International Journal of Combined Research & Development (IJCRD)
eISSN: 2321-225X;pISSN:2321-2241 Volume: 4; Issue: 6; June -2015

Content Based Classification using Object Identification

Dr. Hanumanthappa. M
Professor
Department of Computer Science and Applications
Bangalore University, Bangalore

S. Regina Lourdhu Suganthi
Research Scholar
Department of Computer Science and Applications
Bangalore University, Bangalore

Abstract : Advent of low-cost digital devices have increased Multimedia content extensively in various organizations. The data generated through these devices have massive impact on storage and retrieval. The image and video data captured during various events in organizations are occasionally used after a immediate duration. Classifying and Tagging these images based on the image content and event context would facilitate and expedite the retrieval of these images. The tagged images could then be accessed for use at a later instance based on a query. The content of the image used for tagging could include important personalities and key objects that describe the event context. Object identification based on training data set combined with feature extraction would automatically tag the image set for future retrieval.

Keywords : ***CBIR, HOG, Haar, LBP, ROI***

1. INTRODUCTION

Educational institutions conduct curricular and co-curricular events round the year and the image acquisition activity through the digital devices installed at the event venues is predominant. It is evident that the cost of recording the events is negligible compared to the storage requirement. Data accumulates at a faster rate and becomes a growing challenge for human to process it and to draw inferences. Thus the image data collected at source must automatically be preprocessed, efficiently indexed to create and maintain an archive. The archive of image and video data could then be used for auditing, reporting and future reference. Tagging these images using the event context and image content would facilitate query based retrieval. The event context used for tagging may constitute event name, session details and the like, on the other hand the image content require image processing techniques to identify objects and important personalities with respect to the event context. The event context and the image content could be combined to tag the images and classify the images into predetermined groups.

The paper is organized as follows : Section 2 presents the review of various object recognition techniques proposed by Researchers. Section 3 introduces the proposed technique.

Section 4 presents a detailed discussion of content based image tagging through object identification. The experimental results are presented in Section 5. Finally, the Section 6 concludes with scope of the work.

2. LITERATURE REVIEW

A scene in an event is generally a composition of various objects and persons. To tag these images based on image content, it is quintessential to identify objects that will describe the scene. Person identification has been dealt by the authors using face detection[6][8][9] combined with eigenspace matching and maxmin threshold. To recognize the objects from an image, highly distinctive scale and rotation invariant features are extracted from the training image to train the classifier. In the past decade various Content Based Image Retrieval (CBIR) techniques have been proposed to search the image databases to match the query image[1]. CBIR systems focus on extracting image features such as texture, color, shape and region[2]. Similarity measures or clustering techniques are used to compare the query image with the images in the image repository.

Global as well as Local signatures of an image could be used to characterize an image. Global features and Local features reveal different types of information about an image. Global features can describe the entire object with a single vector. Global features include histograms, shape descriptions, edge descriptions, contour representations, shape descriptors, and texture features[3]. On the other hand, Local features lead to more accuracy due to the fact that the local features are computed at multiple points in the image and are consequently more robust to occlusion and clutter.

To describe local features and structures, Local Binary Pattern (LBP)[4] and its variations have been shown to be simple and robust approach. The initial LBP operator was based on the assumption that texture has two important characteristics namely pattern and its strength. A generic LBP is obtained as sum of thresholded differences weighted by powers of two. In calculating the LBP for a given $N \times M$ image sample ($x_c \in \{0,...,N-1\}$, $y_c \in \{0,...,M-1\}$), the central part is only considered because a sufficiently large neighborhood cannot be used on the borders. Another extension of the original LBP uses uniform patterns to achieve better recognition results in many applications. The limitation of LBP in terms of small spatial area has been alleviated by Multi scale LBP. To capture large scale structures, N LBPs with different P – sampling points and different R – Radii could be combined to obtain information for large scale structures using large spatial area.

Locally normalized Histogram of Oriented Gradient (HOG) descriptors[7] provide edge orientations for pixels within a region after dividing the image into small connected regions. Each region is classified into its angular bin based on the gradient orientation. Then the adjacent regions are grouped into blocks. The normalized histograms of these blocks represent the HOG descriptors. These are widely used to detect objects in an image.

3. FRAMEWORK

The proposed methodology of object identification and classification for event image tagging process is done in stages as stated below.

Step 1 : The set of event images are classified as day1, day2, …, day n using the time stamp.

Step 2 : Each days' events are further grouped into session sets based on the session information.

Step 3 : Each session image set is processed to identify the person of interest or object that describe a scene. Person of interest may include chief guest, guest of honour, Head of the Institution and the like. Object that describe the specifics of an event may include a News Letter in the News Letter Release session or an Institutional Logo in a technical event.

