`

Detection of Cell Phone Usage in Restricted Areas

Abdul Rehman[1] Alam Zaib[2] Shoaib Azmat[2] Shahid Khattak[2]

[2]*Department of Electrical and Computer Engineering*

COMSATS University Islamabad (CUI), Abbottabad Campus

[1]{abdulrehman786302}@gmail.com [2]{alamzaib, shoaibazmat, skhattak}@cuiatd.edu.pk

***Abstract*— The objective of this work is to detect the cell phone and/or camera used by a person in restricted areas. The paper is based on intensive image processing techniques, such as, features extraction and image classification. The dataset of images is generated with cell phone camera including positive (with cell phone) and negative (without cell phone) images. We then extract relevant features by using classical features extraction techniques including Histogram of Oriented Gradients (HOG) and Speeded up Robust Features (SURF).The extracted features are then, passed to classifier for detection. We employ Support Vector Machine (SVM), Nearest Neighbor (K-NN) and Decision tree classifier which are already trained on our dataset of training images of persons using mobile or otherwise. Finally, the detection performance in terms of error rate is compared for various combinations of feature extraction and classification techniques. Our results show that SURF with SVM classifier gives the best accuracy.**

Keywords—Features extraction, Histogram of Oriented Gradients (HOG), Speeded up Robust Features (SURF), Support Vector Machine (SVM), k-Nearest Neighbor (k-NN), and Decision tree

I. INTRODUCTION

Video surveillance systems are extensive and common in many environments. Video surveillance has been a key in maintaining security at airports, banks, dams, and restricted places. Recently, nearly all government's agencies, businesses organizations, and even schools are turning toward video surveillance as a means to increase public security [1]. Typically a video surveillance system encompasses the following tasks: detection, recognition and tracking. The purpose of detection is to locate different things and objects from the images e.g., abandoned bags. Then, the position of object is estimated over time by tracking task. Lastly, the goal of recognition is to describe what is going on in the scene.

Recent works show that a large number of incidents have happened due to usage mobile of phone in restricted areas. Due to its impact on safety and property, several states and countries have validated regulations to ban mobile phone usage in restricted areas. The study in [2, 16] describes a precise detection of cell phone by different methods such as phoning gesture recognition and from the lips of driver by using car camera. In these works histogram of oriented gradients (HOG) was used for feature extraction and support vector machine (SVM) for classification. The only background considered is the internal view of the car. According to works in [7], cell phone can precisely detected by localizing the driver's face region within the front windshield image using the deformable part model (DPM). Different features were used such as bag-of-visual-words (BOW), vector of locally aggregated descriptors (VLAD) and Fisher vectors (FV). Although the problem of cell phone detection has been pursued by few researches in the context of driver using the cellphone, e.g., [2, 7], however, it has not been addressed in general context i.e. for any place where use of cell phone is prohibited.

In our work, we propose a method which can detect the cell phone precisely in any restricted area. Detecting cell phone from images is a challenging task owing to their variable position of handling and color of cell phone body. Hence, the foremost requirement is that of robust feature set that allows the cell phone to be distinguished cleanly, even in jumbled backgrounds under difficult illumination. In our work we use various combinations of state-of-the art feature extraction e.g., speeded up robust features (SURF), HOG, and classification techniques namely, SVM, k-nearest neighbor (KNN) and decision tree (DT). Our results show that SURF combined with SVM gives the best accuracy in cell phone detection.

II. PROPOSED METHODOLOGY

Here we consider the general flow of the various processes that are being performed in order to achieve the desired objective. The proposed approach consists of two main steps: training of classifier and testing of classifier. The general flow of cell phone detection from images is illustrated in Fig.1.

For training purposes, we built our own data set using different mobile phone cameras. The dataset images are resized to [150 150] pixels in order to reduce the processing time. The whole dataset is divided in two categories with the positive images and negative images for binary classification based on two different labels.

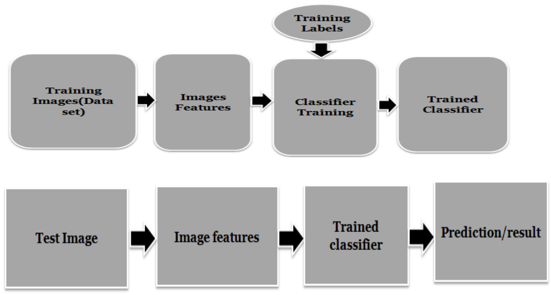

Fig.1. Flow chart for (a) the training of classifier and (b) the testing of classifier

Then, we extract features from whole dataset by applying different feature extracting techniques e.g. HOG, SURF. Based on the feature vector extracted from images, we train classifier as shown in Figure1a. Once the classifier training phase is completed, we take the given the test image and extract features in the form of feature vector which is passed to the trained classifier. Then, the classifier predicts the label of the test image to be either negative or positive as shown in fig.1.

