Middle-East Journal of Scientific Research 24 (4): 1057-1062, 2016
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2016.24.04.23189

Image Object Retrieval Using Distribution of Mixed Shape Descriptors

S. VanithaSivagami and K. Muneeswaran

Department of Computer Science and Engineering,
Mepco Schlenk Engineering College, Sivakasi, India

Abstract: Most of the real world objects are characterized by their shape. Computing the shape signature to truly reflect the object of interest is proceeded through the edge detection, extraction of relevant features from the border pixels. In this paper, we propose a distribution of mixed shape features (gradients, angle and radius) for image representation and retrieval of images. These features outperform the recent state of art methods like Bag of SIFT features and the Spatial Pyramid features. The experiments were conducted using the Caltech 101 dataset.

Key words: Distribution of mixed shape features · Visual Words · Spatial Pyramid · Caltech dataset

INTRODUCTION

Human brain works faster than computer system to identify salient image regions for image classification and fast retrieval. Mostly the visual observation of the environment by the human is based on his/her object of interest. In order to make the computer system to function very similar to human brain, human behavior can be simulated for identifying most effective salient regions. Many salient region based detection methods are reported in the literature [1, 2] used in applications based on image classification and retrieval. In the last two decades, the work on image retrieval was carried out based on the primitive features and its derivatives. However the focus was only on syntax based techniques. Recent applications like face recognition, Architectural and engineering design, Crime detection and prevention, Medical Diagnosis etc. dictate the retrieval based on semantic features insisting the bridge between the image understanding and the low level features (color, texture).

In the semantic based image retrieval, the visual features are derived from the primitive features whose values are clustered to form a visual word. In other words, a dictionary or a code book of visual words is constructed. This Bag of Visual Words (BoVW) [3, 4] plays a significant role in adding meaning to the features. The objective of constructing the visual words is to have minimal code book size and maximal exemplar representation. Hence each image's signature is represented using a Bag of Code or Visual words. Recently Scale Invariant Feature Transform (SIFT) features are dominantly used features which are invariant to scale, orientation and translation of images. However the construction of visual word is time consuming. In the proposed work we have constructed the distribution of edge orientations, angle and radius of the object from its Centroid as Distribution of Mixed Shape features for representing the single object image. Fig. 1 shows the sample shape based images from the Caltech 101 data set which is used in the proposed work.

Fig. 1: Sample Images from Caltech 101 dataset

The rest of this paper is as follows : Section II discusses the related works existing in this domain. Section III details about the proposed distribution of mixed shape features technique. Section IV highlights how the proposed technique is used for object retrieval. Section V analyzes the experimental results by comparing the proposed work with existing techniques. The concluding remarks are added in Section VI.

Corresponding Author: S. Vanitha Sivagami, Department of Computer Science and Engineering,
Mepco Schlenk Engineering College, Sivakasi, India.

Related Works: In machine learning and computer vision, Shape-based detectors conform to the shape details of the images such as borders, straight lines, or arcs to capture the location of points of interest. They are used mainly to retrieve the target object such as rigid and disjoint objects using shape characteristics. Edges were used as shape descriptors[5] where the first- and second-order digital derivatives such as the Sobel, Prewitt, Laplacian, Laplacian of a Gaussian (LoG) and Canny operators were used for the detection of edges in an image. Sparse oriented edge maps [6] were found at locations having high edge energy. Corners were used as point of interest and the features based on them such as Harris Laplace Detectors, Hessian Laplace detectors [7] gave good results in image understanding and retrieval. Histogram of the image and Histogram of Gradient (HOG) play an important role for representing images [3, 8]. Since these HOG features capture the edge information, they become the significant shape feature and they are invariant to small amount of translation and rotation. This is applicable when the translation is smaller than the orientation bin size.

A fast method to compute histograms was proposed by Fatih Porikli [9] where it started from an origin point and traversed through remaining points along a scan line. At each point the histogram is integrated using the previously processed histograms of the neighbors. Histogram of target regions can be computed from the histograms of their corner points. A multi-texton histogram was constructed to represent the image [10] for efficient retrieval of the images.

Most of the earlier image retrieval was performed by combining color, texture and shape features [11], where a small set of dominant colors were used as the color feature after Color Quantization. Texture features were extracted using steerable filter decomposition. Pseudo-Zernike moments of an image were used for shape features. The combination of the color, texture and shape features provided a set of features for the image retrieval. The SIFT features is another set of features extracted from the image on important points in the image [12].

