International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 5, May 2014

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

Comparative Analysis of Human Detection Methods Using Combination of Hough Circle Transform and Descriptors

Avita Verlekar[1], Dr. H. G. Virani[2]

M.E. Student, Department of Electronics and Telecommunication Engineering, Goa College of Engineering, Ponda, India[1]

Professor, Department, Department of Electronics and Telecommunication Engineering, Goa College of Engineering, Ponda, India[2]

Abstract: Crowd detection and counting is an important task in video surveillance. In this paper a method of human detection and counting is implemented by combination of Hough Circle Transform and Descriptors. The Descriptor model uses three descriptors for identifying distinguishing features of head-neck-shoulder region. Comparative analysis is done and performance evaluation of the various methods is carried out.

Keywords: Human Detection; Hough Circle Transform (HCT); Descriptor Model (DM); Background Subtraction

I. INTRODUCTION

The analysis of crowd dynamics is an interesting topic in computer vision. A fundamental task in crowd analysis that enjoys wide spectrum of applications is to automatically count the number of people in crowd. One of the key application areas of crowd counting is public safety and security. There are many algorithms for human detection and the algorithms either follows part based or full body based detection with other feature considerations. Human head is one body part that can be robustly detected even in crowded scenes, therefore in this paper a method for human detection and counting based on head detection is implemented.

The paper is organized as follows: In section II we discuss some of the related work in this field; in section III an overview of the descriptor model implemented is discussed .In section IV the results of using the descriptor model are discussed. Section V gives the conclusion

II. RELATED WORK

There are various approaches proposed for human detection and counting. In counting by detection [4], instances of pedestrian are detected through scanning the image space using a detector trained with local image features. An alternative approach is counting by clustering [15], which assumes a crowd to be composed of individual entities, each of which has unique yet coherent motion patterns that can be clustered to approximate the number of people. Another method is inspired by the capability of human beings, in determining density at a glance without numerating the number of pedestrians in it. This approach is known as counting by regression [16], which counts people in crowd by learning a direct mapping from low-level imagery features to crowd density.

One approach of human detection is by segmenting the foreground into individuals [4]. Zhao et al. previously discussed head detection technique to generate initial human hypotheses. They use a Bayesian framework, in which each person is localized by maximizing a posterior probability over location and shapes, to match 3D human shape models with foreground blobs. They handle the inter-object occlusion in 3D space.

In [6, 7], SURF features are extracted and only those which are in motion are considered. The number of SURF features is mapped to the people count using a regression function. Another method involves exhaustively searching images using a scanning window. Human or non-human classification of each window depends on the colour, shape or motion features. Texture-based methods such as histograms of oriented gradient (HOG) use edge features and support vector machine (SVM) to detect humans [5].

Sim, Rajmadhan, and Ranganath [9] use the Viola-type classifier cascades for head detection. The results from the first classifier are further classified by the second classifier, which is based on colour bin images. Zhao and colleges [10] detect heads from foreground boundaries, intensity edges, and foreground residue.

In another approach [3] an SVM classifier is used to detect human beings based on finding people's head by searching for circular patterns through a 2D correlation using a bank of annular patterns. In [11] a region based tracking approach using skeleton graph is used for head detection. The skeleton graph is extracted from the foreground mask obtained using background subtraction.

In [8] interest point detector based on gradient orientation feature is used to locate regions similar to the top of head region from gray level images. This method may require background subtraction and can even work on single frames. In [1], [2] Hough circle transform is used for detecting the head region of human.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 5, May 2014

III. OVERVIEW OF METHOD

The descriptor system consists of pre-processing the video input followed by extraction of the head-neck-shoulder region for head detection. Three descriptors are used in searching the human head in the frame and finally a weight based decision is used to get the final count of people in the frame.

The HCT system consists of pre-processing of the video input followed by canny edge detection and clustering based HCT algorithm for head detection. Finally head count of the humans present in the scene is obtained. The following subsections describe each of the blocks in detail.

A. Pre processing

The video frames are converted to grayscale and thereafter subjected to optimal global thresholding.This is followed by filtering in order to remove effects of hidden noise. Further morphological operations are carried out on the frames in order to get quality input for the next stage which is head detection. For the HCT method the input image must be an edge image and hence Canny edge detection is used. It was found to give better results as compared to other edge.

