Counting of People in a crowded Image

Amit Joshi[1] | Chetan Zade[2] | Divyank Shukla[3] | Meet Bhimjiyani[4]

[1,2,3,4]*(UG Scholars, Department of Computer Engineering, KKWIEER, Maharashtra, India, amits2194@gmail.com)*

Abstract— *Crowd counting and management are a area of quite interest in field of computer vision. The operation of crowd counting can be useful in many filed like security, statistical researches. The data collected from crowd counting will also be helpful for further management of the crowd,.This paper uses mainly Head and shoulder parts of the human body to detect the presence of the human in a given region of interest. Instead of using a single detector to detect the images, use of multiple part detectors is proposed, which are learned by a boosting method. The results from all the part detectors are combined together to give out a single results i.e an estimated crowd count. Use of part detectors also helps tackle the problem of crowd occlusion to some level.The main advantage of part detectors is that even though complete human face and structure are not visible, they can still detect most of the partially visible parts.There are many papers for the purpose of humans detection and counting, but most of them are for videos. They also don't deal with added problems like occlusion. Using the part based approach, we try to address this problem.*

Keywords— Computer Vision,Crowd counting,Image Processing

I. INTRODUCTION

Robust detection of humans is important in many applications such as security, visual surveillance, statistical information, media agencies. Detection and tracking of humans in video sequence has made considerable progress. These methods mainly rely of segmentation and foreground detection to detect and track multiple blobs using background substraction. This method is simple and effective, however its only good when camera is stationary and the change in the illumination conditions are gradual.This is rarely true in real life scenarios. Under such conditions, direct detection of humans may solve the problem. Also detection of humans in static conditions has its own applications.

Most human detections systems rely on a high resolution face to be visible or entire human shape to be visible.In this paper we focus on problem of detecting many, possibly inter-occluded humans form static images where complete face may not always be visible, or not recognized easily. We limit the position of camera to approximately frontal or rear positions, where the camera pan angel does not exceed 45 degrees. We test our results on our own dataset of images collected from the internet.

Detection of human body is complicated in comparison to other objects like vehicles as human body has many varying features changing from person to person.

The number of poses in which a person appears in front of the camera also changes. Image appearance is also highly influenced on clothing, hair and skin color of the human in the image, to overcome these difficulties at-least partially, we use multiple part detectors. We combine the results of all the part detectors to calculate probability of existence of human in the image. As we are working on a crowded scene, head, head-shoulder will only be the part that we will be working on as all the other part of the body will most probably be hidden behind other humans.

1.1 Related Work

The framework for object detection proposed by Viola and Jones has proved very efficient for the face detection problem. The basic idea of this method is to select weak features, Haar wavelets, by AdaBoost to build a cascade structured detector. Viola et al. report that applied to human detection, this approach does not work well using the static Haar features. They augment their system to use simple motion

features to achieve much better performance. Some part-based representations have also been developed. Mohan et al. divide human body into four parts, head-shoulder, legs and left/right arms. They learned SVM detectors using Haar wavelet features. Mikolajczyk et al. divide human body into seven parts, face/head for frontal view, face/head for profile view, head-shoulder for frontal and rear view, head-shoulder for profile view, and legs. For each part, a detector is learned by following the Viola-Jones approach applied to SIFT like orientation features. Their method achieved better results on the MIT dataset than . Neither of these papers, however, addresses the situation of crowded scenes where humans may be occluded by others.

1.2 Outline Of Our Approach

Our approach uses multiple part based representation of human face which we use as a primary detection feature. We suggest an approach for dividing face into multiple parts and creating cascades for the different parts differently. The main advantage of this approach is that this helps in dealing with occlusion.It is very rare that in a crowded scene, face of all the humans in an image are visible completely. Many of the times, only parts of the face are visible. Another advantage of the part based approach is that changes in the camera direction and lighting are more tolerant and this gives better results. This also helps dealing in multiples poses of human heads.

The part based approach is divided in two parts: location and detection of the part based features and then their combination to calculate the probability of presence of a human face in the detected area. For the first part we use part based detectors, and for the second part we use the complete face, head shoulder detectors. The head shoulder features are more invariant to human variations like skin colors and clothing.In this paper we mainly deal with presence of multiple humans in a crowd, where most of the crowd is facing front and complete faces of all the humans is not visible. The cameras location might be facing directly orthogonal or facing downwards with a slight tilt(30 degrees-45 degrees)

Detecting humans is much more of a complex task than the detection of cars or some well defined objects. Even when we constrain the pose to frontal and sideways facing humans, the possible variations in the poses are very large. Add to this occlusion and illumination changes, this becomes a very complex task. The humans appearance in the image is also changed depending upon clothes and hairstyles. We use part based learning features which are combination of haar like features and LBP features to achieve a mean between speed as well as accuracy. Combining the output of these part detectors we can get a much better approximation of the count of humans in a given image.

We use part based detectors for detection of occluded humans. The learning of the part detectors is done boot approach , an advanced version of Voila and Jones framework.

We use Gentle Ada Boost to improve the results of our project,instead of the discrete-valued AdaBoost algorithm; histogram weak classifiers replacing the threshold weak classifiers; and upgrading the cascade structured detector to the nesting structured detector.

