# EDGELET BASED HUMAN DETECTION AND TRACKING BY COMBINED SEGMENTATION AND SOFT DECISION

[1]Ms. K.Bhuvaneswari, [2]Prof.H.Abdul Rauf,
[1]Lecturer, Department of CSE, VLB Janakiammal college of Engineering and Technology
[2]Dean, Department of CSE, VLB Janakiammal college of Engineering and Technology
Email. bhuvana_k4@yahoo.com Ph.9952404622

Abstract- Human detection and tracking from video stream is important for many applications. Existing detection methods were based on skin-color segmentation or gray level face detection to detect the human. In this paper human detection is based on a silhouette oriented feature called edgelet feature. The system automatically detects and tracks possibly partially occluded humans from a single camera, which is stationary. The discriminative classifiers of objects of a known class are learnt and applied to the video sequence frame by frame. The output of the detection module is a soft decision which consists of a set of detection responses of different confidence levels. The combined detection responses provide the observations used for tracking. The responses of a multiple view detection system are taken as the observation of the human hypotheses. Trajectory initialization and termination rely on the confidences computed from the detection responses. Finally the human is tracked by mean shift style tracker. The system tracks human with interobject and scene occlusions with static or non-static backgrounds. Edgelet features are suitable for human detection as they are relatively invariant to clothing differences.

Keywords: edgelet, human detection, human tracking, segmentation, trajectory

## I. INTRODUCTION

Object detection and tracking is a fundamental problem of computer vision research and very important for many real-life applications, such as visual surveillance and human computer interaction. There are three main types of tracking methods. First,.2-D region tracking algorithms focus on the problem of tracking after initialization by assuming that the position of the object is given in the first frame. The algorithms first try to extract some characteristic properties of the object, could be color or salient, from the initialization and then use these properties to track the object in the new frames.

Second, the moving blob tracking algorithms focus on the detection and tracking of moving objects. First motion segmentation is applied and then the moving blobs are tracked based on their appearance and motion [11]. The motion segmentation algorithms are sensitive to abrupt illumination changes, shadows, and reflections.

Third, detection based tracking methods attempt to overcome these limitations by using discriminative methods. Recently, fast development of object detection techniques has resulted in many promising methods for detection of particular
object classes. These object detectors produce good observations for the detection based tracking algorithms. By associating the frame by frame object detection

responses, an idea of when to start or terminate an object trajectory is found.

Consequently, both the detection of people in individual frames as well as the data-association between people detections in different frames is highly challenging and ambiguous. To address this, temporal coherency is exploited, to extract edgelets from a small number of consecutive frames. From the edgelets, models of the individual people are built. As any single person might be detectable only for a small number of frames the extraction of people and edgelets has to be highly robust. At the same time the extracted model of the individual has to be discriminative enough in order to enable tracking and data-association across long periods of partial and full occlusions.

## II.BACKGROUND

The previous efforts use skin-color segmentation [1, 2, 3, 4, and 5] or gray-level face detection [2, 6] to find the humans. Some others detect humans by background subtraction [3, 4, and 7]. However the applications of these methods are limited, since the underlying assumptions are not always valid. For example, faces are not visible from rear view. Template matching is used as a shape constraint on color-based segmentation. The two steps process in an iterative manner.

The previous work [9], proposed a human detection method by combining body part detection

responses. For each part a cascade detector is learned by boosting edgelet features based weak classifiers.

## III.IMPLEMETATION

### 3. SYSTEM MODEL

System model contains two modules Detection Module, Tracking Module. Detection module combines responses from various classifiers.

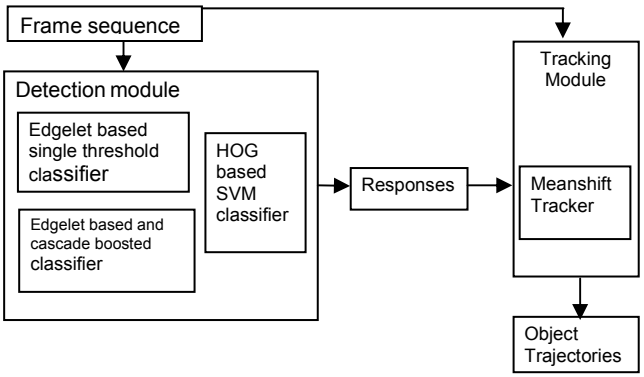


**Fig .1. System diagram**

Fig.1. shows the detection module, tracking module and various classifiers are used to detect the partially occluded human in the meeting room. Level1, Level 2, Level3 responses are obtained from HOG based SVM classifier, Edgelet based boosted classifier, and Edgelet based single threshold classifier. Tracking Module initializes and terminates the trajectory and growth is also a main concern.

