Object Counting in Video Sequences

Chirag I. Patel and Ripal Patel

***Abstract*—This paper addresses the problem of counting thenumber of object in an image frame. This paper presents ahuman detection model, that is designed to work with object. The system proposed does learning through templates. The modelmakes use of Haar based features to form templates performsmatching of Haar-transformed images. Object can be detectedirrespective of the texture and color of their clothing as well asorientation.**

***Index Terms*—Object detection, hear templates, object counting**

I. Introduction

This paper attempts to provide a Wavelet based humandetection system. Human beings are non rigid objects and as such deteting them is a hard problem, due to the variouspossible combinations that arise out of clothes being worn, there texture, the orientation of the individual.To overcome this, we need a systems that is invariant to thecolour differences, this is made possible by using Haar transforms. These have the property that they extract informationfrom a given image, which is invariant to the absolute colourand makes use of only color changes. The problem of handlingmultiple orientatios can be tackled by having a sufficientlylarge database of object in different orientations. Having alearning system simplifies the task of adding more templatesas and when needed to handle new cases that may arise. Multi resolution Haar transform were found for human templates and Pyramidal search was caried out to match humanbeings. Human detection and counting has numerous advantages in real life problems. Some of the major applicationsare• Human Intrusion detection. • Tracking usage of Resource / Preferences of object. (Unobtrusive monitoring) • Optimizing working of road crossing signals. • Getting a rough count of the number of object in anenclosed area (malls & bank).

II. Previous Work

Oren et al [1] present a wavelet based technique for pedestrian detection.It is based on Haar wavelets and template matching. The detection is restricted to frontal and rear views of pedestrians. The training set is static. Viola Jones [2] presents a method of speeding up Haar transform using integral imaging.

III. Methodology

We are finding number of people in a frame as object detection and classification problem i.e. given an image to determine at what locations object of a particular classis present in the image. In this case we need to defineHUMAN class which represents all people. This class isdefined using a set of templates. In the learning phase, adatabase of these templates has to be created using groundtruth. Thesetemplates are then matched in test images to detectand count people. The method of creating templates is basedon "Haar Wavelet Features" and is mainly inspired by the workof Oren et al [1].

Manuscript received June 12, 2012; revised July 22, 2012.

The authors are with the Computer Science Department, Nirma Institute of Technology, Ahmedabad, India (e-mail: chirag453@gmail.com, ripalpatel315@gmail.com)

IV. Algorithm

A. Build Template Database(Learning)

In order to build a template database, we used Background Subtraction to identify Foreground objects. The foreground objects were marked in the original inputusing coloured rectangles. and simultaneously foreground portion of the input image was cropped and saved.

Once Background subtraction has run on all the input image set, we use the UI tool generated for Template generation. Main purpose of UI was to assist the system in identifying False Positives, and help the system in learning from TruePositives only, so that invalid data did not get into the templates database.

The UI application displays the image with foreground object marked with rectangles. The user is then shown a separate window having ability to show the cropped images. In this window we can select the cropped imageswhich are TRUE positives and label them.

Labeling is the process of assigning a class to each cropped image, there can be multiple classes. We can have a class from frontal views of object, side views,object on cycle, etc. Note that this feature can be used to extend the system to detect and count non human objects as well.

Once the cropped images are either rejected or labeled (put into appropriate directories) we compute the Haarof the cropped images. Steps followed for finding Haar Transform of an image are as follows:

Convert the image to grayscale using the equation: color = 0.3RED + 0.59GREEN + 0.11BLUE

Depending on the level of the Haar, we find theaverage of two adjoining regions of the gray scaleimage, We then find the difference in the average of the two regions and save it in place of the originalimage pixel

The haar transform can be computed using basisfunctions of a varing number of shapes, howeverwe have restricted our use to only rectangular Haarfeatures - Horizontal, Vertical and Diagonal Haarfeatures. In case of Vertical Haar transform we findthe difference between two adjoining blocks placedin different rows. Horizontal Haar is obtained byfinding the difference between the two adjoiningblocks placed in different columns. Diagonal Haarwas computed in similar way for blocks in different row and column.

We then found the average image of all the Haars,and saved the image so obtained, this was repeatedfor different sized Haars. We computed the Haar forlevels 1 to 3 and saved them. To speed up the computation of Haar features "Integral image represenation" of the image is computed. The method is based on the one described in[2] and general discussion on Haar transforms for2-d images given in [3].

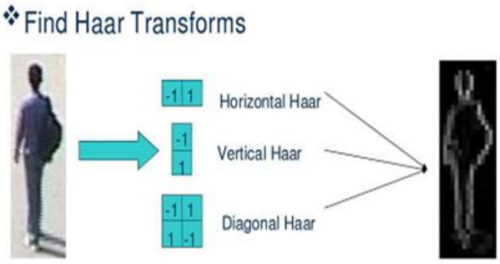

Fig. 1. Finding haar transform

V. Matching

Read input image and find the Haar Transform of the input image using the steps mentioned above.