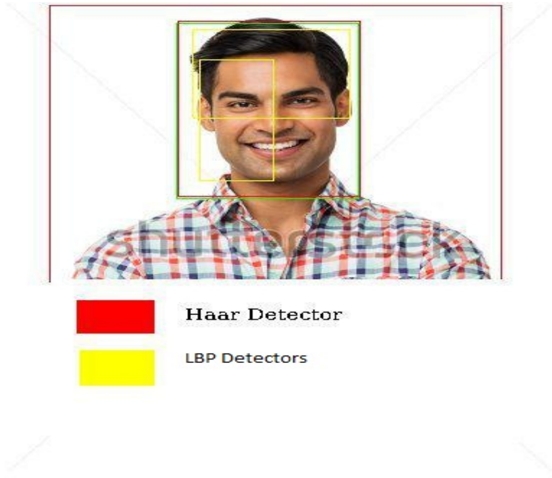
We then combine the results from all the part detectors to form a single count of the humans in the image. Multiples detections of same humans are removed to avoid double counting of same human due to multiple part detectors.

2. PART DETECTORS

We use part detectors for the detection of humans in the image. Part based detectors give better accuracy and much better performance than a single detector. We are mainly looking to train detectors for the front facing and side facing views. The main application of this paper would be to give an approximate count of people in a crowd,Protests, seminar halls, concerts etc. Front and side facing detectors are enough to give better results in these situation. We use a combination of Haar and LBP cascades. Haar cascades are more accurate than the LBP cascades but they are much slower due to considerable amounts of floating point calculations. LBP cascades on the other hand are much faster and slightly less accurate. So we use Haar cascades for the prominent features like the full face and head shoulder detection and LBP cascades for the smaller part detectors like partial faces and side views.

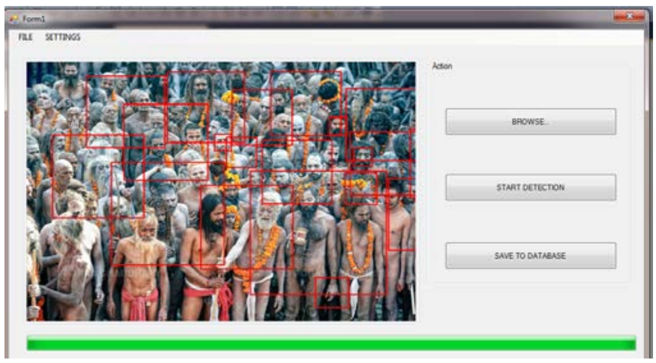
We used training sets on the different datasets collected from the internet and also some additional dataset collected by us. For training of haar cascades we used 1000 positive and 4000 negative images.

For the LBP cascades we used a variation of 500-700 images to obtain satisfactory results.This training was done with the help of Open-CV library, which is a free open source Computer Vision library available on the internet.

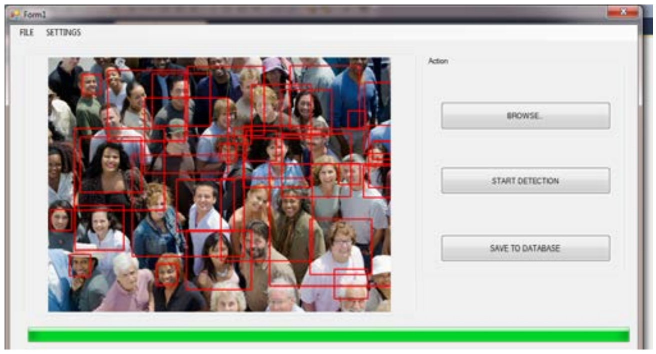


2. COMBINING THE RESULTS

The results of the different cascades provide the count of different detections overall in the image. All these cascades may point to a single person in the image, thus the count may not be correct as same person has been counted multiple times. This may especially happen when person is in a good camera view and more than one cascades detect the person. The extra counts must be substracted to get a correct estimate. This is done by checking if the person has already been detected by other larger cascade before adding him to the final tally. This checking can be done by simple overlapping function which checks if a given region of interest is a sub-region of any other larger cascades. To avoid conflicts between closely grouped humans a gap of 3-5 pixels is taken as a safe margin. If a region is found to be sub-part of any larger region, the count is simply not considered thus giving the correct result of the count.



Crowded scene as Kumbhamela



3. CONCLUSIONS

The results of this process were tested on various images we found on the internet and also on the dataset images. While the accuracy of results vary depending on the quality , resolution and human poses in the image, we were able to get a satisfactory accuracy of 70-80% on most images. This accuracy can be greatly increased by using more cascades which detect the side views as well as the rear views of the human posture, though the time complexity resulting in using so many detectors might get too high. Such systems will provide a good estimate of crowd count if images are taken at regular intervals and sent for processing. The counts obtained this way will help in management, statistics as well as the security of the venue.

REFERENCES

[1]. P. Viola, M. Jones, and D. Snow, "Detecting Pedestrians Using Pattern of Motion and Appearance", ICCV 2003. pp. 734-741

[2]. P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", CVPR 2001. Vol I: 511-518

[3]. B. Leibe, E. Seemann, and B. Schiele, "Pedestrian Detection in Crowded Scenes", CVPR 2005. Vol I: 878-885

[4]. http://homepages.inf.ed.ac.uk/rbf/CAVIAR/

[5]. Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors Bo Wu Ram Nevatia,University of Southern California