An Unsupervised Object Tracking And Detection Using Otsu Algorithm

Sundaresan K, Dinesh Kumar T

[1](Department of Computer Science and Engineering, P.A. College of Engineering and Technology, Coimbatore,Tamilnadu, India, **dineshpacet@gmail.com**
)

[2](Department of Computer Science and Engineering, P.A. College of Engineering and Technology, Coimbatore,Tamilnadu, India, **sundar.pacet@gmail.com**)

Abstract— *Object detection is one of the major goals in computer vision that deals with detecting instances of semantic objects of a certain class. In Video acquisition takes video as an input and splits the number of frames from the given video. Then it removes the noise from the frame using pre-processing and segment the moving object using Otsu algorithm. Region Of Interest can be calculated to perform further analysis for image identification and recognition. Histogram of Oriented Gradients is a feature descriptor used for object detection and using Histogram of Oriented Gradients all the humans in the video can be detected. Random forest is a learning method for classification and regression and by matching the characteristic of the image using random forest classifier particular human can be detected in video. Finally, track the moving object in original video.Object can be detected and tracked based on Otsu algorithm and Random forest. Finally, it detects the image in the video and track the moving object in original video. Based on the characteristic matching of the two images tracking can be performed. Object tracking and detection can be implemented using Matrix Laboratory.*

Keywords— Image Processing, MATLAB, OTSU Segmentation, HOG Extraction

1. Introduction

Image processing is the study of any algorithm that takes an image as input and returns an image as output. Image processing is a method to perform some operation on an image, in order to get an enhanced image or to extract some useful information from it. Image processing is processing of images using mathematical operations by using any form of signal processing. The input can be a series of images, or a video. The output of image processing may be either an image or a set of characteristics. Image processing system includes treating images as two dimensional signals during applying already set signal processing methods. Image processing is among rapidly growing technologies. It forms core research area within engineering and computer science disciplines too. Image processing basically includes the following three steps:

1. Importing the image via image acquisition tools
2. Analyzing and manipulating the image
3. Output result can be altered image or report that is based on image analysis

Common examples of Image processing include

1. Noise Removal
2. Contrast Adjustment
3. Edge Detection
4. Region Detection

1.1 Image Analysis and Computer Vision

Image Analysis involves extracting meaningful information from an image

1. Image segmentation
2. Image matching and comparison
3. Medical diagnosis from an image

Computer Vision strives to emulate the human visual system and interpret our 3-D world from 2-D images or video.

1. Object recognition
2. Motion tracking
3. 3-D shape from multiple 2-D images.

2. OBJECT TRACKING AND DETECTION

The proposed system can be designed to detect and track the object in video by matching the key features of an input image. In the input video noise can be removed and segmented using Otsu algorithm. Region of interest(ROI) can be calculated and all the humans in the video can be detected using Histogram of Oriented Gradients(HOG) feature extraction. By matching the key features of an image with the video particular human in the image can be detected.

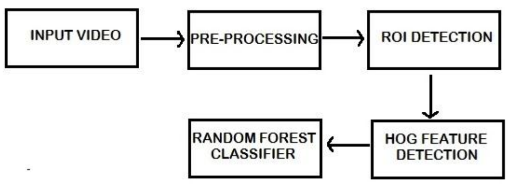

Figure 1 Block diagram

The block diagram of an unsupervised object tracking and detection using Otsu algorithm is shown in figure 1.

2.1 Video acquisition

Video acquisition is a process of getting an input video. Video acquisition is the process of converting an analog video signal produced by a video camera or digital video disc player to digital video. The resulting digital data are computer files referred to as a digital video stream, or more often, simply video stream.

2.2 Pre-processing

The aim of Pre-processing is an improvement of image data that suppress unwanted image data distortions or enhance the image features important for the further processing. If Pre-processing aims to correct the degradation in the image, the nature of a priori information is important

1. Knowledge about objects is not available in advance it can be estimated Knowledge about the nature of the degradation. only very general properties of the degradation are assumed.
2. Knowledge about the properties of the image acquisition device, the nature of noise is sometimes known.
3. Knowledge about objects that are searched for in the image, may simplify the Pre-processing very considerably. If knowledge during the processing.

