Uncalibrated Non-Rigid Factorisation by Independent Subspace Analysis

Sami S. Brandt
IT University of Copenhagen

Hanno Ackermann and Stella Graßhof
Leibniz Universität Hannover

Abstract

We propose a general, prior-free approach for the uncalibrated non-rigid structure-from-motion problem for modelling and analysis of non-rigid objects such as human faces. We recover the non-rigid affine structure and motion from 2D point correspondences by assuming that (1) the non-rigid shapes are generated by a linear combination of rigid 3D basis shapes, (2) that the non-rigid shapes are affine in nature, i.e., they can be modelled as deviations from the mean, rigid shape, (3) and that the basis shapes are statistically independent. In contrast to the majority of existing works, no statistical prior is assumed for the structure and motion apart from the assumption that underlying basis shapes are statistically independent. The independent 3D shape bases are recovered by independent subspace analysis (ISA). Likewise, in contrast to the most previous approaches, no calibration information is assumed for affine cameras; the reconstruction is solved up to a global affine ambiguity that makes our approach simple and efficient. In the experiments, we evaluated the method with several standard data sets including a real face expression data set of 7200 faces with 2D point correspondences and unknown 3D structure and motion for which we obtained promising results.

1. INTRODUCTION

The estimation of structure and motion from image streams is a fundamental problem in computer vision. As an extension to the regular structure-from-motion (SFM) problem, the non-rigid structure-from-motion (NRSFM) problem takes the non-rigidity of the object in consideration in the recovery of structure and motion. The NRSFM problem has received considerable attention during the last two decades and encouraging results have been obtained.

The approaches for NRSFM can be categorised in several ways. From the algorithmic point of view, there are *direct* and *iterative* methods. Starting from the direct methods, the work of Bregler *et al.* [8] can be seen as the starting point for NRSFM research. They proposed an approach where the shape deformations are modelled as a linear com-
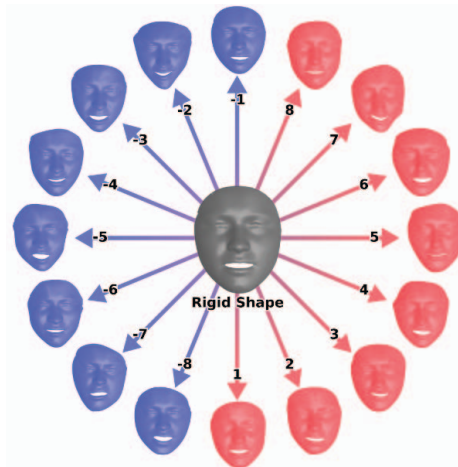

Figure 1. We propose a method that infers the 3D reconstruction, basis shapes, and the underlying affine camera geometry from the 2D projections of a non-rigid object by only assuming an uncalibrated affine camera and *statistically independent* basis shapes.

bination of rigid shape basis that leads to a low-rank model; a heuristic 1D factorisation together with orthogonal constraints were used to recover the camera matrices. This pioneering work was thereafter succeeded by the work of Brand *et al.* [6], who used the heuristic of minimising deformations. Del Bue and Agapito applied additional constraints arising from a stereo rig [11]. Xiao *et al.* constrained the shape basis by assuming that each basis shape is visible unmixed in some frames [26]. Hartley and Vidal proposed a solution for perspective non-rigid structure from motion problem by factoring a multifocal tensor [16].

Regarding the iterative methods, one category is alternation-based methods, such as the trilinear method by Torresani *et al.* [25], the bilinear methods by Paladini *et al.* [22] and Del Bue *et al.* [13] which include projections onto the metric manifold, and the method by Torresani *et al.* [24] which is based on Probabilistic PCA and Expectation Maximisation. Bundle adjustment has been applied, for instance, in [1, 12, 5]. Moreover, Bartoli *et al.* [5] used a coarse-to-fine problem formulation to obtain a robust result. Various works have applied either statistical or physical priors to regularise the non-rigid structure from motion problem. These include priors such as rigidity [12, 5],

smoothness of camera trajectories [15], temporal smoothness [24, 3], deformation locality [6] and type [13].

