Research Journal of Applied Sciences, Engineering and Technology 8(21): 2172-2178, 2014
DOI:10.19026/rjaset.8.1216
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: August 19, 2014 Accepted: September 25, 2014 Published: December 05, 2014

Research Article

Classification of Scenes into Indoor/Outdoor

[1]R. Raja, [2]S. Md. Mansoor Roomi and [2]D. Dharmalakshmi
[1]Department of Electronics and Communication, Pandian Saraswathi Yadav Engineering College, India
[2]Department of Electronics and Communication, Thiagarajar College of Engineering, Madurai-15, India

Abstract: Effective model for scene classification is essential, to access the desired images from large scale databases. This study presents an efficient scene classification approach by integrating low level features, to reduce the semantic gap between the visual features and richness of human perception. The objective of the study is to categorize an image into indoor or outdoor scene using relevant low level features such as color and texture. The color feature from HSV color model, texture feature through GLCM and entropy computed from UV color space forms the feature vector. To support automatic scene classification, Support Vector Machine (SVM) is implemented on low level features for categorizing a scene into indoor/outdoor. Since the combination of these image features exhibit a distinctive disparity between images containing indoor or outdoor scenes, the proposed method achieves better performance in terms of classification accuracy of about 92.44%. The proposed method has been evaluated on IITM- SCID2 (Scene Classification Image Database) and dataset of 3442 images collected from the web.

Keywords: Color model, content based image retrieval, entropy with UV, image annotation, image retrieval, scene classification

INTRODUCTION

With advances in digital imaging, digital photography collection finds an extreme growth of image databases. Retrieving an image in such a large collection of database takes enormous search time. Generally, an image retrieval system can be broadly divided into two approaches namely Annotation-Based Image Retrieval (ABIR) and Content-Based Image Retrieval (CBIR). Keywords are used to annotate the images in ABIR manually. The drawbacks of ABIR are time consumption due to manual image annotation and ambiguity raised by the synonymy and homonymy of the keywords. In CBIR, images are indexed by their own visual contents. The term visual content in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself. In recent years, various retrieval systems have been introduced, such as IBM QBIC (Flickner *et al*., 1995) Systems developed by IBM Almen Research center, Photo Book (Pentland *et al*., 1996) developed by the MIT Media Lab, Visual SEEK (Chang and Smith, 1996) developed at Columbia University and Netra (Ma and Manjunath, 1997) developed at the University of California, Santa Barbara use color and texture features for image retrieval. Other examples of CBIR systems are Virage (Bach *et al*., 1996), SIMPLIcity (Grimson *et al*., 1997), MARS (Huang *et al*., 1996), Candid (Cannon and Kelly, 1995) and Chabot (Ogle and Stonebraker, 1995). Selection of feature extraction techniques and the similarity functions used for retrieving the similar images have significant role in CBIR. Grimson *et al*. (1997) have explored spatial and photometric relationships within and across simple image regions to retrieve images. Alternatively, image annotation technique has been considered as a core research area in CBIR where, a number of attempts have been made to interpret high level semantics using low level visual features extracted from the image regions. Even then, image retrieval based on semantic representation is still an open problem in image and video retrieval community. Scene classification is a special case of image retrieval where the query image consists of different scenes. Scene classification involves automatically labeling an image based on its content like indoor or outdoor scenes. Semantic modeling of the scene and low level features extraction are the two popular approaches in scene classification. Semantic modeling of the scenes can be achieved through describing objects in the scene by extracting local descriptors. In low-level features extraction, features such as texture, color and shape could be either extracted from the entire image or from each partitioned blocks of the image (Bosch *et al*., 2007, 2006).

