Forensic photo/videogrammetry; Monte Carlo simulation of pixel and measurement errors

Jurrien Bijhold and Zeno Geradts
Netherlands Forensic Science Laboratory / Gerechtelijk Laboratorium,
Department Information Technology,
Volmerlaan 17, 2288 GD Rijswijk Netherlands, email bijhold@holmes.nl

ABSTRACT

In this paper, we present some results from a study in progress on methods for the measurement of the length of a robber in a surveillance video image.

A calibration tool was constructed for the calibration of the camera. Standard procedures for computing the lens distortion, image projection parameters and the length of the robber have been implemented in Mathematika. These procedures are based on the use of pixel coordinates of markers on the calibration tool, the robber's head (and, optionally, his feet) and an estimation of his position in the coordinate system that is defined by the calibration tool.

Monte-Carlo simulation is used to compute a histogram of the robber's length, yielding an estimation of minimum and maximum values. In a repeated process pixel and position coordinates are selected randomly from predefined ranges and, using these data sets, the length, position and quality of the fit are computed and stored. The range of the length in the histogram can be made smaller by selecting only those data sets that can meet one or more constraints, e.g. the quality of the fit should be good and the position of the robber should be within the physical limits of the scene. Some experimental results are presented and discussed.

Keywords: forensic, photogrammetry, videogrammetry, Monte Carlo simulation, conditional sampling, surveillance video

1.INTRODUCTION

In the last two years, the Netherlands Forensic Science Laboratory was asked a number of times to identify suspects of robberies in surveillance video images. Comparison of faces appeared to be non-conclusive in most of these cases, due to the small number of facial features that are visible and distinctive. More information can be obtained by performing biometric measurements [1,2]. The length of the robber is often mentioned in wanted signs and can be further used for excluding suspects. This study is focussed on measuring the length of the robber in the video images. When sufficient experience has been gained with measuring length in practice, this study will be continued with measuring other biometrics.

Figure 1. *Two examples of video surveillance images*

Figure 1 shows two examples of surveillance video images of a bank robber that illustrate some difficulties:
- the position of the robber relative to the entrance is unknown when the feet are not inside the view;
- there are not many reference points available in the scene to calibrate the camera;
- there can be a significant lens distortion that should be taken into account.

Methods that depend on measurements on the scene and/or a view of the robber's feet can not be applied in these cases. Another drawback of most existing methods is that they can not cope efficiently with uncertainties in pixel positions and reference measurements. Most error propagation techniques assume that uncertainties are distributed normally. For these reasons new tools and methods for calibrating the camera and performing the measurements have been developed and tested.

2.MATERIALS AND METHODS

2.1 GENERAL DESCRIPTION

The following principles have been utilized:
- correction of lens distortion and calibration of the perspective projection can be separated [3] into different processes. Lens distortion can be considered as not significant or as corrected for when straight edges of objects appear as straight lines in the image. Figure 2 shows an image of a reference board with a raster that was placed in front of the camera for estimating the lens distortion. Accurate positioning of the board is not necessary;
- while the police are at the bank for investigations and to pick up the videotape with the images of the robber, they can move a reference object along the route of the robber through the bank and place a raster board in front of the camera. The reference object that was tested in this study is shown in figure 3. The design of the circle shaped markers allows for an estimation of the center positions with sub-pixel-resolution accuracy. The position of the robber is estimated relative to this reference object. The tape that is in the surveillance system during their visit should also be picked up. Such a procedure saves time and eliminates the need for measuring objects at the scene.
- Monte-Carlo simulation is applied to compute a histogram for the length of the robber from which a minimum and maximum value can be selected. The width of the histogram is determined by the propagation of uncertainties and additional constraints. These constraints can represent prior knowledge such as the practical limits for the position of the robber within a door opening.

