A self-calibration method for a multi stereo-rig system based on robust 3D motion estimation

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

Abstract

This paper discusses self-calibration method for a multi stereo-rig passive face recognition system with overlapping views of individual cameras.

Two cameras fixed inside a rig can be easily calibrated in a factory environment so each rig will be considered individually calibrated. It is incredibly important to estimate relative positions and orientations of the stereo rigs with respect to each other without using any calibration objects. We propose calibration method using point correspondences between all cameras.

In this paper we present a novel more robust method that calculates metric reconstruction. Our method reconstructs 3D point set for every stereo rig from multiple dynamic scene images. Point correspondences are established by tracking points over all images captured simultaneously. So pairing between points is known but data keeping outliers. Then RANSAC technique is applied to reject outliers and finally 3D motion is estimated.

In this paper it is shown that the presented method can be accepted as sufficient in accuracy. Our method can be easily adopted for various numbers of stereo rigs.

Keywords: *Self-calibration, 3D motion estimation, stereo-rig.*

1. INTRODUCTION

Passive 3D face recognition system is the area of intense research over the past decade. A wide range of 3D acquisition technologies, with different cost and operation characteristics exists [1,10].The most cost-effective solution is to use several calibrated 2D cameras fixed inside a rig to capture images simultaneously, and to reconstruct a 3D surface [11] (Fig.1A).

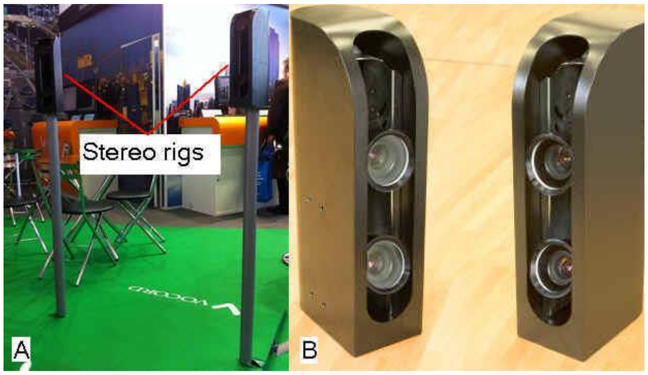

Figure 1: Passive 3D faces recognition system: A – common view; B – stereo-rigs.

The term *stereo rig* is used in this paper to refer to any two-camera system, which comprises a set of two cameras (with overlapping views) that are physically connected together and capture images simultaneously (Fig.1B).

Passive 3D face recognition system using stereo-rigs demands an accurate calibration of the devices which includes, first, intrinsic parameter measurement and estimation of the relative poses of the cameras with respect to each other inside a rig, second, estimation of relative positions and orientations of the stereo rigs with respect to each other. Our paper focuses on the second part of the calibration procedure (estimation of relative positions and orientations of the stereo rigs with respect to each other).

Roughly speaking, there are two groups of calibration methods. The first group is based on using of some object with known geometry (calibration pattern) or moving single feature like an LED [13, 15, 19].

One of such methods can be used for a calibration of two cameras fixed inside each rig. Due to the fact that calibration can be performed in a factory environment, stereo rigs are considered individually calibrated in this paper. Since the goal of this paper is to calibrate a multi-camera system without using any calibration objects, methods from the first group are out of our consideration.

The second and most suitable group of methods is self-calibration. Both the scene shape (3D structure) and the camera parameters (motion) consistent with a set of correspondences between scene and image features are estimated using this group of methods [3, 8, 17, 20].

The main part of calibration literature from the mentioned second group of calibration methods concerns extrinsic calibration. The goal of extrinsic calibration is to determine the 3D rotation and translation (3D motion) parameters relative to a fixed coordinate system. The estimations are based on 2D point features as they appear in an image sequence. Such methods are called structure from motion (SFM) methods. Usually 3D motion estimation methods involve two steps: first, 2D motion estimation that might be represented by 2D displacement vector field and, second, calculation of 3D motion from 2D displacements.

We propose to use 3D displacements for 3D motion estimation in this paper instead of using 2D displacement. We have 2D features as an input of our algorithm. We reconstruct 3D point set for every stereo rig using given calibration. So we receive sets of 3D displacements (keeping outliers) that can be used for estimating of 3D motion between 3D point sets (and consequently between stereo rigs). Such kind of 3D motion estimation methods using 3D points are sometimes called 3D alignment [16].