When analysing, for instance, a set of face images without temporal order, temporal priors cannot be applied. An early prior-free[1] approach for uncalibrated non-rigid structure from motion was proposed by Brandt *et al.* [7] where the shape basis ambiguity was approached by assuming statistical independence between the basis shapes that led to independent subspace analysis (ISA). Dai *et al.*'s solution [10] for prior-free non-rigid structure-from-motion was built upon the observation by Akhter *et al.* [2], namely that even though there is an unresolved ambiguity for shape basis with the standard orthogonality constraints, the 3D shape can be recovered without an ambiguity. Kong and Lucey [21] proposed a prior-free approach where the non-rigid shape is modelled as a compressible basis instead of a low-rank basis. In applications like facial expression analysis, it is also valuable to reconstruct the underlying shape basis and decompose the expressions onto it. This is a drawback for the approach [10] where the shape basis is only implicit and ambiguous. Likewise, Kong and Lucey [21] did not estimate a shape basis but a compressible feature basis.

We propose another prior-free, non-rigid structure-from-motion algorithm based on independent subspace analysis. The assumption hence is that the underlying shape bases live in *statistically independent*[2] subspaces (see also [4]). Statistical independence should not be confused with linear independence used, for instance, by Xiao *et al.* [26] for selecting the shape basis. Remarkably, the statistically independent subspaces, and hence the basis shapes, can be recovered in an uncalibrated, affine setting, thus no calibration information, neither intrinsic nor extrinsic, is required to infer the basis shapes (cf. Fig. 1). This is a *major simplification* of the non-rigid structure-from-motion problem. The basis shapes can be converted to a metric as soon as the cameras are calibrated without affecting the independent subspace structure. Furthermore, in contrast to the method by Brandt *et al.* [7], the our method does not require an exhaustive search over one-dimensional subspace permutations which constitutes a combinatorial problem, nor iterative optimisation for the inhomogeneous projection matrices. In effect, our method is simpler, sounder and scalable.

The contributions of this work are as follows. (1) We propose a straightforward, priorless, direct method for non-rigid structure-from-motion by assuming statistical independence of the basis shapes in an uncalibrated setting. (2) In contrast to many other non-rigid factorisation algorithms built upon the seminal algorithm of Bregler *et al.* [8], we assume that the affine camera matrix, which is shared by all

the basis shapes, will be solely defined by the mean, rigid shape – this is consistent with Independent Subspace Analysis since it has been shown in [17] that this is equivalent to analyse the original or mean corrected observations while the structure of the latter setting is simpler. (3) To recover the shape basis we suggest two alternative ISA algorithms built upon mutual information minimisation: FastISA proposed in [18] and FastICA [19] equipped with our component pooling. The algorithms do not require to exhaustively determine permutations of one-dimensional shape components in contrast to the method in [7]. (4) To recover the block-formed motion matrix after the ISA step, we propose an algebraic, iteratively re-weighted least squares method where only subspace affinities and the shape mixing coefficients are left to be estimated. (5) We propose a non-linear refinement to obtain the final, statistically sound estimates.

2. AFFINE NON-RIGID MODEL

The standard non-rigid factorisation assumes that the non-rigid shape can be represented as a linear combination of the shape bases. That is, 3D points can be expressed as $\mathbf{x}_j^i = \sum_k \alpha_k^i \mathbf{b}_{kj}$, where α_k^i is a scalar. With a *stationary* motion sequence, the model can, without a loss of generality, be assumed affine, i.e., centred around the rigid, mean shape. The 2D projection $\hat{\mathbf{m}}_j^i$ of a 3D point $\mathbf{x}_j^i$ hence is

$$\hat{\mathbf{m}}_j^i = \mathbf{M}^i \mathbf{x}_j^i + \mathbf{t}^i = \mathbf{M}^i \left(\mathbf{b}_{0j} + \sum_{k=1}^K \alpha_k^i \mathbf{b}_{kj} \right) + \mathbf{t}^i, \quad (1)$$

where $\mathbf{M}^i$ is 2×3 projection matrix to the image i, $\mathbf{t}^i$ is the corresponding translation vector, α_k^i, $k = 1, 2, \ldots, K$ are the scalar coefficients, and $\mathbf{b}_{kj}$ refers to the basis shape k; $k = 0$ is the mean rigid shape. The points are indexed by j.

