International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 11 (2015) pp. 29631-29640

http://www.ripublication.com

Comparision of Image Classification Methods on Event Data

Dinesh Kumar. C. K
M. Tech CVIP, Department of Computer Science and Engineering,
Amrita Vishwa Vidyapeetham, Coimbatore
Manjusha. R
Assistant Professor, Department of Computer Science and Engineering,
Amrita Vishwa Vidyapeetham, Coimbatore
Dr. Latha Parameswaran
Professor, Department of Computer Science and Engineering,
Amrita Vishwa Vidyapeetham, Coimbatore

Abstract

The aim of this work is to classify images related to social events. With advances in digital image processing, automated classification of event related images over large categories of dataset is possible only if the name of the event is known. The purpose of the paper is to identify different events in a given set of images with the help of visual content. We compare the performance of Global (GIST) as well as Local Descriptors (SIFT) which is helpful in achieving this classification and retrieval of images. The performances of both the approaches are assessed.

Keywords: Gist Descriptor, SVM classifier, Bag of Words (BoW), SIFT Descriptor, Content Based Image Retrieval (CBIR)

Introduction

In the field of image processing, image classification has been an active research area for the past few decades, the main objective of this is to emulate the human visual system and make sense out of visual objects. An image classification system comprises of several stages, image analysis or feature extraction module, dimensionality reduction or feature selection unit, and/or machine learning system. Figure 1 shows the generalization of such a system.

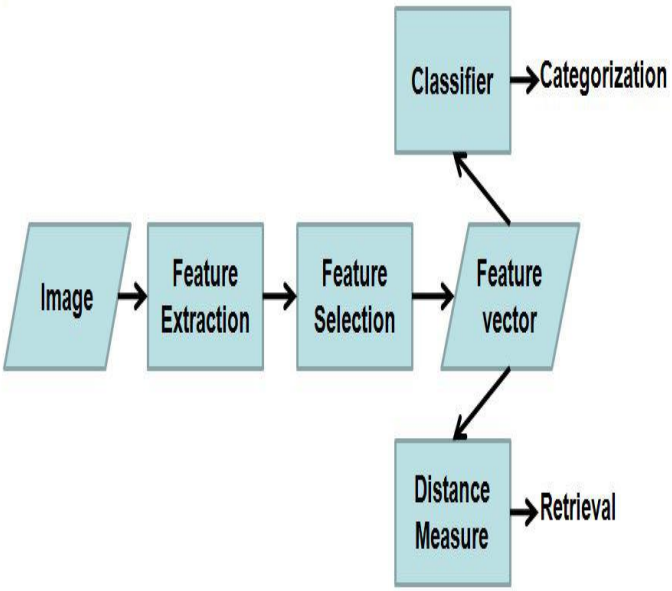

Figure 1: Basic Components of An Image Classification System

The task consists of detecting event- related data and grouping them to the related events. Such a grouping would provide the basis for search applications that support easier discovery, browsing and querying of social events. This work focuses on automatic classification of social events using GIST and SIFT descriptor. We identify images that are "similar" enough to belong to the same event type. Comparision of global descriptor (GIST) and local descriptor (SIFT) is also performed. In the global descriptor approach, GIST descriptor features are computed for the global features. Whereas in the case of local descriptor (SIFT), set of features are computed i.e. local features. These descriptors are then clustered and classified to the class according to its event type and comparision between both GIST and SIFT descriptor is calculated.

Literature Survey

The objective of image classification is to identify and classify the features present in an image according to the class it belongs. Feature extraction is classified as low level, mid level and high level features. Low level features include colour and texture present in the image, middle level feature is the shape and high level feature is the semantic gap of objects. Colour feature is the widely used feature in any image classification technique. Texture and shape features are also widely used in classification of images.

Colour Feature

Colour is the most widely used feature. To extract colour features from an image, we need to select the colour space and use its properties for extraction of features. The most commonly used colour space is RGB colour space. The drawback of RGB

colour space is that it is non-uniform and colour dependent. In the case of HSV colour space, the colour is described by its hue, saturation and brightness value. HSV shows the occurrence of each colour in an image according to its intensity which is similar to the real world colour [5].

HSV Histogram shows the frequency of occurrence of colours in the image according to the intensity values. In this method the colours present in the image are represented by histograms and by matching the histogram the images are classified. Computation is fast but spatial information present is not considered [5].

