# A Low-Rank Constraint for Parallel Stereo Cameras

Christian Cordes, Hanno Ackermann, and Bodo Rosenhahn

Leibniz University Hannover, Germany
{ccordes,ackermann,rosenhahn}@tnt.uni-hannover.de

**Abstract.** Stereo-camera systems enjoy wide popularity since they provide more restrictive constraints for 3d-reconstruction. Given an image sequence taken by parallel stereo cameras, a low-rank constraint is derived on the measurement data. Correspondences between left and right images are not necessary yet reduce the number of optimization parameters. Conversely, traditional algorithms for stereo factorization require *all* feature points in both images to be matched, otherwise left and right image streams need be factorized independently. The performance of the proposed algorithm will be evaluated on synthetic data as well as two real image applications.

## 1 Introduction

Image sequences taken by a stereo camera system are important input to many problems in computer vision. This article proposes a low-rank constraint on the feature trajectories which can be used in applications such as rigid or non-rigid 3d-reconstruction, motion segmentation or trajectory completion.

Given only two, three or four images taken by cameras in general configuration, 3D-reconstructions can be computed using the *epipolar constraint*. If the single-camera sequence consists of more than four images, a commonly employed heuristic is to estimate reconstructions from each two, three or four consecutive image segments, and use these to initialize a *bundle adjustment* [12].

The so-called *factorization algorithm* [11], conversely, is able to compute a 3d-reconstruction from arbitrary many images taken by *affine* cameras[1]. Its popularity stems from its simplicity: a matrix consisting of the feature points is factorized by means of a single *singular value decomposition*.

Generalizations exist to handle missing data [13, 10] and uncalibrated projective cameras [5]. Low-rank constraints were also derived for multi-body [4] and non-rigid [2] 3D-reconstructions. Furthermore, algorithms resting on factorization also exist for other problems such as motion segmentation [7], trajectory completion as well as optical flow estimation [6].

[1] This model requires that the distance between camera and object is large as compared with the variation of depth within the scene. The requirement is necessary for any affine camera model, be it orthographic, weak-perspective, paraperspective or the more flexibel one proposed in [9]. For a comprehensive treatment on affine camera models confer to [8].

A factorization algorithm which estimates a 3d-reconstruction from non-rigidly deforming objects taken by a *convergent* stereo camera was proposed in [3]. However, *during the factorization stage*, this method need consider the two cameras separately if not all correspondences are known across left and right image streams. The stereo constraint is imposed only by means of a subsequent optimization. Given arbitrarily many, static camera rigs, a rank-12 constraint was derived in [1].

Both aforementioned works do not consider the case of missing data which naturally occurs due to tracking failure or scene occlusion. In this work, we consider a stereo setup of *parallel* cameras. In contrast to the algorithm proposed in [3] a low-rank constraint is derived which can be imposed *during the factorization stage*. As compared with the algorithm in [1], the low-rank constraint introduced here is significantly smaller leading to more robust estimates particularly in the presence of missing data.

**The contributions** made in this article can be summarized as follows:

- A *low-rank constraint* is derived assuming a pair of parallely-aligned stereo cameras.
- It can be imposed by means of *matrix factorization*.
- As significantly fewer variables are involved during factorization it is more robust with respect to noise and missing data.
- The proposed solution does not require correspondences between left and right images of the cameras. If available, these can be used to further reduce the degrees of freedom within the model.
- Missing correspondences can be handled.

The proposed solution will be evaluated quantitatively with synthetic data. We demonstrate the versatility of the algorithm by drawing on two real-image sequences. One application draws on rigid 3D-reconstruction while the other achieves trajectory completion given a scene in which several rigid bodies move independently from each other.

In Sec. 2 we will briefly review the factorization algorithm before deriving a low-rank constraint given stereo cameras in Sec. 3. The evaluation on synthetic data is presented in Sec. 4. Results of real-image experiments are demonstrated in Sec. 5. Lastly, we conclude this article with Sec. 6.