3. OTSU SEGMENTATION

Segmentation is often considered to be the first step in image analysis. The purpose is to subdivide an image into meaningful non overlapping regions, would be used for further analysis. It is hoped that the regions obtained correspond to the physical parts or objects of a scene 2-D represented by the image 3-D. In general, autonomous segmentation is one of the most difficult tasks in digital image processing. All the image segmentation methods assume that:

1. The intensity values are different in different regions
2. For each region, it represents the corresponding Object in the scene, the intensity values are similar.

Otsu thresholding chooses the threshold to minimize the infraclass variance threshold black and white pixels. The Otsu method is used to calculates the optimum threshold separating those two class so that their combined spread is minimal. The Otsu method searches for the threshold that minimizes the intra-class variance.

$$\sigma_w^2(t)=\omega_1(t)\sigma_1^2(t)+\omega_2(t)\sigma_2^2(t) \qquad (1)$$

The above equation 1, ωi are the probabilities of the two classes separated by a threshold t and σi2 is the variance of these classes. Otsu shows that minimizing the intra-class variance is the same as maximizing inter-class variance, equation.

$$\sigma_b^2(t)=\sigma^2-\sigma_\omega^2(t)=\omega_1(t)\omega_2(t)[\mu_1(t)-\mu_2(t)]^2 \quad (2)$$

The above equation 2, expressed in terms of class probabilities ω_i and class means μ_i which in turn can be updated iteratively.

The algorithm steps are:

1. Compute the histogram and probabilities of each intensity level
2. Initialize $\omega_i(0)$ and $\mu_i(0)$
3. Step through all threshold values $t = 1$ …. to maximum intensity.
4. Update $\omega_i(0)$ and $\mu_i(0)$
5. Compute the maximum $\sigma_b^2(t)$, which corresponds to the desired threshold.

3.1 Region Of Interest

A region of interest is a subset of an image or a dataset identified for a particular purpose. The dataset could be any of Waveform or 1-D dataset. The ROI is a time or frequency interval on the waveform. A region of interest is a portion of an image that can be used to filter or perform the operations on the image. ROI can be defined by creating binary mask. In the mask image, the pixels that define the ROI are set to 1and all other pixels set to 0.

3.2 HOG Feature Extraction

The Histogram of Oriented Gradients (HOG) is a feature descriptor used in computer vision and image processing for the purpose of object detection.

Gradient computation

Gradient computation is carried out for each cell individually.

$$\nabla f=\nabla\left[\begin{matrix}\frac{\partial f}{\partial x}\\ \frac{\partial f}{\partial y}\end{matrix}\right] \qquad (3)$$

The above equation 3, $\frac{\partial f}{\partial x}$ is the derivative with respect to x and $\frac{\partial f}{\partial y}$ is the derivative with respect to y

Orientation binning

To account for changes in illumination and contrast, the gradient strengths must be locally normalized, which requires grouping the cells together into larger, spatially connected blocks. The HOG descriptor is then the concatenated vector of the components of the normalized cell histograms from all of the block regions. Two main block geometries exist: rectangular R-HOG blocks and circular C-HOG blocks.The vector angle is

$$\varphi_{ij}=\tan^{-1}\left[\frac{y_{ij}}{x_{ij}}\right] \qquad (4)$$

The above equation 4, x_{ij} and y_{ij} are vertical and horizontal of the gradient vector.

$$m_{ij}=\sqrt{x_{ij}^2+y_{ij}^2} \qquad (5)$$

the above equation 5, is length of the gradient vector.

Block normalization

Rather than normalize each histogram individually, the cells are first grouped into blocks and normalized based on all histograms in the block. This block normalization is performed by concatenating the histograms of the four cells within the block into a vector with 36 components.

4. CASCADED RANDOM FOREST CLASSIFIER(CART)

Random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of the individual trees. Random decision forests correct for decision trees' habit of overfitting to the training set.

Random Forest is an ensemble learning method based on Breiman's bagging idea. A Random Forest consists of CART-like decision trees that are independently constructed on a bootstrap sample. Compared to other ensemble learning algorithms. The boosting that build a flat tree structure of decision stumps, a Random Forest uses an ensemble of decision trees, is multi-class capable and has preferable characteristics

1. Similar or better accuracy than AdaBoost.
2. Robust to noise and outliers.
3. Faster training than bagging or boosting.
4. Useful internal estimates: error, strength, correlation and variable importance. A tree is grown using following algorithm
5. Choose n tree samples with M variables from N training samples at random.
6. The remaining samples are used to calculate the out-of-bag error.
7. At each node specify m 0 try << M variables at random based on best split.
8. Completely grow the tree without pruning.

