Calibration maintenance for a four camera acquisition system

Oleg Stepanenko
Vocord Company
Moscow, Russia
oleg.stepanenko@vocord.ru

Abstract

This paper discusses the method of calibration maintenance for four camera acquisition system (the mentioned system will be referenced as the 3D system in this paper). 3D system creates 3D models of human faces. The relative orientation between two vertical stereo pairs could slowly change in time because of a vibration. Thus camera calibration needs to be maintained.

We show how 3D system can maintain calibration without being stopped. The new calibration parameters are being continuously recalculated from new multiple dynamic scene images and previous calibration.

Keywords: *Camera calibration, Stereo, Essential parameters estimation.*

1. INTRODUCTION

3D system consists of four cameras that have to be calibrated (Fig.1). The cameras are synchronized.

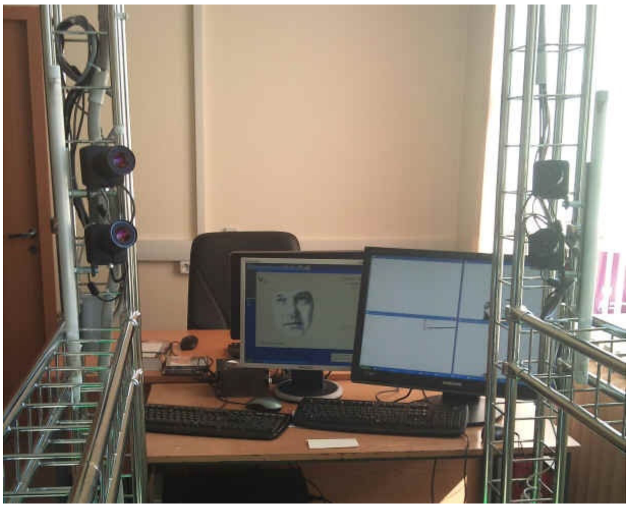

Figure 1: Four camera 3D acquisition system.

Four cameras are coupled in two vertical stereo pairs. Two cameras in each pair are mounted at the common stable basis and have constant orientation in respect to each other.

We had observed that 3D system with two stereo pair is much superior in respect to one that has only one stereo pair. Each stereo pair recovers only its half of the face. Taken from left and right sides the face is fully recovered under wide angle of rotation.

Coupling left and right vertical stereo together however puts another problem: the relative orientation between two vertical stereo pairs may have small alteration because of the vibration of constructive elements which position stereo pairs in space. Thus 3D models of human faces will have alteration in time. In Fig.2 we present a 3D model of human face before the calibration has been changed. In Fig.3 we present a 3D model of same human face after the calibration has been slightly changed. In Fig.3 the face is distorted obviously: nose becomes shorter.

The classical camera calibration [4, 5] is performed by capturing a reference object with a known Euclidean structure (for example, chessboard pattern). This approach can be reasonably used on stages when 3D system is stopped. On those stages 3D system cannot create 3D models of human faces. But 3D system cannot be stopped every time we want to find out if calibration has changed considerably or not. There are also severe conditions in which human traffic may exist all day and night, thus the face recognition system has to operate without break (i.e. at airports).