## 2 Rigid Factorization Algorithm

Given $N$ 3D-points $X_j$, $j = 1, \dots, N$ observed by $M$ affine cameras $P_i$, $i = 1, \dots, M$, the projection $x_{ij}$ of the $j$th point into the $i$th image can be modelled by

$$x_{ij} = P_i X_j. \tag{1}$$

The difference to perspective projection is that equality holds in Eq. (1) whereas the latter implies equality up to scale, only.

Each affine projection matrix $P_i$ can be decomposed into an $2 \times 3$ affine calibration matrix $K_i$, a $3 \times 3$ rotation matrix $R_i$ indicating the orientation of the camera at image $i$, and a 3-vector $t_i$ which implies the position of the camera

$$P_i = K_i \left[ R_i^{-1} \;\; -R_i^{-1} t_i \right]. \tag{2}$$

The homogeneous vectors $X_j$ indicate the $x$, $y$, and $z$-coordinates of the $j$th 3D-point. As model of the affine camera we assume weak-perspective projection. The matrices $K_i$ then are defined by

$$K_i = s_i \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{3}$$

where $s_i$ denotes a scalar.

The projection of all 3D-points into all images can then be formulated as

$$\underbrace{\begin{bmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{M1} & \cdots & x_{MN} \end{bmatrix}}_{W^{2M \times N}} = \underbrace{\begin{bmatrix} P_1 \\ \vdots \\ P_M \end{bmatrix}}_{P^{2M \times 4}} \underbrace{\begin{bmatrix} X_1 & \cdots & X_N \end{bmatrix}}_{X^{4 \times N}} \tag{4}$$

$$\tag{5}$$

Assuming that the cameras are generally oriented and that the 3D-shape is not degenerate, the matrices $P$ and $X$ both have rank 4. This implies that the rank of matrix $W$ cannot be larger than 4.

By means of singular value decomposition, we may therefore factorize $W$ into

$$W = U \Sigma V^\top \tag{6}$$

where all but the largest four singular values on the diagonal of matrix $\Sigma$ are identically zero. This idea was first proposed in [11] and is known as the factorization algorithm.

Consequently, matrices $U$ and $V$ can be truncated to those four vectors corresponding to the four non-zero singular values. Similarly, we truncate $\Sigma$ to be of size $4 \times 4$. With a slight abuse of notation, denote these truncated matrices by $U$, $\Sigma$, and $V$ in the following.

Affinely distorted estimates of $P$ and $X$ can be taken by $U$ and $\Sigma V^\top$, respectively. To obtain undistorted estimates, a correcting matrix $A$ need be determined by affine self calibration similarly to the self calibration step necessary for projective reconstruction.

## 3 Affine Stereo Factorization

### 3.1 A Low Rank Constraint on Parallel Stereo Cameras

Assume that we are given two affine cameras $P^1$ and $P^2$ parallely oriented and with equal distance $c$ to the center in between them. Further assume that this center is located in the origin of the world coordinate system.

If we align the camera orientations with the coordinate axes and take the basis line parallel to the $x$-axis we obtain

$$P^1 = K \begin{bmatrix} I & v \end{bmatrix} \quad \text{and} \quad P^2 = K \begin{bmatrix} I & -v \end{bmatrix} \tag{7}$$

where $I$ denotes the identity matrix and $v = \begin{bmatrix} c & 0 & 0 \end{bmatrix}^\top$.

