W-PnP Method: Optimal Solution for the Weak-Perspective n-point Problem and its Application to Structure from Motion

Levente Hajder
Machine Perception Research Laboratory, MTA SZTAKI
Kende utca 13-17., Budapest, Hungary, H-1111
hajder.levente@sztaki.mta.hu

Keywords: Weak-perspective projection, Calibration, PnP, Structure from Motion

Abstract: Camera calibration is a key problem in 3D computer vision since the late 80's. Most of the calibration methods deal with the (perspective) pinhole camera model. This is not a simple goal: the problem is nonlinear due to the perspectivity. The strategy of these methods is to estimate the intrinsic camera parameters first; then the extrinsic ones are computed by the so-called PnP method. Finally, the accurate camera parameters are obtained by slow numerical optimization. In this paper, we show that the weak-perspective camera model can be optimally calibrated without numerical optimization if the L_2 norm is used. The solution is given by a closed-form formula, thus the estimation is very fast. We call this method as the Weak-Perspective n-Point (W-PnP) algorithm. Its advantage is that it simultaneously estimates the two intrinsic weak-perspective camera parameters and the extrinsic ones. We show that the proposed calibration method can be utilized as the solution for a subproblem of 3D reconstruction with missing data. An alternating least squares method is also defined that optimizes the camera motion using the proposed optimal calibration method.

1 Introduction

The problem of optimal methods in multiple view geometry (Hartley and Kahl, 2007) is a very challenging research issue. This study deals with camera calibration, a key problem in computer vision. There are well-known solutions (Hartley and Zisserman, 2000; Zhang, 2000) to calibrate the perspective camera; these methods give a rough estimate of the parameters first, then refine them using numerical optimization, such as the Levenberg-Marquardt iteration. Optimal camera calibrations using the L_2 norm including the popular Perspective n-point Problem (PnP) were published for the perspective camera only if the intrinsic camera parameters are known (Schweighofer and Pinz, 2008; Lepetit et al., 2009; Hesch and Roumeliotis, 2011; Zheng et al., 2013). The calibration can also be solved under the L_∞ norm (Kahl and Hartley, 2008) as well as the Structure from Motion problem (Ke and Kanade, 2005; Okatani and Deguchi, 2006; Bue et al., 2012); however, the uncalibrated problem has not been optimally solved yet in the least squares sense to the best of our knowledge.

Weak-perspective camera calibration. The optimal estimation of the affine calibration is easy since it is a linear problem as it has been shown in several studies, such as that of Shum et al. (Shum et al., 1995). The weak-perspective (DeMenthon and Davis, 1995) and paraperspective (Horaud et al., 1997) calibration have also been considered, but the proposed algorithms are not optimal since these papers focus on finding the link between para/weak-perspectivity and real projection. Kanatani et. al (Kanatani et al., 2007) also dealt with the calibration of different affine cameras, but they did not consider the optimality itself.

The scaled orthographic calibration can optimally be calibrated as recently discussed in the work of Hajder et al. (L. Hajder and Á. Pernek and Cs. Kazó, 2011). An iteration was proposed by the authors to calibrate the scaled orthographic camera, and it converges to the global minima as proved in (L. Hajder and Á. Pernek and Cs. Kazó, 2011). The orthographic camera is not considered separately, but the method can be used for that purpose as well if the scale of the scaled orthographic camera is fixed. Another possible solution (Marques and Costeira, 2009) for the scaled orthographic calibration is to do an affine calibration and then find the closest scaled orthographic camera matrix to the affine one. However, optimality cannot be guaranteed in this case.

The optimal camera calibration method is proposed for weak-perspective cameras in this paper;

it estimates the camera parameters if $3D$–$2D$ point correspondences are known between the points of a 3D calibration object and corresponding locations on the image. The minimization is optimal in the least squares sense.

Weak-perspective Structure from Motion. The optimal weak-perspective camera calibration is theoretically very interesting, and it has practical significance as well. We show here that the calibration algorithms can be inserted into 3D reconstruction - also called Structure from Motion (SfM) - pipelines as a substep yielding very efficient weak-perspective reconstruction. Mathematically, the problem is a factorization one: the so-called measurement matrix has to be factorized into the matrices containing camera and structure parameters.

The classical factorization method, when the measurement matrix is factorized into 3D motion and structure matrices, was developed by Tomasi and Kanade (Tomasi, C. and Kanade, T., 1992) in 1992. The weak-perspective extension was published by Weinshall and Kanade (Weinshall and Tomasi, 1995). Factorization was extended to the paraperspective (Poelman and Kanade, 1997) case as well as to the real perspective (Sturm and Triggs, 1996) one.

The problem of missing data is also a very important challenge in 3D reconstruction: one cannot guarantee that the feature points can be tracked over the whole image sequence since feature points can appear and/or disappear between frames. The problem of missing data was already addressed by Tomasi and Kanade (Tomasi, C. and Kanade, T., 1992); however, they use only a naive approach which transforms the missing data problem to the full matrix factorization by estimating the missing entries. Shum et al. (Shum et al., 1995) gave a method to reconstruct the objects from range images; their method was successfully applied to the SfM problem by Buchanan et al. (Buchanan and Fitzgibbon, 2005).

The mainstream idea for factorization with missing data is to decompose the rank 4 measurement matrix into affine structure and motion matrices which are of dimension 4. The Shum-method (Shum et al., 1995; Buchanan and Fitzgibbon, 2005) also computes affine structure and motion matrices, but the dimension of those matrices is 3. This problem can mathematically be solved by Principal Component Analysis with Missing Data (PCAMD) as pointed out by mathematicians since the middle 70's (Ruhe, 1974). These methods can be applied directly to the SfM problem as it is written in (Buchanan and Fitzgibbon, 2005). Hartley & Schaffalitzky (Hartley and Schaffalitzky, 2003) proposed the PowerFactorization method which is based on the Power method to compute the dominant n-dimensional subspace of a given matrix. Buchanan & Fitzgibbon (Buchanan and Fitzgibbon, 2005) handled the problem as an alternation consisting of two nonlinear iterations to be solved; they suggested the usage of the Damped-Newton method with line search to compute the optimal structure and motion matrices. Kanatani et al. (Kanatani et al., 2007) showed that the reconstruction problem can be solved without a full matrix factorization. Marques&Costeira (Marques and Costeira, 2009) solved the factorization problem considering the scaled orthographic camera constraints; their method was basically an affine factorization, but the camera matrices were refined based on scaled orthographic constraint at the end of each cycle. An interesting approach was also proposed by Whang et al. (Wang et al., 2008): their so-called quasi-perspective reconstruction fills the gap between affine and perspective approaches.