Colour moments [5] such as first-order moments (mean), second-order moments (standard deviation) and third-order moments (skewness) can be used to represent the colour distributions in an image. Since the feature vector is of low dimension, the results may not be good accuracy.

Texture Feature

Texture features are considered in case of natural scenes. Textures are the visual patterns present in the image. It gives an idea of structural pattern present in the image. Commonly used texture features are Wavelets, Gabor-filter and Co-occurrence matrix.

Gray Level Co-occurrence Matrix (GLCM) is based on conditional probability density function and is widely used for texture feature extraction. The extracted features are stored into the database which includes the position of pixels having similar gray level values which have high dimensionality [12].

Haralick features say about the pattern of textures that are present in an image. Haralick features are used for image classification. Co-occurrence matrix is calculated for each and every image which is computationally expensive [11].

Gabor features are used to extract texture features from images and are used for image classification applications. One advantage of using gabor feature is it can be tuneable i.e., gabor filter used is a wavelet, where the scale and orientation of the filter is tuneable making it useful for texture feature analysis [13].

Shape Feature

Shape of objects present in an image is one of the most significant properties for a cbir system. A good shape representation of an object should be invariant to translation, rotation and scaling. Several techniques used to extract features are fourier descriptors, wavelet descriptor, region-based descriptor and contour-based descriptor. The effective techniques among them are fourier and region-based descriptors [7].

In the case of region based descriptor [7], Moment invariants which are invariant to rotation, scaling and translations. It is fast and computationally inexpensive but the precision is low since the feature vector is of lower dimension.

In the case of fourier descriptor, Oliva and Torralba [6] proposed a shape based feature extraction technique, here similar and stable spatial structure which occurs or exists between the images are classified into the same category. They proposed the GIST descriptor to represent such spatial structures. GIST Descriptor computes the spectral information of the scene using Discrete Fourier Transform (DFT).

Other that these types of descriptors, there exist frequency domain descriptors which can be used for image classification [8]. Some of the Frequency domain descriptors include SIFT, SURF, BRISK etc. Comparing all the descriptors SIFT works better both in terms of rotation invariant and accuracy.

From the above we can conclude that for image classification, GIST and SIFT descriptor works well. Hence we do a comparative study on both the descriptors with different types of images.

Rest of the paper is organised as follows. Section 3 describes the Proposed Method. Experimental results are described in section 4 and results in section 5.

Proposed Method

To discover social event-related images and organize them in event- specific clusters, using descriptors. Descriptor is a piece of stored information that is used to identify an item in an information storage and retrieval system. Descriptors save pertinent information, thus saving the processing time for future queries. Descriptors can save image features that are essential for search and comparison.

Global Descriptor (GIST Features)

Global descriptor represents each image by a single feature vector, representing the features in the image as a whole. No attention is paid to the key points or objects present in the image. Once the image features are computed, similarity between the images is computed using distance metric. GIST Descriptor generally gives an essence of an image. It helps in recognition and classification of scenes and images such as traffic scenes [3], buildings, flowers [4] and so on. It has been studied that the scenes could be recognized and classified using global descriptor, without any information about the objects in the scene. Figure 2 shows the representation of GIST descriptor.

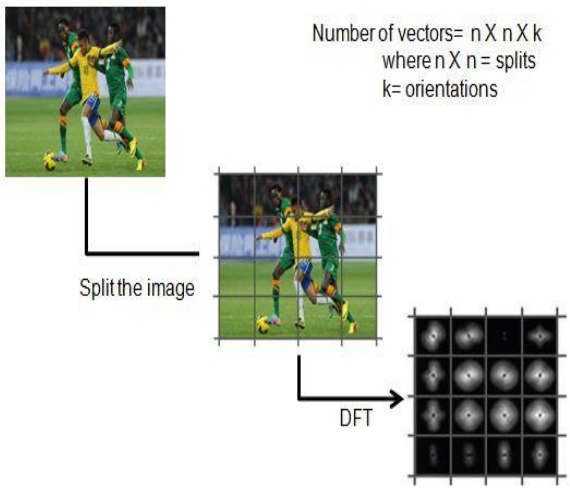

Figure 2: Pictorial representation of GIST Descriptor

Gist Descriptor is better when compared to other image descriptors because no segmentation or grouping operations is done, we can store the features extracted in a vector, easy to handle and works faster than other algorithms. Gist achieved high accuracy in recognizing outdoor scene categories. However the performance of Gist fails drastically in the case of indoor scene recognition.

