Detection of Multiple Humans Using Motion Information and Adaboost Algorithm based on Harr-like Features

JongSeok Lim[1] and WookHyun Kim[1]
[1]*Department of Computer Engineering, Yeungnam University, Korea*
robertlim@yumail.ac.kr, whkim@yumail.ac.kr

Abstract

Robust detection of humans in image sequences is important for many applications. However, if humans are adjacent to each other, it is much more difficult to accurately detect them. In this paper, we propose a method to automatically detect multiple humans using motion information and Adaboost algorithm from a single camera on a mobile or stationary system. In case of mobile system, the ego-motion of the camera is compensated by the corresponding feature sets. The region of interest that moving objects are likely to exist is searched by the projection approach using a difference image between two consecutive images that an ego-motion is compensated. Human detector is learned by boosting a number of weak classifiers which are based on Harr-like features. The proposed approach has been tested to a number of image sequences, and it was shown to detect multiple humans very well.

Keywords*: human detection, ego-motion, Harr-like features, Adaboost*

1. Introduction

Human detection in image sequences has attracted much attention in computer vision research over the past couple of years [1, 2, 3]. This has many practical applications such as video surveillances system, mobile systems etc. In case of the video surveillance system, detecting moving objects is simple since the camera does not move. But, in case of the mobile system, it is difficult to detect moving objects since there is motion by both the camera itself and moving objects in the environment. Those two motions are mixed together in the image sequences. Therefore, in order to robustly detect humans, it should be able to analyze these two motions.

A number of approaches to stabilize camera motions have been proposed by the vision researchers [4, 5, 6]. This paper estimates global motion using estimation of the transformation between two image coordinate systems. Once motion has been identified, human detection is not difficult in case that the number of humans is one or multiple humans are largely apart from each other. However, if humans are adjacent to each other, it is very difficult to segment them accurately. And many recent papers have used machine learning method such as Support Vector Machine (SVM). These methods have not detected humans if they are too close to each other. In order to overcome this problem, we applied an Adaboost algorithm to the region of interest.

In this paper, we propose an approach to detect multiple humans from a mobile or stationary system using a single camera in various environments. Our algorithm has been tested to a number of image sequences with a complicated background. The results show that our method is superior to the other human detection systems.

The remainder of the paper is organized as follows. In Section 2, the proposed algorithm is described in detail. The results obtained from a number of image sequences are presented in Section 3. Section 4 concludes the paper with ideas for future work.

2. Detection of Multiple Humans

The proposed approach finds the region of interest and apply Adaboost algorithm, as is shown in Figure 1. We use motion information to find the region of interest that humans are likely to exist. In case of a mobile system, we find this using the frame difference through the ego-motion compensation from two consecutive image frames. On the other hand, in case of a stationary system, it needs not the ego-motion compensation. Then we detect humans using detector learned from Harr-like features [7]. These features are suitable for human detection as they are relatively invariant to illumination differences. We learn detector by a boosting approach proposed by Viola and Jones [7].

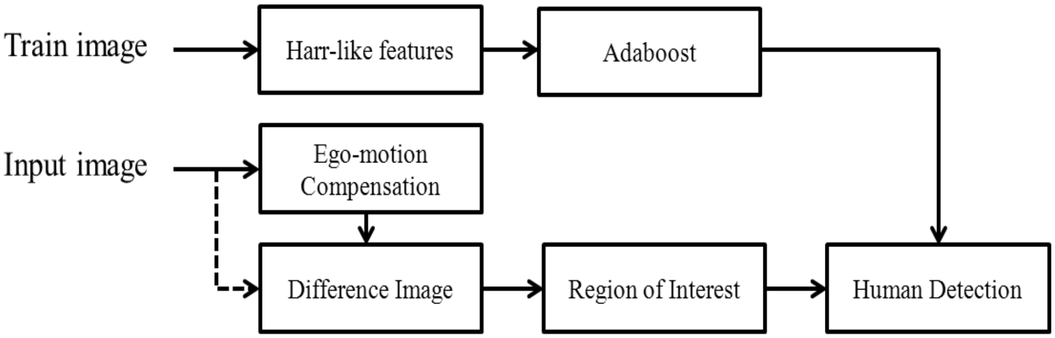

Figure 1. A Schematic Diagram of our Human Detection System

2.1. Moving Objects Detection using Ego-motion Compensation

Typically, the straightforward and fast method used to detect moving objects is to use the frame difference between two consecutive image frames when the input camera is static. However, if the camera moves as when it is mounted on a mobile system, this method is not applicable because objects in the background region are also extracted by the moving camera. In general, there are two independent motions involved in the moving camera environment: motions of moving objects and the camera ego-motion. These two motions are blended into a single image. Therefore, the ego-motion of the camera should be eliminated so that the motion of moving objects can be effectively detected. The detection of moving objects is performed according to frame difference, but the ego-motion of the camera in the previous image is compensated before comparing it with the current image.

In the real world environment, perfect ego-motion compensation is rarely achievable because of various noise sources. Even assuming that the ego-motion compensation is perfect, the difference image would still contain structured noise on the boundaries of objects because of the characteristic of monocular images. We remove these noise terms using the elimination method of an isolated point.