Contribution. The closest work to this paper is proposed by Hajder et al. (L. Hajder and Á. Pernek and Cs. Kazó, 2011). They proved that the scaled orthographic camera can optimally be calibrated by an iterative algorithm and the calibration can be applied in the SfM approach. We deal with the weak-perspective camera model instead of the scaled orthographic one here. *We give a closed-form solution to the calibration problem*, which can be inserted into iterative SfM algorithms similarly to (L. Hajder and Á. Pernek and Cs. Kazó, 2011; Kanatani et al., 2007; Marques and Costeira, 2009) and (Buchanan and Fitzgibbon, 2005). *The novelty here is that all of the steps within the iterations are optimal. Another strength of our method is that it can be proved that the iteration converges to the closest minimum.*

The optimal method proposed here is *interesting theoretically* and *useful practically*. For the latter purpose, we show that the proposed weak-perspective factorization can give good initial values for perspective bundle adjustment (B. Triggs and P. McLauchlan and R. Hartley and A. Fitzgibbon, 2000), and it can be inserted into a 3D reconstruction pipeline.

The main contribution of this paper is threefold: (i) an optimal weak-perspective calibration algorithm (the W-PnP method) is proposed here. The optimal solution is written in closed form given by finding the root of a polynomial with degree 11 [1]; (ii) contrary to the standard PnP methods, the proposed calibration algorithm estimates both the intrinsic and extrinsic camera parameters. It is possible since the application of the weak-perspective projection eliminates the division from the projective equations; (iii) a weak-

[1] However, the root-finding of a 11-degree polynomial can only be carried out by numerical methods according to the Abel-Ruffini theorem.

perspective SfM algorithm is proposed here which is an alternation with two main steps: the 3D structure of the object to be reconstructed as well as the camera motion are calculated optimally. The latter is done by the proposed optimal weak-perspective camera calibration method. The proposition of an alternating-style SfM method is not novel, the main advantage here is the application of weak-perspective projection which makes all supsteps within the iteration optimal. Another mentionable property of our method is that it can cope with missing data.

Structure of paper. In section 2, we introduce basic notations and present formulas to write mathematically the problem. The proposed optimal camera calibration is described in section 3. Then the calibration method is inserted into an alternating-style SfM algorithm. The proposed algorithm is tested on synthesized data (section 5) as well as on coordinates of tracked feature points from real image sequences (section 6). Finally, the paper concludes the research in section 7.

2 Problem Statement

Given the 3D coordinates of the points of a static object and their 2D projections in the image, the aim of camera calibration is to estimate the camera parameters which represent the 3D $\rightarrow$ 2D mapping.

Let us denote the 3D coordinates of the i^{th} point by X_i, Y_i, and Z_i. The corresponding 2D coordinates are denoted by u_i, and v_i. The perspective (pinhole) camera model is usually written as follows

$$\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} \sim C[R|T_{3D}] \begin{bmatrix} X_i & Y_i & Z_i & 1 \end{bmatrix}^T. \quad (1)$$

where R is the rotation (orthonormal) matrix, and T_{3D} the spatial translation vector between the world and object coordinate systems. (these parameters are usually called the extrinsic parameters of the perspective camera) The 'operator $\sim$' denotes equality up to an unknown scale. The intrinsic parameters of the camera are stacked in the upper triangular matrix C (Hartley and Zisserman, 2000).

If the above equation is multiplied by the inverse of camera matrix C, the following basic camera calibration formula is obtained: $C^{-1}\begin{bmatrix} u_i & v_i & 1 \end{bmatrix} \sim [R|T_{3D}] \begin{bmatrix} X_i & Y_i & Z_i & 1 \end{bmatrix}^T$. If the intrinsic parameters stacked in matrix C and the spatial coordinates in $\begin{bmatrix} X_i & Y_i & Z_i & 1 \end{bmatrix}^T$ are known then the calibration problem is reduced to the estimation of the extrinsic matrix/vector R and T_{3D}. This is the so-called Perspective n-point Problem (PnP). There are several

Figure 1: Pixels for different camera models. Scaled orthographic, weak-perspective and affine camera pixels are equivalent to square, rectangle, and parallelogram, respectively.

efficient solvers (Schweighofer and Pinz, 2008; Lepetit et al., 2009; Hesch and Roumeliotis, 2011; Zheng et al., 2013) for PnP, however, estimates for the intrinsic parameters of the applied cameras are usually not presented. We deal with this problem, and it is shown here that the weak-perspective camera calibration is possible without the knowledge of any intrinsic camera parameters.

If the depth of object is much smaller than the distance between the camera and the object, the weak-perspective camera model is a good approximation:

$$\begin{bmatrix} u_i & v_i \end{bmatrix}^T = [M|t] \begin{bmatrix} X_i & Y_i & Z_i & 1 \end{bmatrix}^T. \quad (2)$$

where M is the motion matrix consisting of two 3D vectors ($M = [m_1, m_2]^T$) and t is a 2D offset vector which locates the position of the world's origin in the image.

Contrary to the affine camera model, the rows of the motion matrix are not allowed to be arbitrary for the weak-perspective projection, they must satisfy the orthogonality constraint $m_1^T m_2 = 0$. A special case of the weak-perspective camera model is the scaled orthographic one, when $m_1^T m_1 = m_2^T m_2$. If the affine camera is considered, there is no constraint: the elements of the motion matrix M may be arbitrary.

The difference between the camera models can be visualized by the shapes of the corresponding camera pixels. Affine camera model is represented by a rectangular pixel: the opposite sides are parallel to each other. The weak-perspective model constraints that the adjacent sides are perpendicular, while the length of the sides are equal for the scaled orthographic camera model. The pixels are pictured in Fig. 1.

The optimal calibration of the affine camera in the least squares sense is relatively simple as the projection in Eq. 2 is linear w.r.t. unknown parameters. The solution can be obtained by the Moore-Penrose pseudo-inverse.

The scaled orthographic camera estimation is a more challenging problem. To the best of our knowledge, there is no closed-form solution. Hajder et al. (L. Hajder and Á. Pernek and Cs. Kazó, 2011) proved that the optimal estimation can be given via an iteration. However, their method is relatively slow due to the iteration. **The main contribution of this paper is that the weak-perspective case is solvable**

as a root finding problem of a 11-degree polynomial.