Local Descriptor (Bag-of-words Features)

The idea was inspired from natural language processing applications, where each text document is represented by a histogram of word of occurrences present in the document. In the case of images, "visual" words are extracted from the images which are nothing but the local features present in the image. The key idea is that visual words are represented by SIFT key points. BoW is then created as a histogram of visual words. Figure 3 shows the working of Bag of Words.

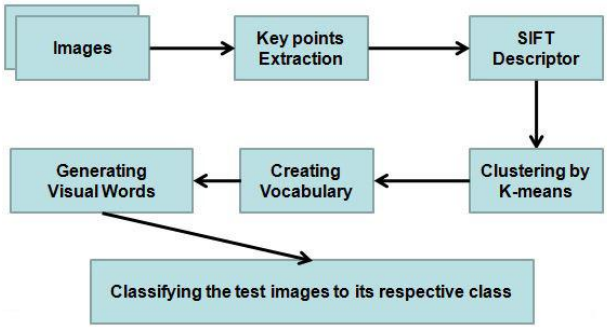

Figure 3: Working of Bag of Words (BoW)

SIFT Features

The SIFT features extracted from an image a collection of frames or key points. SIFT descriptor features are invariant to scaling, rotation and translations. SIFT for an image is performed by convolving Gaussian filters at different scales. Difference of Gaussian (DOG) is taken for the image. Key points are the minima/ maxima of the Difference of Gaussians that occur in multiple scales. Figure 4 shows the working at different levels.

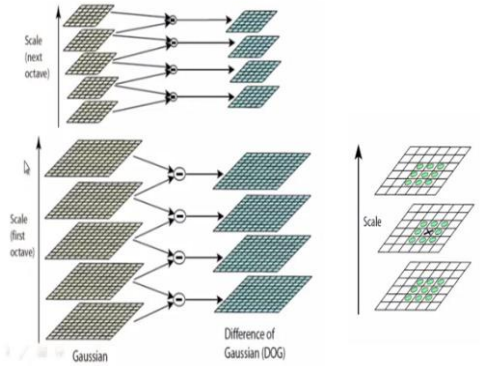

Figure 4: Difference of Gaussian (DOG) performed at different level

Later the key points are localized by interpolation of nearby data for accurate position of key points. Noise may be present at the edges, hence they are convolved with Hessian matrix to eliminate the edge noises. Finally relative orientation and magnitude are computed.

Work Flow

The classification has three stages which are feature extraction, training of the features and Classification. In the case of Global Descriptor, initially the GIST features are extracted from the images and the features are stored in vector form. Next the images are trained to the class which it belongs using SVM classifier. During the testing phase, feature vectors are calculated from the test images and using distance measure the images are classified to the corresponding class. In the case of local descriptor, key points are extracted from the image. SIFT descriptor then vectorizes the feature points in the image. The features are then clustered using k-means clustering. Later a visual word is created which gives a pictorial representation of the cluster to which it belongs. The test image's feature vector is analyzed with the trained class and distance measure says the exact class to with the image belongs.

Experimental Results

Data Set

The dataset used is Media Eval Social Event Detection Dataset [2]. The dataset consists of 1,67,332 images and was acquired from 4,422 unique flickr users. The dataset consists of event annotation of nearly 149 target events. We have considered football event and events where people are present, for which the global and local descriptors are compared. Figure 5 shows examples of events.

Figure 5: Example of Events

In order to analyse the algorithm, 80 positive images and 80 negative images are considered. The positive image says about the event and the negative image says about the images which are not part of football events. 160 images are normalized (resized) to 256 X 256 image size. The resized images are further split into 4 X 4 blocks. Using the GIST image descriptor we extract 512 feature vectors for each image and totally 160 X 512 feature vectors are calculated. During the testing phase we load 20 images which are of both the classes.

Once the features are extracted and vectorized, SVM classifier is used to train the feature vectors to the type of event to which it belong i.e. classify the images according to the class.

Initially SIFT features are computed and key points are found in the image which are the local features present in that particular image. The keypoints are vectorized and quantized into visual words. Figure 6 and figure 7 show the key points and its visual words.