Assuming Gaussian noise, the maximum likelihood solution with respect to the parameters $\mathbf{M}^i, \mathbf{t}^i, \alpha_k^i, \mathbf{b}_{kj}$, $i = 1, \ldots, I, j = 1, \ldots, J, k = 1, \ldots, K$, minimises the cost

$$\sum_{i,j} \|\hat{\mathbf{m}}_j^i - \mathbf{m}_j^i\|^2 = \sum_{i,j} \|\mathbf{M}^i (\mathbf{b}_{0j} + \sum_k \alpha_k^i \mathbf{b}_{kj}) + \mathbf{t}^i - \mathbf{m}_j^i\|^2$$

or equivalently

$$\|\mathbf{W} - \hat{\mathbf{W}}\|_{\text{Fro}}^2, \quad (2)$$

where the translation corrected measurements $\mathbf{m}_j^i - \hat{\mathbf{t}}^i, \hat{\mathbf{t}}^i = \frac{1}{J} \sum_j \mathbf{m}_j^i$, are collected into the matrix $\mathbf{W}$, implying

$$\mathbf{W} \simeq \underbrace{\begin{pmatrix} \mathbf{M}^1 & \alpha_1^1 \mathbf{M}^1 & \cdots & \alpha_K^1 \mathbf{M}^1 \\ \mathbf{M}^2 & \alpha_1^2 \mathbf{M}^2 & \cdots & \alpha_K^2 \mathbf{M}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{M}^I & \alpha_1^I \mathbf{M}^I & \cdots & \alpha_K^I \mathbf{M}^I \end{pmatrix}}_{\triangleq \mathbf{M}} \underbrace{\begin{pmatrix} \mathbf{B}_0 \\ \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_K \end{pmatrix}}_{\triangleq \mathbf{B}}, \quad (3)$$

---

[1]By 'prior-free' we refer to an approach that does not make an assumption, in the Bayesian sense, about the prior distribution of the basis shapes.

[2]Two random variables $\mathbf{X}, \mathbf{Y}$ are statistically independent iff their joint probability density can be factorised such that $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y})$.

where $\mathbf{B}_k = (\mathbf{b}_{k1} \ \mathbf{b}_{k2} \ \cdots \ \mathbf{b}_{kJ})$ and $\mathbf{B}_0$ is the rigid shape. All the shape bases share the same inhomogeneous projection matrix $\mathbf{M}^i$ for image i. From (3) it follows that the noise free measurement matrix has the rank constraint $R \triangleq \text{rank } \hat{\mathbf{W}} \leq 3K + 3$.

The matrix minimising (2) with the rank constraint is obtained by the singular value decomposition of $\mathbf{W} = \mathbf{USV}^{\mathrm{T}}$ by truncating the smallest singular values, keeping the $3K+3$ largest, and removing the corresponding singular vectors. The truncated matrices being $\tilde{\mathbf{U}}$, $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{V}}$ yields

$$\hat{\mathbf{W}} = \underbrace{\left(\frac{1}{\sqrt{J}}\tilde{\mathbf{U}}\tilde{\mathbf{S}}\right)}_{\triangleq \tilde{\mathbf{M}}}\underbrace{\left(\sqrt{J}\tilde{\mathbf{V}}^{\mathrm{T}}\right)}_{\triangleq \tilde{\mathbf{B}}} = \underbrace{\tilde{\mathbf{M}}\mathbf{A}}_{\triangleq \hat{\mathbf{M}}}\underbrace{\mathbf{A}^{-1}\tilde{\mathbf{B}}}_{\triangleq \hat{\mathbf{B}}} = \hat{\mathbf{M}}\hat{\mathbf{B}}, \quad (4)$$

where $\mathbf{A}$ is an unknown affine transformation. To find the estimates for the non-rigid structure $\hat{\mathbf{B}}$ and motion matrix $\hat{\mathbf{M}}$, we need to find the linear transformation $\mathbf{A}$ that (1) separates the statistically independent shape subspaces and (2) recovers the block structure of the motion matrix. Our solution is described in the following section.

3. PROPOSED METHOD

This section describes the proposed method. It consists of the following steps: estimation of the rigid and non-rigid components (Sec. 3.1), independent subspace analysis (Sec. 3.2), block-form motion matrix recovery (Sec. 3.3), and non-linear refinement (Sec. 3.4).