A rigid transformation of the stereo camera system by an rotation $R_i$ and translation $t_i$ amounts to multiplication with

$$H_i = \begin{bmatrix} R_i^\top & -R_i^\top t_i \\ \mathbf{0}^\top & 1 \end{bmatrix} \tag{8}$$

where $\mathbf{0}$ denotes a 3-vector consisting of zeros. For $P_i^1$ and $P_i^2$ we obtain

$$P_i^1 = K_i \begin{bmatrix} R_i^\top & -R_i^\top t_i + v \end{bmatrix} \quad \text{and} \quad P_i^2 = K_i \begin{bmatrix} R_i^\top & -R_i^\top t_i - v \end{bmatrix}. \tag{9}$$

By defining $t_i^1 = -R_i^\top t_i + v$ and $t_i^2 = -R_i^\top t_i - v$ we can simplify Eq. (9) to

$$P_i^1 = K_i \begin{bmatrix} R_i^\top & t_i^1 \end{bmatrix} \quad \text{and} \quad P_i^2 = K_i \begin{bmatrix} R_i^\top & t_i^2 \end{bmatrix}. \tag{10}$$

Let the *joint projection matrix* be

$$P_i = K_i \begin{bmatrix} R_i^\top & t_i^1 & t_i^2 \end{bmatrix}. \tag{11}$$

The projection of the $j$th 3D-point into the $i$th images can be expressed by

$$x_{ij}^1 = P_i \begin{bmatrix} X_j^\top & 1 & 0 \end{bmatrix}^\top \quad \text{and} \quad x_{ij}^2 = P_i \begin{bmatrix} X_j^\top & 0 & 1 \end{bmatrix}^\top. \tag{12}$$

Denote by $W^1$ and $W^2$ the matrices consisting of all feature points of the first and second images, respectively. We now arrive at the affine stereo constraint

$$\begin{bmatrix} W^1 & W^2 \end{bmatrix} = \begin{bmatrix} P_1 \\ \vdots \\ P_M \end{bmatrix} \begin{bmatrix} X_1 & \cdots & X_N & X_1 & \cdots & X_N \\ 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 \end{bmatrix}. \tag{13}$$

We immediately see that the *joint measurement matrix* $W = \begin{bmatrix} W^1 & W^2 \end{bmatrix}$ can have rank 5 at most as both matrices on the right side of Eq. (13) have rank 5 assuming general motion and non-degenerate structure.

As for the factorization algorithm for a single moving camera, we can obtain affinely distorted estimates of motion and structure by singular value decomposition of $W$ into $U$ and $\Sigma V^\top$, respectively.

The model defined by Eq. (13) assumes that no correspondences between the two images taken at the same time are known. Otherwise, the number of parameters can be reduced even further.

### 3.2 Affine Stereo Self Calibration

Given affinely distorted estimates of structure and motion, the *affine stereo self calibration* problem is to determine a $5 \times 3$ matrix $A$ such that each two rows $U_i$ of $U$ are transformed to

$$K_i R_i = U_i A. \tag{14}$$

Letting $Q = AA^\top$ we can eliminate the unknown rotations by squaring both sides

$$K_i K_i^\top = U_i Q U_i^\top \tag{15}$$

Assuming a weak-perspective camera model, we arrive at

$$0 = \left(u_i^1\right)^\top Q\, u_i^1 - \left(u_i^2\right)^\top Q\, u_i^2 \quad \text{and} \tag{16a}$$

$$0 = \left(u_i^1\right)^\top Q\, u_i^2 \tag{16b}$$

where $\left(u_i^1\right)^\top$ and $\left(u_i^2\right)^\top$ denote the vectors corresponding to the first and second rows of $U_i$.

As the rank of matrix $Q$ equals 3, the problem defined by the Eqs. (16) and the rank-3 constraint is nonlinear. However, according to our experience, straight-forward nonlinear minimization converges fast and reliably to a good optimum. The rotation matrices can then be reconstructed by $R_i = U_i A$. The correcting transformation $A$ can be obtained from the eigendecomposition of $Q = VDV^\top$ by taking $A = VD^{\frac{1}{2}}$ as $Q$ is positive semi-definite. To strictly enforce that each $R_i$ is a rotation matrix we can use polar decomposition.