Corresponding Author: R. Raja, Department of Electronics and Communication, Pandian Saraswathi Yadav Engineering College, India

The main contributions of this work are:

- Collection of 3442, indoor and outdoor images to form a database
- Extracting low level features viz
  - Color statistical feature computed from HSV color model
  - Contrast and correlation texture feature using GLCM
  - Entropy feature computed from UV color space
- Training SVM classifier to classify the target scene as an indoor or outdoor

LITERATURE REVIEW

Many approaches have been proposed in the literature for image annotation and scene classification. Scene classification based on modeling of natural scenes using Discriminant Spectrum Templates (DST) was first introduced by Guerin-Dugue *et al*., (1999) through organizing image regions along with three semantic axes, calculated using psychophysical experiments. In their approach, they observed that power spectra of real world images show different energy distribution for each orientation. But this observation fails when the image is affected by noise during acquisition. Vogel and Schiele (2007) employed binary Bayesian classifier to classify images in hierarchical category by using color and texture feature. Gupta *et al*. (2005) proposed a method for indoor/outdoor scene classification using fuzzy C means clustering where the level of accuracy is insufficient. Payne and Singh (2005a) proposed a new approach for scene classification based on analyzing edge straightness in images because indoor images contain man-made objects which have straight edges but outdoor images like tree, mountain do not have straight edges. However, this method fails when the urban outdoor images contain straight edges such as buildings. So, Guo *et al*. (2011) proposed a new method for Indoor/Outdoor image classification using sub block division strategy but the performance of their work is not good for other dataset in terms of classification rate. Color Correlated Temperature feature (CCT) is used by Ghomsheh and Talebpour (2012). They focus only on the ability of color information not edge and texture feature for indoor-outdoor scene classification. Liu *et al*. (2012) presented a novel method using ARP and Gist for scene classification in the expense of computational cost.

PROPOSED METHODOLOGY

In this section, we discuss the classification of given image into indoor/outdoor scenes. Figure 1 shows the overall flow diagram of the proposed methodology in two phases such as training phase and query phase. In training phase, we randomly selected 100 images for each scene categories (indoor and outdoor scenes) from the standard dataset. All the training images are resized into 320×240. Then every resized image is divided into 32×24 non-overlapping rectangular regions. After that low level features are extracted from indoor and outdoor training images. Then two clusters are formed which contain indoor scene features and outdoor scene features. A Word (say indoor or outdoor) is assigned with each cluster centroid to form the dictionary for indoor and outdoor scene categories. The main reason to use dictionary is to reduce the computational complexity and search time. To create the dictionary, the low level features, extracted from the training images are denoted as $f = \{f_1, f_2, f_3 \ldots. f_n\}$ where n is the number of training images and a word is assigned (indoor, outdoor) for each cluster. These keywords are used to automatically annotate the query image at the content level during the testing phase.

Feature extraction: Accuracy of CBIR systems highly depends upon the feature extraction techniques because features carry sufficient information about images. Low level features such as color and textures are recognized as important features to classify an image as indoor or outdoor scene.

Color feature: A color characteristic in an image is often considered as perceptive feature and it is widely

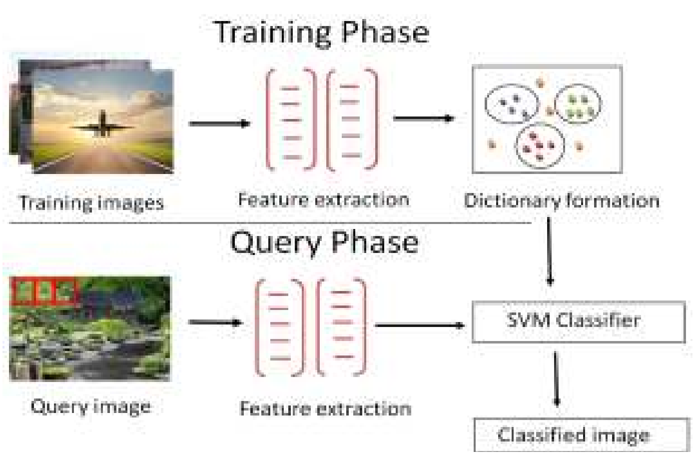

Fig. 1: Overall flow diagram of the proposed methodology

used for scene classification and image retrieval. In real world scenario photography of scenes are commonly affected due to large variation in lighting and viewing conditions. Though many color models like RGB, YCbCr and XYZ are exist, from the literature it has been reported that HSV (Hue, Saturation and Value) color model is invariant to scale and illumination and also more close to human perception. In order to make these color features as scale and rotation invariant, the image is further divided into 32×24 rectangular image regions and for each region the average hue value is calculated. This feature serves as an effective color feature for classification.