Figure 2. *Raster image*

Figure 3. *Calibration tool*

2.2 LENS DISTORTION

The lens distortion is represented by a function that models the deviations from the ideal pinhole camera

$$\begin{pmatrix} x_d / H \\ y_d / V \end{pmatrix} = \begin{pmatrix} x_u / H \\ y_u / V \end{pmatrix} + k r_u^2 \begin{pmatrix} x_u / H - c_x \\ y_u / V - c_y \end{pmatrix}$$

In which:

$$r_u^2 = s_x^2 \left(x_u / H - c_x \right)^2 + \left(y_u / V - c_y \right)^2$$

This function returns the pixel coordinates of the distorted pixel (x_d, y_d) from the coordinates of the undistorted pixel (x_u, y_u), the image size in pixels (H,V) and 4 parameters: k is the strength of the radial distortion, s_x is the distortion aspect ratio, and (c_x , c_y) is the center of distortion with coordinates specified in the range 0-1. The last three parameters are related to the position and orientation of the lens relative to the CCD-chip. Figure 4 demonstrates the working of this distortion model on a rectangle with a negative and a positive value of k.

Figure 4. *Distortion of a rectangle with three different values of k: -0.35, 0, +0.35*

This distortion can be compensated for using the same model. The following procedure was implemented:
* Select edges in the image that should be straight and store for every line the pixel coordinates of two end points and a point in the middle.
* Compute for every edge the distance between the line through the end points and the middle point.
* Compute the sum of the squared distances from every edge as a measure for the distortion.
* Find a distortion transformation of all points that minimizes the distortion measure.

The minimization is performed by the gradient descent method provided by the Mathematika software. Figure 5 shows two graphs representing the edges in the calibration image in figure 2 and the result of the compensating distortion.
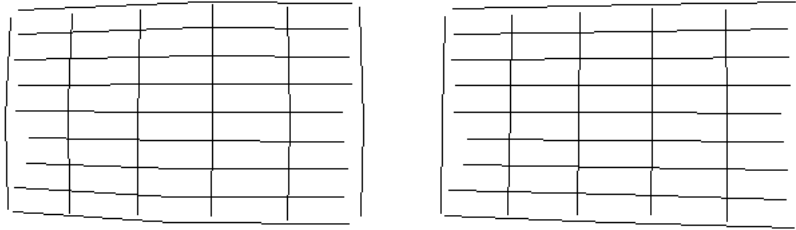

Figure 5. *Two graphs from lines through three points. Left: the graph created from points in figure 2. Right: resulting graph from compensation for lens distortion.*

2.2 PERSPECTIVE PROJECTION

A point with real world coordinates (X,Y,Z) is projected onto an image with pixel coordinates (x,y) using intermediate coordinates (X',Y',Z') by:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{F}{Z'} \begin{pmatrix} as & 0 \\ 0 & s \end{pmatrix} \begin{pmatrix} X' \\ Y' \end{pmatrix}$$

in which *F* is the focal distance, *s* is a scaling factor for the pixel size and *a* is the aspect ratio of the pixel sizes. Since this equation only works in the coordinate system of the camera, the following rotations (alpha, beta, gamma) and translations (T_x, T_y, T_z) have to be performed to align the coordinate systems of the camera and the real world:

$$
\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & Cos(\alpha) & Sin(\alpha) \\ 0 & -Sin(\alpha) & Cos(\alpha) \end{pmatrix} \begin{pmatrix} Cos(\beta) & 0 & -Sin(\beta) \\ 0 & 1 & 0 \\ Sin(\beta) & 0 & Cos(\beta) \end{pmatrix} \begin{pmatrix} Cos(\gamma) & Sin(\gamma) & 0 \\ -Sin(\gamma) & Cos(\gamma) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}
$$

The rotations around the camera axes are called: swing (Z-axis, gamma), pan (Y-axis, beta) and tilt (X-axis, alpha).

This equation can be written in a linear format with 11 parameters that have to be estimated from a calibration using the known real world coordinates of the markers in our calibration tool in figure 3 and the corresponding pixel coordinates:

$$
a_{11}X_i + a_{12}Y_i + a_{13}Z_i + b_x - a_{31}X_i x_i - a_{32}Y_i x_i - b_z x_i = Z_i x_i
$$
$$
a_{21}X_i + a_{22}Y_i + a_{23}Z_i + b_y - a_{31}X_i y_i - a_{32}Y_i y_i - b_z y_i = Z_i y_i
$$

