HUMAN HEIGHT ESTIMATION USING A CALIBRATED CAMERA

István Kispál
SEARCH-LAB Ltd., Budapest, Hungary
Istvan.kispal@search-lab.hu

Ernő Jeges
Department of Measurement and Information Systems, Budapest University of Technology and Economics
jeges@mit.bme.hu

Keywords: soft biometrics, height estimation, intelligent video surveillance, image processing, camera calibration.

Abstract: Due to the increasing number of different unfortunate events recently, the video surveillance systems are becoming more and more sophisticated in multiple ways. The simple recording of video frames is continuously being extended with a wide range of features, both to increase the security by boosting the efficiency and effectiveness of security staff and to allow more accurate and quick reconstruction of the recorded events for subsequent analysis. In both cases the height of the observed people may be of essential importance. Although height has been long used in forensic procedures to narrow the list of possible suspects, it is not distinctive enough to be used in biometric identification. However, by estimating the heights of the tracked subjects on different cameras, it could provide us with an important additional feature making the object tracking over different scenes more robust. In this paper we introduce our method to estimate the height of human subjects tracked on calibrated surveillance camera images.

1 INTRODUCTION

As opposed to classic biometric identification systems, which should use strictly distinctive and permanent features in order to identify people in a large set of possible subjects (e.g. citizens of a country), in an indoor intelligent video surveillance application we can use a much wider range of different features both for object tracking and for identification.

In case of identification we can use some additional features besides the strict biometric ones because we have a much smaller set of possible subjects, as we must deal only with people that entered a certain area; for the same reason there is also a much weaker requirement on the permanence of the used features. This kind of identification is usually referred to as *soft-biometrics*, but we must remark that this term is also associated with the use of soft-computing methods in biometrics. Typical soft-biometric features include skin color, eye color, build, weight, or height (Jain, Dass and Nandakumar, 2004).

In case of object tracking we can solely rely on some features which can be extracted from the shapes moving on the video frames: shapes are simply the areas of the frames that are different from the background. The difference of two shapes can be expressed in form of a distance value defined by the metrics introduced for their features. Using this distance an object tracking system can associate the corresponding shapes to each other both in temporal and spatial domain.

In the temporal aspect of object tracking the goal is to form a chain of shapes extracted from subsequent frames of one single surveillance camera, while the spatial correspondence can be analyzed between two or more cameras observing the same scene, and the goal is to decide which shapes – visible on frames from different cameras taken at the same moment – can be the projections of the same object.

The combination of such spatial and temporal analysis allows us to continuously approximate the position of a moving object in space and to be aware of its route through the observed area. Furthermore, having covered a wider area involving several

scenes observed by independent cameras, we also may need to determine the correspondence of shapes being the projections of an object visible in different times and places. To determine whether the two shapes belong or may belong to the same object or not, the only thing we can rely on are again some well measurable features, which can be extracted from the video camera frames.

Height estimation of objects can be used both for soft-biometrics and as an object tracking feature. In first case we can eliminate some possible subjects having considerably different height than the observed one, and focus on determining more distinctive remote identification features, like color, face or ear shape (Jeges and Máté, 2006), and search for similar subjects in a smaller set of possible candidates. For object tracking it can be used for temporal and spatial correspondence analysis as well or simultaneously for both in case of disjoint cameras (Madden and Piccardi, 2005).

Human height is seldom estimated alone, but is usually used along with other extracted features like the dynamic properties of gait (Abdelkader, Cutler and Davis, 2002), (Wang, Ning, Tan and Hu, 2003), (Collins, Gross and Shi, 2002). Thus extracted parameters are however almost in every case used for identification purposes. Different body models, like the one introduced in (Green and Guan, 2004) are usually used in gesture recognition or pose recovery besides gait based identification. Once we have fitted the human body model to the person visible on the image, we can easily deduce the height from the parameters of the model.

It is interesting to remark that while height estimation tries to eliminate the dynamic properties of gait by canceling out the time variant component of the height of the shape measured through time, gait based identification just strives for extracting these. The exploitation of this latter fact is shown in (Lv, Zhao and Nevatia, 2002), where our intended process is reversed and the camera is calibrated using a walking person. Some other interesting methods of human height estimations can be found in the literature, like the one in (Bernardin, Ekenel and Stiefelhagen, 2006), where the acoustic source localization of the speaking person is used for this purpose.

In this paper we will introduce our simple height estimation method along with the process of calibrating the camera used to observe the subjects. We intend to use the estimated height of the tracked people both for identification and for object tracking purposes in an intelligent video surveillance application, the *Identrace* (Identrace, 2006). As this system performs spatial object tracking analysis as well, the height is measured using a single calibrated camera, and the values measured on different cameras are used to correspond shapes to each other.