3 Optimal Camera Calibration for Weak-perspective Projection: the W-PnP Method

In this section, a novel weak-perspective camera calibration is proposed. The goal of the calibration is to minimize the squared reprojection error in the least squares sense. This is written as

$$\frac{1}{2}\sum_{i=1}^{N}\left|\left|\begin{bmatrix} u_i & v_i \end{bmatrix}^T - [M|t]\begin{bmatrix} X_i & Y_i & Z_i & 1 \end{bmatrix}^T\right|\right|^2, \tag{3}$$

where N is the number of points to be considered in the calibration, and $||\cdot||$ denotes the L_2 (Euclidean) vector norm. As Horn et al. (Horn et al., 1988) proved, the translation vector t is optimally estimated if it is selected as the center of gravity of the 2D points. These are easily calculated as $\tilde{u} = 1/N\sum_{i=1}^{N} u_i$, and $\tilde{v} = 1/N\sum_{i=1}^{N} v_i$.

If the weak-perspective camera model is assumed, the error defined in eq. (3) can be rewritten in a more compact form as

$$\frac{1}{2}\left|\left|w_1^T - m_1^T S\right|\right|^2 + \frac{1}{2}\left|\left|w_2^T - m_2^T S\right|\right|^2, \tag{4}$$

where

$$w_1 = [u_1 - \tilde{u}, u_2 - \tilde{u}, \ldots, u_N - \tilde{u}]^T, \tag{5}$$

$$w_2 = [v_1 - \tilde{v}, v_2 - \tilde{v}, \ldots, v_N - \tilde{v}]^T, \tag{6}$$

$$S = \begin{bmatrix} X_1 & X_2 & \ldots & X_N \\ Y_1 & Y_2 & \ldots & Y_N \\ Z_1 & Z_2 & \ldots & Z_N \end{bmatrix}. \tag{7}$$

If the Lagrange multiplier λ is introduced, the weak-perspective constraint can be considered. The error function is modified as follows

$$\frac{1}{2}\left|\left|w_1 - m_1^T S\right|\right|^2 + \frac{1}{2}\left|\left|w_2 - m_2^T S\right|\right|^2 + \lambda m_1^T m_2 \tag{8}$$

The optimal solution of this error function is given by its derivatives with respect to λ, m_1, and m_2:

$$m_1^T m_2 = 0, \tag{9}$$

$$SS^T m_1 - Sw_1 + \lambda m_2 = 0, \tag{10}$$

$$SS^T m_2 - Sw_2 + \lambda m_1 = 0. \tag{11}$$

m_2 is easily expressed from eq. (10) as

$$m_2 = \frac{1}{\lambda}\left(Sw_1 - SS^T m_1\right). \tag{12}$$

If one substitutes m_2 into eq. (11), and (9), then the following expressions are obtained:

$$\frac{1}{\lambda}SS^T\left(Sw_1 - SS^T m_1\right) - Sw_2 + \lambda m_1 = 0, \tag{13}$$

$$\frac{1}{\lambda}m_1^T\left(Sw_1 - SS^T m_1\right) = 0. \tag{14}$$

If eq. (13) is multiplied by λ, then m_1 can be expressed as

$$m_1 = \left(SS^T SS^T - \lambda^2 I\right)^{-1}\left(SS^T Sw_1 - \lambda Sw_2\right) \tag{15}$$

where I is the 3×3 identity matrix. Remark that the matrix inversion cannot be carried out if the Lagrange multiplier λ is one of the eigenvalues of the matrix SS^T. If the expressed m_1 is substituted into eq. (14), the equation from which λ should be determined is obtained:

$$\frac{1}{\lambda}A^T(\lambda)B^{-T}(\lambda)\left(Sw_1 - SS^T B^{-1}(\lambda)A(\lambda)\right) = 0 \tag{16}$$

where

$$A(\lambda) = SS^T Sw_1 - \lambda Sw_2 \tag{17}$$

$$B(\lambda) = SS^T SS^T - \lambda^2 I \tag{18}$$

$A(\lambda)$ and $B(\lambda)$ are a vector and a matrix that have elements containing polynomials of unknown variable λ. Such kind of vectors/matrices is called *vector/matrix of polynomials* in this study. The difficulty is that matrix $B(\lambda)$ should be inverted. This inversion can be written as a fraction of two matrices. $B^{-1}(\lambda)$ can write as

$$B^{-1}(\lambda) = \frac{\mathrm{adj}\left(SS^T SS^T - \lambda^2 I\right)}{\det\left(SS^T SS^T - \lambda^2 I\right)} \tag{19}$$

where adj$(.)$ denotes the adjoint[2] of a matrix. It is trivial that $\det(B(\lambda))$ is a polynomial of λ, while $\mathrm{adj}(B(\lambda))$ is a matrix of polynomials. This expression is useful since the equation can be multiplied by the determinants of $B(\lambda)$.

If one makes elementary modifications, eq. (16) can be rewritten as

$$\frac{A^T(\lambda)\mathrm{adj}B^T(\lambda)}{\det B(\lambda)}\,\frac{\det B(\lambda)Sw_1 - SS^T\mathrm{adj}B(\lambda)A(\lambda)}{\det B(\lambda)} = 0. \tag{20}$$

It is also trivial that eq. (20) is true if the numerator equals zero If the denominator, the determinant of matrix $B(\lambda)$ equals zero, then the problem cannot be solved; in this case, the 3D points in S are linearly dependent, the points in S form a plane, or a line, or

[2] The transpose of the adjoint is also called the matrix of cofactors.

a single point instead of a real 3D object. The Lagrange multiplier λ is calculated by solving the following polynomial:

$$A^T(\lambda)\text{adj}B^T(\lambda)\left(\det B(\lambda)Sw_1 - SS^T\text{adj}B(\lambda)A(\lambda)\right) = 0. \quad (21)$$

This final polynomial is of degree 11: $A(\lambda)$, and $B(\lambda)$ have terms of degree 1, and 2, respectively. Therefore, adj$(B^T(\lambda))$is of degree 4, while that of $A^T(\lambda)\text{adj}(B^T(\lambda))$ is 5. Since the size of $B(\lambda)$ is 3×3, its determinant has degree $3\cdot 2 = 6$. Other terms are of lower degree, the degree of the final polynomial comes to $5+6=11$.

The roots of the polynomial are 11 real/complex numbers, but only the real values have to be considered. The obtained real values of λ should be substituted into eq. (15) and the obtained m_1 and λ into eq. (12); then the optimal solution is the one minimizing the reprojection error given in eq. (3).

We use Joe Huwaldt's Java Matrix Tool[3] to solve the 11-th order polynomial equation. Our implementation uses the Jenkins and Traub root finder (Jenkins and Traub, 1970), and we found that this algorithm is numerically very stable.

