Multilinear Model Estimation with L^2-Regularization

F. R. Schmidt[1], H. Ackermann[2], and B. Rosenhahn[2]

[1] University of Western Ontario, Canada
fschmidt@uwo.ca
[2] Leibniz University Hannover, Germany
{ackermann,rosenhahn}@tnt.uni-hannover.de

Abstract. Many challenging computer vision problems can be formulated as a multilinear model. Classical methods like principal component analysis use singular value decomposition to infer model parameters. Although it can solve a given problem easily if all measurements are known this prerequisite is usually violated for computer vision applications. In the current work, a standard tool to estimate singular vectors under incomplete data is reformulated as an energy minimization problem. This admits for a simple and fast gradient descent optimization with guaranteed convergence. Furthermore, the energy function is generalized by introducing an L^2-regularization on the parameter space. We show a quantitative and qualitative evaluation of the proposed approach on an application from structure-from-motion using synthetic and real image data, and compare it with other works.

1 Introduction

To detect a model based only on observed images constitutes one of the central tasks in computer vision applications. Problems like structure and motion estimation as well as 3D or even 4D reconstruction can be formulated as a model fitting problem. Assuming temporal coherence leads to a smoothness prior on some variables. In this work we will focus on problems that are given as a *multilinear model* and we will show how the introduction of an L^2-regularizer leads to better solutions which can even be computed more efficiently.

Fitting a model with only a few parameters to observed data is the base of the well understood method of *principal component analysis* (PCA). It is used, for instance, for computation of eigenfaces [13], image matching [20], pose and shape estimation [12], rigid structure and motion (SfM) estimation [19] or non-rigid SfM [2]. *Singular value decomposition* (SVD) is often used to compute a PCA. Since SVD can be computed quite easily, it is very popular for dimension reduction approaches. However, SVD requires all measurements to be known. In many applications in computer vision, for instance structure and motion estimation, points cannot be observed because of occlusions or tracking failures.

In [19, 3], missing observations are dealt with by solving complete sub-sets and propagating these solutions while an EM approache is favored in [18]. Both

types of algorithms work well for low noise or low amounts of missing values, but they fail for realistic problems. A different approach to estimate the multilinear model is the *power factorization* or *NIPALS* approach [22, 11] which minimizes an L^2 energy function. It estimates solutions starting from the complete data, *i.e.* it does not begin on any sub-set of the data.

In [16, 17] Newton and Gauss-Newton approaches were considered which were later generalized to weighted data [9, 4]. To be more robust to errors that are related to missing or corrupted data, different error norms are used in [6, 15, 8]. Another approach to cope with corrupted data is to enforce additional constraints that are specific to the problem. Constraints on individual projection matrices were used in [14], consistency with epipolar geometry was imposed in [1] and the smoothness of camera trajectories was enforced by means of a Kalman filter in [10].

In this work we will minimize the common L^2 energy function of [22, 11, 16, 17, 4] by a gradient descent technique. The difference to power factorization is that this gradient descent jointly optimizes both sets of variables thereby avoiding accidental maximization and other numerical pitfalls.

Furthermore, we include a smoothness prior into the L^2 energy. This leads to the minimization of an energy E that is a convex combination of the L^2 energy E_{data} and the smoothness prior E_{smooth}. At the presence of a strong data term, the smoothness term has only a small effect. Otherwise (due to missing data), the smoothness term takes over control by extra- and interpolating information that are driven by neighboring data. Therefore, our approach is different from a Kalman filter approach in the sense that we do not *indiscriminately* enforce smoothness but only if there is insufficient data. As a result, non-smooth models can be estimated if the data information is very strong. Smoothness on the other hand is stronger in areas of missing data and is weaker in areas that are well defined by the observed measurements. A second difference to the Kalman filter is that we do not process the data sequentially. While the L^2-regularizer depends on a specific temporal order of the observed images, the overall energy functional E depends on all observations at the same time and will not change during the optimization process.