3.1. Factorisation

To facilitate ISA processing and for clarity, we divide the translation corrected measurement matrix into rigid and non-rigid part as follows. We first compute the nearest rigid affine reconstruction by the standard Tomasi–Kanade factorisation [23] that yields the rigid approximation

$$\mathbf{W}_0 = \mathbf{M}_0\mathbf{B}_0, \quad (5)$$

where the inhomogeneous projection matrices, up to an affine transform, are $\mathbf{M}_0 = \frac{1}{\sqrt{J}}\mathbf{U}_0\mathbf{S}_0$ and the mean rigid shape is $\mathbf{B}_0 = \sqrt{J}\mathbf{V}_0^{\mathrm{T}}$. We then subtract the rigid component from the measurement matrix

$$\Delta\mathbf{W} = \mathbf{W} - \mathbf{W}_0, \quad (6)$$

and work with the non-rigid part $\Delta\mathbf{W}$.

Now, by using the remaining constraint $\text{rank } \Delta\mathbf{W} \leq 3K$ for the non-rigid part, we truncate all the singular values, and singular vectors, up to the $3K$ largest that yields

$$\Delta\mathbf{W} \approx \Delta\tilde{\mathbf{W}} \equiv \mathbf{U}'\mathbf{S}'\mathbf{V}'^{\mathrm{T}} = \mathbf{M}'\mathbf{B}', \quad (7)$$

where $\mathbf{M}' = \frac{1}{\sqrt{J}}\mathbf{U}'\mathbf{S}'$ and $\mathbf{B}' = \sqrt{J}\mathbf{V}'^{\mathrm{T}}$.

3.2. Independent Subspace Analysis

By independent subspace analysis (ISA), we map the singular vectors into new vectors that form subsets that will be as statistically independent to each other as possible. Without a loss of generality, our ISA estimate (ISA1) is formed by two steps: (1) estimating the one-dimensional ICA components, and (2) grouping them. By definition, the ICA components minimise the mutual information

$$I(\mathbf{Z}) = \sum_j H(Z_j) - H(\mathbf{Z}) \quad (8)$$

where H refers to differential entropy and $\mathbf{Y} = \mathbf{A}_{\mathrm{ICA}}\mathbf{Z}$ is the random vector of the mixed signals, corresponding to the columns of $\mathbf{B}'$; $\mathbf{Z}$ is the random vector of the sources, and $\mathbf{A}_{\mathrm{ICA}}$ is the mixing matrix. By construction, mixing matrix $\mathbf{A}_{\mathrm{ICA}}$ will be orthogonal, hence,

$$\mathbf{M}'\mathbf{B}' = \underbrace{\mathbf{M}'\mathbf{A}_{\mathrm{ICA}}}_{\triangleq \mathbf{M}_{\mathrm{ICA}}}\underbrace{\mathbf{A}_{\mathrm{ICA}}^{\mathrm{T}}\mathbf{B}'}_{\triangleq \mathbf{B}_{\mathrm{ICA}}} \equiv \mathbf{M}_{\mathrm{ICA}}\mathbf{B}_{\mathrm{ICA}}. \quad (9)$$

where the rows in $\mathbf{B}_{\mathrm{ICA}}$ will be in as statistically independent as possible. Here, we compute the mixing matrix by the FastICA algorithm [19].

The remaining problem is the pooling of the one-dimensional components into groups of three. We use the image population statistics, so that we project the non-rigid measurement matrix $\Delta\mathbf{W}$ onto the orthogonal, $3K$-dimensional ICA basis $\mathbf{B}_{\mathrm{ICA}}$ and estimate the $3K \times 3K$ covariance matrix $\mathbf{C} = \frac{1}{J}\mathbf{B}_{\mathrm{ICA}}\Delta\mathbf{W}^{\mathrm{T}}\Delta\mathbf{W}\mathbf{B}_{\mathrm{ICA}}^{\mathrm{T}} - \frac{1}{4I^2 J}\mathbf{B}_{\mathrm{ICA}}\Delta\mathbf{W}^{\mathrm{T}}\mathbf{11}^{\mathrm{T}}\Delta\mathbf{W}\mathbf{B}_{\mathrm{ICA}}^{\mathrm{T}}$ of these projections. For a statistically independent component pair, the covariance will vanish, *i.e.*, the covariance matrix will show block diagonal structure, as soon as the components are correctly permuted. We thus estimate the ICA component permutation matrix $\mathbf{P}$, and further the orthogonal transformation $\mathbf{A}_{\mathrm{ISA}}^{\mathrm{T}} = \mathbf{P}\mathbf{A}_{\mathrm{ICA}}^{\mathrm{T}}$, by a greedy strategy: in analogy to using Givens rotations, we compute the sequence of optimal pairwise variable permutations that decrease the off-block-diagonal covariation in the covariance matrix.

