Human estimation using a calibrated camera: An Implementation

1st Daivat Bhatt
School of Engineering and Applied Science
daivat.b.btechi15@ahduni.edu.in

2nd Mohit Solanki
School of Engineering and Applied Science
mohit.s.btechi15@ahduni.edu.in

3rd Rahul Savariya
School of Engineering and Applied Science
rahul.s.btechi15@ahduni.edu.in

4th Neel Shukla
School of Engineering and Applied Science
neel.s.btechi15@ahduni.edu.in

5th Jeet Trivedi
School of Engineering and Applied Science
jeet.t.btechi15@ahduni.edu.in

***Abstract*—In this time of necessary and state-of-the-art surveillance devices, identification of human beings and corresponding matching is of utmost importance. One of the primary characteristic of a person is their height. From the paper chosen, we intend to implement the idea described in it.**

***Index Terms*—Height Estimation, Image Processing, Surveillance, Camera Calibration**

I. Introduction

Our project consisted of two parts - human detection and the height estimation. We intended to simulate a scenario observed in surveillance cameras where it was important to identify, track and then make a character-sketch of the identified person using the height. We settled into height as our soft biometric criteria for it was one of the primary and easily identifiable traits of humans. It also helps to broadly categorize people thus helping us to weed out non-matching suspects. Height is one of the features that can be used for object tracking feature. We have, however, not used it.

Pertaining to the paper we have chosen as our reference, we intend to use features of the image to calibrate the camera. These specifications that will be generated will then help us to determine the distance of the object and the corresponding height of the person of interest.

We did attempt to calibrate the camera and produced certain results that will be discussed later in the paper. However, we were not able to implement the estimation of the height according to the paper itself. Our attempts to find different methods are ongoing.

II. Literature Review

The paper we had chosen [1] had a machine learning and pure mathematics-based approach in estimating height.

They have assumed that the subject is standing straight on the ground. In order to determine the height of the object visible on the camera, they determine the distance from the camera. This is done using the view angle of the camera, the altitude of the camera from the floor and the orientation of the optical axis.

The main problem however is that there is significant geometric distortion in the surveillance images. These effects must in some or the other way be compensated.

This, the paper, does using shape segmentation, shadow segmentation to find accurate foot and head points and then compensating for distortion. The part of estimating the height is done later.

The architecture used by the paper is shown below:

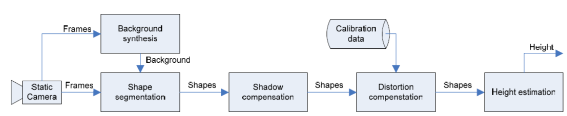

Fig. 1: Architecture used by the reference paper

The algorithm also converts the real world coordinates to that of the camera plane and then that of the image plane. This way, affine transformations applied to the camera both due to intrinsic and extrinsic parameters are measured and the original coordinates in the real world are reverse-engineered.

To compensate for the curvature due to the camera, researchers found out that the curvature of the horizontal line depends only on the tilt angle. Extrapolating this information, they found out that it was the same with the y- coordinates too. Therefore, what was required was a simple mapping of the distorted coordinates represented by the curves using a non-linear transformation.

The researchers used linear regression to fit the distorted curves of the originally straight gridlines to calibrate the radial distortion related intrinsic parameters.

To be able to measure the view angles in radians, the image resolution and the whole view camera angles were also coordinated.

To compensate for the distortion, they first found out a virtual horizontal grid line that had a near enough value of the pixel, as close to the ideal value as it could be. This method was applied to both the horizontal and vertical coordinates. The pseudo code is given below:

```
OriginalX = ModifiedX
do
(OriginalY = ax(ModifiedY) * OriginalX * OriginalX) + ModifiedY
OldX = OriginalX
OldX = OriginalX
OriginalX = ay(ModifiedX) * OriginalY * OriginalY) + ModifiedX
while(abs(OriginalX - OldX) > 0.1)
```

III. Algorithm

The algorithm section will be divided into multiple sections. We will first go over the camera model and then explain the processes used to calibrate the camera.

A. Camera Model

The camera can be modeled as a pinhole camera model, for simplicity's sake. It is but a central projection using the optical center of the camera and an image plane perpendicular to the optical axis of the camera.
The camera projection can then be given as:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K \begin{bmatrix} R & T \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Fig. 2: Matrix Equation

Here, K is the calibration matrix that provides us with intrinsic parameters such as focal length, optical center (principal point)in pixels and the skew parameter.
There are two extrinsic parameters, namely R and T. R is the Rotation Matrix while T is the Translation Vector.