There are two main approaches to the problem of 3D motion estimation. If pairing between 3d points is known, closed form (analytic) solution is suitable [2, 6]. But analytic solution is breaking down in presence outliers (even if we have only one outlier). If pairing between points is unknown, iterative algorithms that start by matching nearby points and then update the most likely correspondence can be used [4].(See [13] for an overview of applications.). Iterative algorithms require good initialization and they are sensitive to overlap and outliers [4]. Some approaches are proposed to deal with 2D feature correspondences selection for robust camera calibration [12, 14, 18, 20]. However, reliable and automatic SFM is a difficult problem so far [18].

The paper is organized as follows. The proposed method is described in Section 2. Experiments on several multi-camera sequences are presented in Section 3. Conclusions are given in Section 4.

2. SELF–CALIBRATED METHOD

2.1 The overview of multi stereo-rig self-calibration method

The overview of multi stereo-rig self-calibration method we propose is shown in Fig.2.

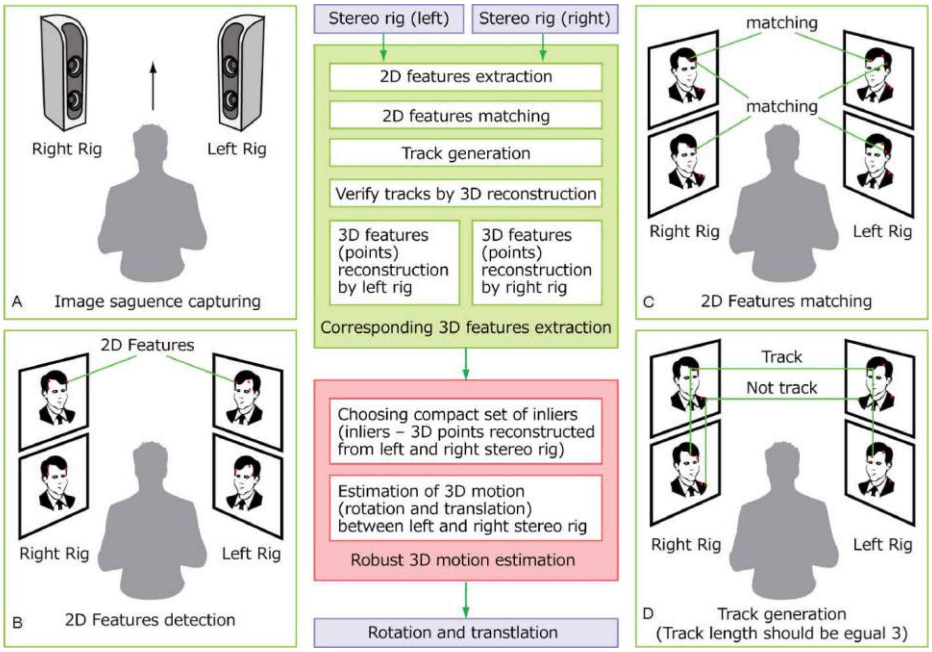
Right Rig
Left Rig
A
Image saguence capturing
Stereo rig (left)
Stereo rig (right)
2D features extraction
2D features matching
Track generation
Verify tracks by 3D reconstruction
3D features (points) reconstruction by left rig
3D features (points) reconstruction by right rig
Corresponding 3D features extraction
matching
matching
Right Rig
Left Rig
C
2D Features matching
2D Features
Right Rig
Left Rig
B
2D Features detection
Choosing compact set of inliers (inliers – 3D points reconstructed from left and right stereo rig)
Estimation of 3D motion (rotation and translation) between left and right stereo rig
Robust 3D motion estimation
Rotation and transtlation
Track
Not track
Right Rig
Left Rig
D
Track generation
(Track length should be egual 3)

Figure 2: The overview of multi stereo-rig self-calibration method: A – captured image sequence, B – 2D features detection, C - 2D features matching, D – track generation

The method mainly includes two stages, corresponding 3D feature extraction and robust 3D motion estimation. To obtain corresponding points, we let a person walks between stereo rigs (Fig.1A).

By analyzing the motion of human body from synchronous video sequences, we can find 2D points suitable for corresponding point (Fig.1B). Matching is performed to find corresponding points on the images captured simultaneously (Fig.1C).

Once we have pairwise matches, next step is to link up matches to form tracks (Fig.1D). Each track can potentially grow up to become eventually a 3D point. Some tracks might be inconsistent. Track should have length equal three in order to be consistent. We remove inconsistent tracks on the feature extraction stage. Then generated tracks are verified by 3D triangulation. On this step reconstructed 3D points should have distance from the appropriate stereo rig within certain reasonable range (for example, from 0.5 m to 1.5 m).

The next stage of our method is robust 3D motion estimation.