The difference to the Gauss-Newton variants of [4] is that we only impose smoothness on one set of variables. In the context of 3D-reconstruction we can thus enforce smooth camera trajectories yet allow for non-smooth surfaces or vice versa. Our experimental evaluation will show that the proposed method performs superior.

Overall, we present the following contributions in this paper:

- A global energy is minimized by gradient descent thus avoiding problems caused by starting from some sub-set of the data.
- The data term is extended by a smoothness term that governs those areas with few measurements.
- We do not enforce smoothness on all variables indiscriminately, but only smooth selectively. Thus partially non-smooth solutions can be obtained.
- We will demonstrate the proposed algorithm for simulated and real data.

This paper is organized as follows. In Section 2, we derive the gradient descent algorithm and discuss the advantages compared to power factorization. Section 3 generalizes the functional to include a smoothness prior. A quantitative analysis with synthetic data is conducted in Section 4. Real image experiments with challenging sequences are conducted in Section 5. In the same section we briefly discuss future work. Section 6 provides a summary.

2 Energy Minimization Formulation

In this section, we will formulate the multilinear model estimation as an energy minimization method and derive the gradient of this energy functional. We will then discuss in which sense a gradient descent deviates from the popular *power factorization* [11]. After presenting these two approaches, we will in Section 3 introduce a generalization that incorporates an L^2-smoothness term into the here presented energy functional.

First let us start with the general problem of multilinear model estimation. To this end, we have a set of observations that are encoded in a $m \times n$ matrix W. This can be understood as n observations of dimension m. The idea of a multilinear model is it now to incorporate the knowledge that the observations do not form an n-dimensional but rather an r-dimensional subspace with $r \ll m$.

Hence, we have r model vectors $x_1, \ldots, x_r \in \mathbb{R}^m$ and $y_1^\top, \ldots, y_r^\top \in \mathbb{R}^n$ that form a left and right base of the r-dimensional model space. Let $x_{i,k}$ and $y_{k,j}$ denote the kth coordinate of vector x_i and $y_j^\top$. Every element W_{ij} of W can then be written as a linear combination of x_i and $y_j^\top$ and we obtain for $W_{ij} = \sum_{k=1}^{r} x_{i,k} \cdot y_{k,j}$. If we now put the vectors x_i and $y_j^\top$ into the $m \times r$ matrix X and the $r \times n$ matrix Y, we receive the following equation

$$W = X \cdot Y. \tag{1}$$

In the perfect noiseless case, W has rank r, but since measurements are usually perturbed by noise, matrix W can also exhibit ranks which are larger than r. In practice, Equation (1) can thus not be solved exactly and is often reformulated as the following least squares problem:

$$\min_{X \in \mathbb{R}^{m \times r}, Y \in \mathbb{R}^{r \times n}} \|W - X \cdot Y\|_{\text{fro}}^2 \tag{2}$$

where the Frobenius norm $\|A\|_{\text{fro}} := \sqrt{\sum_{i,j} a_{i,j}^2}$ is the canonical norm on matrices. In order to solve this problem, we can simply use the SVD of $W = Q_1 \Sigma Q_2^\top$. This results in the solution $X = Q_1 \Sigma^{\frac{1}{2}}$ for the left subspace and $Y = \Sigma^{\frac{1}{2}} Q_2^\top$ for the right subspace, respectively.

As SVD requires each entry of W to be known, for most real computer vision problems it is not applicable. If some entries of W are unknown, we also have a visibility mask $V \in \{0,1\}^{m \times n}$ which encodes the information whether the entry w_{ij} is a valid observation ($v_{ij} = 1$) or not ($v_{ij} = 0$). Equation (2) then becomes

$$E_{\text{data}}(X, Y) := \frac{1}{2} \|(W - X \cdot Y) \odot V\|_{\text{fro}}^2 \tag{3}$$

where the operator $\odot$ denotes the element-wise product. In Section 3, we will add a smoothness term to this data term in order to obtain better results, but for now we will stick to this data term.