B. Calibrating Camera

The calibration for machine and computer vision have traditionally employed reference grids, the calibration matrix K being determined using images of a checkerboard pattern.
We have used Zhang's calibration method to calibrate one of the video surveillance camera mounted in the campus. To calibrate the camera, we required a planar checkerboard which is placed in front of the camera and then we captured about 15 images (as shown in figure) of checkerboard with different orientations.
Then using Zhang's algorithm the extracted corner points from the checkerboard are used to compute the camera interior and exterior parameters. We have used MATLABs camera calibration toolbox for extracting intrinsic and extrinsic parameters using 15 images as input to the toolbox.

Fig. 3: Original

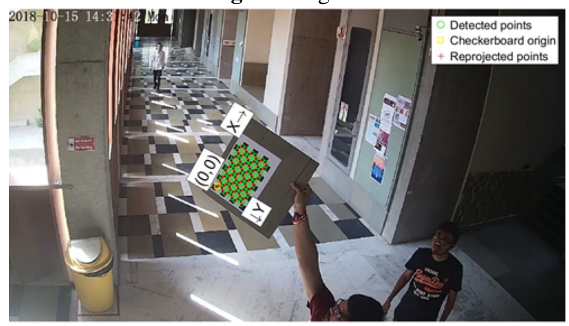

Fig. 4: Detected Image

C. Extrinsic Parameters

We also get extrinsic parameters from calibration of camera i.e. Rotation Matrix and Translation Vector. These camera parameters are used to obtain title angle of the camera and further for height estimation of pedestrian. We have also used method describe in [1] for calculating angle of rotation in x, y and z direction. Equations are given below,

$$R = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix}$$

$$\Theta = sin^{-1}(R_{31})$$
$$\Psi = atan2(\frac{R_{32}}{cos\theta}, \frac{R_{33}}{cos\theta})$$
$$\Phi = atan2(\frac{R_{21}}{cos\theta}, \frac{R_{11}}{cos\theta})$$

Given the rotation matrix R, we can compute the Euler's angles Θ, Ψ and Φ by equating each element in R with the corresponding element in the matrix product $R_x(\phi) \cdot R_y(\theta) \cdot R_x(\psi)$

IV. Result

After implementing the toolbox in MATLAB, we found that the focal length in x and y directions were about 1543.3px and 1559.2px respectively.
When converted into millimeters, they were roughly 2mm each.
We also obtain the Rotation matrix (R) and the Translation vector (T) as:

$$R = \begin{bmatrix} 0.9842 & 0.0605 & -0.1665 \\ -0.0296 & 0.9828 & 0.1821 \\ 0.1764 & -0.1743 & 0.9691 \end{bmatrix}$$

$$T = \begin{bmatrix} 105.2699 & -157.0719 & 1848.1 \end{bmatrix}$$

Also, based on the rotation matrix, we also computed angles in x, y and z directions:

$$\text{Angles (1)} = \begin{bmatrix} 169.8028 & -10.0567 & 178.276 \end{bmatrix}$$

$$\text{Angles (2)} = \begin{bmatrix} -10.1972 & 13.0983 & -1.7235 \end{bmatrix}$$

V. Conclusion

The MATLAB toolbox provided us with enough parameters to obtain intrinsic as well as extrinsic parameters for an image. But it is laden with errors. The parameters are different, sometimes vastly, for two different images even though they were taken from the same camera.

The camera's parameters obtained were also not exact, and was off the mark for certain images. For that particular reason, we had to manually scavenge through the images.

We also proceeded to obtain a crude mathematical relationship between real world coordinates and the image coordinates using the focal length. We then also computed the height using manual pinning of head and foot points. However, the height depended on the distance of object from the camera - something we were not able to obtain from all the processes.

We intend to work on obtaining the distance of the object from the image itself, and then refining the intrinsic and extrinsic parameters generated from the above given formulas. This may help us in creating basic groups of height, namely: Very short, Short, Medium, Tall, Very tall.

Acknowledgment

The authors would express their sincere gratitude to Dr. Mehul Raval for his influence on the course of this work and for his suggestions in the improvement of results. We would also like to acknowledge the assistance provided by Mr. Vandit Gajjar and Mr. Hiren Galiyawala.

References

[1] E. Jeges, I. Kispal, "Human height estimation using a calibrated camera", Oldweb.mit.bme.hu. N. p., 2018. Web. 15 Oct. 2018.

[2] E. Jeges, I. Kispal and Z. Hornak, "Measuring human height using calibrated cameras," 2008 Conference on Human System Interactions, Krakow, 2008, pp. 755-760.doi: 10.1109/HSI.2008.4581536

[3] S. Denman, C. Fookes, Bialkowski and S. Sridharan, "Soft-Biometrics: Unconstrained Authentication in a Surveillance Environment ," 2009 Digital Image Computing: Techniques and Applications, Melbourne, 2008, pp. 755-760.doi: 10.1109/DICTA.2009.38

[4] G. G. Slabaugh, "Computing Euler angles from a rotation matrix," Tech. Rep., 1999. [Online]. Available: http://www.gregslabaugh.name/publications/euler.pdf