Following a short overview of our object tracking framework we will present how the distortion of the camera was modelled and compensated, and our height estimation method will be featured. At the end we will summarize our results and conclude the work.

2 OVERVIEW

In the followings we assume that the subject is standing on the ground, it stands straight and that the horizontal bisecting line of the camera image is horizontal in the real world as well.

In order to determine the height of an object visible on a surveillance camera, we should first determine its distance from the camera. To do this we can use the view angle, under which the bottom of the object (the feet of the person) is visible, the altitude of the camera from the floor and the orientation of its optical axis. If we know the distance, the estimation of height can be done easily from the view angle of the object's top point.

It sounds simple; however the main problem with this approach in our case is that cameras commonly used for surveillance usually involve significant geometrical distortion. Thus we should define a camera model, express and measure the characteristics of the distortion as the parameters of the model and use the above described method on the image on which the effect of the distortion is compensated.

The architecture of our framework is shown in the Figure below. Shapes are segmented from the frames using a continuously synthesized background. To get more accurate bottom and top points of the shape, we should accomplish shadow compensation. Upon having the shapes cleared, the next task is to compensate the effect of the geometrical distortion of the camera, and finally to estimate the height of the shape.

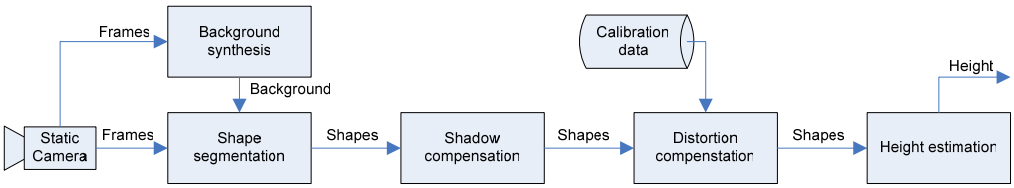

Figure 1: The architecture of out height measurement system.

In the rest of this paper we shall introduce the methods used in the last two modules of the above introduced architecture, namely the compensation of the geometrical distortion and the height estimation process.

3 CAMERA DISTORTION COMPENSATION

3.1 The Camera Model

Cameras are usually modelled with the *pinhole camera model*, which is a central projection using the optical centre of the camera (*O*) and an image plane (that is perpendicular to the camera's optical axis). It represents every 3D world point (*X*) with the intersection point of the image plane and the *OX* line (that is called the image point, noted with *x* in Figure 2).

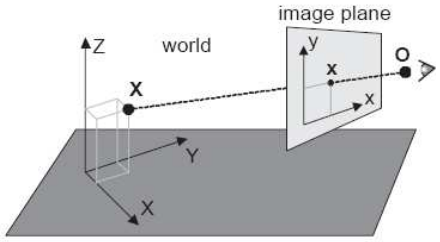

Figure 2: The pinhole camera projection model (Criminisi, 2002).

The pinhole camera projection can be described by the following linear model:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} \mathbf{R} & \mathbf{T} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \qquad (1)$$

where *K* is the calibration matrix, representing the intrinsic camera parameters, *R* is the rotation matrix and *T* is the translation vector. These latter two are the extrinsic camera parameters. Intrinsic camera parameters describe the optical, geometrical and other characteristics of the camera itself, wherever it is placed, while extrinsic parameters describe the position and the orientation of the camera.

However, real lenses seldom project the objects linearly, like the use of *K* in the model above suggests. They may produce a number of different non-linear distortions, but from our point of view the most important is the so-called *radial* (or barrel, pincushion) distortion, because most of the surveillance camera lenses involve them. The effect of this distortion is best perceptible if we put a squared grid in front of the camera; what we can see in the image is a bunch of parabolas rather than parallel lines (Figure 3). So, to get reliable base information to accomplish height estimation, this is the distortion that certainly has to be compensated.

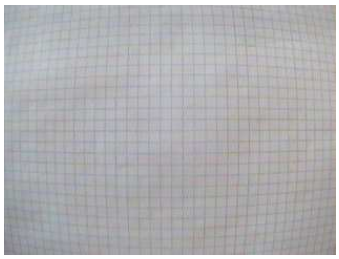

Figure 3: The image of a grid before compensating the radial distortion.

Radial distortion can be modelled with non-linear intrinsic parameters. There is no general closed form solution similar to that used in pinhole model to compensate this non-linearity (Hujka, 2004), so we have chosen to measure it directly in an automatic *calibration* step, and then transform the captured camera images to an ideal form modelled by the pinhole camera model. This second step is called the *compensation* step.

In the followings we will call the coordinates of the 3D points projected according to the pinhole camera model the *ideal image coordinates*. The coordinates of the distorted image originating directly from the camera will be called *distorted image coordinates*.