### 3.1. Edgelet Features

An edgelet is a short segment of a line or a curve. Denote the positions and normal vectors of the points in an edgelet, E, by {ui }k i=1 and {nE i }k i=1, where k is the length of the edgelet shown in Fig. 2. for an illustration.

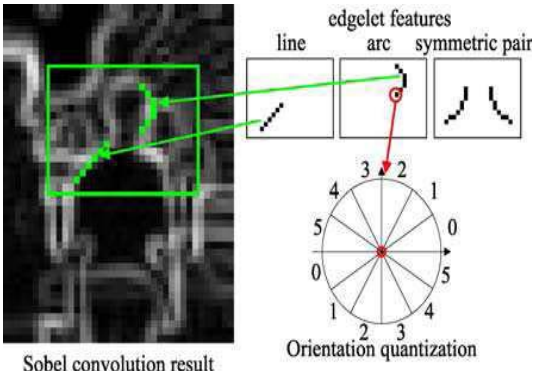


**Fig. 2. Edgelet features**

Given an input image I, it is denoted by MI (p) and nI (p) the edge intensity and normal at position p of I. The affinity between the edgelet E and the image I at position w is calculated by

$$f(E; I,w) = 1/k \sum_{i=1}^{k} MI(ui + w)|\, nI(ui + w), nE I| \quad (1)$$

Where, ui in the above equation is in the coordinate frame of the sub-window, and w is the offset of the sub-window in the image frame. The edgelet affinity function captures both intensity and shape information of the edge.

#### 3.1.1. Object Segmentation

A probabilistic formulation for the segmentation problem is described. At starting point, a refined object hypothesis h= (on, x) taken. Based on this hypothesis, the object is segmented from the background. Whether a certain image pixel p is Fig or ground, is known from the given the object hypothesis. The influence of a given patch e on the object hypothesis can be expressed as

$$P(e,l|\, on,x) = P(on,x|e,l)p(e,l)/p(on,x) = \Sigma f\, P(on,x||l,l)p(l|e)p(e,l)/p(P(on,x|) \quad (2)$$

where the patch votes P(on,x|e,l)p(e,l) are obtained from the codebook. Given these probabilities, information is obtained by a specific pixel by marginalizing over all patches that contain this pixel:

$$P(figure|\, on,x) = \Sigma\, P(p=Figure|on,x,e,l)p(e,l|\, on,x)\, P(p=Figure|on,x,e,l) \quad (3)$$

denoting patch-specific segmentation information, which is weighted by the influence p(e,l| on,x) the patch has on the object hypothesis. Again, patches are resolved by resorting to learned patch interpretations I stored in the codebook:

$$P(p=Figure|on,x) = \Sigma\, \Sigma\, P(p=figure|on,x,e,l)p(e,,l,l|\, on,x) \quad (4)$$

$$pE(e,l)\, I = \Sigma\, \Sigma\, P(p=fig|on,x,l,l)\, pE(e,l)\, I\, P(on,x||l,l)p(l|e)p(e,l)/p(P(on,x|) \quad (5)$$

This means that for every pixel, we build a weighted average over all segmentations stemming from patches containing that pixel. The weights correspond to the patches respective contributions to the object hypothesis. For the ground probability, the result is obtained in an analogue fashion.

### 3.2. DETECTION MODULE

Detection module contains various classifiers .First low cost classifier is applied and the other classifiers are applied. Responses from the various classifiers given to the tracking module to form object trajectories.

#### 3.2.1. Edgelet Based Boosted Classifier

Edgelets are one type of local shape features. These features are suitable for human detection as they are relatively invariant to clothing differences, unlike gray level or color features used commonly for face detection. One edgelet can be seen as a small edge template. For each edgelet in a big feature pool, one weak classifier is built to distinguish objects from background.

Then a boosting algorithm [4] is used to learn tree structured classifiers for multiview objects. Each node of the tree is an ensemble classifier with the cascade decision strategy [11] which makes the detection process efficient.

During the training of the tree classifier, a cascade decision strategy is learned for each branch of the tree. For each node along a branch, a threshold is chosen to accept most positive samples while rejecting as many negative samples as possible. When a target overall false alarm rate is reached, the training procedure is terminated. Because of the cascade decision strategy, most sub windows examined in the image can be discarded by computing only the first few features in the tree. Although this is an efficient strategy, it is aggressive in terms of discarding negative samples.