The Bag of Words (BoW) model was proposed and successfully implemented for document processing [13, 14]. This model did not consider the ordering of words in the document. Later this model was also used in computer vision field for efficient image classification, segmentation, image retrieval [4, 15]. This model used a histogram of significant features in an image to represent the object. Clustering was required so that a discrete vocabulary can be generated from millions (or even billions) of local features sampled from the training data.

Using a spatial pyramid framework [3] BoWs gave better results when multiple resolutions were combined. The feature vectors were quantized into multiple classes. Along with the feature set, their coordinates (X and Y) were also used for segregating and representing the feature set in spatial pyramid form. The spatial pyramid matching was done as a single histogram intersection of "long" vectors formed by concatenating the appropriately weighted BoWs at different resolutions (the entire image, image that was subdivided into 2 x 2 i.e. 4 blocks, image that was subdivided into 4 x 4 i.e. 16 blocks).

Generally, the shape of an object is well described by features like gradients, Centroid, contour of the object. In our proposed work the discriminative shape features such as gradients, radius from the Centroid of the shape, angle at each border pixels are considered. These features represented as Distribution of Gradients (DoG), Distribution of Radius (DoR) and Distribution of Slope Angles (DoSA) are combined and used to form the signature of an image. This distribution of mixed shape features represent the various characteristics associated with the shape for retrieva [16].

Extracting Distribution of Mixed Shape Features: Our proposed work mainly concentrates on the shape based image retrieval and hence Distribution of Mixed Shape Features has been considered by us. Figure 2 shows the phases in the proposed work. The computation of the features from the image involves identifying the gradients, removing the weak edges for the detection of border followed by the computation of statistical mixed shape features, which are presented in the following sub-sections.

Distribution of Gradients Extraction: Our human visual system identifies shape well through edge orientation. Strong edges in an image contribute more to the contour of the object incurred. Gradient of a image ∇I is given by partial derivative of the image in x direction together with the partial derivative of the image in y direction.

$$\nabla I = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right) \tag{1}$$

The magnitude of the gradient image G is calculated as:

$$G = \sqrt{\left(\frac{\partial f}{\partial y} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2} \tag{2}$$

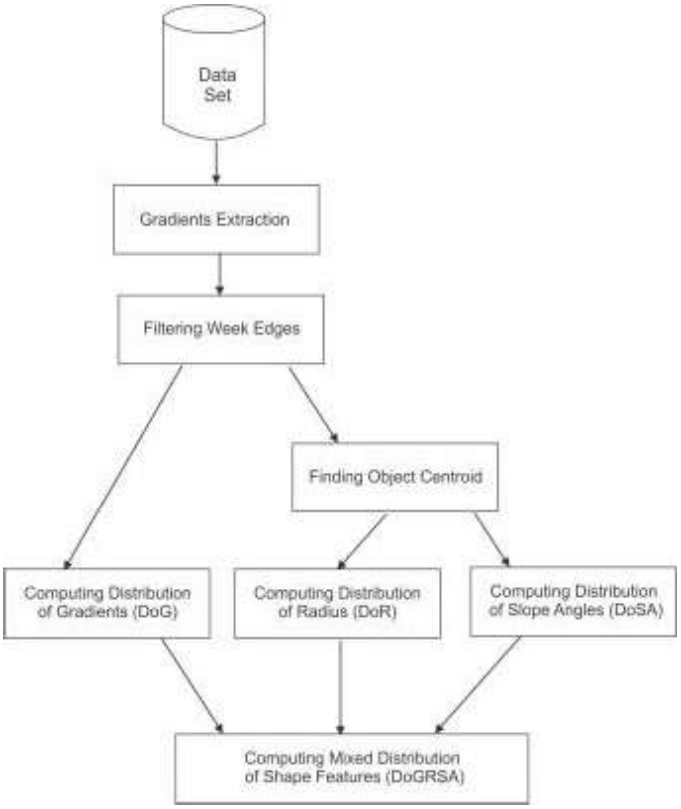

Fig. 2: Flow diagram for the proposed work

The orientation or the phase of the gradients is calculated as:

$$\theta = \tan^{-1}\frac{\left(\frac{\partial f}{\partial y}\right)}{\left(\frac{\partial f}{\partial x}\right)} \tag{3}$$

The phase of the gradients is clustered into K number of bins to form the first distribution of shape features.