To minimize Eq. (3), we can use a gradient descent approach. In order to do this, we have to compute the gradient of E_{data}. The next lemma states that this task is easy in the sense that in only involves elementary matrix operations.

Lemma 1. *The gradient of E_{data} can be computed as*

$$\nabla E_{data} = \begin{pmatrix} \frac{\partial E_{data}}{\partial X} \\ \frac{\partial E_{data}}{\partial Y} \end{pmatrix} = \begin{pmatrix} \left[(XY - W) \odot V\right] Y^\top \\ X^\top \left[(XY - W) \odot V\right] \end{pmatrix}.$$

Proof. We will only show how to compute $\frac{\partial E_{\text{data}}}{\partial X}$. The computation of $\frac{\partial E_{\text{data}}}{\partial Y}$ can be done analogously. Now denote the columns V by v_j. Then, we can write

$$\begin{aligned} \frac{\partial E_{\text{data}}}{\partial y_j} =& \frac{1}{2} \frac{\partial}{\partial y_j} \left[\sum_{j=1}^{n} \left\| (w_j - X \cdot y_j) \odot v_j \right\|^2 \right] \\ =& \frac{1}{2} \frac{\partial}{\partial y_j} \left\| V_j (w_j - X \cdot y_j) \right\|^2 \end{aligned}$$

with the diagonal matrix V_j consisting of the entries of v_j.

$$\begin{aligned} &= \left(X^\top V_j X y_j - X^\top V_j w_j \right) \\ &= X^\top \left((X y_j - w_j) \odot v_j \right) \\ \Rightarrow \frac{\partial E_{\text{data}}}{\partial Y} &= X^\top \left((XY - W) \odot V \right) \end{aligned}$$

□

Since E_{data} is neither convex nor quasi-convex there is no obvious way of finding the global minimum efficiently. In [22, 11], Eq. (3) was minimized by iteratively solving for $\frac{\partial E_{\text{data}}}{\partial X} = 0$ and $\frac{\partial E_{\text{data}}}{\partial Y} = 0$ while keeping the other set of variables fixed. However, this method can get trapped in a local extremum: at every iteration, a potential local extremum at least for one of the two variable X or Y is chosen and thus the vulnerability that a local extremum for E_{data} is found increases dramatically. Of course, we like to believe that this local extremum is at least a local minimum. But this is not true in general. Since E_{data} is not a convex function, every iterative update step can even *increase* the energy that we want to minimize. If for example $\frac{\partial^2 E_{data}}{\partial^2 X}$ is negative definite or even indefinite, the update w.r.t to X will move Y into a local maximum or a saddle-point.

To overcome these problems, we perform a gradient descent approach which jointly optimizes X and Y. After each update of X and Y, X is projected to an orthonormal representation as classical power factorization does. This has several advantages:

1. The gradient descent approach will always decrease and thus we will omit any local maximum.

2. Gradient descent methods tend to not get stuck in saddle-points. This is because the area that will lead neighboring points via gradient descent into the saddle-point form themself a zero set in the definition domain.

3 Introducing L^2-regularization

Many real problems provide further constraints on the model. In this section it will be shown how Eq. (3) can be generalized to include a smoothness prior on the coordinates X. In the context of 3D-reconstruction we want to allow for non-smooth surfaces hence we do not enforce smoothness on the variables Y.

X can be understood as a path in $\mathbb{R}^r$ which corresponds to the temporal coherent observation in $\mathbb{R}^m$ encoded by the rows of the observation matrix W. Therefore, X can be understood as a discrete sub-sampling of the following trajectory:

$$\begin{aligned} c : [0,1] &\to \mathbb{R}^r \\ c\left(\frac{i-1}{m-1}\right) &= (x_{i,1} \cdots x_{i,r})^\top \qquad\qquad \forall i = 1, \ldots, m. \end{aligned}$$

