Object Recognition based on Representative Score Features

Anu Singha*, Mrinal Kanti Bhowmik
Department of Computer Science and Engineering
Tripura University (A Central University), Agartala, 799022, India
anusingh5012@gmail.com*, mrinalkantibhowmik@tripurauniv.in

***Abstract*—In this paper, we present an approach towards object detection and recognition from various environmental conditions such as foggy morning, dust scenarios, and night vision. The goal of the approach is to develop a holistic feature extraction method over object image patch. To categorize objects, the experimental evaluation has prepared through four classifiers. Investigational results with our own collected video sequences are reported to demonstrate the accuracy of the proposed approach.**

Keywords—Akinity; Liability; Representative Score; Object Feature Extraction; Image Patch

I. INTRODUCTION

Object detection and recognition from video sequences has been considered as one of challenging area in computer vision and pattern recognition because of occlusions, cluttered backgrounds, illumination changes, dust, foggy weathers etc. In video sequences, when the frames are processed, one needs computer vision object detection algorithms to detect objects of importance in the scene. Once such objects are detected, recognizing their type requires machine learning algorithms with build-in intelligence. Lipton et al [1] extracted moving targets from a real-time video stream which are applicable to human and vehicle classification. In feature extraction step, after presented holistic feature work of histogram of oriented gradients (HOG) by Dalal et al [2], many other objects classification methods based on multi-feature fusion have been proposed in the last decades, like HOG + local binary pattern (LBP) [3], HOG + color self similarity (CSS) [4], Haar features + histogram of edges [5], thermal-position-intensity-HOG (TPIHOG or TπHOG) [6].

The goal of this paper is to develop a patch feature based method over similarity concept of image block system with regards to (i) the foggy morning (ii) the dust conditions (iii) the imaging modalities with far infrared (FIR) cameras at night time.

The entire processing of the proposed system is illustrated in Fig. 1. In this section, we briefly describe the proposed system. The system observation model is designed to classify objects from video sequences. To do so, objects are initially detected and segmented using an existing Gaussian Mixture Model (GMM) based foreground/background segmentation algorithm. The GMM stage followed by morphological operations is used as post-processing stage to enhance foreground object detection. To detect blob area of segmented object and localized to main video frame, we analyse a bounding box output port to crop object area. Subsequently, Akinity and Liability (briefly described in Section II) based representative (RP) score feature extraction method has been proposed to extract from the refined foreground object and used as features for object type recognition. To classify the objects, we present a statistical test called Friedman Test and a statistical graphical plot that illustrates the diagnostic ability of a binary classifier system called receiver operating characteristic (ROC) curve.

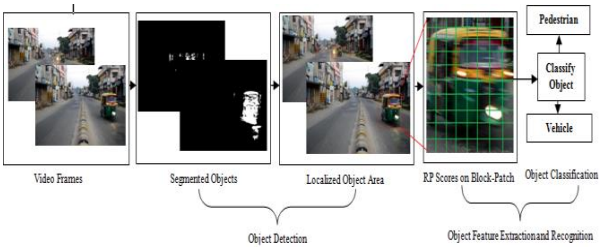

Figure 1: System Overview of our proposed approach

The rest of the paper is organized as follows. Section II describes the proposed framework of feature extraction method and in Section III evaluation of the overall system with experimental outcome. Lastly, concluded in Section IV.

II. REPRESENTATIVE (RP) SCORE BASED FEATURE EXTRACTION

Here, we will go into the details of calculating the RP score on an image patch as feature descriptor. A feature descriptor is a representation of an image or an image patch that simplifies the image by extracting useful information. Typically, a feature descriptor converts an image patch of $p \times q$ to a feature vector of length $n = p \times q$. Clearly, the discriminative feature vectors is not useful for the purpose of viewing the image, it is very useful for tasks such as telling the differences between objects like vehicle, pedestrian, and so on. To illustrate each step, use detected object as a block based image.