For a 3D-reconstruction we further need estimates of $t_i^1$ and $t_i^2$. We can obtain these by

$$\begin{bmatrix} t_i^1 & t_i^2 \end{bmatrix} = U_i\, Q_\perp \tag{17}$$

with $Q_\perp = I - QQ^+$ where the symbol $(\cdot)^+$ denotes the generalized inverse.

Having estimated the motion parameters $K_i$, $R_i$, $t_i^1$ and $t_i^2$, the structure can be inferred by triangulation. A linear solution to this problem is given by

$$\begin{bmatrix} X^1 & X^2 \end{bmatrix} = \begin{bmatrix} P_1 \\ \vdots \\ P_m \end{bmatrix}^+ \begin{bmatrix} W^1 & W^2 \end{bmatrix} \tag{18}$$

If not all entries of $W^1$ or $W^2$ are known, we can use a nonlinear optimization algorithm for matrix completion, see Sec. 4. An excellent guide on affine 3D-reconstruction can be found in [8].

## 4 Experiments on Synthetic Data

We created $N = 146$ 3D-points which were projected $M = 80$ times in two images according to the affine stereo camera model. For each camera six of these images are shown in Fig. 1.

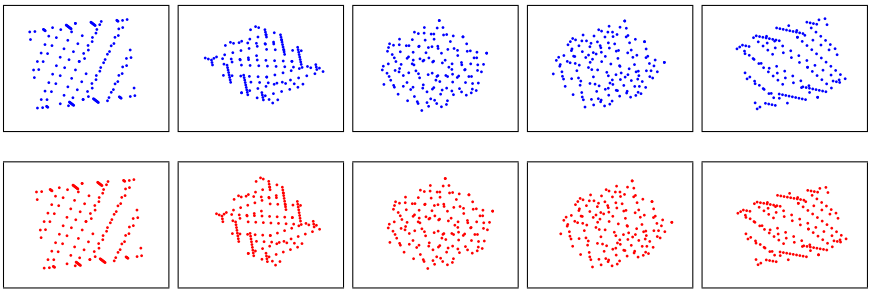

**Fig. 1.** Five images of 146 simulated 3D-points projected into forty images. The upper row corresponds to the left camera of the stereo system, the bottom row to the right.

We simulated occlusion by making all but the measurements along the main diagonal of both $W^1$ and $W^2$ invisible. The amount of unknown data was varied in between 0% and 30% in steps of ten percent. We added normally distributed noise with standard deviations $\sigma = \{0, 1, 2, 3\}$. For each combination of missing data and noise, we performed ten trials, i.e. perturbed the data ten times differently.

For estimating the motion $U$ under missing entries of $W$, we used alternating-least-squares (ALS). The proposed algorithm was compared with the one introduced in [3].

We computed the root-mean-square-error (RMSE) between the visible, unperturbed matrix entries and the estimates. To assess the accuracy of subspace fitting we measured the sum of the canonical angles between the estimated subspace and the noise-free ground truth (SSP error). Lastly, a 3D-error was computed as the average sum of the Euclidean distances between the estimated 3D-points and the ground truth. This error was further normalized by the Frobenius norm of the matrix consisting of the ground truth 3D-points.

Average results of the ten trials are shown in Fig. 2. The plots from left to right correspond to the RMSE, the subspace error and the 3D-error. The solid line indicates the algorithm proposed here, the dashed line the one in [3]. The blue, green, and red colored lines correspond to 10%, 20% and 30% unknown data. As can be seen, the proposed algorithm performs superior.

As the algorithm is based upon ALS iterations, its computational complexity is slightly lower than that of the reference algorithm as the latter requires the larger rank-8 factorization.