As we have already discussed, during height measurement we mainly use the tilt (and pan) angles under which certain 3D points are visible from the camera's point of view. Please note that ideal coordinates are linearly proportional to these angles, so they can be used for height measurement directly. Considering this, the main goal of the distortion compensation is to convert distorted pixel coordinates to ideal pixel coordinates. To be able to accomplish this compensation we first have to determine the parameters of the camera by calibrating it.

3.2 Calibrating the Camera

The curvatures of the grid lines seen in Figure 3 are consequences of the radial distortion, as this image directly shows the transformation of the grid intersection points. If the plane of the grid is perpendicular to the optical axis, the pan and tilt angles under which the intersection points are visible can be calculated from the position of the point in the grid and the distance between the camera and the grid. The image of the grid shows us how these view angles are represented in the distorted camera image, from which the pan and tilt angle of other pixels can be interpolated as well.

Our measures have proved that the curvature of a horizontal line depends only on the tilt angle under which it is actually visible from the camera's point of view and does not depend on any other parameters, like the grid's distance from the camera. This fact is consistent with our understanding of the radial distortion, and implies that the ideal *y*-coordinates of the pixels of a grid line are the same. Similar rules apply to the vertical grid lines, the pan angles and ideal *x* coordinates. All this imply that the radial distortion does not affect the coherent points on the *x* and *y* axes, so the ideal coordinates represented by the gridlines can be mapped easily to the distorted coordinates represented by the curves using an appropriate non-linear transformation as expected.

So, to calibrate the radial distortion-related intrinsic parameters of the camera we fit parabolas to the distorted curves representing the original straight gridlines using linear regression. The axis of these parabolas is always the *x*- or the *y*-axis (so they do not have first order coefficients), and their quadratic coefficient is the function of their zero order coefficient (that is the *x* or *y* ideal coordinate of the points on the parabola). Once we have the values of this function in the grid line points, we can determine the values for the rest of the points with linear interpolation. This way we have defined a *virtual grid line* on the image, and using the parabola corresponding to a pixel we can determine its ideal coordinates.

To be able to measure the view angles in radians later on, we should also know how much radians a 1 pixel distance in the ideal image is representing in both directions. We can calculate these values from the image resolution and the whole view angle of the camera.

Besides the above described intrinsic non-linear parameters, we must also determine the extrinsic parameters of the camera as well. The altitude and the tilt angle of the camera can be of course measured manually, but it is handier to do this in a semi-automatic way based on the image of known-height poles, using for example the method proposed in (Lv, Zhao, Nevatia, 2002).

3.3 Distortion compensation

To convert distorted vertical pixel coordinates to the ideal vertical pixel coordinates we simply choose a virtual horizontal grid line that has an acceptably near value in the pixel's column by a binary search. That way the modified pixel coordinate can be determined with any desired precision. The same method applies for the horizontal coordinates as well.

Although our height measurement system does not require it, the original pixel coordinates can be obtained from the modified pixel coordinates with an iterative approach shown in the following pseudo code:

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

In the pseudo code above the *ax(ModifiedY)* and *ay(ModifiedX)* are the quadratic parameters of the

parabolas. After execution the *OriginalX* and *OriginalY* will store the result.

4 HEIGHT ESTIMATION

From the parameters obtained by the calibration we can calculate the view angle of the top and bottom points of the person shown in the camera image. As the h_0 altitude of the camera and the φ tilt angle of the camera are known, and supposing that the person is standing on the floor, the *h* height of the person can be calculated from the following expression:

$$\frac{\tan(\alpha_1 - \varphi)}{h_1} = \frac{\tan(\varphi - \alpha_0)}{h_0}, \qquad (2)$$

where $h=h_0+h_1$, so *h1* is the difference between the height of the subject and the camera altitude, α_0 and α_1 are the tilt angles of the person's top most and bottom most points relative to the camera's optical axis. According to this the height of the subject is:

$$h = h_0\left(1 + \frac{\tan(\alpha_1 - \varphi)}{\tan(\varphi - \alpha_0)}\right). \qquad (3)$$

The height measurement method described above is intended to be used in our intelligent video surveillance system measuring the heights of persons visible on the camera images. This raises some additional problems, as the measured height varies over time, involving some periodicity because of the gait, and moreover the amplitude of this periodic deviation also varies over different distances from the camera. We tackled this by compensating the periodic deviation using a sliding windowed average of the last 10 measured heights, but some more sophisticated solutions can refine the gained results further.