A very important remark is that in the case, when the coordinates in vectors w_1 and w_2 are noise-free, it is possible that λ equals zero. Then the camera vectors m_1 and m_2 can be computed as $m_1 = \left(SS^T\right)^{-1} Sw_1$ and $m_2 = \left(SS^T\right)^{-1} Sw_1$.

Minimal solution. For PnP algorithms, the minimal number of points for the algorithms is also an important issue. The proposed optimization method is based on reprojection error: each point adds two equations to the minimization. The camera matrix consists of eight elements: six for camera pose and scales, two for offset. The pose gives 3 Degrees of Freedom (DoFs), vertical and horizontal scales are two DoFs, while the offset yields another two parameters. In summary, *the problem has 7 DoFs and they can be estimated from at least four $3D \rightarrow 2D$ point correspondences.*

4 Structure from Motion with Missing Data

We describe here how the previously discussed optimal calibration method can be applied for the factorization (SfM) problem. Our method allows the points to appear and/or disappear; thus, it can handle the missing data problem.

[3]Available at http://thehuwaldtfamily.org/java/Packages/MathTools/MathTools.html

The proposed reconstruction method is an alternating least squares algorithm to minimize the reprojection error defined as follows

$$\left|\left|H \odot \left(W - [M|t]\begin{bmatrix} S \\ \mathbf{1}^T \end{bmatrix}\right)\right|\right|_F^2, \quad (22)$$

where M is the motion matrix consisting of the camera parameters in every frame, and structure matrix S contains the 3D coordinates of the points (points are located in the columns of matrix S). Operator '$\odot$' denotes the so-called Hadamard product[4], and H is the mask matrix. If H_{ij} is zero, then the j^{th} point in the i^{th} frame is not visible. If $H_{ij} = 1$, the point is visible.

Each cycle of the proposed methods is divided into the following main steps:

1. **W-PnP-step**. The aim of this step is to optimally estimate the motion matrix $M = [M_1^T, M_2^T, \ldots, M_F^T]^T$, and translation vector $t = [t_1^T, t_2^T, \ldots, t_F^T]^T$ if S is fixed, where the index denotes the frame number. It is trivial that the estimation of these submatrices are independent from each other if the elements of the structure matrix S are fixed. The optimal solution is given by W-PnP method defined in Section 3. Note that missing data should be skipped in the estimation.
2. **S-step**. The goal of S-step is to compute the structure matrix S if the elements of the motion matrix and the translation vector are fixed[5]. The 3D points represented by the columns of the structure matrix must be computed independently (they are independent from each other). Missing data should be considered during the estimation of course. It is a linear problem w.r.t. the coordinates contained by structure matrix S; the optimal method can be obtained using the Moore-Penrose pseudo-inverse as described in (Shum et al., 1995).

The proposed algorithm iterates the two steps until convergence as overviewed in Alg. 1. The convergence itself is guaranteed since both steps decrease the non-negative reprojection error defined in Eq 22. The proposed factorization method requires initial values of the matrices. The key idea for initializing the parameters is that the factorization with missing data can be divided into full matrix factorization of submatrices. If there is overlapping between submatrices, then the computed motion and structure submatrices can be merged if they are rotated and translated with the appropriate rotation matrices and vec-

[4]$A \odot B = C$ if $c_{ij} = a_{ij} \cdot b_{ij}$.

[5]This task is usually called triangulation. This term comes from stereo vision where the camera centers and the 3D position of the point form a triangle.

Algorithm 1 Summary of weak-perspective factorization

$M^{(0)}$,$t^{(0)}$,$S^{(0)} \leftarrow$ Parameter Initialization
$k \leftarrow 0$
repeat
  $k \leftarrow k+1$
  $M^{(k)}$,$t^{(k)} \leftarrow$ W-PnP-Step(H,W,$S^{(k-1)}$)
  $S^{(k)} \leftarrow$ S-Step(H,W,$M^{(k)}$,$t^{(k)}$)
until convergence.

tors, respectively. We use the method of Pernek et al. (Pernek et al., 2008) for this purpose.

Algorithm 2 Skeleton of Scaled Orthographic Camera Calibration

repeat
  $w_3 \leftarrow$ Completion(R,t,S,$scale$)
  R,t,$scale \leftarrow$ Registration(S,w_1,w_2,w_3)
until convergence.

Comparison with scaled orthographic factorization. The scaled orthographic camera calibration (L. Hajder and Á. Pernek and Cs. Kazó, 2011) is overviewed in Alg 2. The main idea of the calibration is as follows: the measured 2D coordinates are completed with a third coordinate that is simply calculated by reprojecting the spatial coordinates with the current camera parameters. Then the registration-step refines the camera parameters, and the completion and registration steps are repeated until convergence. Hajder et al. (L. Hajder and Á. Pernek and Cs. Kazó, 2011) proved that this iteration converges to the global optimum and this convergence is independent of the initial values of the camera parameters. The completion is simple, easy to implement, however, it is very costly as the calibration algorithm is iterative, closed-form solution is not known.

An alternating-style SfM algorithm can also be formed using the scaled orthographic camera model as it is visualized in Alg 3. It has more steps than the weak-perspective SfM method (Alg. 1) as the completion of the 2D coordinates is required after every other steps.

Comparison with affine factorization. As it is discussed before, the estimation of affine camera parameters is a linear problem. There are several methods (Shum et al., 1995; Buchanan and Fitzgibbon, 2005) dealing with affine SfM factorization as well. They are relatively fast, but the accuracy of those is lower compared to the scaled orthographic and weak-perspective factorization as the affine camera model enables shearing (skew) of the images that is not a realistic assumption. Remark that the skeleton of the affine SfM methods is the same as that of weak-perspective one defined in Alg. 1.

Algorithm 3 Summary of scaled orthographic factorization

$M^{(0)}$,$t^{(0)}$,$S^{(0)} \leftarrow$ Parameter Initialization
$\tilde{H}$, $\tilde{W}^{(0)}$, $\tilde{M}^{(0)}$, $\tilde{t}^{(0)} \leftarrow$ Complete(H, W,$M^{(0)}$,$t^{(0)}$, $S^{(0)}$)
$k \leftarrow 0$
repeat
  $k \leftarrow k+1$
  $\tilde{M}^{(k)} \leftarrow$ Registration($\tilde{H}$,$\tilde{W}^{(k)}$,$S^{(k-1)}$)
  $\tilde{W}^{(k)} \leftarrow$ Completion(W,$\tilde{H}$,$\tilde{M}^{(k)}$,$S^{(k-1)}$)
  $S^{(k)} \leftarrow$ S-Step($\tilde{H}$,$\tilde{W}^{(k)}$,$\tilde{M}^{(k)}$)
  $\tilde{W}^{(k)} \leftarrow$ Completion(W,$\tilde{H}$,$\tilde{M}^{(k)}$,$S^{(k)}$)
until $\left|\left|\tilde{H} \odot \left(\tilde{W}^{(k)} - \left[\tilde{M}^{(k)}|t^{(k)}\right] \left[\begin{array}{c} S^{(k)} \\ 1 \end{array}\right]\right)\right|\right|_F^2$ converges.