In which the parameters a_{ij} and b_i represent a matrix that combines rotation and scaling and the translation, respectively. The eleven unknown parameters can be solved for by writing this equation for a number of points i in a vector-matrix notation:

$$
\begin{pmatrix} X_1 & Y_1 & Z_1 & 0 & 0 & 0 & -X_1 x_1 & -Y_1 x_1 & 1 & 0 & -x_1 \\ 0 & 0 & 0 & X_1 & Y_1 & Z_1 & -X_1 y_1 & -Y_1 y_1 & 0 & 1 & -y_1 \\ & & & & & & & & & & \\ & & & & & & & & & & \\ X_i & Y_i & Z_i & 0 & 0 & 0 & -X_i x_i & -Y_i x_i & 1 & 0 & -x_i \\ 0 & 0 & 0 & X_i & Y_i & Z_i & -X_i y_i & -Y_i y_i & 0 & 1 & -y_i \end{pmatrix} \begin{pmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{21} \\ a_{22} \\ a_{23} \\ a_{31} \\ a_{32} \\ b_x \\ b_y \\ b_z \end{pmatrix} = \begin{pmatrix} Z_1 x_1 \\ Z_1 y_1 \\ \\ \\ Z_i x_i \\ Z_i y_i \end{pmatrix}
$$

This equation can be written as:

$$ Ap = b $$

A solution for p that yields a least squared error, is given by [4]:

$$ p = \left(A^T . A \right)^{-1} A^T b $$

This solution can only be obtained when at least 6 points have been used to set up the equation. The covariance matrix for the parameters in p is given by $(A^T.A)^{-1}$. This matrix can be used for computing error propagation in case the pixel estimation errors are normally distributed.

2.3 LENGTH MEASUREMENTS

The linear format equations that are used for calibrating the perspective projection are also used for computing the length of a robber. After calibration, the perspective parameters a and b have become constants and real world coordinates (X,Y,Z) are now the unknowns. We distinguish two cases: images with a view of head and feet and images that only show the upper half of the robber's body. In all cases we assume that the robber is standing right up on the floor. Real world coordinates of feet and head are now: (X,0,Z) and (X,L,Z) respectively, in which L is the robber's length. In a view of head and feet four

equations can be set up with estimated pixel coordinates of head and feet which is more than sufficient to solve for X,Z and L. In this study we have used only three equations to keep the experiments simple. In a view without feet, only two equations can be set up with the pixel coordinates of the head. In this study we use an estimate for the Z-coordinate for solving L and X.

2.4 MONTE CARLO SIMULATION

During the computations many estimations have to be made. All these estimations produce errors that are distributed in some way. The estimates of the center pixel positions of the markers could be the most accurate when interpolation is used. The estimation of the pixel position of the ground level between feet is probably the least accurate. The distribution of estimation errors in the X or Z coordinate is unknown. In case there are physical limitations like the space available in a door opening or the space in front of a desk, It can be argued that the distribution has one or two tails. In the Monte Carlo Simulation a value for a coordinate is selected from a predefined distribution. In our experiments we tested a situation where the estimated pixel coordinates are discrete and homogeneously distributed in a 5x5 template at the point of interest. Further, we assume that the X or Z-coordinate is distributed homogeneously in a range, specified by a minimum and maximum value.

The Monte Carlo Simulation is used to compute a histogram of the length by the following process:
1. Enter distribution parameters for the pixel coordinates of every point that is used in the computations
2. Select randomly values for all coordinates from the distributions
3. Compute the lens distortion and transform all pixel coordinates to compensate the lens distortion
4. Compute the perspective projection parameters
5. Select randomly a value for the X or Z coordinate when the image do not show the robber's feet
6. Compute the length and upgrade the histogram
7. Return to step 2 as long as the histogram changes significantly.

2.5 CONDITIONAL SAMPLING

During the Monte Carlo simulation combinations of pixel coordinates can be selected that are not very likely because e.g. they yield a large value for a distortion measure: D, or a large value for a measure of the quality of the fit in the calibration: Q, or a value for the X or Z coordinate that is physically impossible. The resulting values for the robber's length could be excluded from the histogram by performing some tests on every set of randomly selected coordinate values.

3.EXPERIMENTS

3.1 NUMERICAL PROCEDURES

All the experiments are based on three images that were acquired with the O2-Cam from Silicon Graphics. Figure 5 shows one of our electronic engineers in our video laboratory. The camera was fixed at the ceiling and pointed downwards to get a camera view that is typical in our video investigations. The images in figure 2 and 3 are used for computing the parameters of the lens distortion and the perspective projection. Note that the position of the calibration tool is not exactly equal to the position of the marker between the feet in figure 2.