Texture feature: Texture aids in identifying objects of interest or region of interest irrespective of the source of the image. Along with color feature, also the texture feature is considered to improve the classification accuracy. This paper uses Gray Level Co-occurrence Matrix (GLCM). Gray Level Co-occurrence Matrices (GLCM) estimates image properties by second-order statistics from the joint probability $P_{d\theta}$ computed between pairs of a pixel separated by a distance d in the direction θ. Four textural features such as Energy (E), Contrast (C), Correlation (Cor) and Homogeneity (H) are extracted from the co occurrence matrix defined by the Eq. (1) to (9) as follows:

$$E = \sum_i \sum_j P_{d\theta}(i,j)^2 \tag{1}$$

where, $P_{d\theta}\ (i,j)$ number of occurrences of the pair of gray levels i and j which are at a distance d in the direction θ:

$$C = \sum_i \sum_j (i-j)^2 P_{d\theta}(i,j) \tag{2}$$

$$H = \sum_i \sum_j \frac{P_{d\theta}(i,j)}{1+|i-j|} \tag{3}$$

$$Cor = -\sum_i \sum_j \frac{(i-\mu_x)(i-\mu_x)P_{d\theta}(i,j)}{\sigma_x \sigma_y} \tag{4}$$

where,

$$\mu_x = \sum_i \sum_j iP_{d\theta}(i,j) \tag{5}$$

$$\mu_y = \sum_i \sum_j jP_{d\theta}(i,j) \tag{6}$$

$$\sigma_x = \sum_i \sum_j (i-\mu_x)^2 P_{d\theta}(i,j) \tag{7}$$

$$\sigma_y = \sum_i \sum_j (i-\mu_y)^2 P_{d\theta}(i,j) \tag{8}$$

In this study, we have used contrast and correlation with the displacements of (1 0), (-1 0), (-1 1) and (-1 -1) to compute the textural features of the image.

Computation of entropy feature from LUV color model: To make our proposed system more efficient, entropy feature is computed from the LUV color model of the given image. Since entropy represents texture randomness in color space it helps to differentiate outdoor scenery from indoor in terms of color and edge details. To calculate this feature, initially image is converted into LUV color space which is again divided into 32×24 blocks. Then only from the chrominance (UV) components entropy feature is calculated by using the Eq. (9):

$$H(s) = -\sum p_i \log_2 p_i \tag{9}$$

where, p_i-Probability distribution. Since this 2D entropy calculated from the U, V color image contains both color and texture information, it is more useful for effective classification.

SVM (Support Vector Machine) classifier: Low level features are extracted from the test image in query phase using the same feature extraction techniques discussed in training phase. Extracted low level features are then classified using SVM classifier. In general, SVM classifier is considered as suitable classifier for real world classification problems. The main reason for choosing SVM is more robust against noisy data, less computational complexity and provides good performance in higher dimensional feature space.

SVM is a discriminative classifier formally defined by a separating hyper plane. In other words, given labeled training data (supervised learning), new samples are classified based on an optimal hyper plane that gives the largest minimum distance to the training samples.

The hyper plane is defined as:

$$f(x) = \beta_0 + \beta^T x \tag{10}$$

where, β is weight vector, β_0 is Bias and x is training samples closest to the margin called as support vectors. Many hyper planes can be produced to separate the support vectors. Among the entire possible hyper plane, an optimal hyper plane is chosen by using the equation as shown in Eq. (11):

Fig. 2: Sample images of indoor scenes (upper row) and outdoor scenes (bottom row) from IITM-SCID dataset

Fig. 3: Sample images of indoor scenes (upper row) and outdoor scenes (bottom row) from our dataset

$$\beta_0 + \beta^T x = 1 \tag{11}$$

In this study, to categorize an image into indoor or outdoor scene, binary class SVM with linear kernel function is used.

EXPERIMENTAL RESULTS

In this section, performance analysis and experimental results of our proposed method has been presented.