Step 1 (Pre-processing): Before dividing into block based image, the images at multiple scales should have a fixed aspect ratio. In our case, the object images need to have an aspect ratio of 3:4 as they can be 96x128. To illustrate this point, suppose we have a large image of size 720x475. We have to detect and bound box the objects and select an object image of size, say, 100x200. This selected object is cropped out from an image and resized to 96x128. Now we are ready to calculate the RP score based feature descriptor for this cropped object image.

Step 2 (Representation of Block Patch): In this step, the object image is divided into number of 4x4 blocks and RP score is calculated for each 4x4 block i.e. patch. One of the important reasons to use a feature descriptor to describe a patch of an image is that it provides a compact representation. A 4x4 patch contains 16 pixel values. As total 96x128 object image divided into 4x4 blocks which adds up to

* Corresponding Author

2161-377X/18/$31.00 ©2018 IEEE
DOI 10.1109/ICALT.2018.00106

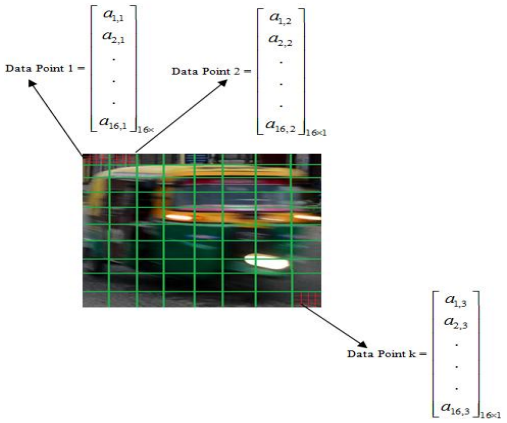

Figure 2: Division of a patch into blocks

$\frac{96\times128}{4\times4}=24\times32=768$ number of blocks. This way, the block-wise representation will give the representation of more compact or reduce in size.

Then each block or patch are taken as a data point as shown in Fig. 2. For example, a 4x4 patch have total 16 pixel values taken in a vector of size 16x1. Therefore, there are total $k=768$ data points.

Step 3 (Calculation of RP Score): Initially, we estimate the similarity between each data points that will generate a similarity based square matrix of size *k* x *k* from *k* data points. The similarity matrix gets as input a gathering of real valued 1 minus Euclidean distance between data point d_i and d_j (*DSim(i,j)*) i.e. maximum similarity towards 1 and minimum similarity towards 0.

The RP score basically estimated from combining scalar value of *Akinity* and *Liability*. The concept of these two terms is briefly described as follows. The term 'Akinity' derived from word Akin, which indicates a most appropriate similar patch that has maximum analogous characteristics among all patches. Fig. 3 shows the *Akinity(i,j)* sent from data point *i* to candidate point *j* (black points) i.e. candidate point *j* is to serve

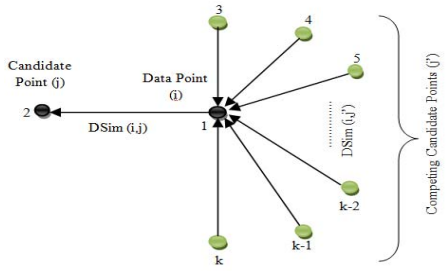

Figure 3: Akinity *a(i,j)* sent from data point *i* to candidate point *j*

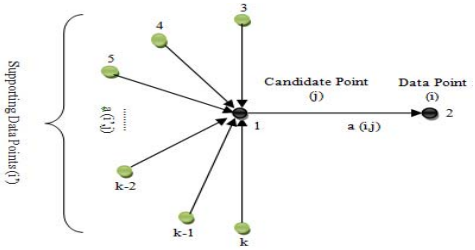

Figure 4: Liability *l(i,j)* sent from candidate point *j* to data point *i*

as the most similar candidate for data point *i*, while other candidate point *j'* (green points) will compete for data point *i*. While the rest of candidate points will consider as competitors for holding that data point, we need to consider how much superior the candidate point than other competing points. To answer this, we have subtracted the largest of the similarities among competing candidate points *j'* as

$$a(i,j)=DSim(i,j)-\max_{j'\neq(i,j)}\{DSim(i,j')\} \quad \text{if } i\neq j \tag{1}$$

$$a(j,j)=\max_{j'\neq j}\{a(j,j')\} \quad \text{if } i=j \tag{2}$$

For $i=j$, the Akinity $a(j,j)$ is set to maximum of akinity values from all the estimated values for $i\neq j$ defined as (2). That means, how appropriate it would be for data point i is chosen as an akin itself.