With this formulation, we can now introduce the L^2-regularization on c via $E_{\text{smooth}}(c) = \frac{1}{2}\int_0^1 c'(t)^2 \,\mathrm{dt}$ which becomes for its discrete representation X the following backward difference:

$$E_{\text{smooth}}(X) = \frac{1}{2}\frac{1}{m-1}\sum_{s=1}^{r}\sum_{i=2}^{m}(x_{i,s} - x_{i-1,s})^2. \tag{4}$$

By weighting the importance of the smoothness term over the data term by a non-negative number λ, we can formulate the *multilinear model estimation with L^2-regularization* as minimizing the following energy function:

$$E(X,Y) = E_{\text{data}}(X,Y) + \lambda \cdot E_{\text{smooth}}(X) \tag{5}$$

As in Section 2 we want to minimize this energy via a gradient descent approach. In order to do this, we have to compute the gradient of E_{smooth}. It turns out that also this gradient can be computed by easy matrix operations:

$$\frac{\partial E_{\text{smooth}}}{\partial x_{i,s}} = \frac{1}{m-1}\left(-x_{i-1,s} + 2x_{i,s} - x_{i+1,s}\right) \tag{6}$$

Instead of matrix multiplication as in Lemma 1, we only need to compute a simple linear combination of neighboring rows in the matrix X. Combining Equation (6) with Lemma 1, we can find a minimum of E by projected gradient descent.

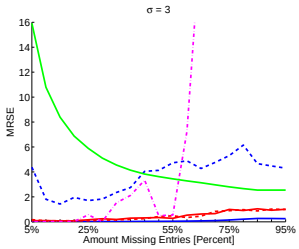

Fig. 1. Mean root square error (MRSE) of 10 trials between the estimated matrix and the ground truth. The solid blue line indicates the proposed method, the dashed blue line power factorization, the solid (dashed) red line Kalman-EM with (without) specified variance, the solid green line nuclear norm minimization (NNM), and the magenta dash-dotted line the regularized Gauss-Newton scheme.

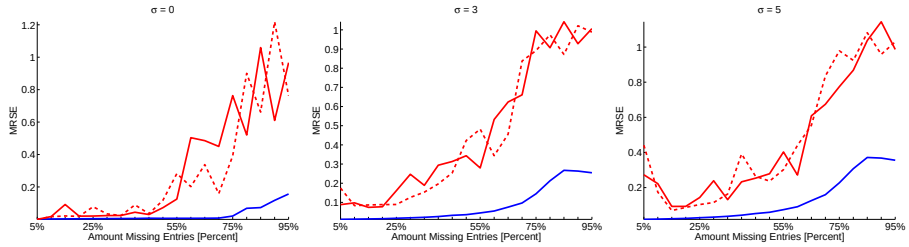

Fig. 2. Average mean root square error (MRSE) with ground truth data of 10 trials. The solid blue line indicates the proposed method, and the solid (dashed) red line Kalman-EM with (without) specified variance.

4 Evaluation for Synthetic Data

For experimental evaluation, we draw on an application from structure from motion: it was shown in [19] that feature trajectories of a rigid body over several images taken by *affine* cameras are constrained to span a low-dimensional linear subspace. Due to the incomplete trajectories, centroids cannot be computed, thus the standard rank-3 constraint used in [19] must include the unknown center, hence generalizes into a rank-4 constraint [1, 18] in Eq. (2). We simulated 200 3D-points distributed on a cylindrical surface. The points were translated, rotated, and projected onto 20 images. For projection, an affine camera model was used, thus avoiding non-Gaussian noise induced by estimating an incorrect model. We experimentally determined that the functional obtains a global minimum for $\lambda = 3 \cdot 10^8$.

The proposed method was further compared with the power factorization (NIPALS) algorithm, the Kalman-filtering EM approach, a method which minimizes the nuclear norm [5][3], and a Tikhonov-regularized Gauss-Newton scheme from [4]. Power factorization, the Gauss-Newton scheme, and the proposed gradient descent were randomly initialized 50 and the best result taken. The Kalman-

[3] The code is provided at *svt.caltech.edu*.