2.1.1. Ego-motion compensation: The ego-motion of the camera can be estimated by feature tracking between images [4]. When the camera moves, two consecutive images, I^t (the current image) and I^{t-1} (the previous image) cannot be compared directly since these coordinate systems are different. Thus, it is necessary to compensation for ego-motion, which is a transformation from the image coordinate of I^{t-1} to that of I^t. Transformation can be estimated using both a set of features in I^t and a set of corresponding features in I^{t-1}. We adopt the Harris corner detector [8] for feature set selection. The feature selection algorithm generates features (f^{t-1}) by running on image (I^{t-1}). The feature tracking algorithm by the Shi-Tomasi [9] is applied to find the corresponding set of features (f^t) in the subsequent image (I^t).

Once correspondence is generated, the ego-motion of the camera can be estimated using a transformation model. We used a bilinear model among the various transformation models because it is a nonlinear transformation model that can estimate most ego-motion of the camera regardless of the length of the interval between consecutive images. A bilinear model used in our experiments is as follows:

$$\begin{bmatrix} f_x^t \\ f_y^t \end{bmatrix} = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 & a_7 \end{bmatrix} \begin{bmatrix} f_x^{t-1} & f_y^{t-1} & 1 & f_x^{t-1} f_y^{t-1} \end{bmatrix}^T \quad . \tag{1}$$

where $(f_x^{t-1} \ , f_y^{t-1})$ is features generated in the previous image and $(f_x^t \ , f_y^t)$ is features tracked those in the current image. We calculate parameters of the transformation model using this.

For ego-motion compensation, image I^{t-1} is converted using the transformation model before being compared to image I^t. For each pixel (x,y), the difference image between two consecutive images is computed using the compensated image as follows:

$$I_{di}(x, y) = \left| \left(I^{t-1} \left({T_{t-1}^t}^{-1}(x, y) \right) - I^t(x, y) \right) \right|. \tag{2}$$

where T_{t-1}^t is a transformation model.

2.1.2. Preprocessing procedure: The difference image generated in the previous step is transformed into a binary image to find the region of interest efficiently. The binary image is greatly affected by the threshold value. If a threshold value is very small, then unnecessary noise sources are formed in the image space. These noise terms lead to serious trouble in the next step. On the other hand, if a threshold value is very big, then nothing or only a small part of objects remains in the image space. In this case, the region of interest becomes the entire image area. Thus, determining a suitable threshold value is very important. We use a value (e.g. 40) obtained by many experiments.

The binary image has a lot of unnecessary noise that is interfering with the detection of the region of interest. To remove noise terms, we use the area elimination method. This counts the number of pixels existing within the window area while moving the sliding window of n×n at first. And then if the number is smaller than a threshold value, all pixels in the window area is eliminated. In this paper, we use the sliding window of 25×30 and the threshold value of 72.

2.1.3. Finding the region of interest: In general, the region of interest that moving objects are likely to exist is searched by the projection approach using the image generated in the previous step. The projection method generates the histogram by the counting of the pixel values greater than 0 in the horizontal and vertical direction. The formulation is as follows:

$$H_x = \sum_{y=0}^{n-1} I_{nd}(x, y). \tag{3}$$

$$V_y = \sum_{x=0}^{m-1} I_{nd}(x, y). \tag{4}$$

where H_x and V_y is the projection histogram for horizontal and vertical direction respectively. $I_{nd}(x,y)$ is the noise eliminated image, x is from 0 to m-1 and y is from 0 to n-1. m and n is height and width of the image.

This projection histogram is used to find the region of interest to detect humans in the Adaboost algorithm. We utilize an aspect ratio and shape information of a human body while finding the region of interest. The results are represented by a bounding box. The number of the bounding box varies depending on the type of moving objects. If humans are adjacent each other, the size of the bounding box is very big. Otherwise if they are far from each other, the size of that is small and the number of that is several.

2.2. Human Detection using Adaboost Algorithm

In the previous section, we found the region of interest. This is used to detect humans using Adaboost algorithm. This algorithm generates a human detector that is learned from Harr-like features [7]. To detect humans, we use Adaboost algorithm proposed by Viola and Jones [7]. By default, a strong classifier to build a reliable human detector can be formed by a linear combination of weak classifiers. Such classifiers are trained with a learning algorithm to good classification results at a small computational cost.

Detection proceeds in two stages: first, the region of interest is searched using motion information and so on; second, humans are detected by our classifier based on the sub window within the region of interest. If the size of the region of interest is above 2/3 of the input image, the detection conducts from the minimum window (24×48) to the maximum window (120×240) with the scale factor of 1.2 within that area. If that is small, the detection conducts in the horizontal direction that exist the region of interest. Thus, our method has a relatively lower computational time than the previous method.

3. Experimental Results

To train the detector, a set of human and non-human training images were used. The human training set consisted of 3547 humans scaled and aligned to a base resolution of 64×128 pixels. The non-human images used to train the detector come from 3748 images which were manually inspected and found to not contain any humans.