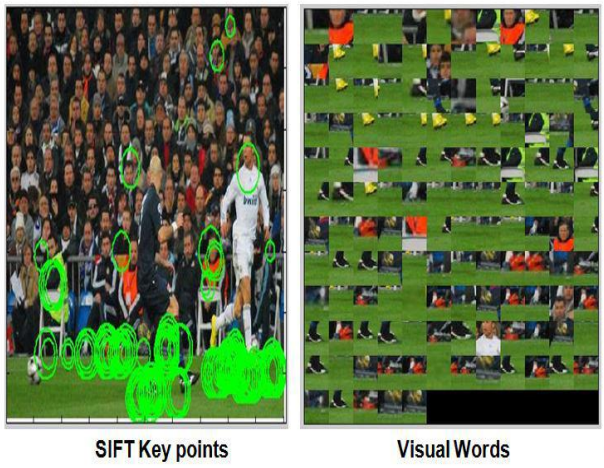

Figure 6: Key Points and Its Corresponding Visual Word

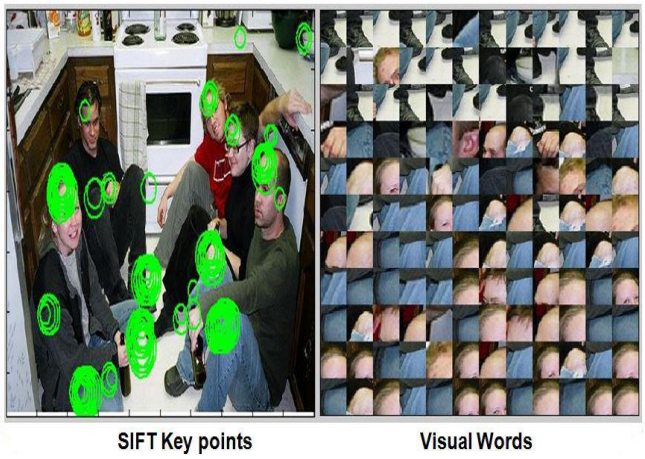

Figure 7: Key Points and Its Corresponding Visual Word

The frequency of each visual word is recorded in a histogram. The final feature vector is the concatenation of all the histograms of the images in the dataset.

Evaluation Measure

After successfully classifying the events we must use evaluation measure to check the accuracy of the classification. Accuracy of the system is calculated by comparing the test images with respect to the ground truth.

Accuracy is calculated by

$$ACCURACY = \frac{TRUE\ NEGATIVE + TRUE\ POSITIVE}{(TOTAL\ TEST\ IMAGES)} \quad (1)$$

It was tested on two different events, outdoor event data and indoor event data. The outdoor event data includes different games and for the case of indoor event data we have images with human face. It is seen that the GIST accuracy is about 77.1% for the games data. In the case of face data, the recognition rate was found to be 64.29%. Similarly precision and recall was found out to be 0.81 for event 1 and 0.72 for event 2.

In the case of BoW accuracy is about 52.8% in the case of games data. But for the case of face data, the accuracy is found out to be 80.56%. Similarly the precision and recall was 0.70 for event 1 and 0.87 for event 2.