We additionally experiment another ISA algorithm (ISA2) that yields an estimate for the orthonormal, subspace separation matrix $\mathbf{A}_{\mathrm{ISA}}^{\mathrm{T}}$. It is principally the FastISA algorithm [18] that has been developed from the FastICA [19] algorithm with the difference that the statistical independence of individual components is not assumed but instead between vectors residing in different subspaces. This is in analogy to assuming sparsity or group sparsity of multivariate signals. However, the approach has been reported local hence relatively sensitive to intialisation. Moreover, the construction of FastISA is based on an *ad hoc* probability density model that may compromise its statistical performance. To cope with the locality, we compute FastISA from multiple initialisations and take the estimate that maximises the likelihood of the solution with the density assumption.

3.3. Recovery of the Block Structure

By a blind subspace separation method, the independent subspaces can be recovered only up to an unknown linear transform for each independent subspace, since the energy of the independent components cannot be recovered [17]. In other words, after ISA, we need to estimate the 3×3 mapping $\mathbf{D}_k$ from the rigid shape coordinate system onto coordinate system of the independent subspace k.[3] Let $\mathbf{D}$ be the block diagonal matrix containing all the K subspace affinities in the respective blocks. We may then write

$$\mathbf{M}_{\text{ISA}}\mathbf{B}_{\text{ISA}} = \underbrace{\mathbf{M}_{\text{ISA}}\mathbf{D}}_{\triangleq\hat{\mathbf{M}}}\underbrace{\mathbf{D}^{-1}\mathbf{B}_{\text{ISA}}}_{\triangleq\hat{\mathbf{B}}} \equiv \hat{\mathbf{M}}\hat{\mathbf{B}} \qquad (10)$$

that also maps the motion matrix $\mathbf{M}_{\text{ISA}}$ into the block-form matrix. To compute an algebraic estimate for $\mathbf{D}$, we use the assumption that each shape basis component, including the rigid shape, share a common affine projection matrix to each view, and minimise

$$\min_{\mathbf{D},\alpha} \sum_{i,k} \|\mathbf{M}_k^i \mathbf{D}_k - \alpha_k^i \mathbf{M}_0^i\|_{\text{Fro}}^2 \qquad (11)$$

subject to $\|\mathbf{D}_k\|_{\text{Fro}}^2 = 1$ for $k = 1, 2, \ldots, K$, where $\mathbf{M}_k^i$ is the 2×3 block of $\mathbf{M}_{\text{ISA}}$, indexed by k and i. The matrix $\mathbf{M}_0^i$ is the inhomogeneous affine projection matrix i in $\mathbf{M}_0$ in (5). The estimate can be found by iteratively reweighted least squares as detailed in Appendix A.

3.4. Non-linear Refinement

Since (11) is an algebraic criterion, we finally make a non-linear refinement to minimise the reprojection error, or

$$\min_{\mathbf{D},\alpha} \|\Delta\mathbf{W} - \mathbf{M}_0^\alpha \mathbf{D}^{-1}\mathbf{B}_{\text{ISA}}\|_{\text{Fro}}^2 \qquad (12)$$

where $\mathbf{M}_0^\alpha$ is defined by the matrix $\mathbf{M}_0$ repeated K times, and having the scalar weights α_k^i on the corresponding 2×3 blocks, c.f. (3). Using the fact that the rows of $\frac{1}{\sqrt{J}}\mathbf{B}_{\text{ISA}}$ are orthonormal,

$$\|\Delta\mathbf{W} - \mathbf{M}_0^\alpha \mathbf{D}^{-1}\mathbf{B}_{\text{ISA}}\|_{\text{Fro}}^2$$

$$= \left\|\frac{1}{J}\Delta\mathbf{W}\mathbf{B}_{\text{ISA}}^{\text{T}}\mathbf{B}_{\text{ISA}} - \mathbf{M}_0^\alpha \mathbf{D}^{-1}\mathbf{B}_{\text{ISA}}\right\|_{\text{Fro}}^2 +$$