Source code. The proposed weak-perspective SfM algorithm is implemented in Java and will be available after publication.

5 Tests on Synthesized Data

Several experiments with synthetic data were carried out to study the properties of the proposed methods. Three methods were compared: (i) **SO** Scaled Orthographic factorization (L. Hajder and Á. Pernek and Cs. Kazó, 2011), (ii) **WP** proposed Weak-Perspective factorization, and (iii) **AFF**: Affine factorization (Shum et al., 1995).

We have examined three properties as follows.

1. Reconstruction error: The reconstructed 3D points are registered to the generated (ground truth) ones using the method of Arun et al.(Arun et al., 1987). This registration error is called reconstruction error in the tests. The charts show the improvement of the method (in percentage) w.r.t. the original Tomasi-Kanade factorization (Tomasi, C. and Kanade, T., 1992).
2. Motion error: The row vectors of the obtained 3D motion matrix can be registered to that of the generated (ground truth) motion matrix. This registration error is called motion error here. The charts show the improvement in percentage similarly to visualization of the reconstruction error.
3. Time demand: The running time of each algorithm was measured. The given values contain every step from the parameter initialization to the final reconstruction.

To compare the affine method (Shum et al., 1995) listed above with the other two rival algorithms, the computation of the metric 3D structure was carried out by the classical weak-perspective Tomasi-Kanade factorization (Tomasi, C. and Kanade, T., 1992). The $2F \times 4$ affine motion was multiplied by the $4 \times P$ affine structure matrix, and a full measurement matrix was obtained. Then this measurement matrix was factorized by the Tomasi-Kanade algorithm (Tomasi, C. and Kanade, T., 1992) with the Weinshall-Kanade (Weinshall and Tomasi, 1995) extension.

All of the rival methods were implemented in Java. The tests were run on an Intel Core4Quad 2.33 GHz PC with 4 GByte memory.

5.1 Test Data Generation

Generation of moving feature points. The input measurement matrix was composed of 2D trajectories. These trajectories were generated in the following way: (i) Random three-dimensional coordinates were generated by a zero-mean Gaussian random number generator with variance σ_{3D}. (ii) The generated 3D points were rotated by random angles. (iii) Points were projected using perspective projection. [6] (iv) Noise was added to the projected coordinates. It was generated by a zero-mean Gaussian random number generator as well; its variance was set to σ_{2D}. (v) Finally, the measurement matrix W was composed of the projected points. (vi) Motion and structure parameters were initialized as described in Sec. 4. For each test case, 100 measurement matrices were generated and the results shown in this section were calculated as the average of the 100 independent executions.

Generation of mask matrix. The mask generator algorithm has three parameters: (i) P: Number of the visible points in each frame, (ii) F: Number of the frames. (iii) O: offset between two neighboring frames. The structure of the mask matrix is seen in Fig. 2. Each point appears and disappears only once. If a point has already disappeared it will not be visible again in the sequence.

5.2 Test Evaluation

General remarks. The charts basically show that the **SO** algorithm outperforms the other methods in every test case as it is expected. This is evident since the scaled orthographic projection model is the closest one to real perspectivity. This is true for the reconstruction error as well as the motion error. The second place in accuracy is given to the proposed weak-perspective (**WP**) method which is always better than the affine one, but slightly less accurate than the **SO** method.

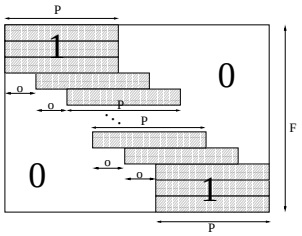

Figure 2: Structure of mask matrix. Vertical and horizontal directions correspond to the frames and points, respectively. If an element is zero then the corresponding feature is not visible in the pointed frames. This type of mask matrices simulates the realistic case when the features appear and disappear only once.

Examining the charts of time demand, it is clear that the fastest method is the affine (**AFF**) one; however, the affine algorithm can be very slow as discussed during real tests later if there is a huge amount of input data. It is because a full factorization (Tomasi, C. and Kanade, T., 1992) must be applied after the affine factorization to obtain metric reconstruction, and this can be very slow due to the Singular Value Decomposition. This SVD-step can be faster if only the three most dominant singular values and vectors are computed (Kanatani et al., 2007). Unfortunately, the Java Matrix Package (JAMA) which we used in our implementation does not contain this feature. As shown in (Buchanan and Fitzgibbon, 2005), there are several methods which implement affine reconstruction. Pernek et al. have shown earlier (Pernek et al., 2008) that the fastest method of those is the so-called Damped-Newton algorithm, which is significantly faster than our affine implementation.

The main conclusion of the tests is that there is a tradeoff between accuracy and time demand. The **SO** factorization is the most accurate but slowest one, while the affine is fast but less accurate. The proposed WP-SfM algorithm is very close to **SO** and **AFF** algorithms in accuracy and running time, respectively.

Error versus noise (Figure 3) The methods were run with gradually increasing noise level. The reconstruction error increases approximately in a linear way for all the methods. Therefore, the improvement is approximately the same for all noise levels as the error of the reference factorization (Tomasi, C. and Kanade, T., 1992) increases with regard to noise as

[6] We tried the orthographic projection model with/without scale as well, the results had similar characteristics. Only the fully perspective test generation is contained in this paper due to the page limit.

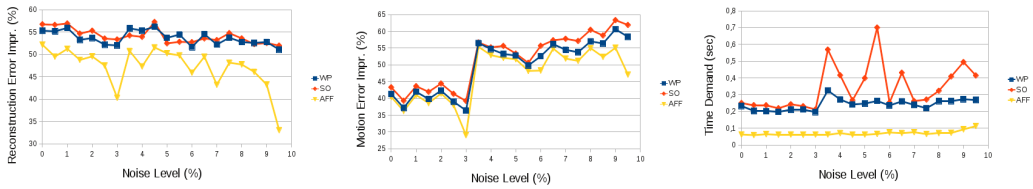

Figure 3: Improvement of reconstruction and motion errors (left charts) and time demand (right) w.r.t. 2D noise.

well. The test sequence consisted of 20 frames, and $P = 100$ was set. The missing data ratio was 30.6%. The noise level was calculated as $100\sigma_{2D}/\sigma_{3D}$.