III. FEATURES EXTRACTION AND CLASSIFICATION

Feature extraction is a low-level image processing operation. It is usually performed as the first operation on an image, and examines the whole image to see if there is a feature present in any location of the image. The two features that we have used are explained below:

A. Histogram Oriented Gradient (HOG)

HOG is a global feature extraction technique, used to detect objects in an image. The working of HOG technique is as follows:-, firstly, HOG calculates the directional changes (i.e., derivatives) in x and y direction of an image which are known as gradients [3]. Then it divides the whole image in to blocks of size (8x8) and then each block into cells of sizes (4x4), (8x8) and (16x16). The visualization plot of feature vectors from these cell sizes shown in Fig. 2.

The gradients are obtained by sliding the window, over the whole image. By sliding the window some parts of image overlaps and it gives better results. HOG splits every cell in to nine angular bins from 0^o to 180^o, of 20^o each [3]. Thus HOG uses angular bins to reduce 8x8=64 values to only nine values. Also the HOG normalizes the gradients for brightness and contrast changes. The normalized histogram is called block histogram and the set of normalized histogram is known as HOG descriptor.

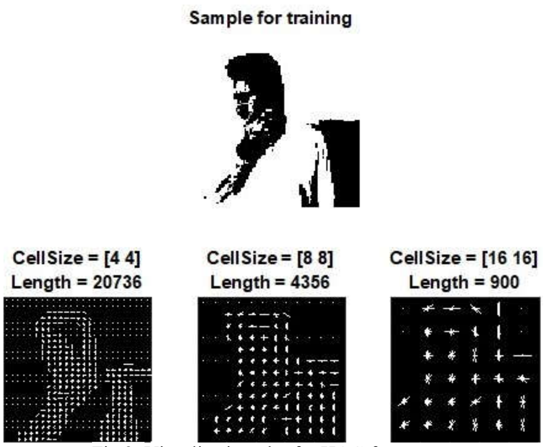

Fig.2. Visualization plot for HOG feature vectors

Selecting the best cell size for feature vector on the basis of visualization shows that a cell size of (16x16) does not encode much shape information, while a cell size of (4x4) encodes a lot of shape information but increases the dimensionality of the HOG feature vector significantly.

A good compromise is achieved by (8x8) cell size [3]. This cell-size setting encodes enough spatial information to visually identify the shape while limiting the number of dimensions in the HOG feature vector, which helps speed up the training. In practice, the HOG parameters should be varied with repeated classifier’s training and testing in order to identify the optimal parameter settings.

B. Speeded Up Robust Features (SURF)

SURF is a feature detector and descriptor. It is commonly used for the detection of objects. SURF is a speeded up version of scale invariant feature transform (SIFT). SURF is based on approximating logarithm with box filter [4]. The main purpose of using approximation of logarithm with a box filter is to easily obtain the convolution with box filter using integral images. For scale and location, SURF also depends on determinant of hessian matrix. In our work, we create a visual vocabulary, or bag of features, by extracting feature descriptors from representative images of each category [6]. This bag of features object defines the features, or visual words, by using the k-means clustering algorithm on the feature descriptors extracted from training dataset. The algorithm iteratively groups the descriptors into *k* mutually exclusive clusters. The resulting clusters are compact and separated by similar characteristics. Each cluster center represents a feature, or visual word. Overall, 500 visual words are extracted from training dataset by using bag of features and are encoded into feature vector.

The histogram of this feature vector is shown in Fig. 3. Another advantage of using SURF is that it provides greater scale invariance.

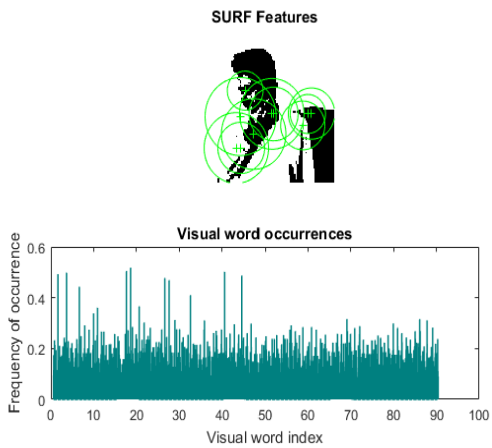

Fig.3. Result of Visual words and occurrences

C. Classifiers

Classification is the process of predicting the class of given test image. Classes are sometimes called as targets/ labels or categories. In our case we have only two classes; a person with mobile or without mobile. On basis of these two labels classifier predicts the label by analyzing the feature vector of the test image. In our work, three types of classifiers namely SVM, DT and K-NN have been used to classify the data set. A short description of classifiers is given below.