Pixel positions for all relevant points in images 2,3 and 6 have been estimated only once and were entered manually in lists for processing. All routines for data processing have been written in Mathematika. The source file is free and can be obtained by writing to one of the authors. Random selection of pixel coordinates from a distribution was replaces by adding discrete white noise with values: (-1,0,1) to the pixel coordinates in the list. All the routines were tested using virtual data from virtual objects and cameras.

Figure 6. *Test image that was used in all experiments for length measurements. There are markers on his head and between his feet.*

While the procedures were being tested, it appeared that the computation of the distortion parameters was a time consuming process. For this reason, a table was computed with 50 sets of distortion parameters by Monte Carlo Simulation. In all other experiments the computation of distortion parameters from randomly selected pixel coordinates was replaced by picking randomly a set of distortion parameters from this table.
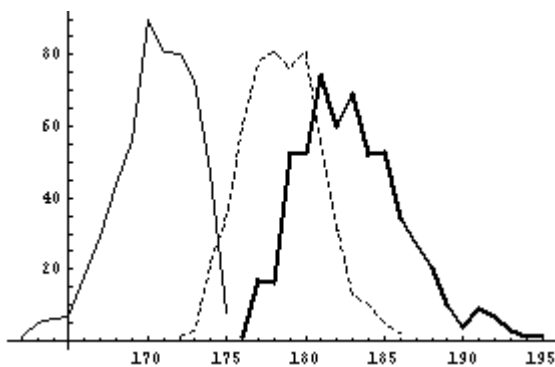
3.2 INFLUENCE OF LENSE DISTORTION

Figure 7. *Histogram of the height of the top marker from 1000 data samples. Thin line: no compensation for lens distortion. Broken line: lens distortion from raster image. Thick line: lens distortion form edges in the scene.*

The first experiment was performed to test the influence of compensating lens distortion. Figure 7 shows three histograms for the length of the calibration tool using the pixel coordinates of the marker at the top and the marker on the ground in front of the tool. All histograms have been computed from 1000 data samples. The thin line shows the histogram obtained by leaving out the compensation for the lens distortion. The broken line and thick line show a shift of this distribution caused by including the compensation of the lens distortion based. The thin line is computed using the raster image in figure 3 and the thick line is computed using the edges that are available in the image of the scene. The actual height of the top marker was 179.5 cm. This figure demonstrates that the raster image yields more precise results since the range of the histogram is smaller.

3.3 INFLUENCE OF CONDITIONAL SAMPLING

Then, experiments were performed to demonstrate the influence of conditional sampling on the shape of the histograms. Figure 8 shows three histograms for the length of the engineer computed from 1000 data samples and different conditions.

The broken line shows the result of testing the quality of the calibration. The average square distance between the actual marker pixel positions and the back projected pixel positions had to be smaller or equal to one. This constraint appeared to have a small influence on the size of the tails. The thin line is obtained by an additional test of the X-coordinate. This value had to be in the range of –5 cm to +5 cm. The tick line was obtained by an additional test of the Z-value. This value had to be positive, since the images 3 and 6 show that the tool is in front of the engineer. This result shows clearly that conditional sampling can improve our results at the cost of an increasing number of data samples that are required.
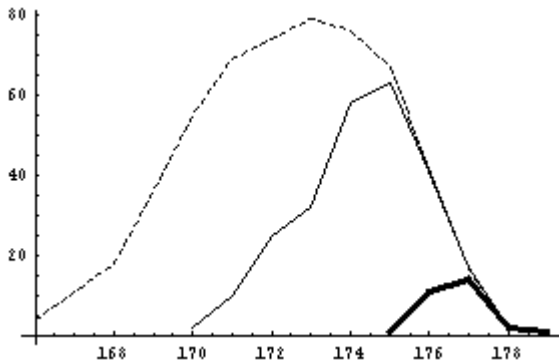

Figure 8. *Histogram of the engineer's length from 1000 data samples. Broken line: samples with good quality of calibration fit. Thin line: samples with X-coordinate within –5 cm and +5 cm, Thick line: samples with Z-coordinate > 0.*

These results were not completely satisfactory, since the actual length of the engineer was 178.5 cm. The actual length is in the tail of the first two histograms, but it was expected to find this value in the peak of each histogram. Additional experiments with images of the calibration tool on other positions showed that this deviation was proportional with the distance between the positions of the engineer and the tool. Experiments with test data from virtual images showed that the calibration tool should be positioned in such a way that the robber is within the range of the markers.