EM-algorithm was executed twice: once with specified variance (see below), once with a generic variance of 1. For the nuclear norm minimization (NNM) there are several parameters to specify. We set them to values which are very conservative according to the authors.

Occlusion was simulated by randomly removing parts at the beginning and the end of trajectories. We thus had trajectories only visible on a more or less narrow band on the diagonal of W increasing the difficulty [4]. We increased the amount of invisible data from 5% until 95% in steps of 5%. Visible measurements were perturbed with normally distributed noise with standard deviations $\sigma = \{0, 3, 5\}$. For each combination of noise and missing observations, we simulated 10 different realizations of W, *i.e.* perturbed and sampled its entries, and computed average errors and computation times.

Figure 1 shows the average Frobenius error per pixel between the the estimated matrix and the ground truth, *i.e.* a mean root sum of squares error (MRSE). The noise level was $\sigma = 3$. The solid blue line indicates the proposed approach, the dashed blue line, the solid red line Kalman-EM with known variance and the dashed red line Kalman-EM without known variance. The green solid line indicates NNM. Lastly, the magenta dash-dotted line indicates the Tikhonov-regularized Gauss-Newton scheme of [4]. The NNM approach usually converged to solutions of rank larger than 4. Since the physical model requires rank 4, we then truncated the estimated left and right subspaces which caused large errors. We varied its parameters yet could not find a more successful combination. The Gauss-Newton method performed poorly for large amounts of missing data. Other variants from this box achieved similar results. Both Kalman-approaches (KF) and the proposed solution both achieve low errors. Our approach performs superior to all other methods including power factorization.

The plots in Figure 2 compare both KFs and our method. The left plot corresponds to $\sigma = 0$, the middle to $\sigma = 3$, and the right to $\sigma = 5$. The blue error curves look similar for $\sigma = 3$ and $\sigma = 5$, yet differ slightly. For noise-free data, all three methods achieve similar errors if less than 40% of the matrix is known. For larger sampling ratios, the proposed algorithm performs more than twice as good. For noisy data, the proposed method is between 2.5 and more than 14 times more accurate.

5 Real World Applications

In this section we show successful application of the proposed method to two real image sequences containing large noise and even a few outliers. While regularized energies have already been applied to 3D reconstruction [7, 21], the problem that we address here is different from prior work. In [7, 21], camera calibration including intrinsic and extrinsic parameters is known, while the current work considers unknown calibration information. Furthermore, regularization is not applied to the 3D-points. Instead, we regularize the camera path.

[4] Due to the random occlusion, trajectories have to be permuted properly to make the band-diagonal structure of W visible.

The scene shown in Fig. 3(a) shows a corner of a historic building. A total of 2000 trajectories were observed over 60 images with 68.6% missing features. While there are no obvious outliers, noise is very large. Four images of the 3D-reconstruction are shown in Fig. 3(b). The color of the pixel in the image it was first observed in was assigned to each 3D-point. The overall reconstruction looks reasonable, only the depth of the scene is underestimated. This error is due to the affine camera model which cannot handle significant scene depth compared with the the distance to the camera.