2.2 3D motion estimation algorithm based on RANSAC

The basic 3D alignment algorithms presented in literature are sensitive to outliers in the data [16]. As 3D tracks (and appropriate points correspondences) automatically extracted from images will almost always contain false matches, robustness with respect to outliers is very important. In this section, we will describe algorithm for this.

2.2.1 Problem definition

Given a set of point correspondences $=\{(P_{1,j},P_{2,j})\in \mathrm{P}^3\times \mathrm{P}^3 | 1\le j\le m\}$ measured in two Cartesian coordinate systems (left stereo rig, right stereo rig) find the rigid transformation (rotation and translation) R,T, between the two systems so that for corresponding points P_1 (from left coordinate system) and P_2 (from right coordinate system) we have: $P_1 = RP_2 + T$.

To achieve robustness with respect to false correspondences, the well-known (adaptive) RANdom SAmple Concensus (RANSAC) approach [7, 9] can be applied. RANSAC is a very generic method for rejecting outliers. Here, we will describe a robust motion estimation algorithm based on RANSAC and one of the basic 3D alignment algorithms [6].

2.2.2 RANSAC based algorithm

The motion of point $P_1$from the left rig coordinate system can be expressed as

$P_1 = [X_1, Y_1, Z_1]^T = RP_2 + T$, where P_2 is the appropriate point in the right rig coordinate system, the orthogonal matrix R describes rotation and vector T describes translation of P_2. We assume that camera geometry is described by perspective projection with intrinsic camera parameters

$K \overset{\mathrm{def}}{=} \begin{pmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}$, where f_x and f_y are the effective focal lengths, s is the skew parameter, and $(c_x, c_y)^T$ is the principal point.

RANSAC repeatedly samples a small subset $D_s \subseteq D$ containing M point correspondences, and generates a hypothesis for the solution using only the sample D_s.

M has to contain at least four non-planar 3D points. Each of the multiple solution candidates generated by one of the appropriate algorithm [6] can be treated as a single hypothesis E. Each hypothesis E is describes appropriate motion (R_E, T_E) which is in turn a candidate of a solution of the whole task.

Each hypothesis is evaluated by counting how many 3D point correspondences are consistent with it. A correspondence $(P_1, P_2) \in$ is considered consistent with a hypothesis E if a suitable error measure $dE(P_1, P_2)$ (error measure is presented further), is below a certain threshold β.

We propose to use the following error measure $dE(P_1, P_2) = \left|\frac{1}{M}\sum_{(P_{1,j},P_{2,j})\in D_S} |P_{1,j} - RP_{2,j} - T| - |P_1 - RP_2 - T|\right|$.

The set of all consistent correspondences $S = \{(P_1, P_2) \in D | dE(P_1, P_2) < \beta\}$ is called the support set of hypothesis E. The hypothesis with the largest support set S_L found during the iterations could be returned as the result of RANSAC.

Usually, in the final step, the result is estimated from S_L. However this approach is based on the assumption that S_L is outlier-free, which in general cannot be guaranteed. So in our algorithm the result is estimated from S. As the influence of noise is typically lower when estimating from a large set of data (as opposed to the very small samples D_s) we sampled a subset $D_s \subseteq D$ containing more than 4 point correspondences (as usual – 10 point correspondences).

Let us assume the data D contains a proportion ε of outliers. The probability of getting at least one outlier-free sample D_s is $p_1 = 1 - (1 - (1 - \varepsilon)^M)^N$, where N denotes the number of RANSAC iterations. In order to get at least one outlier-free sample $D_s \subseteq D$ with a probability of (at least) p_1, we hence need to perform at least $N > \log_{1-(1-\varepsilon)^M}(1 - p_1)$ iterations. Typically, the proportion of outliers ε is unknown. The number of iterations of the RANSAC algorithm can optionally be adapted on-line basing on the following approach [2]. Let us assume the largest support set S_L is founded during previous iterations. It can be used to derive an upper bound for ε[2]: $\varepsilon \leq \frac{|S_L|}{|D|}$.

Hence, the required number of iterations is [2]:

$$N_{S_L} \overset{\text{def}}{=} \left\lceil \frac{\log(1-p_1)}{\log\left(1-\left(1-\frac{|S_L|}{|D|}\right)^M\right)} \right\rceil.$$

If the proportion of outliers is very high, however, N_{S_L} might always stay very large leading to a very big running time. In order to enforce a certain limit on the running time, we specify a maximum number of iterations in the beginning and make sure that N not increased by using the following adaptation rule [2]: $N := min(N, N_{S_L})$.

When we have 3D motion (R and T) calculated we can easily calculate position and orientation of all cameras relatively to each other.