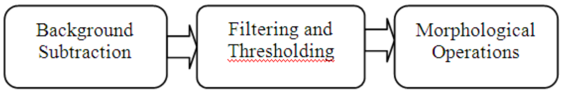


Fig 1(a): Pre processing

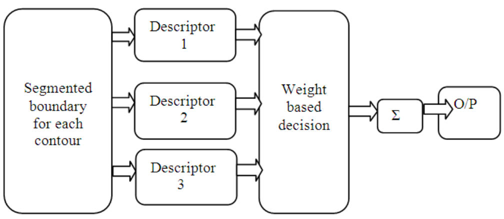


Fig1 (b): Human Head Detection – Descriptor Model

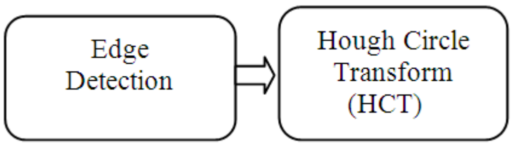


Fig1 (c): Human Head Detection - HCT [1]

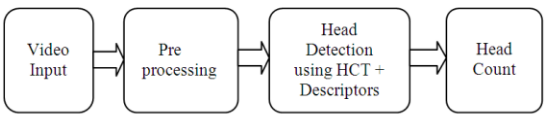


Fig 1(d): Process Flow

B. Hough Circle Transform

HCT method is chosen because it gives robust detection in the presence of noise and partial occlusion. HCT requires three parameters for completely defining a circle: (xcenter, ycenter, radius).In HCT, the computation required to perform the Hough transform is reduced by making use of gradient information which is often available as output from an edge detector. In the case of the Hough circle detector, the edge gradient tells us in which direction a circle must lie from a given edge coordinate point.

The HCT uses a 3D array (accumulator) with the first two dimensions representing the coordinates of the circle and last third specifying the radii. The values in the accumulator array are increased every time a circle is drawn with desired radii over every edge point. The accumulator (which keeps count of how many circles pass through coordinates of each edge point) proceeds to a vote to find the highest count. The coordinates of the centre of the circles in the image are the coordinates with the highest count.

C. Descriptor Model

In this work three descriptors [12] are used to classify the human beings from other non- human objects. Each descriptor is assigned a particular weight.

1) Descriptor 1: Ω_1 (Head Neck Shoulder dimensions of model)

This shape based descriptor is defined by its minimum and maximum dimensions in X and Y directions in the frames. The object of interest is enclosed in a bounding box and then further processing on this is done to extract the head-neck-shoulder region. The extracted data is compared with a threshold (obtained experimentally) to determine the presence of a human in the frame.

2) Descriptor 2: Ω_2 (Convexity of model)

This shape based descriptor is based on the convex hull of the set of boundary points obtained from the extracted contour. The convex hull of the set of boundary points of the contour is the enclosing convex polygon with the smallest possible area.

3) Descriptor3: Ω_3 (Curvature of model)

This descriptor classifies human based on the information of curvature of human head-neck-shoulder portion. At each point in the boundary of the contour, curvature is estimated and based on these curvature values obtained for each of the boundaries of the contour the number of local minima's are studied. Based on this experimental analysis, a threshold is obtained and decision is taken whether a human being is present in the scene or not.

D. Implementation

In Combination Type1 (CBM1), HCT and the three descriptors are given equal weight, that is ¼ .Human is detected only if output is greater than threshold of ¾.

In Combination Type2 (CBM2), both HCT and Descriptor algorithms are run on entire video and counting is done by dividing frame height into 3 regions and taking count only from selected regions of video frame in each method. This is done to reduce errors which take place at entry and exit of person from the frame.

In Combination Type3 (CBM3), a mask is used before the HCT and Descriptor algorithms are run. This is done to limit the search area and thus improve processing time by HCT.

E. Performance Measures

a) Mean Relative Error (MRE) = $\frac{1}{N}\sum_{i=1}^{N}|G(i)-T(i)|$

where N = number of frames for test sequence

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 5, May 2014

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

G(i) = detected person in i'th frame
T(i) = true number of persons in i'th frame

b) $Precision = \frac{tp}{tp+fp}$

c) $Recall = \frac{tp}{tp+fn}$

d) $F-Measure = \frac{2Precision * Recall}{Precision + Recall}$

e) $Accuracy = \frac{tp+tn}{tp+fn+fp+tn}$

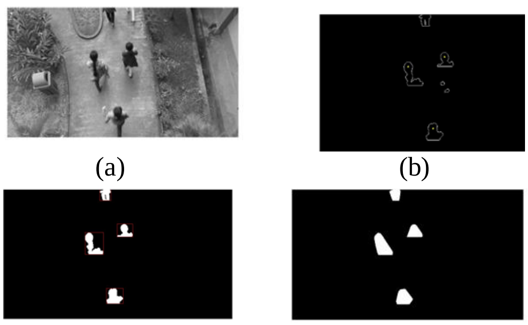
f) $(Pfa) = \frac{fp}{number\ of\ test\ frames}$

Pfa = Probability of False Alarm
tp = True Positive indicating correct detection of human
fp = False Positive indicating false detection of human
fn = False Negative indicating human missed when actually present
tn = True Negative indicating correct absence of human.