On the other hand, Liability indicates how much a chosen point is answerable. Fig. 4 shows the liability $l(i,j)$ sent from candidate point j to data point i. The earlier concept *Akinity* is concentrated on a data point at a time where the candidate points are rival for occupancy to rein that data point. Through akinity when a candidate point is decided for a data point, liability update the fact that data point i pick candidate j as the most appropriate point which has been decided by other supporting data points. It is set as the average sum of the akinity values received by candidate point j from other supporting data points i'

$$l(i,j)=\frac{1}{k-1}\sum_{i'\neq(i,j)}a(i',j) \quad \text{if } i\neq j \tag{3}$$

$$l(j,j)=\max_{i'\neq i}\{l(i',j)\} \quad \text{if } i=j \tag{4}$$

To estimate the RP score, the akinity and liability values can be combined for diagonal positions as

$$r(i,j)=a(i,j)+l(j,j) \quad \text{if } i=j \tag{5}$$

A. Depiction of Features

Let us look at one pedestrian image of 4x4 blocks with RP score values in Fig. 5. We have noticed that the object area in the patch image is very separable. The RP scores over object are smaller than the background area. The reason would be the background area has much smoother texture than the object area. The similar rate of smoother area will always give higher similarity score than dynamic changed texture of the object area.

B. Statistical Evaluation

In order to statistically analyse the RP score based features, we have conducted a non-parametric significance assessment known as the Friedman Test (FT)camera Ready for number of data points. Unlike two-way analysis of variance, it is a test for whether the columns (independent variables) are different after adjusting for possible row (dependent variables) differences. In our case, the test is based

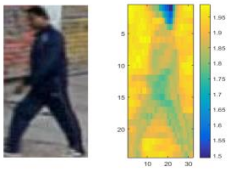

Figure 5: An image patch of corresponding 4x4 blocks with RP score features

on an analysis of variance using the RP score of the data points or blocks across object categories. We want to examine whether object has an effect on the perceived RP score required to perform a classification session. The dependent variable is "RP scores of each block" and the independent variable is "object type", which consists of two groups: 'Pedestrian', and 'Vehicle'. As a null hypothesis, it is assumed that there is no differences between the RP score values of these two object categories. The alternative hypothesis considers an existent difference between the RP score values of these categories. Use the *P*-value to determine whether any of the differences which are statistically significant against the null hypothesis. All of maximum *P*-values reported in Table I are less than $\alpha = 0.0001$ which is strong evidence against the null hypothesis. We have defined the significant level according to *P*-value as 'extremely significant', and 'statistically significant'.

III. Experimental Analysis

To collect objective data for validating the proposed method, visual and thermal video sequences on object detection performance were used. In this paper, the system experimented on our own collected videos comprising various environmental scenes. The data were examined the relationship between object recognition performance and three types of environmental scenes (i) the foggy morning (ii) the dust conditions (iii) the night time. There two modalities of images used: one for visual video sequences to capture the foggy, and dust; other one for infrared video sequences to capture the night scenarios. For visual scenes, we have used NIKON D5100 VR KIT Camera, and for infrared scenes, we have used FLIR E60 camera of 320 x 240 IR resolution with thermal sensitivity $<0.05^0$C. The experiment consists of three video clips: two NIKON system video clips, and one FLIR system video clip. The video clips were recorded on Agartala (India) city routes that represent a range of road types where pedestrian and vehicle sufferers that are attributable.