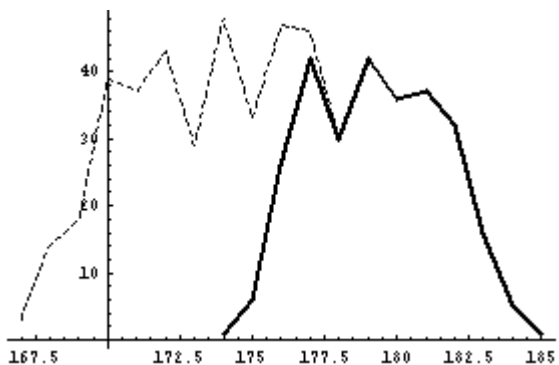## 3.4 RESULTS FROM A VIEW WITHOUT FEET

Figure 9. *Histogram of the engineer's length in a view without feet from 1000 data samples. Broken line: no data selection, Z-coordinate is randomly selected from (–50 cm to +50 cm). Thick line: samples with X-coordinate within –5 cm and +5 cm.*

In the last experiments, we tested the equations for the length from the pixel coordinates of the marker on the head and an estimate for the Z-coordinate. Figure 9 shows three histograms computed from 1000 data samples. The range for the Z-coordinate was chosen to be (–50 cm, +50 cm). The thin line shows the result obtained without constraints. This histogram has wider ranges than the histograms in figure 8. The thick line shows the result of testing the value of the X-coordinate. Again, this value had to be within the range of –5 cm and +5 cm.

4. DISCUSSION AND CONCLUSIONS

From the results in this study the following remarks and conclusions are given:

- Lens distortion can not be neglected in these experiments. The raster image appeared to give the best results. Experiments with methods that utilize all pixels on edges in the image of a scene[3] should be performed to demonstrate that a raster image is necessary in practice or not.
- Sub pixel resolution estimation of the marker positions could improve the results for cases where the robber is outside the ranges of the markers on our calibration tool. This should be demonstrated in other experiments.
- The calibration tool should be redesigned to allow for easy transportation
- Other constraints for conditional sampling could come from known distributions of biometric measures like e.g. the size of the head, the width of the shoulders or the distance between the eyes. This should also be demonstrated by experiments.
- For the purpose of conditional sampling robust measures for the lens distortion and the quality of the calibration fit could be used in stead of the commonly used squared value measures.
- Monte Carlo Simulation might be not the most efficient way to compute a distribution function for the length, but it has the advantage that the method can be rather easily explained in court. Most other methods focus on the computation of minimum or maximum values for the length, and do not give information about the probability of these extreme values.
- The histogram from Monte Carlo Simulation looks like a probability distribution. A theoretical study will be carried out to find out in which aspects they are equal and in which aspects they differ.

5. ACKNOWLEDGEMENTS

6. REFERENCES

1. Lena Klasen and Olov Fahlander, "Using videogrammetry and 3-D image reconstruction to identify crime suspects", SPIE Vol. 2942, "Investigative Image Processing" , editors: L. Rudinamd and S. Bramble , pp.162-169, Boston, 1996
2. Stephen C. Jensen and Leonid Rudin, "Measure: an interactive too for accurate forensic photo/videogrammetry", SPIE Vol. 2567, "Investigative Image Processing" , editors: L. Rudinamd and S. Bramble , pp.73-83, San Diego, 1995
3. Frederic Devernay and Olivier Faugeras, "Automatic calibration and removal of distortion from scenes of structured environments", SPIE Vol. 2567, "Investigative Image Processing" , editors: L. Rudinamd and S. Bramble , pp.62-72, San Diego, 1995
4. J. Stoer and R. Burlisch, "Introduction to numerical analysis", (New York: Springer, 1980)