3. EXPEROMENTAL RESULTS

The algorithm described above was tested with synthetic and experimental data. Synthetic data allows us to study the algorithm with respect to 3D image noise and to assess the conditions under which reliable results are expected.

We used two types of experimental data: calibration points and real points. Calibration points are obtained from the images of 3D calibration object (chessboard pattern). Since the sizes of this pattern are known, we can use standard camera calibration algorithm and compare the results obtained with our self-calibration algorithm with standard camera calibration algorithm. Calibration points are so accurate that the motion parameters obtained with this data and with the standard calibration algorithm may be considered as the ground-truth.

The synthetic data consists of a hundred 3D points. The points in the set were chosen randomly from a uniform distribution within a cube of size 750x750x750 mm centered about the origin. The synthetic data was formed by adding noise to the individual points and transforming them to a new location. Then a percentage of outliers was injected in the point set. The noise added to each component was uncorrelated, isotropic and Gaussian, with zero mean and variable standard deviation. For each noise level one hundred trials were performed. The average response of algorithm over this hundred trials was used to compute several different error statistics for the calculated transformations. For absolute accuracy error estimation in this article we propose mean 3D position error. It is given by $e_{mean} = \frac{1}{m}\sum_j^m \left|P_{1,j} - \hat{R}P_{2,j} - \hat{T}\right|$, where $\hat{R}$ is rotation estimation, $\hat{T}$ is translation estimation.

3.1.1 Noise sensitivity analysis

The noise added to the 3D points is Gaussian with standard deviation varying from 0.05 to 2.5 mm. A percentage of outliers that was injected in the dataset was established as 25%.

We studied the behavior of the algorithm as a function of 3D point noise while s percentage of outliers is fixed.

Fig.3A shows mean 3D position error with various standard deviation.

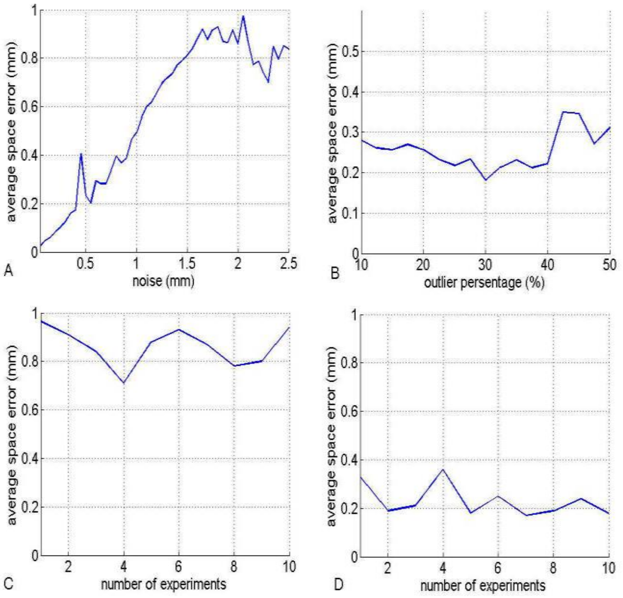

Figure 3: Mean 3D position error: A - synthetic data for different standard deviation of noise; B – synthetic data for different outlier percentage; C – real data: images of moving human body; D – real data: images of static chessboard pattern.

From fig.3A we can deduce the stability of self-calibration method against noise on the poses of the input 3D features.

3.1.2 Outliers sensitivity analysis

Percentage of outliers that was injected in the dataset was varying from 0 to 50 %. The noise added to the 3D points is Gaussian with standard deviation that was established as 1.0 mm.

Fig.3B shows mean 3D position error with various percentages of outliers. From fig.3B we can deduce the quality of the estimation is independent of the outlier percentage.

3.1.3 Experiments with real data

Real experiments were conducted using image sequences of a moving person. Two stereo-rigs were calibrated by using chessboard pattern. Feature points of the body were found, filtered and reconstructed. Motion of these points was estimated using the proposed method.

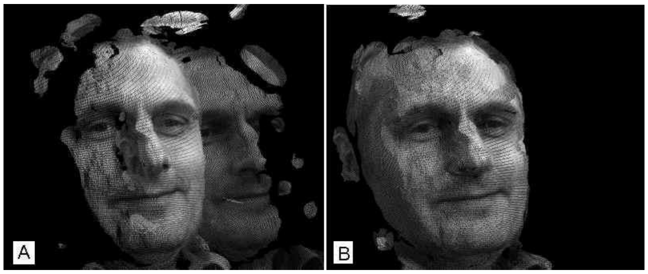

Figure 4: 3D image of a head: A - two stereo rigs are not calibrated so two 3D images of the head are not aligned. B – after self-calibration. Two 3D images of the head are aligned after self-calibration has performed using the proposed method.