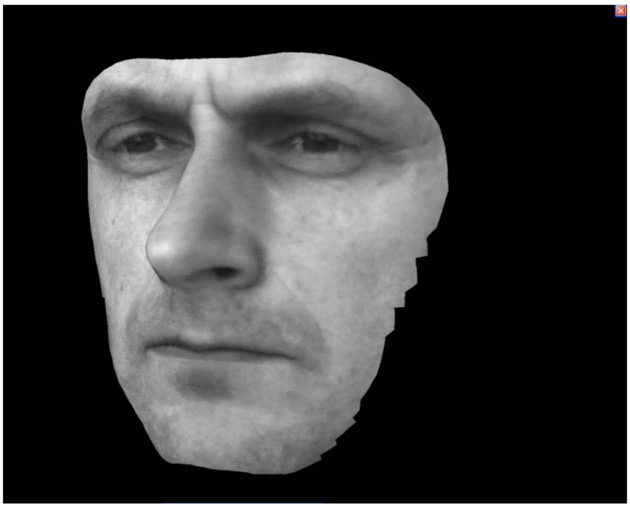

Figure 2: 3D model of human face before camera calibration has been changed

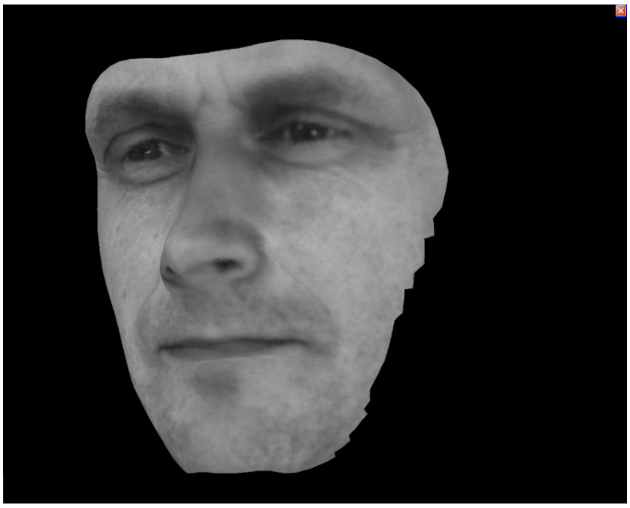

Figure 2: 3D model of human face after camera calibration has been changed (the shape of nose gets a distortion)

A question arises: how new camera calibration can be obtained from dynamic scene images and previous calibration?

We use the following features of the mentioned problem: cameras in pairs keep relative orientation, the observed scene is dynamic.

2. NOTATION AND PROBLEM DEFINITION

2.1 Notation

The pinhole camera model is used here. A 3D point has coordinates $M = [x, y, z]^T$ in a world coordinate system. The retinal image coordinates of a 3D point are $m = [u, v]^T$. World coordinates and retinal image coordinates are related by

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \text{ or } s\tilde{m} = \mathbf{P}\tilde{M},$$

where s is a scale coefficient, $\mathbf{P}$ is a 3×4 perspective projection matrix. We use tilde sign to denote the augmented vector $\tilde{x}$ (adding 1 as its last element) of a vector x.

Matrix P can be written as $P = \mathbf{A}[\mathbf{R}\,\mathbf{t}]$ with $= \begin{bmatrix} \alpha_u & c & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$, where $\mathbf{A}$ is a 3×3 matrix of intrinsic parameters, and $(\mathbf{R}, \mathbf{t})$ is a displacement (rotation and translation) from the world coordinate system to the camera coordinate system. $\mathbf{R}$ is a 3×3 rotation matrix. $\mathbf{t} = [t_x, t_y, t_z]^T$ is a translation vector. Translation may be presented in a form of skew-symmetric matrix: $[\mathbf{t}]_\times = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & t_x \\ -t_y & t_x & 0 \end{bmatrix}$.

One of the cameras (top camera from one pair) is chosen as reference camera and has $\mathbf{R} = \mathbf{I}$ and $\mathbf{t} = \mathbf{0}$, where $\mathbf{I}$ is 3×3 identity matrix and $\mathbf{0}$ is zero 3×1 vector . The parameters in $\mathbf{A}$ are obtained through calibration procedure [4, 5].

2.2 Problem definition

We consider multiple perspective images of a dynamic scene, and have to determine the relation between the multiple images and the camera pair displacement. This arises from a situation described further. Four cameras take an image sequence. We assume the images are projections of a moving human (to be specific, top part of human body), the cameras in pairs are calibrated (i.e. their intrinsic parameters are known), cameras in pairs have a known displacement (orientation and translation).

3. ALGORITHM

3.1 Main steps

The main steps of the algorithm are:

Step 1: Establish two sets of pixel correspondences between camera pairs at multiple points of time in a dynamic scene. One set is established between two images that are taken by top cameras. The other set is established between two images that are taken by bottom cameras (Fig.4). Select a set that is bigger.

Step 2: Establish pixel correspondences between two images that are taken by top and bottom cameras from each pairs at multiple points of time in a dynamic scene (Fig.5).

Step 3: Triangulate 3D point sets from pixel correspondences that have been established on the step 1 and 2. Thus we have two sets of 3D points that are triangulated from one pair and the other pair.

Step 4: Match 3D points from sets that are mentioned above in step 3. Form set of pixel correspondences from 3D matched points.