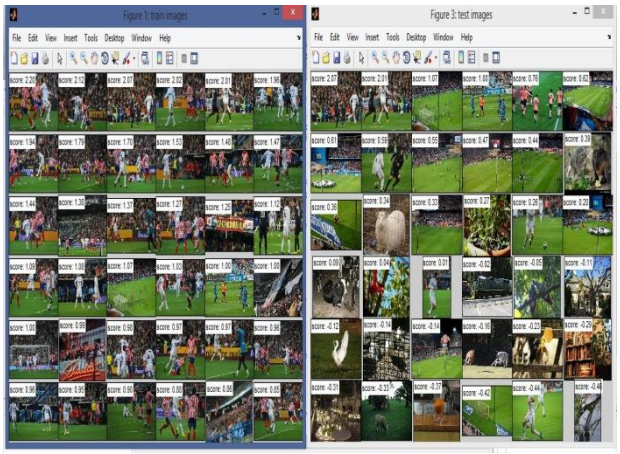

Figure 8: Training and Test images

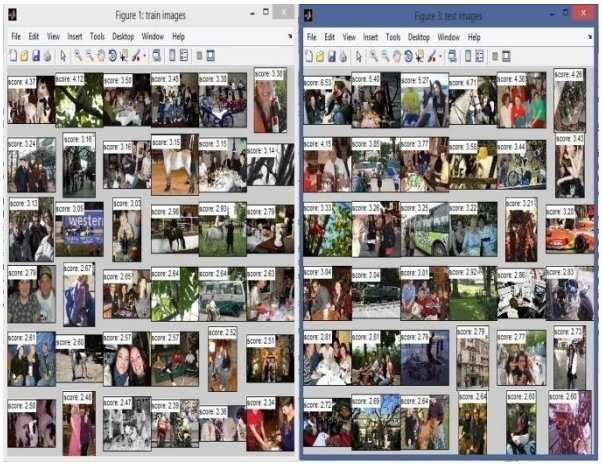

Figure 9: Training and Test images

Table 1: Comparision of both the methods

	GAMES	FACE
TRAINING IMAGES	160	160
TESTING IMAGES	20	20
GIST ACCURACY	77.1%	64.29%
GIST AVERAGE PRECISION	0.81	0.72
BoW ACCURACY	52.8%	80.56%
BoW AVERAGE PRECISION	0.70	0.87

Conclusion

Here classification of social event related images based on GIST descriptors features and SIFT descriptor is performed. Our preliminary results show that the GIST descriptor alone is sufficient for the purpose of classification of outdoor event related images. In the case of indoor events, Bag of Words (BOW) outperforms GIST Descriptor. From this we can conclude that in the case of indoor images Bag of Words performs better than GIST. For outdoor images GIST performs better than Bag of Words. One of our future goals would be to combine both the global and local features into a single feature and measure its outcome.

References

[1] Markus Brenner, Ebroul Izquierdo, "Event-driven retrieval in collaborative photo collections" 14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), 2013.

[2] Timo Reuter, Symeon Papadopoulos, Giorgos Petkos, Vasileios Mezaris, Yiannis Kompatsiaris, Philipp Cimiano, Christopher de Vries, Shlomo Geva, "Social Event Detection at MediaEval 2013: Challenges, Datasets and Evaluation".

[3] Ivan Sikiric, Karla Brkic, Sinsa Segvic, "Classifying Traffic Scenes Using The GIST Image Descriptor" Croatian Computer Vision Workshop, CCVW, 2013.

[4] S. Arivazhagan, R. Newlin Shebiah, S. Aarthi, M. Abinaya, S. Sindu Priya, "Flower Species Classification using Statistical and Gist Descriptor", Journal of Emerging Trends in Computing and Information Sciences, July, 2014.

[5] Keyuri M. Zinzuvvadia, Bhavesh A. Tanawala, Keyur N. Brahmbhatt, "A Survey on Feature Based Image Retrieval Using Classification and Relevance Feedback Techniques", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, January, 2015.

[6] Oliva, Torralba, "Building the gist of a scene: the role of global image features in recognition", Visual Perception, Progress in Brain Research, vol. 155, 2006.

[7] Harpreet Kaur, Kiran Jyoti, "Survey of Techniques of High Level Semantic Based Image Retrieval", International Journal of Research in Computer and Communication Technology, Vol. 2, January, 2013.

[8] Rekhil M. Kumar, SreeKumar K, "A Survey on Image Feature Descriptors", International Journal of Computer Science and Information Technologies, Vol. 5, 2014.

[9] Chih-Fong Tsai, "Bag-of-Words Representation in Image Annotation: A Review", ISRN Artificial Intelligence, Vol 2012, 2012.

[10] Ritendra Datta, Dhiraj Joshi, Jia Li, James Z. Wang, "Image Retrieval: Ideas, Influences, and trends of the New Age", ACM Computer Surveys, Vol. 40, No. 2, Article 5, April 2008.

[11] Robert M. Haralick, K. Shanmugam, It'sHak Dinstein, "Texture Features for Image Classification", IEEE Transactions on Systems, Man and Cybernetics, 1979.

[12] P. Mohanaiah, P. Sathyanarayana, L. GuruKumar, "Image Texture Feature Extraction Using GLCM Approach", International Journal of Science and Research Publications, Vol. 3, May 2013.

[13] S. Mangijao Singh, K. Hemachandran, "Content-Based Image Retrieval using Color Moment and Gabor Texture Feature", International Journal of Computer Science Issues, Vol. 9, September, 2012.