## 5 Experiments on Real Images

### 5.1 Application 1: Rigid 3D-Reconstruction

Figure 3 shows five out of 74 images of a sequence of a rigid scene. The images in the upper row are taken by the left camera, those in the bottom row by the right

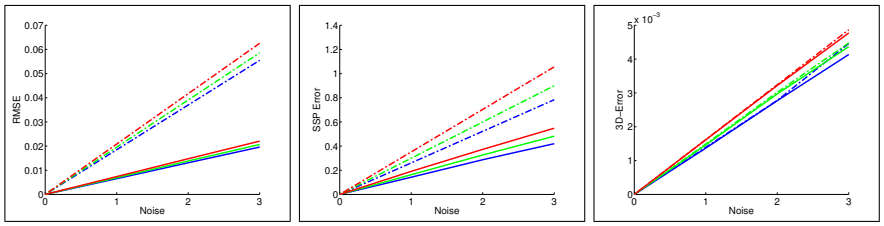


**Fig. 2.** From left to right: root-mean-square-error (RMSE); subspace error (SSP, sum of canonical angles); normalized 3D-error. The solid line indicates the proposed algorithm, the dashed line the one in [3]. The blue, green and red lines correlate to 10%, 20% and 30% missing data.

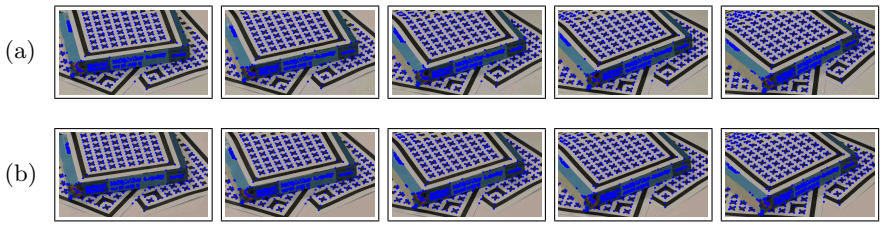


**Fig. 3.** Five out of 74 images of a sequence of a rigid scene. A total of 2112 3D-points were tracked through the images. Starting from the first image, each 15th image is shown. The joint measurement matrix has 32% unknown entries. The upper row shows the images taken by the left camera, the bottom right the images to the right camera.

camera. A total of 2112 feature trajectories was followed through the images. The joint measurement matrix has 32% unknown entries.

The images weren taken by a HDC-Z10000 stereo camera with focal length set to $28mm^2$ The object was about $4m$ apart from the camera and measured approximately $30cm$ in diameter. The optical axes were set such that a 3D-view appeared on the camera screen.

Some feature points are located at lines and do not move rigidly. We thus perform a simple outlier rejection. First, the standard deviation of the image-to-image motion vectors is computed. We then execute the alternating-least-squares method and remove trajectories whose estimated motion vectors differs from the known motion vectors by more than two standard deviations. These two steps are iterated until no more outlying trajectories are detected.

Six views of the 3D-reconstructions of 1658 trajectories are shown in Fig. 4. The different planes are perpendicular, and the repetitive patterns of the 2D-points is well reflected by the 3D-points.