To obtain a prediction of high detection rate, ignore the series of thresholds of the cascade decision strategy and learn an overall threshold for each branch of the tree. The threshold is chosen to accept all positive samples and reject as
many negative samples as possible. However, as the decision is made at the leaf nodes of the tree, all features in the classifier need to be computed to classify one sub-window. The classification results of the boosted classifier with the cascade decision strategy are the second level responses.

### 3.2.2. HOG Based SVM Classifier

During training, the false alarms and the successful detected samples from the cascade boosted classifier are collected and used to learn the SVM classifier. During detection, first the cascade boosted classifier is applied to the whole image, and then the positive responses are sent to the SVM classifier for verification.

## 3.3. TRACKING MODULE

Tracking module first performs data association to find the responses from the various classifiers and then associate it with the hypothesized objects. Tracking algorithm has three main components: trajectory initialization, growth and termination.

### 3.3.1. Data Association

The task of data association is to match the detection responses with the human hypotheses. Suppose at time t, have n hypotheses H1 ...Hn whose predictions at time t+1 are rt+1, 1 …. rt+1,n , and at time t+1 have m responses are st+1,1 …. st+ 1,m. First compute an m × n affinity matrix A of all (rt+1,i , st+1,j) pairs and ( i.e A(i,j) ) is an affinity matrix score between rt+1,i and st+1,j Then in each step, the pair, denoted by ( k , l) with the largest affinity is taken as a match and the k th row and the l th column of A are deleted. This procedure is repeated until no more valid pairs are available.

One detection response is represented by a 4-tuple, {p, s, c, f}, where p is the image position, s is the size, c is an appearance model, and f is a real-valued classification confidence. But here c is a color histogram. The classification confidence f is the weighted sum of all weak classifiers outputs for the boosted classifier or the distance to the classification boundary for the SVM classifier. The object hypotheses have the same representation as the detection responses. Similarly [3], the affinity between two detection responses, r1 and r2 are defined by

$$A(r1,r2)=Apos\ (p1,p2)\ Asize\ (s1,s2)\ Aappr\ (c1,c2) \quad (6)$$

Where Apos, Asize and Aappr are affinity measure based on position, size, and appearance respectively. Apos and Asize are modeled by Gaussian functions, and Aappr is modeled by Bhattachayya distance.

$$Apos\ (p1, p2) = \gamma pos \exp(-(x1-x2)2 / \sigma 2\ x) \exp(-(y1-y2)2 / \sigma 2\ y) \quad (7)$$

$$Asize\ (s1, s2) = \gamma size \exp(-(s1-s2)2 / \sigma 2\ s) \quad (8)$$

$$Aappr\ (c1, c2) = B\ (c1, c2) \quad (9)$$

Where B (c1, c2) is the Bhattachayya distance between two color histograms γpos and γsize.

### 3.3.2. Trajectory Initialization

The basic idea of our initialization strategy is to start a trajectory when enough evidence is collected from the detection responses. Due to the correlation between neighboring frames, if the detector outputs a false alarm at certain position in one frame, the probability is high that a false alarm will appear around the same position in the next frame. This is called persistent false alarm problem. However, suppose we have found T consecutive responses, {r1 …. rT} corresponding to one object hypothesis H, still the probability of H being a false alarm should be an exponentially decreasing function of T. Then model it as $e^{-\lambda init \sqrt{T}}$. The confidence of initializing a trajectory for H is defined by

$$InitConf\ (H; r1 \ldots. rT) = 1/(T-1) \sum_{t=1}^{T-1} A\ (rt+1, st+1) \cdot (1- e^{-\lambda init \sqrt{T}}) \quad (10)$$

The first term in the left side of equation is the average affinity of the T responses, and the second term is based on the detector's property. Our trajectory initialization strategy is, if InitConf (H) is larger than a threshold, TH, a trajectory is started from H, and H is called a confident trajectory. Otherwise, H is called a potential trajectory. A trajectory hypothesis H is represented as a 3-tuple, {rt 1….T, D, C} where {rt} is a series of responses, C is the appearance model, and D is a dynamic model. C is the average of the appearance models of all detection responses, and D is modeled by a Kalman filter for constant speed motion.