Dataset description: In this study two datasets have been used to evaluate the proposed methodology. Firstly we consider a standard dataset as IITM-SCID2 (Fig. 2) which consist of 907 low-quality and low-resolution images (442 indoor, 465 outdoor), which are subdivided into 393 training and 514 testing images. Along with this standard dataset, to show the efficiency of our algorithm we also built our own dataset (3442 images) by collecting indoor and outdoor photographic images from internet which contains 18 classes (2011 images) of indoor scenes such as airport, classroom, auditorium, library, living room, kitchen, temple etc. and 13 classes (1431 images) of outdoor scenes such as building, highways, beach, forest, mountains etc. Figure 3 shows sample images from indoor scenes (upper row) and outdoor scenes (bottom row) from our dataset.

Results and discussion: To test the performance of the proposed algorithm three measures have been considered viz., Accuracy, Precision and Recall. These are calculated from the confusion matrix which is constructed according to the features classification as shown in Table 1 and 2. The measures have been computed by using the following formulas given in Eq. (12) to (14):

$$precision = \frac{t_p}{t_p + f_p} \tag{12}$$

$$recall = \frac{t_p}{t_p + f_n} \tag{13}$$

$$accuracy = \frac{t_p + t_n}{t_p + f_p + t_n + f_n} \tag{14}$$

where,
t_p = True positive rate
t_n = True negative rate

Table 1: Confusion matrix obtained from the images collected from internet using SVM classifier

	Target class	
Actual class	Indoor scene (accuracy %)	Outdoor scene (accuracy %)
Indoor scene	89.28	10.72
Outdoor scene	4.39	95.61
Overall accuracy		92.44

Table 2: Confusion matrix obtained from IITM-SCID2 dataset using SVM classifier

	Target class	
Actual class	Indoor scene	Outdoor scene
Indoor scene	84.68	15.32
Outdoor scene	10.65	89.35
Over all accuracy		87.01%

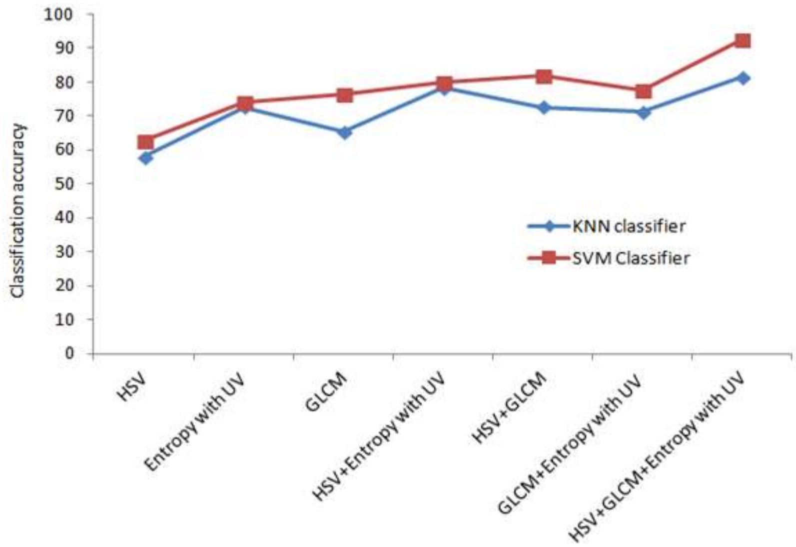

Fig. 4: Low level feature relevance for our dataset

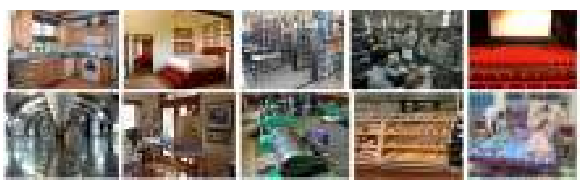

Fig. 5: Sample images of correctly classified indoor scenes from our dataset

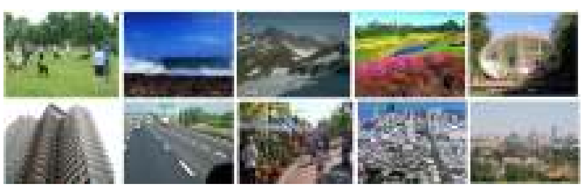

Fig. 6: Sample images of correctly classified outdoor scenes from our dataset

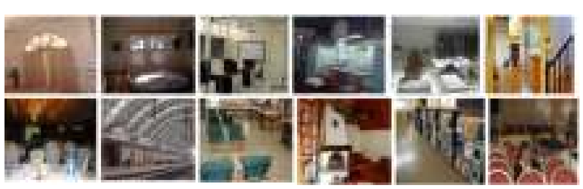

Fig. 7: Sample images of correctly classified indoor scenes from IITM-SCID2 dataset