Step 4 : Face Identification technique from computer Vision has been coupled with eigen faces to train the training data set. Features are extracted from each face and are compared with the trained data for matching by applying a maxmin threshold. If a match is found, the image is tagged with the date, event name and the name of the guest.

Step 5 : To detect objects whose aspect ratio do not vary much, Cascade Object Detector is used to train the classifier by sliding a window over the image in stages. The ROIs that are of different in size and orientation are provided to the classifier.

Step 6 : The statistical texture features are extracted from the ROIs used for training the classifier. The average measure of these features are computed to refine the results.

Step 7 : Test images are processed one at a time to detect the trained object. The detected object windows may include some false positives. These false positive windows (windows that are wrongly identified as positives) are further processed using statistical texture features detected in Step 6 to accurately identify the trained object. If an image is identified to include the trained object, it is tagged with the object name.

Step 9: The images that are tagged in steps 4 and 7 can be ranked based on the priority tags. These tags on images would undoubtedly facilitate easy future reference.

The following figure (Figure 1) illustrates the image tagging process:

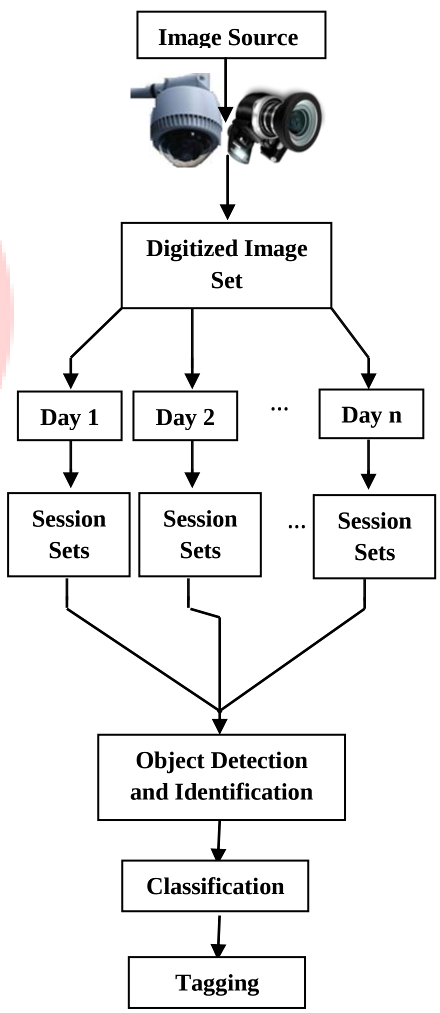

Figure 1 : Image Tagging Process

4. PROCESS FLOW

Image processing applications are generally domain specific. In organizations, captured images during various events are used mainly for documentation and for printing annual newsletters, magazines and the like. Thus the purpose of tagging the images here is for creating an archive based on image content and context in order to enable easy and efficient retrieval based on a simple query.

4.1 Classification of the image set into sub classes

The time stamp of the image plays crucial role in the top most hierarchy. Session durations coupled with time stamp could be used to classify the entire event image set into sub classes of pre-determined labels.

4.2 Content Based Classification

In the next level of classification hierarchy, the image content is used to provide contextual information. The objects that describe a particular event attribute could be used to train the classifier. Objects such as logo, sign board, face, car and the like, where the aspect ratio of the object do not vary significantly, cascade object detector could be used to detect the objects.

4.2.1 Training the Classifier

Cascade Object Detector is used to train the objects of interest. The classifier is trained in stages using a technique known as Boosting. The classifier labels the regions by sliding a window of a specific size through the image and by marking them as positive or negative. If a region is labeled as positive at a particular stage then that region is passed on to the next stage. The ultimate goal of training the classifier is to have minimum false negative rate.

The classifier uses two image sets of samples for training. One set consisting of images with the regions of interest specified in every image to act as positive samples and the other set with images where the object of interest is not present in order to be used as negative samples. The number of stages of training could be increased to achieve accuracy. The three features used by the object detector are Histogram of Oriented Gradients (HOG), Haar and Local Binary Patterns (LBP).

4.2.2 Extracting the Texture features

The high false positive rate in the output of the classifier is addressed by extracting texture details from the ROIs of the positive samples. Using a distance measure and an appropriate threshold, the false positive regions are further processed to yield either true positive or true negative.

4.2.3 Processing the Test images

Images that are to be tagged are processed by the object detector one at time. The regions that are detected by the cascade object detector are reported as bounding boxes. These bounding boxes contain some false positives. Thus each of the bounding boxes are further processed by extracting the texture details of the region. The average texture data from the positive samples used for training the classifier is computed and compared with the extracted texture feature of each bounding region using a distance measure and an appropriate threshold to accurately detect and declare only true positives.