Fig.4A shows motion that exists between two 3D images of a head reconstructed using images from different rigs since extrinsic calibration between stereo rigs was not known. Then 3D motion between point sets was estimated using the proposed method. Fig.4B shows that two 3D images of the head are aligned after 3D motion estimation.

Fig.3C and Fig.3D show mean 3D position error in various experiments with real data. Fig.3C shows mean 3D position error for 3D motion calculated using proposed method and points detected and matched on images of moving human body. Fig.3C shows mean 3D position error for 3D motion calculated using classical calibration method [19].

From fig.3C and fig.3D we can deduce that accuracy of the presented method is worse than accuracy of common calibration techniques but it can be accepted as sufficient for some practical purposes. For example, proposed method can be used when using of calibration objects is undesirable or impossible.

4. CONCLUSION

A new self-calibration algorithm has been proposed for obtaining 3D motion parameters of a multi stereo-rig system over time without using any particular calibration apparatus. The idea is to use previously valid stereo-rig calibration parameters and image point matches to perform an alignment of two 3D paired point sets that contains outliers. The method is tested in both artificial data and real video sequences. The results show that our method is robust in datasets with up to 50% of outliers. The advantage of our approach is the fact that no calibration objects are needed to perform metric calibration of the multi stereo rig system as most reference approaches demand.

The proposed method was evaluated against standard method for multi-rig calibration and proved to have acceptable in accuracy.

5. REFERENCES

[1] *Mostafa Abdelrahman, Asem M. Ali, Shireen Y. Elhabian, Ham Rara, Aly A. Farag "A passive stereo system for 3D human face reconstruction and recognition at a distance", CVPR Workshops, 2012.*

[2] *K.S. Arun, T.S. Huang, S.D. Blostein Least-squares fitting of two 3-D point sets. IEEE Trans Pattern Anal Machine Intell 9:698–700, 1987.*

[3] *F.Bajramovic "Self-Calibration of Multi-Camera Systems . PhD thesis", Friedrich Schiller University of Jena ,2010.*

[4] *D. Chetverikov, D. Stepanov, P. Krsek "Robust Euclidean alignment of 3D point sets: the Trimmed Iterative Closest Point algorithm" , Image and Video Computing (2005), vol.23 (3), p.p.299–309.*

[5] *D. Colbry "Human Face Verification by Robust 3D Surface Alignment. PhD thesis", Michigan State University ,2006.*

[6] *D.W. Eggert, A. Lorusso, R.B. Fisher "Estimating 3-D rigid body transformations: a comparison of four major algorithms", Machine Vision and Applications (1997) vol. 9, p.p. 272–290.*

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

[8] A. *Fusiello "Uncalibrated Euclidean reconstruction: a review", International Journal of Image and Vision Computing, vol.* 18, No.6-7 (2000), p.p.555-*563*

[9] *Hartley, R. and Zisserman, "A. Multiple View Geometry in Computer Vision (2nd ed.)",Cambridge: Cambridge University Press, 2003.*

[10] *Akihiro Hayasaka, Takuma Shibahara, Koichi Ito, Takafumi Aoki, Hiroshi Nakajima, Koji Kobayashi "A Passive 3D Face Recognition System and Its Performance Evaluation", IEICE Trans.Fund. v. E91-A, p. 1974-1981, 2008.*

[11] *Svetlana V. Korobkova, Archil Tsiskaridze "Face recognition system using 2D and 3D information fusion" , Proceedings of Graphicon 2011, pp. 153—156, 2011.*

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

[13] *F. Pedersini, A. Sarti, S. Tubaro "Accurate and simple geometric calibration of multi-camera systems", Signal Processing vol. 77, No.3 (1999), p.p. 309-334.*

[14] *N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3d. In SIGGRAPH '06: ACM SIGGRAPH 2006 Papers, pages 835–846, 2006.*

[15] *T. Svoboda "Quick guide to multi-camera self-calibration. Technical report", Computer Vision Lab, Swiss Federal Institute of Technology, Zurich, 2003.*

[16] *R. Szeliski "Computer Vision: Algorithms and Applications". Springer, 2010.*

[17] B. *Triggs "Autocalibration from Planar Scenes". Proceedings of the European Conference on Computer Vision (ECCV), vol. 1(1998),p.p. 89–105*

[18] *X. Zhang, Y. Zhang, X. Zhang, T. Yang, X. Tong, H. Zhang A convenient multi-camera self-calibration method based on human body motion analysis. Proceedings of the Fifth International Conference on Image and Graphics, ICIG 2009, China, p.3-8, 2009.*

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

[20] *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