$$+ \left\|\Delta\mathbf{W}\left(\mathbf{I} - \frac{1}{J}\mathbf{B}_{\text{ISA}}^{\text{T}}\mathbf{B}_{\text{ISA}}\right)\right\|_{\text{Fro}}^2$$

$$= \left\|\frac{1}{\sqrt{J}}\Delta\mathbf{W}\mathbf{B}_{\text{ISA}}^{\text{T}} - \mathbf{M}_0^\alpha \mathbf{D}^{-1}\right\|_{\text{Fro}}^2 +$$

$$+ \left\|\Delta\mathbf{W}\left(\mathbf{I} - \frac{1}{J}\mathbf{B}_{\text{ISA}}^{\text{T}}\mathbf{B}_{\text{ISA}}\right)\right\|_{\text{Fro}}^2, \qquad (13)$$

---

[3]In the calibrated case, $\mathbf{D}_k$ would be the 3×3 rotation $\mathbf{R}_k$ between the rigid and the non-rigid shape basis. Here, however, $\mathbf{D}_k$ is a general 3×3 matrix, constrained to unity norm to fix the arbitrary scale of the solution.

where the latter term does not depend on $\mathbf{D}_k$ and α_k^i and can be dropped. That yields an equivalent bilinear problem

$$\min_{\mathbf{D},\alpha} \left\|\frac{1}{\sqrt{J}}\Delta\mathbf{W}\mathbf{B}_{\text{ISA}}^{\text{T}} - \mathbf{M}_0^\alpha \mathbf{D}^{-1}\right\|_{\text{Fro}}^2 \qquad (14)$$

that we minimise by standard non-linear least squares.

4. EXPERIMENTS

4.1. Torressani's Shark Dataset

For the first experiment, we use Torressani's synthetic Shark data set [24]. It is a degenerate data set with $K = 1$. Moreover, the original measurement matrix ($I = 240$, $J = 91$) has rank 5 after the translation correction. Hence, the deformation basis is degenerate. This implies a non-unique reconstruction as there will be a 3-parameter-family of solutions for even a single 3D shape basis. We compared the proposed method against the pseudoinverse method proposed by Dai *et al.* [10], as well as their Block Matrix Method, and Kong and Lucey's priorless compressible method [21]. Since our method is affine and the reconstruction will be known only up to an unknown affine transformation, it will not be meaningful to compare the results in the 3D space. Instead, we compare the reprojections onto the image plane between the methods. The results are shown in 3, and in Tab. 1.
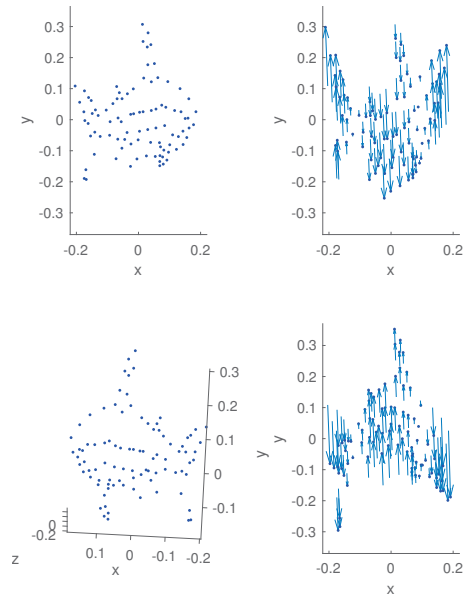

Figure 2. Affine reconstruction on Torressani's Shark dataset for which the projections onto the image plane can be modelled by a single degenerate (planar) 3D shape basis. (Left column) rigid affine 3D Shape from two different directions; (right column) illustration of the estimated non-rigid ISA component on both sides around the rigid shape.

Table 1. Relative reprojection error, reported as the inverse SNR, for the tested NRSFM approaches on different data sets.

Inverse SNR [%]	PI (Dai *et al.* [10])	BMM (Dai *et al.* [10])	Kong&Lucey [21]	Proposed ISA
Shark	3.5	0.33	160	**0.12**[†]
Balloon	0.11	**0.012**	1.2	0.12[†]
Face LS3D-W	0.025	0.024	0.93	**0.014**[†]
Face Binghamton	*	*	9.3	**0.015**[‡]

[†] by ISA1 variant; [‡] by ISA2 variant; * no result within 24 hours.