1. Support Vector Machine (SVM)

SVM is a machine-learning algorithm widely used for classification and regression purposes. SVM plots data items in features dimension having each feature value in suitable coordinate. The classification is done so as to separate two classes by hyper line. In our work, we used binary SVM classifier [8]. A 2-dimensional plot of features vector of HOG feature extraction is shown in Fig. 4. From the figure 4 a clear separation between both classes can be seen.

2. Decision Tree

A classification tree object represents a decision tree with binary splits for classification. The object contains the data used for training, so it can also compute reconstitution predictions. It has a flow chart or tree like structure as shown in fig. 6. Internal node denotes a test on an attribute. Branch represents an outcome of the test. Leaf nodes represent class labels or class distribution. In our work, we have two classes, so we used a binary tree in our work.

3. K-Nearest Neighbor (K-NN)

K-NN used nearest neighbor algorithm for both classification and regression. In K-NN classifier, k is the input parameter, which decides the closest neighbors in feature space. K-NN classifier works by votes of the nearest neighbors.

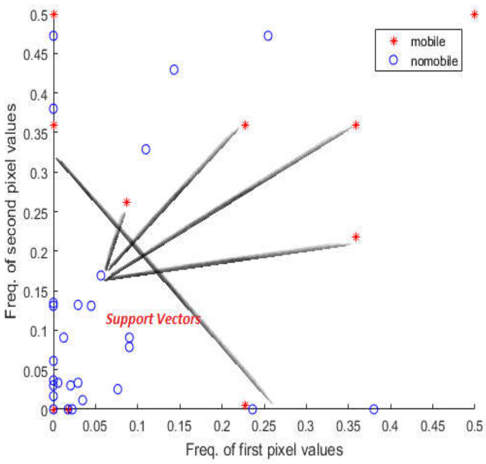

Fig.4. Two-dimension scatter plot showing separation of two classes

Nearest neighbor uses distance functions like Euclidean, Manhattan, Minkowski functions [5]. The distance functions are as follows:

$$\text{Euclidean} \sqrt{\sum_{i=1}^{k} (x_i - y_j)^2} \qquad (1)$$

$$\text{Manhattan} \sum_{i=1}^{k} |x_i - y_j| \qquad (2)$$

$$\text{Minkowski} \sum_{i=1}^{k} [((|x_i - y_j|)^q]^{\frac{1}{q}} \qquad (3)$$

IV. RESULTS

The software part of our work is based on MATLAB R2018, and the hardware specifications for our system are Intel core i5-3320M CPU @ 2.4 GHz processor with 6GB RAM. Dataset contain 1000 images that are equally divided as positive (500 with mobile phone) and negative images (500 without mobile phone). As shown in fig. 5, the SVM classifier accurately predicts the label for a negative test image. Here, a feature vector is extracted from test image based on cell size (8x8) by using HOG which is then passed to SVM classifier for prediction. As seen in fig. 6, SVM classifier accurately predicts the label on a positive test image.

In fig. 7 SVM classifier accurately predicts the result on positive test image. In this case feature vector is extracted from the test image by using SURF based visual words. Then feature vector is passed to SVM classifier for prediction. In fig. 8 the SVM classifier accurately predicts the label on negative image.

Fig.5. Testing a positive image using HOG

Fig.6. Testing a negative image using HOG

We have also used some other classifiers such as Decision tree and K-NN with aforementioned features extraction techniques, but they were less accurate as compared with SVM. Besides, there are some issues like background light intensity changes and the redundant features exist in image. We implemented different connected components algorithm to get rid of their redundant features.

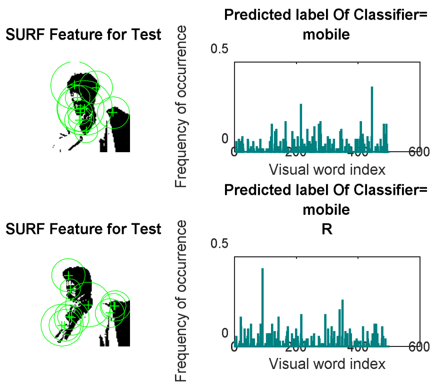

Fig.7. Testing of image using SURF

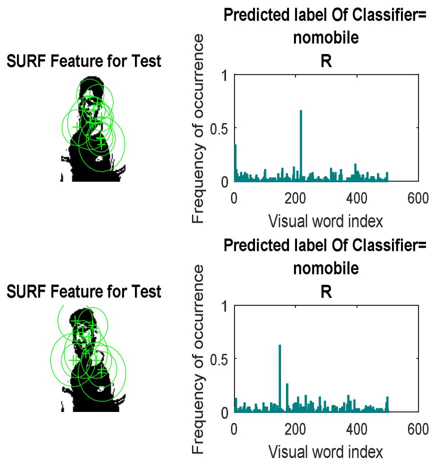

Fig.8. Testing of image using SURF

Then, the combination of feature extraction with classifier, the obtained results are shown in Table 1. The highest accuracy of 91% is achieved for the SURF-SVM combination of features and classifier.