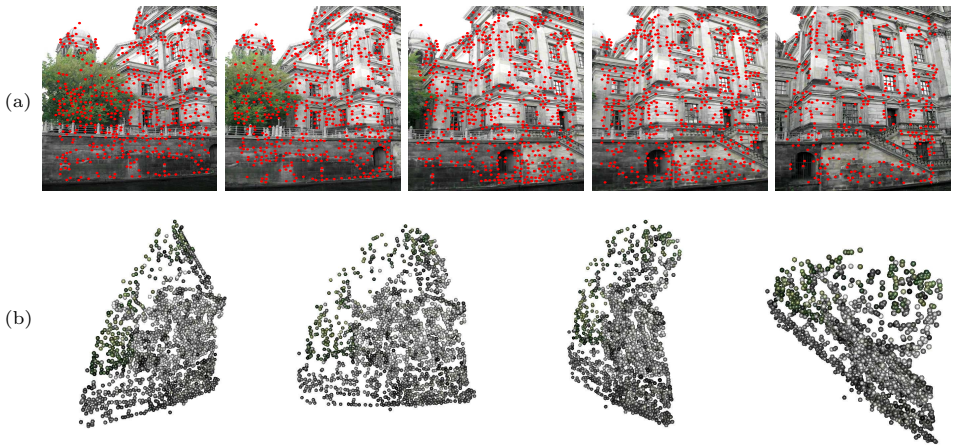

Fig. 3. (a) Five images of a 60 image sequence with 2000 trajectories. 68.6% of the data is missing and there is large noise. Red points indicate the feature found in each image. (b) Four views of the reasonable 3D-reconstruction.

The second sequence consists of 672 trajectories over 10 images[5]. A single image is shown in Fig. 4, left. A total of 57.7% of the data matrix is unknown. Since there are several outliers present in the data, we adopted a RANSAC approach on minimal subsets.

Four images of this 3D-reconstruction are shown in the left images of Fig. 4. The ground plane is not rectangular with the wall of the house, and the right side is heavily distorted. Considering the affine camera model, the reconstruction is reasonable.

The achieved results look reasonable considering the affine camera model and the fact that the shown squences have significant depth variation whereas the assumption is that all 3D-points have similar depths. Approaches for projective or Euclidean bundle adjustment can achieve better results yet require good initializations which can be provided by the proposed algorithm. Furthermore, such software packages are much more complex than the proposed algorithm.

It is known that the L^2 error metric defined by Eq. (3) is not entirely suitable for 3D-reconstruction. Nonetheless, the L^2 metric is quite general and can

[5] This sequence is provided at *http://www.robots.ox.ac.uk/∼vgg*

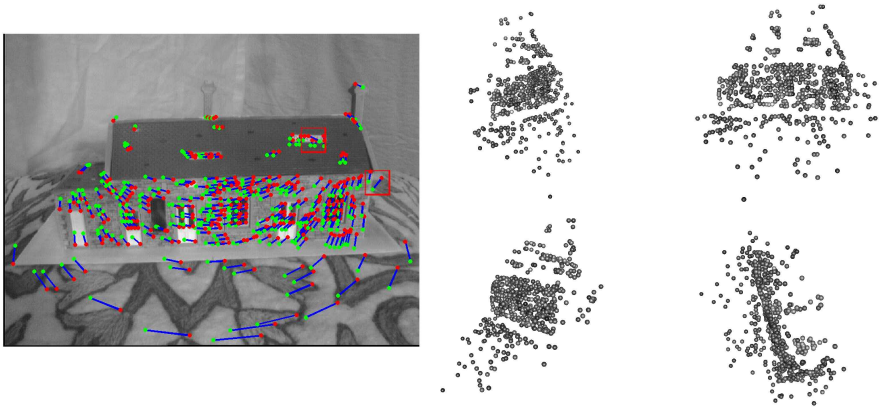

Fig. 4. Left: One image of a 10 image sequence with 672 trajectories and 57.7% unknown features. Red points indicate features in the current image, green points correspondences in the next image. Red boxes indicate apparent outliers. Right: Four views of the 3D-reconstruction. Overall, it looks reasonable. The angle between ground plane and house is not orthogonal due to the strong perspective distortion of the points close the to camera which the affine camera model cannot handle.

be directly applied to many other problems [13, 20, 12]. For SfM, we therefore like to interpret the used error as an approximation of the prefered metric. Future work will focus on studying more descriptive errors which better suit 3D-reconstruction. For the general problem of multilinear model, we would still advocate the Frobenius error because it is a very general error which is consistent with the proposed L^2-regularizer.

6 Conclusion

In this work, a factorization algorithm for partially known matrices was presented. It uses a globally invariant energy function which was generalized to include a smoothness prior. This prior penalizes non-smooth models only if the data term is locally insufficient. Using the generalized energy functional, a gradient descent method was derived. Using simulated data, we showed that this algorithm is more accurate than all other methods even if significant parts of the matrix are unknown. Using real image data, reasonable 3D-reconstructions were presented. The proposed solution can be used to initialize a bundle adjustment. Although structure and motion estimation was presented as application the proposed algorithm is general and can be applied to any PCA problem [13, 20, 12, 19, 2].