The test indicated that the **SO** algorithm outpowered the rival ones, and the **WP** method was better than the affine one as expected; however, **SO** needs the most time to finish its execution, thus the fastest method is the affine one.

Error versus number of points (Figure 4) P increased from 40 to 180 (the missing data rate decreased from approx. 80% to 20%). The noise level was 5%, and the sequence consisted of 100 frames. The conclusion was similar to the previous test case: the most accurate model was given by the **SO** algorithm, the second one was from the **WP** method. The difference was not significant in either accuracy or execution time.

Error versus number of frames (Figure 5) F increased from 10 to 46. The corresponding missing data ratio increased from 10% to 80%.The noise level was 5%, and $P = 100$. In each test case, the most accurate algorithm was the one consisting of the scaled orthographic camera model, but this was also the slowest one as expected. The accuracy of the weak-perspective factorization is better than the affine one after both structure and motion reconstruction.

5.3 Parameter Initialization for Bundle Adjustment

As discussed above, the affine, weak-perspective and scaled orthographic SfM method can estimate the 3D structure of the tracked points. In this chapter, we are examining how obtained 3D points can be used as initial parameters for perspective reconstruction. The 3D coordinates are perspectively projected. The applied perspective reconstruction itself is the SBA implementation [7] of the well-known bundle adjustment (B. Triggs and P. McLauchlan and R. Hartley and A. Fitzgibbon, 2000) method.

When the structure matrices have already been computed, the estimation of the 3×4 projection matrices is a camera calibration problem. In our test, the normalized Direct Linear Transformation (DLT) algorithm (Hartley and Zisserman, 2000) was applied (it is also known as the 'six-point method'). The projection matrix was then decomposed into camera intrinsic and extrinsic parameters.

We compared the initial parameters of the three compared method. BA cannot guarantee that global optimum is reached through estimation; it is interesting that BA after the *weak-perspective, scaled orthographic and affine parameter initialization usually gives the same results*. The time demand of the two methods differs a bit: the weak-perspective (**WP**) and scaled orthographic (**SO**) methods usually help BA to yield faster convergence than affine (**AFF**) parameterization. We also applied the classical Tomasi-Kanade (**TK**) algorithm (Tomasi, C. and Kanade, T., 1992) for parameter initialization, and that yielded the slowest BA convergence. Moreover, its results were usually less accurate than those of the other three algorithms (**AFF**,**SO**,**WP**); therefore, it seems that BA usually converges to local minima if the initial parameters are obtained by Tomasi-Kanade factorization. Time demand (msec) in our test sequences are listed in Table 1. There is not significant difference between the case when the scaled orthographic or proposed weak-perspective factorization is applied in order to compute initial parameters for perspective BA. Therefore the overall running time of **WP** method is smaller as the **WP** factorization is faster than the **SO** one.

The conclusion of the parameter initialization test is that the weak-perspective algorithm gives the fastest results since the time demand for factorization itself is faster than that of rival methods, while the speed of the BA algorithm is approximately the same in the case of **WP** and **SO** parameter initialization; the BA method usually converges to the same 3D reconstructions.

6 Tests on real data

We tested the proposed algorithm on several real sequences.

'Face' sequence. Our first test sequence consisted of 331 images of a quasi-rigid human face

[7] http://users.ics.forth.gr/∼lourakis/sba/

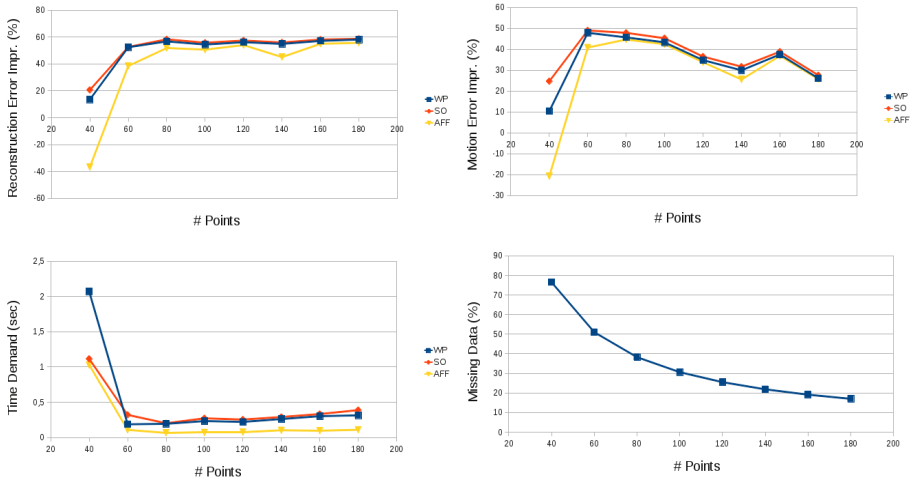

Figure 4: Improvement of reconstruction and motion errors (top charts) and time demand (bottom left) w.r.t. number of points. Bottom right chart shows the ratio of missing data.

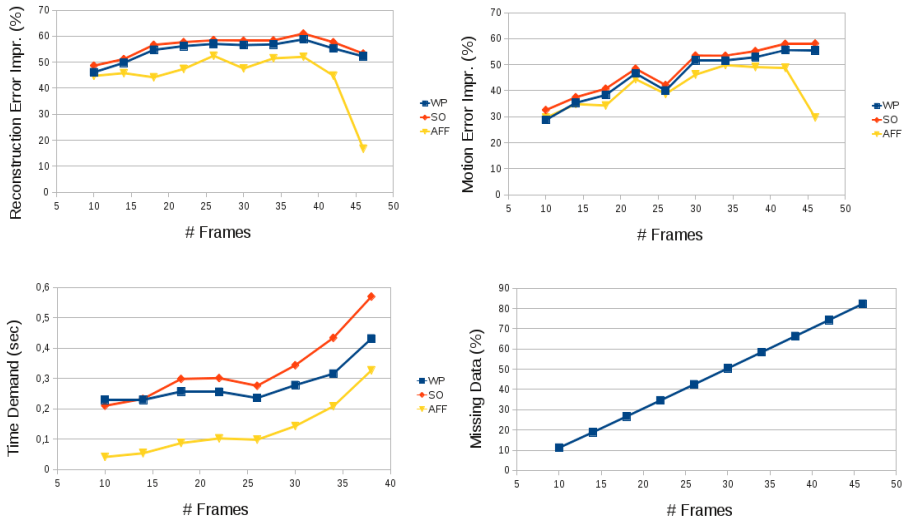

Figure 5: Improvement of reconstruction and motion errors (top charts) and time demand (bottom left) w.r.t. number of frames. Bottom right chart show the ratio of missing data.