IV. EXPERIMENTAL RESULTS

Analysis was performed on a video of moving people (having frame rate 29frames/sec) for a total of 300 frames. The 'x' in the HCT output indicates the detected head.

Fig 2 shows the results for frame 246 for Combination Type1 where head count is correctly detected as 3 and Fig 3 shows the human count frame wise for this method.



(a) (b) (c) (d)

Fig 2. Combination Type1 output (a) Grayscale Image (b) HCT Output (c) Descriptor 1 (d) Descriptor 2

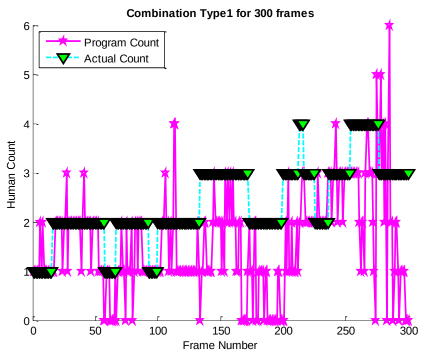
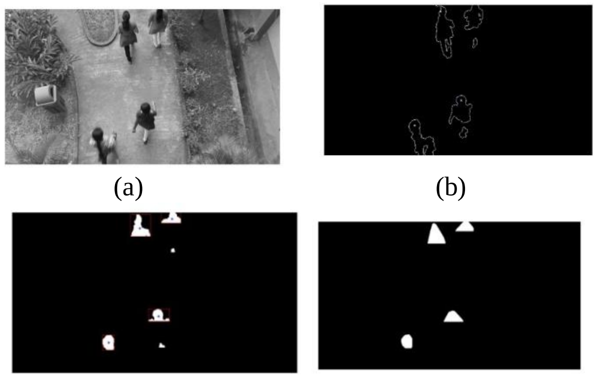


Fig 3 Graph of Human Count v/s Frames for Combination Type1

The analysis for frame 220 is shown in Fig 4 for the Combination Type2 method where HCT count is 2 and descriptor model count is 1 thereby giving a total count of 3 humans. Fig 5 shows the human count frame wise for this method.



(a) (b) (c) (d)

Fig 4 Combination Type2 output (a) Grayscale Image (b) HCT Output (c) Descriptor 1 (d) Descriptor 2

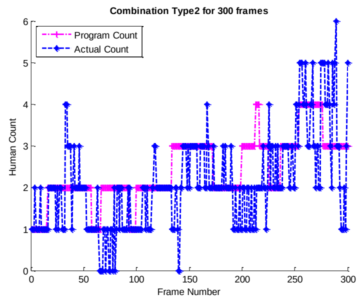
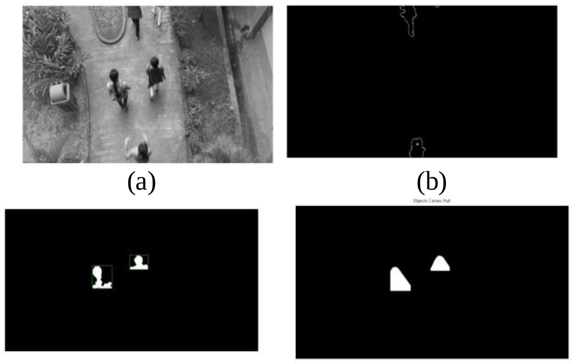


Fig 5 Graph of Human Count v/s Frames for Combination Type2



(a) (b) (c) (d)

Fig 6 Combination Type 3 output (a) Grayscale Image (b) HCT Output Descriptor 1 (d) Descriptor 2

The analysis for frame 241 is shown in Fig 6 for the Combination Type3 method. The HCT head count for the masked upper and lower region is 1 whereas descriptor model count for middle masked region is 2, thus giving a total head count of 3 in the frame. Fig 7 shows the human count frame wise.