For experiment purpose, 64 frames were collected for each scenario after some frames were excluded due to the objects being very close to the border of the frames or not appearing in the frame. Of the 64 frames, 100 object images are used to extract features. The approximately equal amount of object images are collected from two types of objects i.e. pedestrian and vehicle. In order to check the effectiveness of our proposed features in classification performance, we compare four classifiers, namely, decision tree, linear discriminant analysis (LDA), naive bayes, and nearest neighbour. Fig. 6 shows the ROC curve by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The accuracy of the classification depends on how larger value of area under the ROC curves i.e. AUC. An area of 1 represents a perfect classification. From Fig. 6, it can be seen that the kNN classifier achieves a higher AUC values for foggy, and dust data. In case of infrared data (night vision), the naïve bayes classifier achieve higher AUC value.

TABLE I. LIST OF *P*-VALUES FROM FRIEDMAN TEST FOR EACH IMAGE SET WHICH CONSISTS OF A VEHICLE AND A PEDESTRIAN

Image Set	*P*-value	Significance
1	0.0481068278885200	Statistically Significant
2	3.77887111305396e-17	Extremely Significant
3	1.91855845259336e-17	Extremely Significant
4	0.0114120363860015	Statistically Significant
5	1.49382698267167e-37	Extremely Significant
6	0.0438049944147666	Statistically Significant
7	3.50969550511088e-13	Extremely Significant
8	1.79929047756131e-33	Extremely Significant
9	3.27797034180604e-10	Extremely Significant
10	9.48396288479757e-08	Extremely Significant

IV. Conclusion

The contribution of this paper is to develop a patch feature based method over similarity concept of image block system. To do so, objects are initially detected using an existing Gaussian Mixture Model (GMM) foreground subtraction algorithm. Then the Akinity and Liability based Representative (RP) Score features over patches of detected object are extracted which have been used for object type recognition. To classify the objects, we present a statistical Friedman test as well as analyses the performance over own collected video clips in circumstances like fog, dust, and night.

Acknowledgement

The work presented here is being conducted in the Biometrics Laboratory of Department of Computer Science and Engineering of Tripura University (A central university), Tripura, Suryamaninagar-799022.

References

[1] A. J. Lipton, H. Fujiyoshi, and R. S. Patil, "Moving target classification and tracking from real-time video," In Proceedings: Applications of Computer Vision (WACV'98), IEEEXplore, pp. 8-14, 1998.

[2] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," In Proceedings: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005.

[3] X. Wang, T.X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," In Proceedings: IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 2009.

[4] S. Walk, N. Majer, K. Schindler, and B. Schiele, "New features and insights for pedestrian detection," In Proceedings: IEEE Conference on Computer Vision and Pattern Recognition (CVPR'10), San Diego, CA, USA, 13–18 June 2010.

[5] D. Gerónimo, A. Sappa, D. Ponsa, A. López, "2D-3D based on-board pedestrian detection system," Comput. Vis. Image Underst., Vol. 114, pp. 583–595, 2010.

[6] A. González, Z. Fang, Y. Socarras, J. Serrat, D. Vázquez, J. Xu, and A. M. López, "Pedestrian Detection at Day/Night Time with Visible and FIR Cameras: A Comparison," Sensors, MDPI, Vol. 16, 820, 2016.

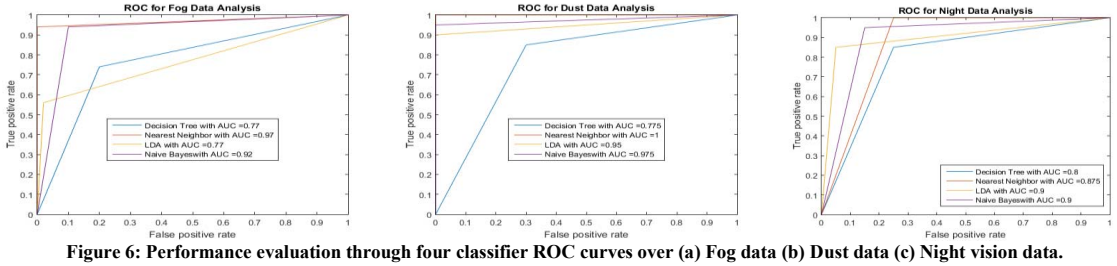

Figure 6: Performance evaluation through four classifier ROC curves over (a) Fog data (b) Dust data (c) Night vision data.