A completed Random Forest consists of several classification trees $1 \leq t \leq T$ in the class probabilities, estimated by majority voting, are used to calculate the sample's label y(x) with respect to a feature vector x. Random Forest Classifier is ensemble algorithm. In next one or two posts exploring such algorithms. *Ensemble algorithms* combines more than one algorithms of same or different kind for classifying objects. For example, running prediction over Naive Bayes, SVM and Decision Tree and then taking vote for final consideration of class for test object.

Random forest classifier creates a set of decision trees from randomly selected subset of training set. It then aggregates the votes from different decision trees to decide the final class of the test object. Random Forests grows many classification trees. To classify a new object from an input vector, put the input vector down each of the trees in the forest. Each tree gives a classification, and say the tree "votes" for that class. The forest chooses the classification having the most votes.

Each tree is grown as follows

1. If the number of cases in the training set is N, sample N cases at random - but with replacement, from the original data. The sample will be the training set for growing the tree.
2. If there are M input variables, a number m<<M is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. The value of m is held constant during the forest growing.
3. Each tree is grown to the largest extent possible. There is no pruning.

In the original paper on random forests, it was shown that the forest error rate depends on two things

1. The correlation between any two trees in the forest. Increasing the correlation increases the forest error rate.
2. The strength of each individual tree in the forest. A tree with a low error rate is a strong classifier. Increasing the strength of the individual trees decreases the forest error rate.

Reducing m reduces both the correlation and the strength. Increasing it increases both and in between is an "optimal" range of m - usually quite wide. Using the oob error rate a value of m in the range can quickly be found. The is the only adjustable parameter to random forests is sensitive.

The features of random forest are

1. It runs efficiently on large data bases.
2. It can handle thousands of input variables without variable deletion.
3. It gives estimation of variables are important in the classification
4. It generates an internal unbiased estimate of the generalization error as the forest building progresses.
5. It has an effective method for estimating missing data and maintains accuracy in a large proportion of the data are missing.
6. It has methods for balancing error in class population unbalanced data sets.
7. Generated forests can be saved for future use on other data.
8. Prototypes are computed that give information about the relation between the variables and the classification.
9. It computes proximities between pairs of cases that can be used in clustering, locating outliers, give interesting views of the data.
10. The capabilities of the above can be extended to unlabeled data, leading to unsupervised clustering, data views and outlier detection.
11. It offers an experimental method for detecting variable interactions.

5. RESULT

The below Figure 2 represents input video. The video consists of many people and the desired people can be detected and tracked.

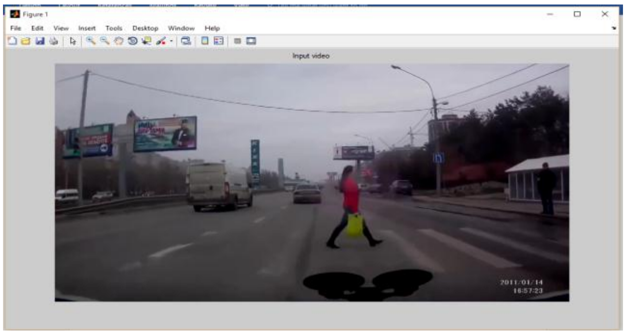

Figure 2. Input video

Then from the image calculate the Region Of Interest (ROI) detection, used to filter or perform some operation on image. Chance for the multi-image arrival, hence apply the Histogram of Oriented Gradients is used to detect all the humans in the video. The below figure 3 shows the implementation of HOG methodologies in the given image.

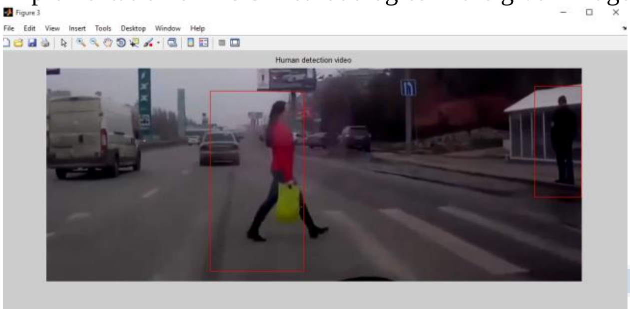

Figure 3 Human detection video

The above Figure 6 shows the tracking of particular person. By using random forest, the particular person is detected by matching the characteristic matching of the input image.