5 RESULTS

First we tested our implementation by manually selecting the top and bottom points of some known height still objects in the camera image. We have done measures on 10 different known height objects from various distances and in various part of the camera image. An adjustable height pole was used as a known height object, and 10 different measures were done for each height from different distances and with different horizontal view angles. The standard deviation of all 100 measures was about 3.1 cm and the maximum deviation from the real height was 5.5 cm. It was not surprising that the relative error increased with the distance between the camera and the subject, because the pixel size is proportional to the view angle, which means smaller resolution from larger distances. Speaking in numbers, with the camera and lens used in our experiments a one-pixel difference corresponds to about 0.00104π radians, which means about 1.63 cm resolution ambiguity for one pixel from distance of five meters.

The difference we experienced between the measured and the real heights of the tracked people was mainly caused by the inaccurate measuring of calibration parameters such as the altitude of the camera or the vertical orientation.

After the manual tests we embedded the developed height estimation method to the object tracking subsystem of our *Identrace* intelligent video surveillance system. This subsystem detects and tracks moving objects on surveillance camera images. The detected objects were segmented and their highest and lowest points were determined. We used a sliding windowed average calculation function to calculate the mean value of the various measured heights of the tracked object. The precision was tested with 3 known height moving persons in realistic videos. The results are show in the table below.

Table 1: The results of height estimation tests.

Data	Subj1	Subj2	Subj3	Total
Actual height (cm)	171	190	185	→
The average of 5 measures (cm)	169.2	188.6	184.2	→
The standard deviation (cm)	3.8	4.2	4.9	4.3
Maximal deviation from the real height	8	7	5	6.7

Regarding our experiences, during the tests the main source of inaccuracy was the false determination of the top and the bottom points of the tracked subjects. The lowest point of the image of a man walking towards the camera is somewhere around his toes for example, but this point is considerably closer to the camera than the top of his

head, which is his highest point. The inefficiency of our shadow compensation method caused the other common problem with the determination of these points, because some shadow area was often segmented together with the moving object. As we were focusing on the accuracy of our height measurement method, we have chosen image sequences with perfect illumination, in which this effect of shadows was not determinant.

Figure 2: Screenshot of height estimation in Identrace with an actual estimation of height of a tracked person.

As you can see in Figure 4, the contour of the moving objects along with the estimated height of these objects is shown in centimeters in our test application. The value after the label 'Mean' means the average value of the measured heights, while the number after the label 'Curr.' means the currently measured height

4 CONCLUSIONS

The first tests of our proposed solution have proved that the method is accurate enough to allow its application in intelligent video surveillance, where people can be remotely identified and their identity can be continuously tracked using the height on video surveillance camera images.

ACKNOWLEDGEMENTS

The research is ongoing as part of the Integrated Biometric Identification Systems project (identification number 2/030/2004), with the support of the National Research and Development Programme (NKFP) of the National Office for Research and Technology (NKTH), Hungary.

REFERENCES

Anil K. Jain, Sarat C. Dass and Karthik Nandakumar, 2004. Can soft biometric traits assist user recognition? *Proceedings of SPIE Vol. 5404, pp 561-572.*

Ernő Jeges, László Máté, 2006. Model-based human ear identification. *World Automation Congress, 5th International Forum on Multimedia and Image Processing (IFMIP).*

C. Madden and M. Piccardi, 2005. Height measurement as a session-based biometric for people matching across disjoint camera views. *IAPR.*

Chiraz BenAbdelkader, Ross Cutler, Larry Davis, 2002. View-invariant Estimation of Height and Stride for Gait Recognition; *Proceedings of ECCV Workshop on Biometric Authentication, p155.*

Liang Wang, Huazhong Ning, Tieniu Tan, Weiming Hu, 2003. Fusion of Static and Dynamic Body Biometrics for Gait Recognition. *Proceedings of the 9th IEEE International Conference on Computer Vision, Vol 2, p1449.*

Robert T. Collins, Ralph Gross and Jianbo Shi, 2002. Silhouette-based Human Identification from Body Shape and Gait; *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition, p0366.*

R. D. Green and L. Guan, 2004. Quantifying and Recognizing Human Movement Patterns from Monocular Video Image. *TCSV.*

Lv, Zhao, Nevatia, 2002. Self-Calibration of a camera from video of a walking human. *Proceedings of the 16th International Conference on Pattern Recognition, p562-571.*

Keni Bernardin, Hazim Kemal Ekenel and Rainer Stiefelhagen, 2006. Multimodal Identity Tracking in a Smartroom; *3rd IFIP Conference on Artificial Intelligence Applications and Innovations (AIAI).*

Identrace – Intelligent Video Surveillance system. http://www.identrace.com/

Antonio Criminisi, 2002. Single-view Metrology: Algorithms and Applications. *24th DAGM Symposium, vol 2449, p224-239.*

Petr Hujka, 2004. Model of Geometric Distortion Caused by Lens and Method of its Elimination. *ElectronicsLetters.com, 1/4/2004.*