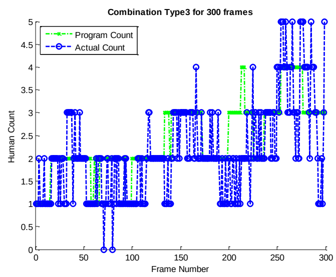


Fig 7 Graph of Human Count v/s Frames for Combination Type3

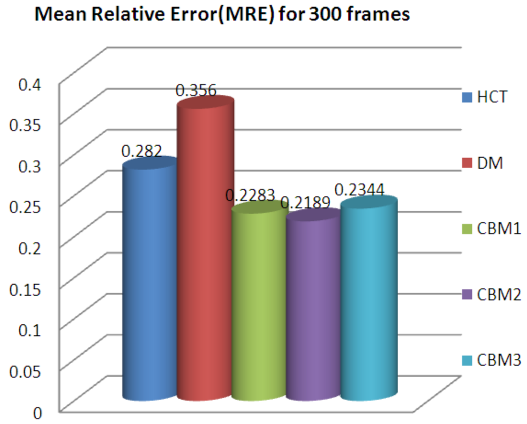
The methods were analyzed for different performance parameters, the results of which are tabulated below.

TABLE I
PERFORMANCE EVALUATION

Performance Parameters	Human Detection Methods				
	HCT	*DM*	*CBM1*	*CBM2*	*CBM3*
Mean Relative Error	0.282	0.356	0.2283	0.2189	0.2344
Average Time per frame (sec)	2.51	0.103	2.25	2.19	1.74
Precision	0.887	0.7861	0.932	0.955	0.9416
Recall	0.727	0.8125	0.8167	0.9133	0.9064
F-Measure	0.799	0.8221	0.8711	0.9338	0.9236
Accuracy	0.872	0.798	0.9025	0.941	0.9307
Probability of False Alarm Pfa	0.106	0.153	0.09	0.063	0.08

V. CONCLUSION

A method of human detection and counting using combination of HCT and Descriptors in different ways has been implemented in this paper. Experiments performed on the video validate the effectiveness of the approach. Combination of HCT and Descriptors does increase the performance of the human detection and counting system. Each combined method gives a higher precision, above 90%, as compared to individual algorithms. Also the mean relative error is lesser than in the case of HCT or DM. In the combination methods, the processing time is less for CBM3 since the search area is reduced for HCT. However the false alarm rate is found to be on slightly higher side. In our future work we would like to reduce the false detections, and incorporate additional methods to reason out the same.



REFERENCES

[1] Avita Verlekar, Dr. H.G. Virani, "Clustering Based Head Count Using Hough Circle Transform", IJRTS February (2014)
[2] Hong Liu, Yueliang Qian and Shouxun Lin" Detecting Persons using Hough Circle Transform in surveillance video", VISAPP 2010 - International Conference on Computer Vision Theory and Applications
[3] Huazhong Xu, Pei Lv, Lei Meng, "A People Counting System based on Head-shoulder Detection and Tracking in Surveillance Video", Proceedings of International Conference On Computer Design And Applications(ICCDA) (2010)
[4] Wu, B., Nevatia, R.: Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors. International Journal of Computer Vision 75(2), 247–266 (2007)
[5] TN. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in CVPR, 2005, pp. 886893.
[6] D. Conte, P. Foggia, G. Percannella, F. Tufano, and M. Vento, "An effective method for counting people in video-surveillance applications." in VISAPP, 2011, pp. 67– 74.
[7] D. Conte, P. Foggia, G. Percannella, and M. Vento, "A method based on the indirect approach for counting people in crowded scenes," in AVSS, 2010, pp. 111–118.
[8] Venkatesh Bala Subburaman Adrien Descamps Cyril Carincotte," Counting people in the crowd using a generic head detector",(2013)
[9] Sim, C.-H., Rajmadhan, E., & Ranganath, S," Using color bin images for crowd Detection",. In IEEE International Conference on Image Processing (ICIP).(2008)
[10] Zhao, T., Nevatia, R., & Wu, B,"Segmentation and tracking of multiple humans in crowded environments",. IEEE Transactions on Pattern Analysis and Machine Intelligence,(2008)
[11] K. E. Aziz, D. Merad, B. Fertil, and N. Thome, "Pedestrian head detection and tracking using skeleton graph for people counting in crowded environments," in 2th IAPR Conference on Machine Vision Applications (MVA2011), 2011.
[12] Subra mukherjee , Karen Das," Omega Model for Human Detection and Counting for application in Smart Surveillance System", (IJACSA), Vol. 4, No. 2(2013)
[13] Comparative Analysis of Descriptor Model and Clustering based Hough Circle Transform Human Detection System, Proceedings.ICSIE 2014 Chennai.
[14] Li, M., Zhang, Z., Huang, K., Tan, T.: Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection. In: International Conference
[15] Rabaud, V., Belongie, S.: Counting crowded moving objects. In: IEEE Conference on Computer Vision and Pattern Recognition pp. 705–711 (2006)
[16] Chan, A., Vasconcelos, N.: Counting people with low-level features and bayesian regression. IEEE Transactions on Image Processing 21(4), 2160–2177 (2012).