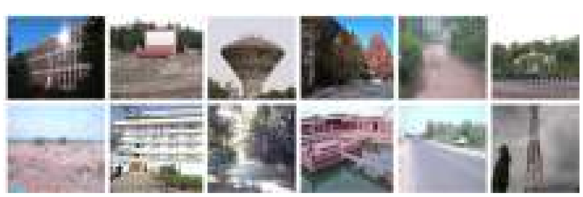

Fig. 8: Sample images of correctly classified outdoor scenes from IITM-SCID2 dataset

f_p = False positive
f_n = False negative

Table 1 shows the confusion matrix of the proposed method using SVM classifier. The most apparent behavior of the classifier is the fact that the confusion matrix shows a strong correlation between the class size and the classification result. This method achieves overall accuracy, precision and recall as 92.44, 95.31 and 89.28%, respectively experimented on more than 3442 images.

From the Table 2, it is observed that the proposed method gives classification accuracy about 87.01%, precision as 95.04% and recall as 84.61% for IITM-SCID2 dataset. Compared to the performance of the proposed method in our own dataset, overall classification rate is reduced in IITM-SCID2 dataset because of low-quality and low-resolution images.

Figure 4 shows the classification performance of the proposed method with various combinations of HSV, GLCM and Entropy with UV features on created dataset. From this plot we observed that the combination of HSV, GLCM and Entropy with UV perform well than all other combinations. SVM and KNN (K-Nearest Neighbor) classifiers are used to categorize the scene. SVM classifier works better than the KNN classifier because SVM is robust against noisy data.

Figure 5 and 6 shows the correctly classified indoor and outdoor scenes from the newly built dataset which are collected from web.

Figure 7 and 8 shows the correctly classified indoor and outdoor scenes from IITM-SCID2 dataset.

The proposed method is compared with the several existing methods as shown in Table 3. Payne and Singh (2005b) proposed an algorithm using edge straightness based on KNN classifier and rule based classifier to

Table 3: Comparison of proposed methodology with existing methods

Algorithm	Classification accuracy (%)
Szummer and Picard (1998)	90.20
Roomi *et al*. (2013a)	81.55
Payne and Singh (2005a)	71.01
Payne and Singh (2005b)	65.50
Proposed methodology	92.44

distinguish the indoor and outdoor scenes with the classification rate of 65.50%. However this method fails when outdoor scene contains urban images like buildings, city, etc. Roomi *et al*. (2013b, 2012) and Szummer and Picard (1998) proposed an algorithm based on low level features but these features fails to get higher classification rate. In our proposed method, using the simple low level features itself we achieved higher classification rate of 92.44%. From the table it is observed that the proposed methodology shows superior performance than the others.

CONCLUSION

This study proposed an automatic scene categorization algorithm based on SVM classifier. In this study, simple low level features are derived from local regions as well as from global region of the image. These features perform well for categorizing the given test image as indoor or outdoor scene. The major advantage of the work is, by using minimal features and minimal training images using a benchmark classifier; good classification accuracy has been obtained. The proposed method has been tested over IITM-SCID2 and newly built dataset and the results are tabulated. Through the performance measures like accuracy, precision and recall the efficiency of the proposed method is proved. The proposed method works better than the existing methods in terms of producing higher classification rate in the expense of less computational time to categorize the scene as an indoor or outdoor scene.

REFERENCES

Bach, J., C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey and R. Jain, 1996. Virage image search engine: An open framework for image management. P. SPIE, 2670: 76-87.

Bosch, A., X. Munoz and A. Zisserman, 2006. Scene classification via pLSA. Proceeding of the European Conference on Computer Vision, pp: 517-530.