Filtering Weak Edges: Among many edge points detected, only few of them contribute to the contour of the objects. Hence only significant edges are retained and weak edges are filtered out using a threshold *Th* as per the equations (4) and (5):

$$Th = H_{mag} * c, \quad where \quad H_{mag} = \max(G) \tag{4}$$

and c is the user defined constant. Now the weak edges can be filtered if their gradient magnitude is less than the threshold and the edge image I_e given by:

$$I_e = \begin{cases} 1 & G \geq Th \\ 0 & else \end{cases} \tag{5}$$

Finding Object Centroid: After filtering the weak edges, the contour of the object is defined by the significant edges. The centroid of the object is calculated as:

$$x_c = \frac{1}{N}\sum_{i=1}^{N} x_i, \quad y_c = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{6}$$

where x_c and y_c are the co-ordinates of the object's centroid and N is the number of strong edge pixels and x_i,y_i are the x and y coordinate of the i^{th} strong edge pixel.

Computing Distribution of Radius: The radial distance between each strong edge pixel 'i' and the centroid of the object is computed as:

$$r^i = \sqrt{(x_c - x_i)^2 + (y_c - y_i)^2} \tag{7}$$

But some of these gradients can be present in the left side of the centroid whereas some others can be in the right side of the centroid. To consider these directions

Fig. 3: Finding radius of gradients for different objects in Caltech dataset (a)Objects in Caltech dataset (b) strong edges of objects (c)Selected radius (marked as red lines) from centroid of the object (marked as green square)

also in the object's shape definition, a special sign is introduced. The sign of the radius is included as +1, if the gradient is to the left side of the centroid and it is included as -1, otherwise.

Hence the radius is considered as a signed integer, holding both the distance and direction of each gradient from the centroid of the object. Some selected radial lines for some sample images with gradients are shown in the third column of Figure 3.

These radius values are quantized into R bin histogram forming the feature vector of size R.

Computing Distribution of Slope Angles: The next important feature we have considered is the Distribution of Slope Angles (DoSA). This feature takes the angle between the radial line of the edge pixel and the x axis which is represented by the slope of the line joining the centroid (x_c, y_c) and each of the strong edge pixels (x_i, y_i) as:

$$angle^i = \tan^{-1}\left(\frac{y_i - y_c}{x_i - x_c}\right) \quad (8)$$

The distribution of the angle of slopes for all the strong edge pixel is used to form S dimensional feature vector.

The features from DoG, DoR, DoSA are combined to form the Distribution of Mixed Shape Descriptors (DoGRSA). The size of the feature vector for the given image is K + R + S.

Image Retrieval Using Mixed Shape Features: For image retrieval we have considered images including only single object. The distribution of mixed shape descriptors were extracted from the set of training images and stored in a database as a set of Feature Vectors (FVs). Then for each query image a similar process is followed. By comparing the FVs, similar images from the training set are retrieved and ranked.

Table 1: Performance compared with existing techniques

S.No	Techniques Used	Accuracy (%)
1.	Bag of SIFT Features using KMeans clustering [4,15]	38.25
2.	Bag of SIFT features with Fuzzy C Means clustering [4,15]	47.98
3.	Spatial Pyramid of Bag of SIFT features [3]	59.19
4.	Proposed Distribution of Mixed Shape Features	66.5

Since the FVs are computed using the statistical measures, the histogram intersection distance is used to compare the similarity between two FVs with the expression given in equation (9).

$$d(h_A, h_B) = 1 - \frac{\sum_{i=1}^{n} \min(h_A(i), h_B(i))}{\sum_{i=1}^{n} h_B(i)} \quad (9)$$

Here n is the total number of bins (K + R + S) in DoGRSA. If this distance is minimum, then the FVs compared are similar, thereby representing images of same class.

Experimental Results and Discussion: We conducted experiments using a subset of Caltech 101 dataset [16] which includes a total of 101 categories where each category includes 40 to 800 images. The total number of images that we have considered for our experimentation is 423 from 8 different categories (Aero plane, Bonsai, Chair, Elephant, Sunflower, Laptop, Motorbike and Gramophone). The size of each image is roughly 300 *200 pixels. These experiments were conducted with 50% training and 50% testing images.