### 3.3.3. Trajectory Growth

After a trajectory is initialized, the object is tracked by two strategies, data association and meanshift tracking. For a new frame, first for all existing hypotheses, their corresponding responses are considered. If there is a new response matched with a hypothesis H, then H grows by data association, otherwise a meanshift tracker [7] is applied. The basic idea of meanshift is to track a probability distribution. In the appearance model, C, the dynamic model, D, are combined and the detection confidence, f to build likelihood map which is then fed into the meanshift tracker. Let P appr (u) be the appearance probability map. As C is a color histogram, Pappr (u) is the bin value of C, shown in Fig.4.A dynamic probability map, P (u), where u represents the image coordinates, is calculated from the dynamic model D shown in Fig 5. Suppose, at one frame the set of the original responses of the edge detector is {rj}, then the detection probability map Pt (u) is defined by

$$Pt(u) = \sum j: u \xi Reg(rj)\, fj + ms \qquad (11)$$

Where Reg (rj) is the image region corresponding to rj, fj, is a real-valued confidence of rj, and ms is a constant corresponding to the missing rate. The original response is used here, because of possible errors in the clustering algorithm.

**Fig .3.Appearance probability map**

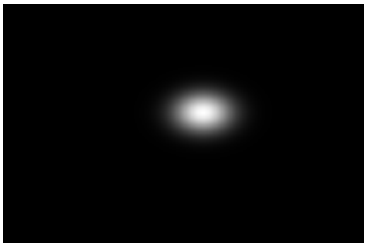

**Fig .4.Dynamic probability map**

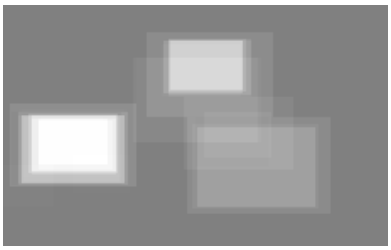

**Fig .5. Detection probability map**.

### 3.3.4. Trajectory termination

The unmatched first level detection responses are used to initialize new trajectories. If at one frame, a first level detection response does not correspond to any existing trajectory, then start a new potential trajectory. If in the succeeding consecutive frames, first level responses are matched then, compute an initialization confidence [5]. If the confidence is larger than a threshold, it is a confident trajectory. The trajectory termination criterion is similar to that of initialization. If in consecutive frames it is not possible to find the matched first or second level responses for one object hypothesis, and then compute a termination confidence.

If in consecutive T time steps, no detection responses are found for an object H, compute a termination Confidence of H by

$$EndConf(H; r1 \ldots. rT) = 1/(T-1) \sum A(rt+1, st+1) . (1- e\, end\sqrt{T}) \qquad (12)$$

Here that the responses rt are obtained from the meanshift tracker, not from the detection module. If EndConf (H) is larger than a threshold, St, hypothesis H is terminated, then it is a dead trajectory, otherwise it is an alive trajectory Initialization and termination strategies are very similar to that of [6], here detection responses of multiple confidence levels are used.

## IV.EXPERIMENTATION AND RESULTS

Here some experimental results are presented. Detection and tracking modules are implemented separately. Frame is acquired from the real time video input connected to the system. In detection module, humans are detected by finding the edges. Various edge detectors are used to improve the classification efficiency.

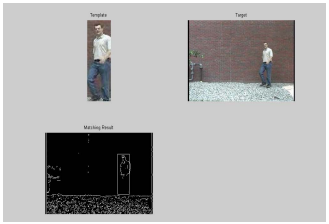

**Fig.6.Training samples**

Various Filters (low pass, high pass, Gaussian…) are applied to enhance the required edges from the frame to avoid false alarm rate and improve detection response.

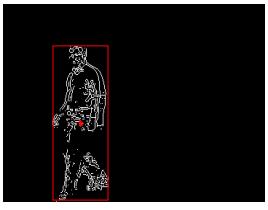

**Fig.7.Sample tracking result.**

Table 1 shows the values for Detection rate and the corresponding number of frames. False alarm rate is decreased by increasing the detection rate.

**Table.1 Detection Rate**

| S.No | Detection Rate | No. of Frames |
|---|---|---|
| 1. | 24.2 | 1 |
| 2. | 25.6 | 2 |
| 3. | 45.1 | 4 |
| 4. | 48.7 | 5 |
| 5. | 54.1 | 6 |
| 6. | 58.2 | 6 |
| 7. | 65.3 | 7 |
| 8. | 68.2 | 8 |
| 9. | 70.4 | 9 |
| 10. | 72.7 | 10 |

The detection rate increases if the number of input frames is increased. The detection rate for the given input frames are displayed in the Fig.8.