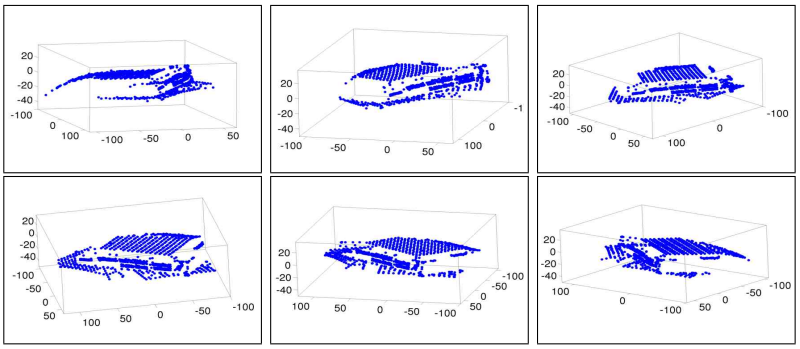

**Fig. 4.** Six views of the 3D-reconstruction from the data shown in Fig. 3.

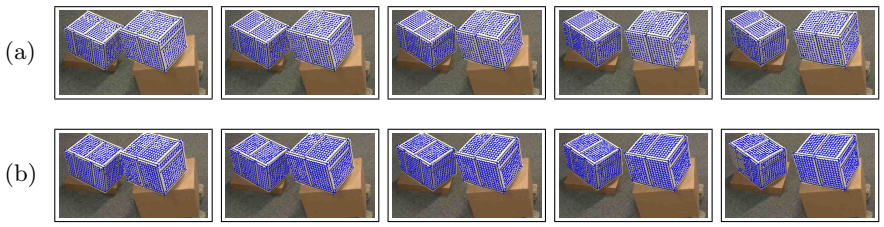


**Fig. 5.** Five out of 251 images of a sequence with two rigid bodies moving independently. Starting from the first image, each 50th image is shown. A total of 5605 3D-points were tracked. The joint measurement matrix has 34% unknown entries. The upper row shows the images taken by the left camera, the bottom images to the right camera.

### 5.2 Application 2: Trajectory Completion

Figure 5 shows five out of 251 images of a sequence in which two rigid bodies move independently. The images in the upper row are taken by the left camera, those in the bottom row by the right camera. A total of 3967 trajectories was found in the sequence. The joint measurement matrix has 38% unknown entries.

The images were also taken by a HDC-Z1000 stereo camera. The two boxes were approximately $2m$ in front of it. The depth variation within the scene is larger than $0.5m$ hence the assumption required by the affine camera is strongly violated.

As both bodies move independently, each of the two sets of trajectories spans a 5-dimensional subspace. Therefore, we can perform a rank-10 factorization of the joint measurement matrix. Although some outliers were present in the data, we did not perform any filtering.

[2] This amounts to a focal length of $320mm$ in terms of a $35mm$ sensor).

Using these estimates, we impute missing data and compare the results with a regular factorization. The latter needs to process both image streams independently, as not all 2D-points are matched across the left and right image streams.

As not all correspondences are known *between* the two image streams, a regular matrix factorization needs to process each camera stream independently. Results of the estimated 2D-points using a rank-8 alternating-least-squares on the data to the left camera stream are shown in Fig. 6. The two images correspond to the first and last frames shown in Fig. 5(a). As can be seen, many points are placed randomly. The shown results are representative for the other images since many points move randomly in all images of the sequence.

Figure 7 shows the completed trajectories using the proposed algorithm. The images in the upper row correspond to the left camera stream shown in Fig. 5(a), those in the bottom row to the right stream of Fig. 5(b). The three images in each row of Fig. 7 show the estimates corresponding to the first, third and fifth frame shown in Fig. 5. As can be seen, the proposed algorithm estimates missing feature points correctly throughout the complete sequence.

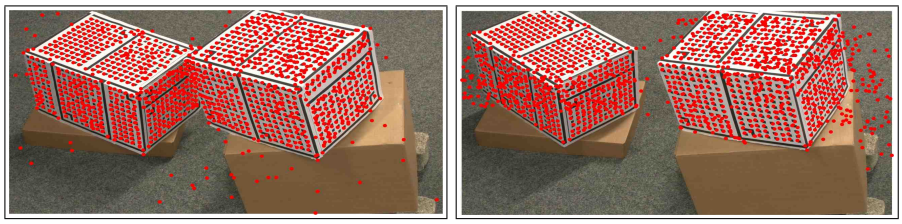

**Fig. 6.** Results of a rank-8 matrix factorization (cf. [3]) on the data to the left camera stream. The two images correspond to the first and last image of the sequence shown in Fig. 5(a). As can be seen, many points are erroneously estimated.

