Human Height Estimation

Deep Patel, Maunil Vyas, Aashima Yuthika, Divya Patel

School of Engineering and Applied Science, Ahmedabad University, India

***Abstract*—Surveillance cameras are omnipresent in today's world. This has a lot of advantages especially, in finding individuals for a variety of reasons. Use of hard biometrics for person identification is an accurate but intrusive. Due to this reason, soft biometrics has gained a lot of popularity. In this project we explore the height of a person as a soft biometric feature for identification. We are dividing the height of people into three categories – short, medium and tall. The estimation is done using a calibration matrix and detecting the person in the video sequence.**

***Index Terms*—Computer Vision, Camera Calibration, Soft Biometrics, Histogram of Gradients(HoG), Segmentation**

I. Introduction

Advances in image quality and that of digital cameras have supported various computer vision (CV) applications. It is now possible to take certain measurements and decisions based on videos and images. CV has influenced modern video surveillance systems significantly. Soft biometric features are one of the vital parameters for such applications. One such soft biometric feature is the human height estimation problem, where one wants to estimate the height of human from given video sequences.

There are various ways to tackle this problem. The most popular way is to determine the intrinsic and extrinsic camera parameters by camera calibration. Mainly, these methods rely on estimating vanishing points from the walking humans [1]. As these techniques are more sensitive towards the noise, they are likely to fail more often. There are other ways too, which use the nonlinear regression for the camera calibration [2]. These methods perform comparatively well and often give high accuracy results. But, they required constraint environment such as static backgrounds with fixed camera position.

Here in this paper, we have tried to categorise the human height into three categories. We have taken an Indian scenario and thus, divided the height values accordingly. 1) Short (below 165 cm) 2) Medium (160 cm to 182 cm) and 3) Tall (Above 179 cm). To tackle this problem we have used two approaches. First, by computing the camera calibration parameters using nonlinear regression [2] techniques followed by the human segmentation, the second approach is a more generic one, where we are using HoG features combined with an SVM Classifier to detect the humans followed by the height estimation by a priori knowledge of the depth. Mainly, for the segmentation we have used frame differencing and GrabCut [3] techniques.

TABLE I: Height Categories, Indian Scenario

Range	Category
Below 165 cm	Short
160 cm to 182 cm	Medium
Above 179 cm	Tall

II. Literature Review

- **A Simplified Nonlinear Regression method for Human height Estimation in Video Surveillance by Li et. al., EURASIP, 2015 [2]**
  This paper uses a camera calibration method using non-linear regression to estimate the human height. Here, the authors consider a static background with a fixed camera position to estimate the human height. Moreover, they claimed that as most of cameras for surveillance purposes are installed in high positions with slightly tilted angles, only the focal length, tilting angle and camera height are retained during the calibration. The paper claims high accuracy but, there is a flaw, the authors manually segment the humans from a video sequences. This motivate us to apply their concept with our own segmentation mechanism to make the estimation system robust.
- **Estimating Vanishing Points from walking human [1]**
  This paper uses three vanishing points to estimate the human height. The approach is a bit skeptical as estimating the vanishing points is usually the bottleneck; It is extremely sensitive to the noise and even negligible inaccuracy in a vanishing point can cause huge inaccuracy in the height estimation.

III. Approaches

A. Using Camera Calibration

Approach 1 In this approach, we first detect the person in the video sequence using an SVM Classifier trained with HoG features. After the person identification is done, we have used the head and the foot points to calculate the real-world height of the person using equation (13) in Li et al [2]. The Bounding Box estimated for a person wasn't tight, in the sense that the height of the bounding box was, in fact, more than that of the person. This resulted in improper height estimates. To tackle this, we have used the GrabCut algorithm on the detected bounding box (SVM + HoG). However, the segmentation was not desirable, and generated a lot of false positives.

Approach 2 Additionally, we have also tried the frame differencing idea to segment the foreground and

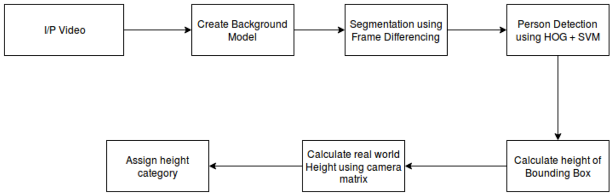

Fig. 1: The Concluded Approach for estimation, Approach 2

background followed by the HoG feature detection. Once, we figure out the bounding box, we calculate its height, and at the end we have used the calibrated parameters to calculate the real height, which helps to categories it in one of the group.

B. Without Camera Calibration, Only Focal length is estimated

Approach 3 Here, we have used the depth information of the person, i.e., how far a person is from the camera. This entailed calculating the focal length of the camera using a marker with known dimensions (width, height, and distance from the camera). Next, we calculate the real-world height of the person at a point where we know the distance of a person from the camera. After this, in the consecutive frames, we calculate the distance of a person from the camera to get a depth estimate and how the pixel height changes with the depth. This was a very naïve approach for height and depth estimation and was tested on videos shot in a constrained setting.