Classifiers	**Feature Detection Techniques**	**Average Accuracy**	**Running Time**
SVM	HOG	84%	18.922sec
Decision Tree	HOG	56%	34.018sec
K-NN	HOG	61%	23.858sec
SVM	SURF	91%	150.23sec

Table 1: Summary of combination feature extraction and classification techniques

V. CONCLUSION

In this work, we used various feature extraction and classification techniques for detecting cell phone in restricted areas. By comparing different classifiers, we have seen that the accuracy achieved from SURF with SVM classifier has the highest value. However, turning to other side, the speed achieved from HOG with SVM classifier was high but at the cost lower accuracy.

REFERENCES

[1] Shukla, Vipin & Singh, Gaurav & Shah, Pratik. (2013). Automatic Alert of Security Threat through Video Surveillance System.

[2] “Detecting Driver Use of Mobile Phone Based on In-car Camera” by Dan Wang, Mingtao Pei and Lan Zhu publish in 2014 Tenth International Conference on Computational Intelligence and Security.

[3] Histogram of oriented gradient for human detection by Navneet dalal, Bill Triggs publish in IEE computer society conference on computer, 2005, 10.1109/CVPR.2005.177

[4] Herbert Bay , Andreas Ess , Tinne Tuytelaars , and Luc Van Gool . Speeded-Up Robust Features (SURF).2006.

[5] J. Maillo et al., "kNN-IS: An Iterative Spark-based design of the k-Nearest Neighbors classifier for big data", *Knowledge-Based Systems*, vol. 117, pp. 3-15, 2017.

[6] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray.Visual categorization with bags of keypoints. In Workshop on statistical learning in computer vision, ECCV, volume 1,pages 1–22, 2004.

[7] Artan, Yusuf & Bulan, Orhan & Loce, Robert & Paul, Peter. (2014). Driver Cell Phone Usage Detection From HOV/HOT NIR Images. 10.13140/2.1.4011.8409.

[8] Srivastava, Durgesh & Bhambhu, L. (2010). Data classification using support vector machine. Journal of Theoretical and Applied Information Technology. 12. 1-7.

[9] Kishi, Rodrigo & Trojahn, Tiago & Goularte, Rudinei. (2018). Temporal Video Scene Segmentation By Fused Bags-of-Features. 173-180. 10.1145/3243082.3243109.

[10] Yang, F & Lu, H & Zhang, W & Yang, G. (2012). Visual tracking via bag of features. Image Processing, IET. 6. 115-128. 10.1049/iet-ipr.2010.0127.

[11] Siraç Özerdem, Mehmet & BAMWENDA, Julius. (2019). Recognition of static hand gesture with using ANN and SVM. DÜMF Mühendislik Dergisi. 10. 561-568. 10.24012/dumf.569357.

[12] Joseph, Siji & Pradeep, Arun. (2017). Object Tracking using HOG and SVM. International Journal of Engineering Trends and Technology. 48. 321-325. 10.14445/22315381/IJETT-V48P257.

[13] Li, Zhenbo & Niu, Bingshan & Peng, Fang & Li, Guangyao & Yang, Zhaolu & Wu, Jing. (2018). Classification of Peanut Images Based on Multi-features and SVM. IFAC-PapersOnLine. 51. 726-731. 10.1016/j.ifacol.2018.08.110.

[14] Fang, Yuming & Ding, Guanqun & Yuan, Yuan & Lin, Weisi & Liu, Haiwen. (2018). Robustness Analysis of Pedestrian Detectors for Surveillance. IEEE Access. PP. 1-1. 10.1109/ACCESS.2018.2840329.

[15] D. G. Lowe, Object Recognition from local scale–invariant features, ICCV (International Conference on Computer Vision), 1999.

[16] Berri, Rafael & Silva, Alexandre & Parpinelli, Rafael & Girardi, Elaine & Arthur, Rangel. (2014). A Pattern Recognition System for Detecting Use of Mobile Phones While Driving. VISAPP 2014 .