang S.D. (2006) *ACM SIGMOD international conference on management of data.*527-538.

[2] Ping Guo, Tao Wan and Jin Ma (2011) *Lecture Notes in Computer Science*, 6761, 562-570

[3] Chen Y., Bi J., Wang J.Z. (2006) *IEEE Trans Pattern Anal Mach Intell,* 28(12), 1931-1947.

[4] Dooly D.R., Zhang Q., Goldman S.A., Amar R.A. (2003) *J Mach Learn Res* 3(1), 651-678.

[5] Szummer M., Picard R.W. (1998) *International Workshop on Content-Based Access of Image and Video Databases,* 42-44.

[6] Vailaya A., Figueiredo M., Jain A., Zhang H.J. (2001) IEEE Transactions on Image Processing, 10, 117-130.

[7] James Z. Wang, Jia Li, Gio Wiederhold (2001) IEEE Trans. on Pattern Analysis and Machine Intelligence, 23(9), 947-963.

[8] Bi J., Chen Y., Wang J.Z. (2005) *CVPR,* 1121-1128.

[9] Yixin Chen and James Z. Wang (2004) *Journal of Machine Learning Research*, 5, 913-939.

[10] Lokesh Setia, Hans Burkhardt (2006) DAGM-Symposium, 294 -303.

[11] Lazebnik S., Schmid C., Ponce J. (2006) *CVPR* (2) 2169-2178.

[12] Liu Y., Perronnin F. (2008) *CVPR*.

[13] Paolo Piro, Sandrine Anthoine, Eric Debreuve and Michel Barlaud (2009) *Advances in Multimedia Modeling*, 5371 227-238.

[14] Dietterich T.G., Lathrop R.H., Lozano-Perez T. (1997) *Artificial Intelligence Journal*, 89.

[15] Oded Maron, Tomás Lozano-Pérez (1998) *Conf. on Advances in Neural Information Processing Systems* 10, 570-576.

[16] Wang J. and Zucker J.D. (2000) 17*th Int'l Conf. on Machine Learning*, 1119-1125.

[17] Ma W.Y. and Manjunath B. (1997) *IEEE Int'l Conf. on Image Processing*, 568-571.

[18] Shi J. and Malik J. (2000) *IEEE Transactions on PatternAnalysis and Machine Intelligence*, 22(8), 888-905.

[19] Hartigan J.A. and Wong M.A. (1979) *Applied Statistics*, 28, 100-108.

[20] Daubechies I. (1992) *SIAM*.

[21] Manjunath B.S. and Ma W.Y. (1996) *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 18(8) 837-842.

[22] Zhang D.S., Lu G. (2000) *First IEEE Pacific-Rim Conference on Multimedia,* 392-395.

[23] Jianping Gou, Taisong Xiong and Yin Kuang (2011) *Jjournal of Computers*, 6(5), 833-840.

[24] Edgar G.A. (1995) *Measure, topology, and fractalgeometry* (3*rd print*).

Table 3- Confusion Matrix for Categorization

Percent	Africa	beach	buildings	buses	dinosaurs	elephants	flowers	horses	mountains	food
Africa	60	2	10	10	12	2	0	0	4	0
beach	2	54	26	2	0	2	0	0	0	14
buildings	18	8	54	4	2	4	6	0	0	4
buses	8	2	2	80	6	0	0	2	0	0
dinosaurs	0	0	0	0	100	0	0	0	0	0
elephants	10	22	16	0	6	44	0	0	0	2
flowers	12	0	2	8	0	0	56	12	10	0
horses	20	0	0	14	12	0	10	40	2	2
mountains	2	2	0	0	0	0	0	4	92	0
food	4	28	24	8	0	6	0	0	2	28