Table 1 shows the performance of the proposed technique compared with Bag of SIFT Features and Spatial Pyramid of Bag of SIFT features where the dictionary or code book is constructed by clustering all the available visual features. The experiments were conducted using KMeans clustering, Fuzzy C Means clustering for the clustering purposes and listed in the Table 1. The proposed work shows improved performance.

In the existing techniques the dictionary size is 75 where as in the proposed method, the optimal number of bins for the maximum accuracy is found (by empirical method) to be 24 (K =5 for DoG, R =7 for DoR and S=12 for DoSA). The average time taken for extracting Mixed Shape features for an image is very less compared to the SIFT. In the proposed method, it takes an average time of **0.012** seconds for the feature extraction as against the value of **0.44** seconds in the case of SIFT features showing the computational gain of **37** (0.44 / 0.012) for each image. The reason is that for other methods using the code book or dictionary, there is an additional over head in terms of computational complexity for the code book construction though it is done offline. All these discussions show that the proposed feature extraction technique shows improvements in accuracy and reduced computation time for feature extraction.

CONCLUSION

Thus in our proposed work, we have experimented with the state of art feature extraction techniques and our own proposed techniques showing the improvement in the accuracy. The tuning parameter for all the methods based on Bag of Visual Words is the size of the dictionary/code book, number of bins in the distribution. Also the quality of the features extracted is a matter of interest for the given set of images. There is no unique method or technique which gives the universal solution for all the kind of images for the retrieval applications. Due to the clustering process in the code book construction over a very large number of visual features, the number of images considered in our proposed work has been completed for a subset of images. Further work is under progress with the deployment of the distributed computing environment (Hadoop) over the large number of images.

REFERENCES

1. Manipoonchelvi, P. and K. Muneeswaran, 2014. Region-based saliency detection. IET Image Processing, 8(9): 1-9.
2. Cheng Ming-Ming, Guo-Xin Zhang, N.J. Mitra and Xiaolei Huang, 2011. Global contrast based salient region detection. IEEE International conference on Computer Vision and Pattern Recognition, pp: 409-416.
3. Lazebnik, S., C. Schmid and J. Ponce, 2006. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2: 2169-2178.

4. Otavio, A., B. Penatti, Eduardo Valle and Ricardo da S. Torres, 2011. Encoding spatial arrangement of visual words. CIARP'11 Proceedings of the 16th Iberoamerican Congress conference on Progress in Pattern Recognition, Image Analysis, Computer Vision and Applications, 7042: 240-247.
5. Gaung Hai Liu and Jing Yu Yang, 2013. Content Based Image Retrieval using Color Difference Histogram. Pattern Recognition, Elsevier, 46(1): 188-198.
6. Fergus, R., P. Perona and A. Zisserman, 2005. A Sparse Object Category Model for Efficient Learning and Exhaustive Recognition. IEEE Conference on Computer Vision and Pattern Recognition, 1: 380-387.
7. Mikolajczyk Krystian and Cordelia Schmid, 2004. Scale & Affine Invariant Interest Point Detectors. International Journal of Computer Vision, 60(1): 63-86.
8. Dalal, N. and B. Triggs, 2005. Histograms of oriented gradients for human detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1: 886-893.
9. Porikli Fatih, 2005. Integral Histogram : A fast way to extract Histogram in Cartesian spaces, IEEE International conference on Computer Vision and Pattern Recognition, pp: 829-836.
10. Liu Guang-Hai, 2010. Image retrieval based on multi-texton histogram. Pattern Recognition, 43: 2380-2389.
11. Xiang-Yang Wang, Yong-Jian Yu and Hong-Ying Yang, 2010. An effective image retrieval scheme using color, texture and shape features. Computer Standards and Interfaces, Elsevier, 33(1): 59-68.
12. Lowe David, G., 2004. Distinctive image features from scale-invariant key points, Int. Journal of Computer Vision, 60(2): 91-110.
13. Boureau Y. Lan, F. Bach, Y. LeCun and J. Ponce, 2010. Learning mid-level features for recognition, IEEE Conference on Computer Vision and Pattern Recognition.
14. Sivic, J. and A. Zisserman, 2003. Video Google: A text retrieval approach to object matching in videos, International Conference on Computer Vision, 2: 1470-1477.
15. O'hara Stephen and Bruce A. Draper, 2011. Introduction to the Bag of Features Paradigm for Image Classification and Retrieval. arXiv: 1101.3354, Cornell University.
16. http://www.vision.caltech.edu/Image_Datasets/Caltech101/.