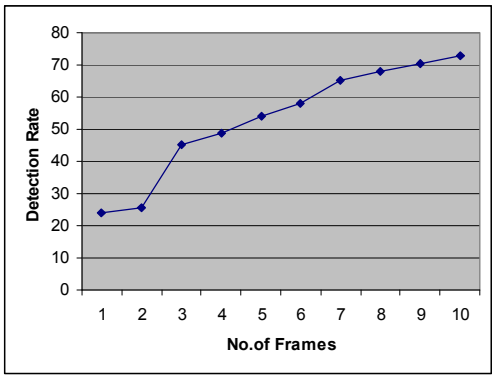


**Fig.8 Detection Rate**

If the detector outputs a false alarm at certain position in one frame, the probability of false alarm is high, that a false alarm will appear around the same position in the next frame. Table 2 lists the False Alarm Rate and number of frames.

**Table 2 False Alarm Rate**

| S.No | False Alarm Rate | No. of Frames |
|---|---|---|
| 1. | 0.28 | 1 |
| 2. | 0.26 | 2 |
| 3. | 0.24 | 3 |
| 4. | 0.23 | 5 |
| 5. | 0.12 | 4 |
| 6. | 0.13 | 7 |
| 7. | 0.17 | 6 |
| 8. | 0.11 | 9 |
| 9. | 0.10 | 8 |
| 10. | 0.09 | 10 |

The False Alarm Rate for the given number of frames is given in the following Fig.9.

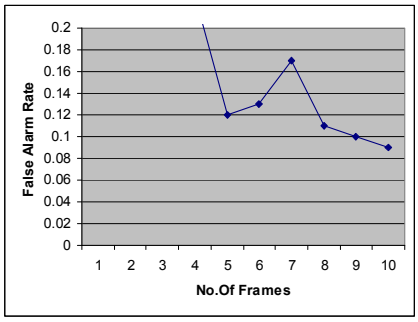


**Fig.9 False Alarm Rate**

Detection rate is the trade off between the number of missed detections and false alarm rate. False Alarm Rate decreases with increased number of frames given as input.

## V.CONCLUSION AND FUTURE SCOPE

Edgelet features are best for human detection because they are relatively invariant to reflections and shadows. To increase the problem space of the tracking algorithm a new scheme named Continuous Probability Distribution, a soft decision is used in this project. The Edgelet based Human Detection and tracking methodology is developed and experimented. The experimentation shows better performance to detect and track human in gray level or color based images. Cascaded decision strategy is used rather than single decision strategy to detect the human. From the experimental results, the proposed system has low false alarm rate and achieves a high tracking accuracy

In Future Work the learning process can still be enhanced using bagging algorithm instead of boosting algorithm. Motion consistency based on the Lucas-Kanade feature tracker can be used in future work instead of the affinity based information. To detect human an offline object model is used, in future online object model can be used for learning of classifiers.

## REFERENCES

[1] B. Wu and R. Nevatia (2007) Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet based Part Detectors, International Journal of Computer Vision.1-6.

[2] F.Wallhoff, M. Zobl, G. Rigoll, and I. Potucek: Face Tracking in Meeting Room Scenarios Using Omnidirectional
Views.ICPR'04. Vol IV: 933-936.

[3] C. Huang, H. Ai, Y. Li, and S. Lao (2007) High Performance Rotation Invariant multiview Face Detection, IEEE Transactions on PAMI, 29(4) 671-686

[4] Davis, L., Philomin, V. and Duraiswami, R. 2000. Tracking humans from a moving platform. ICPR, vol. IV, pp. 121–128.

[5] B.Wu, and R. Nevatia. Tracking ofMultiple Humans inMeetings. In V4HCI workshop, in conjunction with CVPR 2006. 5, 6, 7

[6] B. Wu, and R. Nevatia. Tracking of multiple, Partially Occluded Humans based on Static Body Part Detection. In CVPR 2006.

[7] R. Feraud, O.J. Bernier, Jean-Emmanuel Viallet, and Michel Collobert, "A Fast and Accurate Face Detector Based on Neural Networks", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 23, No. 1, 2001, 42-53

[8]B. Wu, and R. Nevatia. Tracking of Multiple, Partially Occluded Humans based on Static Body Part Detection. In CVPR 2006.

[9] J. Garofolo, C. Laprum, M. Michel, V. Stanford, and E. Tabassi. The NIST Meeting Room Pilot Corpus. In: Proc. Of Language Resource and Evaluation Conference. 2004.

[10] I. McCowan, S. Bengio, D. Gatica-Perez, G. Lathoud, F.Monay, D. Moore, P. Wellner, and H. Bourlard. Modeling Human Interaction in Meetings. in Proc. IEEE Int. conf. on Acoustics, Speech and Signal Processing (ICASSP), Hong Kong, April 2003.

[11]Brostow, G.J. and Cipolla, R. 2006. Unsupervised bayesian detection of independent motion in crowds. CVPR,
vol. I, pp. 594–601.