References

1. Ackermann, H., Rosenhahn, B.: Trajectory Reconstruction for Affine Structure-from-Motion by Global and Local Constraints. In: CVPR (June 2009)

2. Brand, M.: Morphable 3d models from video. In: IEEE Computer Vision and Pattern Recognition (CVPR). pp. 456–463. Kauai, Hawaii, USA (2001)
3. Brand, M.: Incremental Singular Value Decomposition of Uncertain Data with Missing Values. In: ECCV. pp. 707–720. Copenhagen, Denmark (June 2002)
4. Buchanan, A., Fitzgibbon, A.: Damped Newton Algorithms for Matrix Factorization with Missing Data. In: CVPR. pp. 316–322. Washington, DC, USA (2005)
5. Cai, J.F., Candès, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM Journal on Optimization 20(4), 1956–1982 (2010)
6. Candès, E.J., Recht, B.: Exact matrix completion via convex optimization. Foundations of Computational Mathematics 9(6), 717–772 (2009)
7. Cremers, D., Kolev, K.: Multiview stereo and silhouette consistency via convex functionals over convex domains. IEEE TPAMI (2010)
8. Eriksson, A., van den Hengel, A.: Efficient computation of robust low-rank matrix approximations in the presence of missing data using the l1 norm. In: CVPR. San Francisco, USA (2010)
9. Gabriel, K., Zamir, S.: Lower rank approximation of matrices by least squares with any choice of weights. Techonometrics 21(4), 489–498 (1979)
10. Gruber, A., Weiss, Y.: Factorization with uncertainty and missing data: Exploiting temporal coherence. In: NIPS. Vancouver, Canada (December 2003)
11. Hartley, R., Schaffalizky, F.: PowerFactorization: 3D Reconstruction with Missing or Uncertain Data. In: Australia-Japan Advanced Workshop on Computer Vision (June 2002)
12. Hasler, N., Ackermann, H., Rosenhahn, B., Thormählen, T., Seidel, H.P.: Multilinear pose and body shape estimation of dressed subjects from image sets. In: CVPR. San Francisco, USA (June 2010)
13. M. Turk, A.P.: Face recognition using eigenfaces. In: CVPR (1991)
14. Marques, M., Costeira, J.: Estimating 3d shape from degenerate sequences with missing data. CVIU 113(2), 261–272 (2009)
15. Peng, Y., Ganesh, A., Wright, J., Ma, Y.: Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. In: CVPR. pp. 763–770. San Francisco, USA (June 2010)
16. Ruhe, A.: Numerical computation of principal components when several observations are missing. Tech. rep., Dept. Information Processing, University of Umeda, Umeda, Sweden (April 1974)
17. Ruhe, A., Wedin, P.: Algorithms for separable nonlinear least squares problems. Society for Industrial and Applied Mathematics Review 22(3), 318–337 (1980)
18. Sugaya, Y., Kanatani, K.: Extending Interrupted Feature Point Tracking for 3-D Affine Reconstruction. In: ECCV. pp. 310–321. Prague, Czech Republic (May 2004)
19. Tomasi, C., Kanade, T.: Shape and motion from image streams under orthography: A factorization method. IJCV 9(2), 137–154 (November 1992)
20. Ullman, S., Basri, R.: Recognition by linear combinations of models. IEEE TPAMI 13, 992–1006 (1991)
21. Vu, H.H., Keriven, R., Labatut, P., Pons, J.P.: Towards high-resolution large-scale multi-view stereo. In: CVPR. Miami (Jun 2009)
22. Wold, H.: Estimation of principal components and related models by iterative least squares. In: Krishnaiah (ed.) Multivariate Analysis. pp. 391–420 (1966)