IV. Results

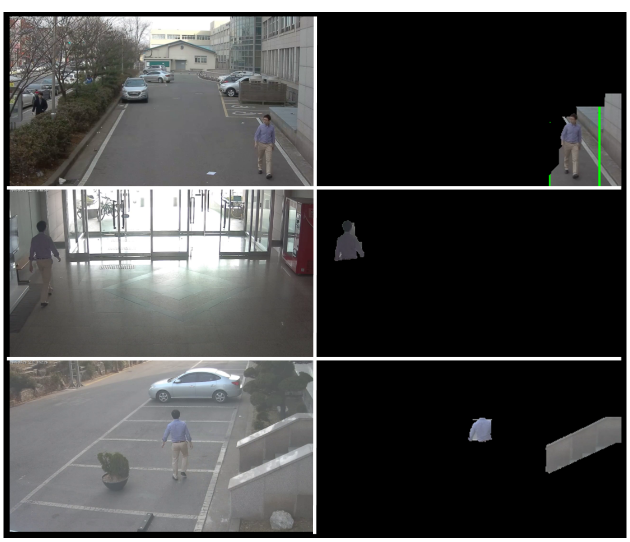

Fig. 2: HoG detection followed by the GrabCut, Approach 1

As we can see in Fig. 2. The GrabCut fails in the segmentation and thus, we could not able to identify the height accurately. Also, without GrabCut and using simple HoG + SVM we can able to get good results as we can see in table 2. But, again here we are not sure that every time the bounding box fits the person exactly, many times it may happen that when a person is not near the vicinity, the HoG will not consider it. Thus, the approach become futile in such situations.

Camera	Estimated Height (in cm)	Category
1	181	Medium
2	133	Short
5	178	Medium
8	176	Medium
10	173	Medium
11	168	Medium
12	182	Tall

TABLE II: Estimated height of the subject using the first approach in different video sequences shot using difference cameras. The true height of the subject is 174.5 cm, Approach 1 without GrabCut

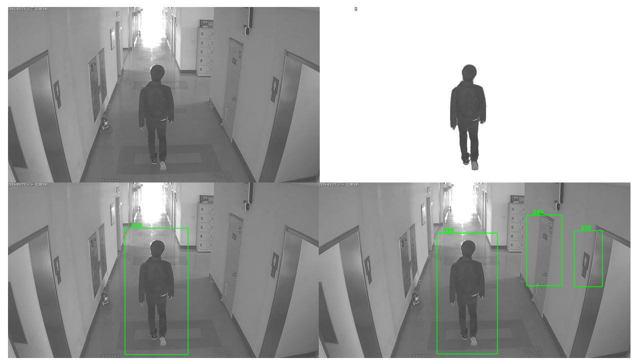

Fig. 3: This is the Success Case, where segmentation works fine, Approach 2

Fig. 4: This is the Failure Case, Approach 2

As we can see in Fig. 3. provides the best result for our approach and Fig 4 provides the worst results. Moving clockwise in those images from top left : 1). Original, 2). Segmented using our method 3). Person Detection and image height measurement using HOG only 4). Person detection and image height measurement using our Segmentation with HOG. Clearly, approach 2 outperforms the approach 1 in suitable conditions.
We have also carried out video shooting at SEAS campus and tried the approach 3. But, the results are not satisfiable.

V. Conclusion

In this paper, we have tried to use three approaches to estimate the human height. Mainly, we have extended the traditional camera calibration idea and utilised the segmentation to make the height estimation task a wholly automated one. In the end, in video surveillance, specifically in a constraint environment, having camera parameters are sufficient to get the human height with a reasonable accuracy. But, by analysing various segmentation techniques, we would like to conclude that proper segmentation of human plays a vital role in the current work and that limits the accuracy too.

References

[1] B. Micusik and T. Pajdla, "Simultaneous surveillance camera calibration and foot-head homology estimation from human detections," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 1562–1569.

[2] S. Li, V. H. Nguyen, M. Ma, C.-B. Jin, T. D. Do, and H. Kim, "A simplified nonlinear regression method for human height estimation in video surveillance," *EURASIP Journal on Image and Video Processing*, vol. 2015, no. 1, p. 32, 2015.

[3] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," in *ACM transactions on graphics (TOG)*, vol. 23, no. 3. ACM, 2004, pp. 309–314.

[4] A. Bovyrin and K. Rodyushkin, "Human height prediction and roads estimation for advanced video surveillance systems," in *Advanced Video and Signal Based Surveillance, 2005. AVSS 2005. IEEE Conference on*. IEEE, 2005, pp. 219–223.

[5] P. Tome, J. Fierrez, R. Vera-Rodriguez, and M. S. Nixon, "Soft biometrics and their application in person recognition at a distance," *IEEE Transactions on information forensics and security*, vol. 9, no. 3, pp. 464–475, 2014.

[6] S. X. Yang, P. K. Larsen, T. Alkjær, B. Juul-Kristensen, E. B. Simonsen, and N. Lynnerup, "Height estimations based on eye measurements throughout a gait cycle," *Forensic science international*, vol. 236, pp. 170–174, 2014.

[7] G. Edelman, I. Alberink, and B. Hoogeboom, "Comparison of the performance of two methods for height estimation," *Journal of forensic sciences*, vol. 55, no. 2, pp. 358–365, 2010.