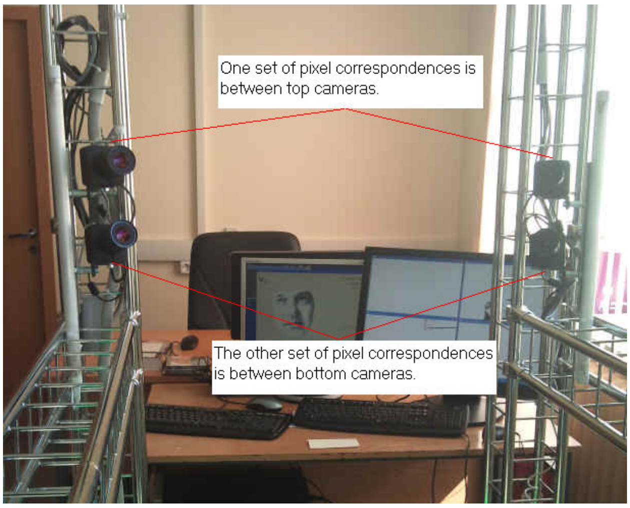
One set of pixel correspondences is between top cameras.
The other set of pixel correspondences is between bottom cameras.

Figure 4: Two sets of pixel correspondences that are establishing at the step 1

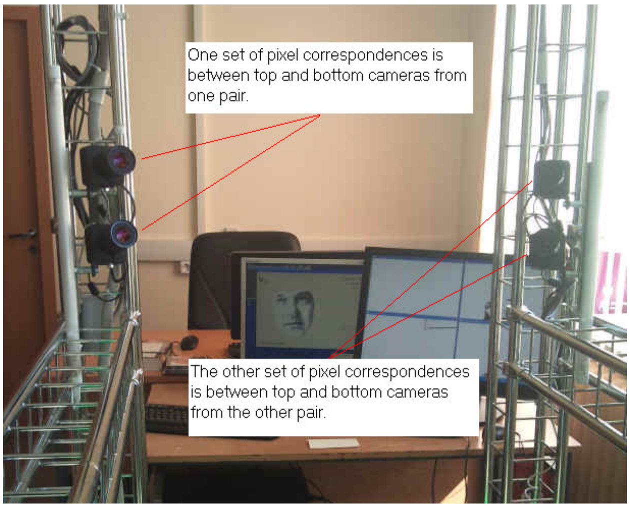
One set of pixel correspondences is between top and bottom cameras from one pair.
The other set of pixel correspondences is between top and bottom cameras from the other pair.

Figure 5: Two sets of pixel correspondences that are establishing at the step 2

Step 5: Estimate the essential parameters with 8-point algorithm [3]. The obtained matrix is denoted by $\boldsymbol{E} = [\mathbf{t}]_\times \mathbf{R}$. Recover the displacement parameters $\mathbf{t}$ and $\mathbf{R}$ from $\mathbf{E}$. This displacement takes place between top cameras from different pairs. One of those top cameras is the reference camera.

Step 6: Refine the displacement parameters.

3.2 Detection of false matches on step 4

In order to detect false matches on step 4 we have to establish correspondences between two sets of 3D points. The criterion being minimized is the 3D Euclidean distance between points. Suppose we have i^{th} and j^{th} 3D points from the sets of points that has been triangulated from top and bottom cameras of one pair (coordinates are M_i) and the other pair (coordinates are M_j). The Euclidean distance from point M_i to point M_j is $d(M_i, M_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$.

After matching procedure has ended we have n 3D point correspondences: $\{(M_i, M_j)\}$. For each of the mentioned correspondences we can determine $\{(m_i, m_j)\}$, where m_i are pixel coordinates of M_i on the image of top camera from one pair, m_j are coordinates of M_j on the image of top camera from the other pair.

Thus we have two sets of matched 3D points and matched pixel correspondences that are taken by two cameras from the different pairs.

3.3 Estimate the essential parameters with 8-point algorithm on step 5

Due to the presence of incorrect correspondences (outliers), essential parameters estimators must be robust. The Random Sample Consensus – RANSAC methods [1] have become the methods of choice for outlier removal in essential parameters estimation [2]. We use RANSAC-like techniques on step 5. We start by selecting (at random) a subset of k correspondences, which is then used to compute the essential parameters estimation. The cost function of the full set of correspondences is then computed. The cost function expresses numbers of inliers that are within a certain neighborhood form their predicted epipolar lines. The random selection process is repeated S times, and the sample set with largest number of inliers is kept as the final solution.