Now perform Pyramidal search on the transformed image using all the Templates loaded at start-up. Matching is done using Pearson's coefficients.

We maintain a global Threshold for the match to be true. When a match is found we mark the area in the input image using colored rectangles.

In order to prevent multiple matches in same area we maintain an array of matched regions.

We added the newly matched regions to this array to keep track of the regions that need not be searched again.

Every time we find a match we keep track of it using aglobal variable.

VI. Results

The method used for detecting object gave a variable accuracy across image frames. As we try to detect object in frames further away in time from the one in which we collected the templates, we see that the number of False negatives increases. This could be on account the change in posture of the object across the frame sequence.

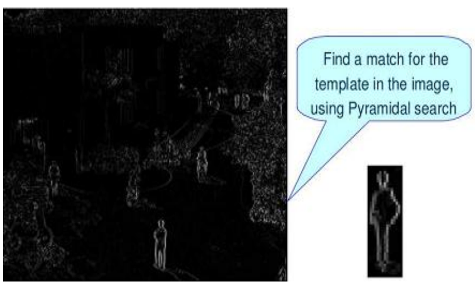

Fig. 2. Matching process using Haar transform

VII. Limitation

People of small sizes cannot be detected reliably If threshold background subtraction is lowered to form smaller templates, smaller images are detected. However, this result in generation of large number of false detections Detection fails if a template of that shape is not present in the database. As the detection is done by template matching, the detection will fail if appropriate template is unavailable in the database

Fig. 3. Matched people

TABLE I: Matching Results

Image	True Positiv e	False Positiv e	Actual People	Accuracy
LHCVedio1_002000.jpg	6	0	11	54.55
LHCVedio1_002001. jpg	5	0	11	45.45
LHCVedio1_002002 .jpg	4	0	11	36.36
LHCVedio1_002003 .jpg	4	0	11	36.36
LHCVedio1_002004 .jpg	4	0	11	36.36
LHCVedio1_002005 .jpg	3	1	11	27.27
LHCVedio1_005000 .jpg	3	1	10	30
LHCVedio1_005001 .jpg	2	0	10	20
LHCVedio1_005002. Jpg	2	1	10	20
LHCVedio1_005003. Jpg	1	0	10	10
LHCVedio1_006000.jpg	6	0	12	50
LHCVedio1_006000. Jpg	2	0	12	16.67
LHCVedio1_006000. Jpg	1	0	12	8.33
LHCVedio1_007000. Jpg	5	0	11	45.45
LHCVedio1_007001. Jpg	4	0	11	36.36
LHCVedio1_007002. Jpg	2	0	11	18.18
Total	54	3	175	30.86

VIII. Occlusion

If a person gets occluded by a shape the detection fails as it distorts the Haar features.

IX. Future Direction

Need to build a larger database of templates efficient matching algorithm is required. In a video frame sequence the number of object does not change drastically in adjacent frames, therefore the search space could be reduced by racking object across frames.

X. Conclusion

The basic working of Haar based matching and people counting were implemented, however the accuracy was founding Table I .GUI for template database creation to be very low. This can be improved by increasing number of templates. The code for matching needs to be speeded up based on heuristic approach or a probabilistic model.

REFERENCES

[1] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio, “Pedestrian detection using wavelet templates,” in *Computer Vision and Pattern Recognition*, 1997. Proceedings, 1997 IEEE Computer Society Conference. San Juan, Puerto Rico: IEEE Computer Society Press, pp.193–199, vol. 2, 1997.

[2] M. J. P. Viola, “Fast and robust classification using asymmetric adaboost and a detector cascade,” September 2001.

[3] E. J. Stollnitz, T. D. DeRose, and D. H. Salesin, “Wavelets for computer graphics: A primer, part 1,” *IEEE Computer Graphics and Applications*, vol. 15, no. 3, pp. 76–84, 1995.

Chirag I. Patelis pursuing Ph.D. in Computer Science & Engineering from Nirma Institute of Technology, Ahmedabad, Gujarat, India. He received B.E. in Information Technology Engineering from A.D.Patel Institute of Technology, New VallabhVidyanagar, Gujarat, India, in 2006, and the M.Tech. in Computer Science & Engineering from Nirma Institute of Technology , Ahmedabad, Gujarat, India in 2009. His research interests are in video processing, statistical image processing and 3D vision.

Ripal Patel obtained the B.E. degree in Electronics & Communication Engineering from A.D.Patel Institute of Technology, New VallabhVidyanagar, Gujarat, India, in 2006, and the M.E. in electronics & communication Engineering from Dharamsinh Desai University, Nadiad, Gujarat, India in 2009. Her research interests are Computer Vision, Texture Classification, Video Processing and Image Registration