To test our algorithm performance, we used a set of 2321 images which were not used in training. Those images contain roughly above 5000 humans except that it is difficult to determine human. The sizes of the humans considered vary from 24×48 to 120×240. We have performed experiments on a 1.5Ghz Pentium Ⅳ CPU.

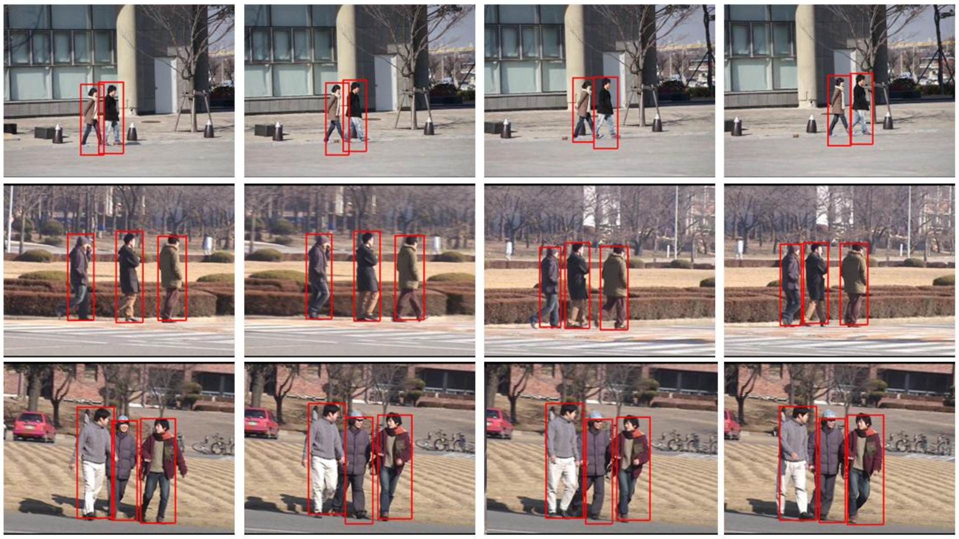

Figure 2. The Detection Results of our Human Detector on a Number of Test Images from the Test Set

We compare our algorithm with the other systems for the correct detection rate. We obtain a detection rate of 93.4% and our algorithm is superior to the other systems. Figure 2 shows the detection results of our human detector on some test images from test set.

4. Conclusions

In this paper we have presented an approach to detect multiple humans using motion information and Adaboost algorithm. The key point of our approach is to find the region of interest to obtain a relatively low computational time. To do this, we detect moving objects using a frame difference image. In case of mobile system, the ego-motion of the camera is estimated using corresponding feature sets. To detect humans within or surroundings the region of interest, we have used the Adaboost classifier learned by the Harr-like features.

This paper presented a set of various experiments with a complicated background and showed excellent experiment results. Our human detector obtained the detection rate of 93.4% and compared to the other systems to evaluate performance.

Acknowledgments

This research was supported by the Yeungnam University research grants in 2010.

References

[1] Y. Byeongju, A. Taeki, L. Wonjae, S. Youngjun and H. Yousik, “Image Surveillance System using Intelligence”, The Journal of IWIT. 9, 5 **(2009)**.

[2] H. Taewoo and S. Yongho, “Emergency Situation Detection using Images from Surveillance Camera and Mobile Robot Tracking System”, The Journal of IWIT. 9, 5 **(2009)**.

[3] B. Leibe, E. Seemann and B. Schiele, “Human Detection in Crowded Scenes”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, **(2005)**.

[4] A. Censi, A. Fusiello and V. Roberto, “Image Stabilization by Features Tracking”, Proceedings of the International Conference on Image Analysis and Processing, **(1999)**.

[5] S. Srinivasan and R. Chellappa, “Image Stabilization and Mosaicking using the Overlapped Basis Optical Flow Field”, Proceedings of IEEE International Conference on Image Processing, **(1997)**.

[6] M. Irani, B. Rousso and S. Peleg, “Recovery of Ego-motion using Image Stabilization”, Proceedings of the IEEE Computer Vision and Pattern Recognition, **(1994)**.

[7] P. Viola and M. Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features”, CVPR, **(2001)**.

[8] C. Harris and M. J. Stephens, “A Combined Corner and Edge Detector”, Proceedings of the 4th Alvey Vision Conference, **(1998)**.

[9] J. Shi and C. Tomasi, “Good Features to Track”, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, **(1994)**.

Authors

JongSeok Lim received the B.S. degree in physics from Kyemyung University in 1991 and the M.S. degree in Computer Science from Daegu Catholic University in 1996 and the Ph.D. degree in Computer Engineering from Yeungnam University in 2004. He is currently visiting professor with Yeungnam University. His research interests include computer vision, image processing, neural network, etc.

WookHyun Kim received his B.S. and M.S. degrees in electronic engineering from Kyungbuk National University in 1981, and 1983, respectively. He received his Ph.D. degree from University of Tsukuba in 1993. He is currently a professor of Department of Computer Engineering at Yeungnam University. His research interests include computer vision, image processing, pattern recognition, etc.