Assuming that the set of correspondences may contain up to a portion ε of outliers, the probability that one of S samples is good is given by $\mathrm{P} = 1 - (1 - (1 - \varepsilon)^{\mathrm{k}})^{\mathrm{S}}$. In our implementation, we determine $\varepsilon = 25\%$, k = 35, P =0.999, thus S=200000. The algorithm can be speed up considerably by means of CUDA technology.

The standard 8-point algorithm [3] is used to estimate the essential matrix.

3.4 Refine the displacement parameters on the step 6

The nonlinear minimization on step 6 is done with the Levenberg-Marquardt algorithm. The criterion being minimized is the sum of squared reprojection errors. The Levenberg-Marquardt algorithm is one of the most popular methods for iterative minimization, when cost function to be minimized is of this type [2]. The optimization is carried out for all displacement parameters.

In Fig.6 we present a 3D model of the same human face after calibration has been recovered. In Fig.6 a shape of the face is recovered: nose gets former shape.

4. CONCLUSION

In this paper we have used a general scheme of displacement estimation from multiple calibrated images [2, 3, 6] in the field of a four camera acquisition system.

Steps 1, 2, 3, 4 of our algorithm were developed specifically for our field of investigation and we haven't found any references to such methods implementation in literature.

3D system has to be accurately calibrated. But 3D system has not to be stopped every time when we want to do calibration procedure with reference object with a known Euclidean structure.

Due to the scene being dynamic we have matched points from cameras that fill the observed scene fairly uniformly.

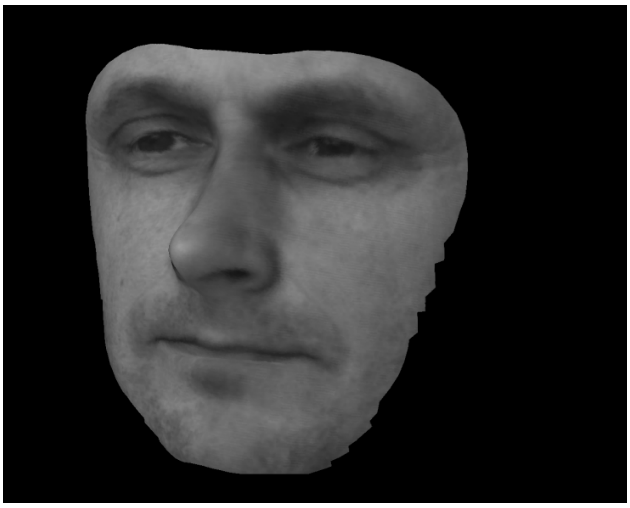

Figure 6: 3D model of human face after camera calibration has been recovered (nose gets former shape)

The new camera calibration can be obtained from dynamic scene images and previous calibration. Our experimental results suggest that our method can be applied when displacement alteration is rather small. Experience has shown that it was enough to run our algorithm once at the middle of the day. Nevertheless classical camera calibration must be done as soon as 3D system may have a break in its work (for example, at the end of the day or before a new day).

5. REFERENCES

[1] *M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. CACM, 24(6):381–395, June 1981.*

[2] *R. Hartley and A. Zisserman Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge, UK, 2000.*

[3] Y. Ma, S. Soatto, J. Kosecka and Shankar Sastry. *An Invitation to 3D Vision: From Images to Models. Springer Verlag, December 2003.*

[4] R. Y. Tsai. *A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf cameras and lenses. IEEE Journal of Robotics and Automation, 3(4): 323-344, Aug. 1987.*

[5] Z. Zhang. *A flexible new technique for camera calibration.* IEEE Transactions *on Pattern Analysis and Machine Intelligence*, Vol.22, No.11, pages 1330-1334, 2000

[6] Z. Zhang. *Motion and Structure From Two Perspective Views: From Essential Parameters to Euclidean Motion Via Fundamental Matrix. Journal of the Optical Society of America* , Vol.14, no.11, pages 2938-2950, 1997

About the author

Oleg Stepanenko (Ph.D, Associate Professor) is a scientist at Vocord Company, Department of Advanced Developing. His contact email is oleg.stepanenko@vocord.ru.