Table 1: Time demand of Bundle Adjustment. There is not significant difference between the scaled orthographic (**SO**) and weak-perspective (**WP**) values.

Test Sequence	TK	WP	SO	Aff
versus noise	1628.35	**986.12**	989.805	1033.27
versus frames	1649.63	598.93	**582.22**	693.77
versus points	985.65	452.525	**444.7**	450.4375

as visualized in the left two plots of Fig. 6. We computed a two-dimensional Active Appearance Model (Matthews and Baker, 2003) (AAM) that contained 44 feature points of the face. The tracking was done by a modified implementation of GreatYao library. The missing ratio in this example is 0% since the AAM model computation estimates all the points in all the frames. The proposed weak-perspective algorithm successfully computed the 3D coordinates of the AAM feature points as pictured in the right part of Fig. 6 (the points are triangulated and the whole model is textured based on one of the original image). We tried the scaled orthographic reconstruction method as well, but the affine model was not run, because there are no missing elements in the data, thus the classical Tomasi-Kanade factorization (Tomasi, C. and Kanade, T., 1992; Weinshall and Tomasi, 1995) can be carried out. The threshold ε of the stopping criterion was set to 10^{-5} for both the scaled orthographic, and the weak-perspective methods. The time demand of the proposed algorithms was 35 secs, while the scaled orthographic one finished its computation in 49 secs.

Figure 6: 2 out of 331 original image (top) and two views (bottom) of the reconstructed 3D model of 'Face' sequence.

'Dino' sequence. The 'Dino' sequence, downloaded from the web page of the Oxford University[8], consisted of 36 frames and 319 tracked points. The measurement matrix had a missing data ratio of 77%. Input images are visualized on the left images of Fig. 7. The reconstructed 3D points were computed by the proposed SfM method. The time demands of that was 26 seconds (the affine and scaled orthographic SfM methods have computed the reconstruction in 6 and 34 seconds, respectively). The results are plotted in the right part of Fig. 7.

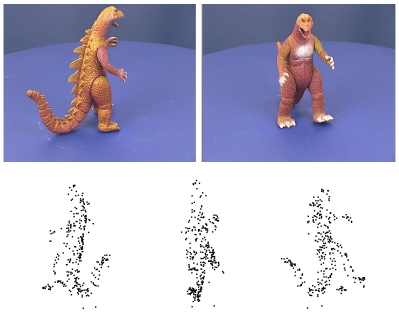

Figure 7: Results on 'Dino' sequence: Top: 2 out of 36 original image and (bottom) reconstructed point cloud captured from three views.

Another interesting examination is to compare the quality of the reconstructed 3D models; the points themselves seem very similar, but the camera positions differs significantly. We compared those after factorization by the original Tomasi-Kanade method to affine, weak-perspective and scaled orthographic improvement as visualized in Fig. 8. The quality of the original factorization method (top-left image) is very erroneous since the cameras should be located at regular locations of a circle. The improvements are significantly better. As expected, the scaled orthographic reconstruction (bottom-right image) serves better quality, the proposed weak-perspective (bottom-left) is slightly worse, but it serves acceptable results; the affine refinement (top-right plot) is also satisfactory.

The visualization of the camera optical centers for non-perspective cameras was not trivial. The pose of the cameras were obtained by the factorizations, but the focal length could not be estimated. For this reason, the focal length was set manually.

'Cat' sequence. We tested the proposed algorithm on our 'Cat' sequence. The cat statuette was rotated on a table and 92 photos were taken by a common commercial digital camera. The regions of the statuette in the images were automatically deter-

[8] http://www.robots.ox.ac.uk/∼amb/

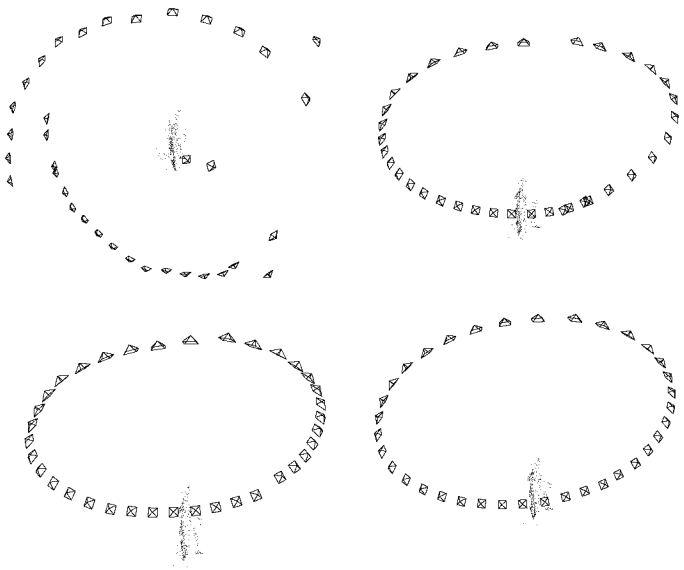

Figure 8: Reconstructed 'Dino' model with estimated cameras. Top-left: Original Tomasi-Kanade factorization. Top-right: Affine factorization. Bottom-left: Weak-perspective factorization (proposed method). Bottom-right: Scaled orthographic factorization. The cameras should be uniformly located around the estimated point cloud of the plastic dinosaur. The difference between weak-perspective and scaled orthographic camera parameters is not significant.

mined.

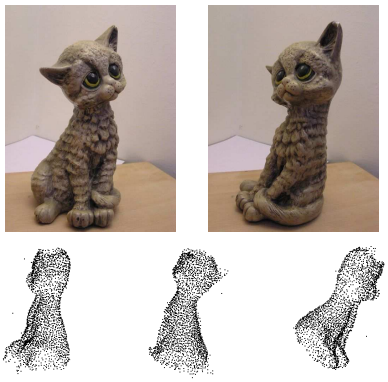

Figure 9: Two images (top) of sequence 'Cat' and the reconstructed points from three views (bottom).

Feature points were detected using the widely-used KLT (Tomasi, C. and Shi, J., 1994) algorithm, and the points were tracked by a correlation-based template matching method. A features point was labeled as missing if the tracker could not find its location in the next image, or the location was not inside the automatically detected region of the object. The measurement matrix of the sequence consisted of 2290 points and 92 frames. The missing data ratio was 82%, that is very high.

The 3D reconstructed points are visualized on the right plots of Fig. 9. We tested every possible method and compared the time demand of the methods: the running times of the affine, scaled orthographic, and weak-perspective factorization were 484, 199, and 99 seconds, respectively.