Bosch, A., R. Martı and X. Munoz, 2007. Which is the best way to organize/classify images by content? Image Vision Comput., 25: 778-791.

Cannon, M. and P. Kelly, 1995. Query by Image Example: The CANDID Approach. Los Alamos National Laboratory White Paper.

Chang, S.F. and J.R. Smith, 1996. VisualSEEk: A fully automated content-based image query system. Proceeding of the 4th ACM International Conference on Multimedia. New York, pp: 87-98, DOI: 10.1145/244130.244151.

Flickner, M., W. Niblack and H. Sawhney, 1995. Query by image and video content: The QBIC system. IEEE Comput., 28(9): 23-32.

Ghomsheh, A.N. and A. Talebpour, 2012. A new method for indoor-outdoor image classification using color correlated temperature. Int. J. Image Process. (IJIP), 6(3).

Grimson, E., P. Lipson and P. Sinha, 1997. Configuration based scene classification and image indexing. Proceeding of IEEE Computer Soc. Conferences on Computer Vision and Pattern Recognition, pp: 1007-1013.

Guerin-Dugue, A., J. Herault, A. Oliva and A. Torralba, 1999. Global semantic classification of scenes using power spectrum templates. Challenge of Image Retrieval CIR, Newcastle, UK.

Guo, F., L. van Gool and L. Bossard, 2011. Indoor outdoor image classification. B.A. Thesis, Department of Computer Science, ETH Zurich.

Gupta, L., V. Pathangay, A. Patra, A. Dyana and S. Das, 2005. Indoor versus outdoor scene classification using probabilistic neural network. EURASIP J. Adv. Sig. Pr., 2007: 094298.

Huang, T.S., S. Mehrotra and K. Ramchandran, 1996. Multimedia Analysis and Retrieval System (MARS) project. Proceeding of 33rd Annual Clinic on Library Application of Data Processing-Digital Image Access and Retrieval.

Liu, W., S. Kiranyaz and M. Gabbouj, 2012. Robust scene classification by gist with angular radial partitioning. Proceeding of 5th International Symposium on Communications Control and Signal Processing (ISCCSP, 2012), pp: 1-6.

Ma, W.Y. and B.S. Manjunath, 1997. NeTra: A toolbox for navigating large image databases. Proceeding of IEEE International Conference on Image Processing (ICIP'97), 1: 568-571.

Ogle, V.E. and M. Stonebraker, 1995. Chabot: Retrieval from relational database of images. IEEE Comput., 28(9):40-48.

Payne, A. and S. Singh, 2005a. Indoor vs. outdoor scene classification in digital photographs. Pattern Recogn., 38(10): 1533-1545.

Payne, A. and S. Singh, 2005b. A benchmark for indoor/outdoor scene classification. In: Singh, S. *et al*. (Eds.), ICAPR 2005. LNCS 3687, Springer-Verlag, Berlin, Heidelberg, pp: 711-718.

Pentland, A., R.W. Picard and S. Sclaroff, 1996. Photobook: Content-based manipulation of image databases. Int. J. Comput. Vision, 18(3): 233-254.

Roomi, S.M.M., R. Raja and D. Kalaiyarasi, 2012. Semantic modeling of natural scenes by local binary pattern. Proceeding of International Conference on Machine Vision and Image Processing (MVIP, 2012), pp: 169-172.

Roomi, S.M.M., R. Raja and D. Kalaiyarasi, 2013a. Outdoor scene classification using invariant features. Proceeding of 4th National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG, 2013), pp: 1- 4.

Roomi, S.M.M., R. Raja and D. Kalaiyarasi, 2013b. Classification and retrieval of natural scenes. Proceeding of the 4th International Conference on Computing, Communications and Networking Technologies (ICCCNT, 2013), pp: 1- 8.

Szummer, M. and R.W. Picard, 1998. Indoor-outdoor image classification. Proceeding of the IEEE Workshop on Content Based Access of Image and Video Database. Bombay, India, pp: 42-51.

Vogel, J. and B. Schiele, 2007. Semantic modeling of natural scenes for content-based image retrieval. Int. J. Comput. Vision, 72: 133-157.