5 . EXPERIMENTAL RESULTS

Cascade Object Detector of the Computer Vision toolbox of MATLAB software has been used to detect objects. Images captured in a 2 day seminar organized at an educational institution consisting of 704 images have been taken as image data set for processing. The focus is to tag the images by identifying the podium logo. Using the time stamp, the image set is first classified into two sets. One set with 247 images taken on day 1 and another set with 457 images taken on day 2.

To train the classifier the following data sets are used :

- Positive samples : One set of 18 images with the logo object specified as ROIs are chosen (shown in figure 2)
- Negative samples : To achieve accuracy negative samples have to be more and hence, a set of 111 images without the logo object are chosen (shown in figure 3)

Figure 2 : Positive Samples with logo object as ROI

Figure 3 : Negative Samples without logo Object

The number of training stages must also be more to have greater accuracy, and hence, the classifier is trained in five stages. HOG features are useful in detecting objects with overall shape. Since the object of interest logo too has a definite shape, HOG features have been extracted from the positive samples and are used to train the classifier.

The texture features from the 18 positive samples are extracted and the mean of few of the moments are computed for each color components (RGB).

The test data consisting of all the 704 images have used. The detector marks the detected regions in each of the images with bounding boxes to represent the positives (Shown in figure 4).

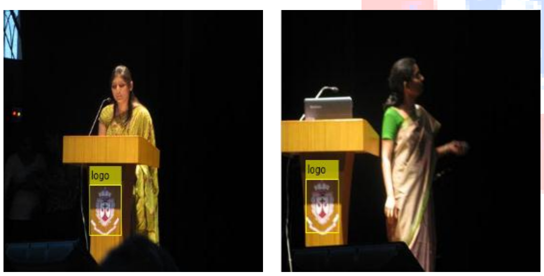

Figure 4 : Detected Region in the Test Image

The detected regions that include false positives are shown in figure 5.

Figure 5 : Detected False Positive Regions in the Test Image

Thus before labeling each region as positive and to reject false positives, the procedure is further refined. From each detected region the component wise texture features are extracted and compared with component wise mean texture features of the positive samples using a distance measure and an appropriate threshold. This refined procedure results in perfect detection with no false positives.

The detected images are classified into pre-determined groups based on the content of the image through person, object identification and the context of the event. The images are then tagged with the classified group id.

6. CONCLUSION

Image processing applications are domain specific. The problem under consideration is to automatically classify the image set that are taken during various events in organizations and to tag them. This tagged archive could then be retrieved at any point of time using simple queries and could be used in preparing annual documents, magazines and at time even to retrieve historic event images. Face detection based tagging discussed in earlier work help in identifying person of interest. This paper focused on object based tagging. The two aspects could be combined to classify the images into different pre-determined classes. The challenges that are to be addressed include : i) Objects that are occluded ii) Objects that are not orthogonal to the focus and iii) Objects that are far from the focus (Shown in Figure 6). These issues could be addressed by enhancing the detection by considering objects in parts.

Figure 6: Occluded, non-orthogonal, distant from focus

7. REFERENCES

[1] A.Kannan, V.Mohan, N.Anbazhagan ,“ Image Clustering and Retrieval using Image Mining Techniques”, IEEE International Conference on Computational Intelligence and Computing Research, 2010.

[2] D.R. Martin, C.C. Fowlkes, and J.Malik, “Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004.

[3] David G. Lowe ,“Distinctive Image Features from Scale-Invariant Keypoints”, International Journal of Computer Vision 60(2), 91–110, 2004.

[4] Junding Sun, Guoliang Fan, Xiaosheng Wu, “New Local Edge Binary Patterns For Image Retrieval ”, ICIP 2013, 4014 - 4018.

[5] Ole Helvig Jensen , ” Implementing the Viola-Jones Face Detection Algorithm”, thesis, 2008, ISBN 87-643-0008,ISSN 1601-233X.

[6] Paul Viola, Michael J. Jones, Robust Real-Time Face Detection, International Journal of Cumputer Vision 57(2), 2004.

[7] Qiang Zhu, Yeh, M., Kwang-Ting Cheng, Avidan, S.,” Fast Human Detection Using a Cascade of Histograms of Oriented Gradients”, Computer Vision and Pattern Recognition, IEEE Computer Society Conference,, Volume:2, pg :1491-1498, 2006.

[8] Yi-Qing Wang, “An Analysis of the Viola-Jones Face Detection Algorithm”, Image Processing On Line, 4 (2014), pp. 128–148.

[9] Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. “ Face recognition: A literature survey. ACM Computing Surveys”, 2006,35, 399–459.