7 Conclusion

We have presented the optimal calibration algorithm for the weak-perspective camera model here. The proposed method minimizes the reprojection error of feature points in the least squares sense. The solution is given by a closed-form formula. We have also proposed a SfM algorithm; it is an iterative one, and every iteration consists of two optimal steps: (i) The structure matrix computation is a linear problem, therefore it can be optimally estimated in the least squares sense, while (ii) the camera parameters are obtained by the novel optimal weak-perspective camera calibration method. The introduced SfM approach can also cope with the problem of missing feature points.

The proposed SfM algorithm was compared to the affine (Shum et al., 1995) and scaled orthographic (L. Hajder and Á. Pernek and Cs. Kazó, 2011) methods. It was shown that our method is significantly more accurate than the affine one, and usually faster than the scaled orthographic SfM algorithm due to the optimal weak-perspective calibration. We successfully

applied the novel method to compute the initial parameters for bundle adjustment-type 3D perspective reconstruction.

The Java implementation of our weak-perspective SfM algorithm can be downloaded from the web [9].

Acknowledgement. This work was supported in part by the project SCOPIA Development of software supported clinical devices based on endoscope technology (VKSZ_14-1-2015-0072) financed by the Hungarian National Research, Development and Innovation Fund (NKFIA).

REFERENCES

Arun, K. S., Huang, T. S., and Blostein, S. D. (1987). Least-squares fitting of two 3-D point sets. *IEEE Trans. on PAMI*, 9(5):698–700.

B. Triggs and P. McLauchlan and R. Hartley and A. Fitzgibbon (2000). Bundle Adjustment – A Modern Synthesis. In *Vision Algorithms: Theory and Practice*, pages 298–375.

Buchanan, A. M. and Fitzgibbon, A. W. (2005). Damped newton algorithms for matrix factorization with missing data. In *Proceedings of the 2005 IEEE CVPR*, pages 316–322.

Bue, A. D., Xavier, J., Agapito, L., and Paladini, M. (2012). Bilinear modeling via augmented lagrange multipliers (balm). *IEEE Trans. on PAMI*, 34(8):1496–1508.

DeMenthon, D. F. and Davis, L. S. (1995). Model-based object pose in 25 lines of code. *IJCV*, 15:123–141.

Hartley, R. and Kahl, F. (2007). Optimal algorithms in multiview geometry. In *Proceedings of the Asian Conf. Computer Vision*, pages 13–34.

Hartley, R. and Schaffalitzky, F. (2003). Powerfactorization: 3d reconstruction with missing or uncertain data.

Hartley, R. I. and Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Hesch, J. A. and Roumeliotis, S. I. (2011). A direct least-squares (dls) method for pnp. In *International Conference on Computer Vision*, pages 383–390. IEEE.

Horaud, R., Dornaika, F., Lamiroy, B., and Christy, S. (1997). Object pose: The link between weak perspective, paraperspective and full perspective. *International Journal of Computer Vision*, 22(2):173–189.

Horn, B., Hilden, H., and Negahdaripourt, S. (1988). Closed-form Solution of Absolute Orientation Using Orthonormal Matrices. *Journal of the Optical Society of America*, 5(7):1127–1135.

Jenkins, M. A. and Traub, J. F. (1970). A Three-Stage Variables-Shift Iteration for Polynomial Zeros and Its Relation to Generalized Rayleigh Iteration. *Numer. Math*, 14:252263.

Kahl, F. and Hartley, R. I. (2008). Multiple-view geometry under the linfinity-norm. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(9):1603–1617.

[9] http://web.eee.sztaki.hu/Factorization.zip

Kanatani, K., Sugaya, Y., and Ackermann, H. (2007). Uncalibrated factorization using a variable symmetric affine camera. *IEICE - Trans. Inf. Syst.*, E90-D(5):851–858.

Ke, Q. and Kanade, T. (2005). Quasiconvex Optimization for Robust Geometric Reconstruction. In *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision*, pages 986–993.

L. Hajder and Á. Pernek and Cs. Kazó (2011). Weak-Perspective Structure from Motion by Fast Alternation. *The Visual Computer*, 27(5):387–399.

Lepetit, V., F.Moreno-Noguer, and P.Fua (2009). Epnp: An accurate o(n) solution to the pnp problem. *International Journal of Computer Vision*, 81(2):155–166.

Marques, M. and Costeira, J. (2009). Estimating 3d shape from degenerate sequences with missing data. *CVIU*, 113(2):261–272.

Matthews, I. and Baker, S. (2003). Active appearance models revisited. *International Journal of Computer Vision*, 60:135–164.

Okatani, T. and Deguchi, K. (2006). On the wiberg algorithm for matrix factorization in the presence of missing components. *IJCV*, 72(3):329–337.

Pernek, A., Hajder, L., and Kazó, C. (2008). Metric Reconstruction with Missing Data under Weak-Perspective. In *BMVC*, pages 109–116.

Poelman, C. J. and Kanade, T. (1997). A Paraperspective Factorization Method for Shape and Motion Recovery. *IEEE Trans. on PAMI*, 19(3):312–322.

Ruhe, A. (1974). Numerical computation of principal components when several observations are missing. Technical report, Umea Univesity, Sweden.

Schweighofer, G. and Pinz, A. (2008). Globally optimal o(n) solution to the pnp problem for general camera models. In *BMVC*.

Shum, H.-Y., Ikeuchi, K., and Reddy, R. (1995). Principal component analysis with missing data and its application to polyhedral object modeling. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(9):854–867.

Sturm, P. and Triggs, B. (1996). A Factorization Based Algorithm for Multi-Image Projective Structure and Motion. In *ECCV*, volume 2, pages 709–720.

Tomasi, C. and Kanade, T. (1992). Shape and Motion from Image Streams under orthography: A factorization approach. *Intl. Journal Computer Vision*, 9:137–154.

Tomasi, C. and Shi, J. (1994). Good Features to Track. In *IEEE Conf. Computer Vision and Pattern Recognition*, pages 593–600.

Wang, G., Wu, Q. M. J., and Sun, G. (2008). Quasi-perspective projection with applications to 3d factorization from uncalibrated image sequences. In *CVPR*.

Weinshall, D. and Tomasi, C. (1995). Linear and Incremental Acquisition of Invariant Shape Models From Image Sequences. *IEEE Trans. on PAMI*, 17(5):512–517.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Trans. on PAMI*, 22(11):1330–1334.

Zheng, Y., Kuang, Y., Sugimoto, S., Åström, K., and Okutomi, M. (2013). Revisiting the pnp problem: A fast, general and optimal solution. In *ICCV*, pages 2344–2351.