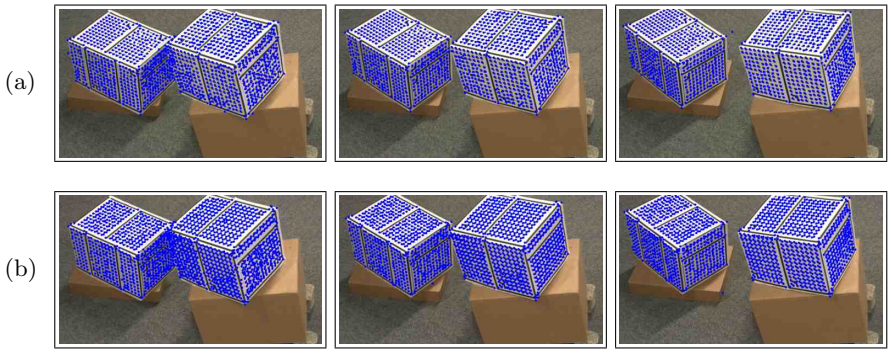


**Fig. 7.** Estimated 2D-points for the left (a) and right (b) camera streams shown in Fig. 5. The three images correspond to the first, third and fifth images shown in Fig. 5.

## 6 Summary and Discussion

Given image streams taken by a parallely-aligned affine stereo camera system, this article introduced a low-rank constraint which trajectories across both images of both streams need to satisfy.

Conversely, existing algorithms for stereo cameras need to factorize both streams independently if not all correspondences are known between *the two image streams*. In other words, the stereo constraint cannot be considered.

The viability of the derived low-rank constraint was evaluated using synthetic data. Furthermore, two different applications using real images demonstrated that the algorithm is indeed able to estimate high-quality results. The introduced constraint does not only apply to rigid data but can be readily generalized to multi-body or non-rigidly deforming scenes.

## References

1. Angst, R., Pollefeys, M.: Static multi-camera factorization using rigid motion. In: International Conference on Computer Vision (ICCV). pp. 1203–1210 (2009)
2. Bregler, C., Hertzmann, A., Biermann, H.: Recovering non-rigid 3d shape from image streams. In: IEEE Computer Vision and Pattern Recognition (CVPR). pp. 690–696. Hilton Head, SC, USA (2000)
3. Bue, A.D., de Agapito, L.: Non-rigid stereo factorization. International Journal of Computer Vision (IJCV) 66(2), 193–207 (2006)
4. Costeira, J.P., Kanade, T.: A multibody factorization method for independently moving objects. International Journal of Computer Vision (IJCV) 29(3), 159–179 (Sep 1998)
5. Heyden, A., Berthilsson, R., Sparr, G.: An iterative factorization method for projective structure and motion from image sequences. Image Vision Comput. 17(13), 981–991 (1999)
6. Irani, M.: Multi-frame optical flow estimation using subspace constraints. In: International Conference on Computer Vision (ICCV). pp. 626–633 (1999)
7. Kanatani, K.: Motion segmentation by subspace separation: Model selection and reliability evaluation. International Journal of Image and Graphics 2(2), 179–197 (2002)
8. Kanatani, K., Sugaya, Y.: Factorization without factorization: complete recipe. Tech. Rep. 1&2, Okayama University, Japan (March 2004)
9. Kanatani, K., Sugaya, Y., Ackermann, H.: Uncalibrated factorization using a variable symmetric affine camera. In: 7th European Conference on Computer Vision (ECCV). pp. 147–158 (2006)
10. Ruhe, A., Wedin, P.: Algorithms for separable nonlinear least squares problems. Society for Industrial and Applied Mathematics Review 22(3), 318–337 (1980)
11. Tomasi, C., Kanade, T.: Shape and motion from image streams under orthography: a factorization method. International Journal on Computer Vision (IJCV) 9(2), 137–154 (November 1992)
12. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment – a modern synthesis. In: Proceedings of the International Workshop on Vision Algorithms: Theory and Practice. pp. 298–372 (2000)
13. Wold, H.: Estimation of principal components and related models by iterative least squares. In: Krishnaiah (ed.) Multivariate Analysis. pp. 391–420 (1966)