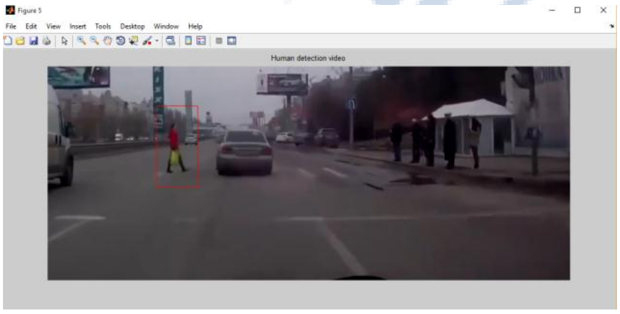

Figure 4 Tracking the image in video

Finally, the above Figure 5 shows the output video that particular person is detected with respect to the input image.

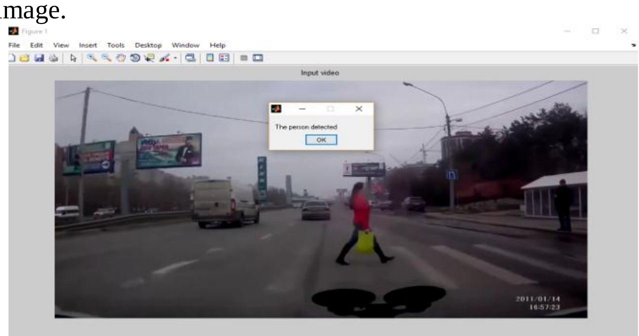

Figure 5 Output video

As final point of view , the desired person was find out from the given video through Ostu algorithm.

6. CONCLUSION

The proposed method is used for object detection and tracking in video surveillance and management. It introduces images termed as vector-featured images that record moving regions in videos. By utilizing these images, to detect moving objects and accumulate unmoving regions directly obtained from video. The generation of vector featured images and the effectiveness of moving object detection and tracking using the featured images in several experiments. The algorithm is able to detect moving objects without prior knowledge of the shape of the object.

The proposed method is further improved to detect and track the object for given skin and color. The performance of the object recognition system depends on the features used and the classifier employed for recognition. To propose a novel feature extraction method for extracting global features and obtaining local features from the region of interest.

7. REFERENCE

[1] Babu, R. et al. (2016) 'A survey on compressed domain video analysis techniques', vol 75(2), pp. 1043-1078.

[2] Babu, R. et al. (2004) 'Video object segmentation: A Compressed domain approach', vol 14(4), pp. 462-474.

[3] Bjontegaard, G. et al. (2003) 'Overview of the H.264|AVC Video Coding Standard', vol (13), pp. 560-276.

[4] David power and Weisheng Li (2017) 'Multiple object tracking using motion vectors from compressed video'.

[5] Favalli, L. et al. (2000) 'Object tracking for retrieval applications in MPEG-2', vol 10(3), pp. 427-432.

[6] Fei, W. and Zhu, S. (2010) 'Mean shift clustering - based moving object segmentation in the h.264 compressed domain', vol 4(1).

[7] Kas, C. and Nicolas, H. (2009) 'An approach to trajectory estimation of moving objects in the H.264 Compressed domain', pp. 318-329.

[8] Lan Dong, et al. (2006) 'Object Tracking in Compressed Video with Confidence Measures', pp. 753-756.

[9] Ling-Yu Duan, et al. (2003) 'Robust moving video object segmentation in the MPEG compressed domain', vol (3), pp. 933.

[10] MacQueen, J. B. (1967) 'Some Methods for classification and Analysis of Multivariate Observations', pp. 281-297.

[11] Plakas, K. and Trucco, E. (2006) 'Video Tracking: A Concise Survey, Oceanic Engineering', 31(2), pp. 520-576.

[12] Yokoyama, T. et al. (2009) 'Motion Vector Based Moving object Detection and Tracking in the MPEG Compressed Domain', pp. 201-206.

[13] You, W. et al. (2012) 'Real-time detection and tracking of multiple objects with partial decoding in H.264|AVC bitstream domain', vol (7244), pp. 72440D-72440D-12