It can be seen from Tab. 1 that the proposed method (ISA2) achieves the lowest reprojection error, measured by the inverse signal-to-noise-ratio. The other prior free methods do not perform well for this dataset due to the degeneracy, especially, the compressible method failed completely, as also reported in [21]. In spite of the degeneracy of the data set, our method was able to pinpoint the major mode of deformation that is a vector field normal to a reference plane, as Fig. 2 illustrates. Here, since only one deformation subspace was considered, the pooling step was trivial.

4.2. Balloon Deflation

For the second test data set, we use the balloon deflation from the NRSfM Challenge 2017 [20]. It is a simulated data set with $I = 51$ projections generated by reprojecting real tracked 3D data points ($J = 211$) by a virtual, perspective camera having a circular camera trajectory. By using an affine camera model, we can thus only achieve an approximation of the ground truth camera geometry. We then estimated the result using the reference approaches and our ISA methods. We assumed five deformation modes ($K = 5$). The results are in Tab. 1, and Fig. 3 and 6.

For this data set, the Block Matrix Method of [10] gave the best result, whereas the Pseudoinverse and the proposed ISA approach obtain similar scores. Fig. 6 illustrates the estimated, statistically independent modes. The third component most clearly indicates the size change, whereas the other modes represent different kinds of non-linear shape deformations. The mode covariance matrix (Fig. 6b) shows that the highest correlations are concentrated onto the diagonal, hence, the independence assumption is fair. Nonetheless, there are some off-block-diagonal-correlations that most likely contributed to the higher score.

4.3. Face LS3D-W Dataset

For the third experiment, we use the LS3D-W data set [9] consisting of matched feature points for 7200 human faces with various expressions. Each face contains 68 2D feature points that were automatically found and matched, as described in [9]. The faces were in random orientation and order so no temporal smoothness could be applied. We compare our method (ISA2) against Dai's [10] and Kong and Lucey's [21] methods. Dai's method is computationally most demanding due to the size of the database: the computation of the result took about 2 CPU days, Kong and Lucey's about 6 CPU hours. In contrast, an ISA estimate

could be computed in about twenty CPU minutes. The results are shown in Tab. 1, and in Figs. 4 and 7.

From the results, it can be seen that the proposed method gave the best numerical results with almost a half of the inverse SNR when compared to either of the methods by Dai *et al.* [10]. When looking at the reprojections, it can be seen that their approach had more difficulties in reproducing the fine structure of the mouth (see columns 2, 4, 6, 8, and 9 in Fig. 4) than the proposed method. Each estimated basis shape, shown in Fig. 7a, demonstrate a clear semantic interpretation. From the mode correlation matrix (Fig. 7b) it can be seen that the strongest off-block-diagonal covariance is between the the first and fourth basis shape. One can also note that the lips are slightly distorted in both modes that suggest that there is in fact a statistical dependency between the modes while the statistical independence assumption yields an accurate approximation for the shapes and poses in the data set.

4.4. Binghamton 3D Facial Expression Dataset

The data set [27] contains 25 shapes of 100 subjects with 7 different expressions (*neutral*, *happy*, *disgusted*, *fear*, *angry*, *surprised*, *sad*) recorded by a 3D-face scanner. All the expressions, except the neutral, were recorded in four different strengths. The subjects had varying ethnic background and their age range was from 18 to 70 years. A total of 56% of the subjects were female and 44% male. We obtained 7308 3D-correspondences between the shapes by non-rigid registration [14]. 2D correspondences, simulated from the 3D correspondences, were used as the input for the experiment. Results are shown in Tab. 1, and Fig. 8. A comparison with the baseline algorithms by Dai *et al.* [10] was not possible since both methods did not converge within a reasonable amount of time. The result by Kong and Lucey [21] was modest probably due to the fact there was no temporal structure in the data. Our method (ISA1) was able to produce an accurate fit, as the low inverse SNR demonstrates. Also, as can be seen in Fig. 8, the estimated shape basis was able to capture the major structure variations in the database, including those related to person expression changes. From the mode covariance marix (Fig. 8n), it can be seen that the covariance was concentrated onto the diagonal while the statistical dependences are not as strong as with the LS3D-W data set.
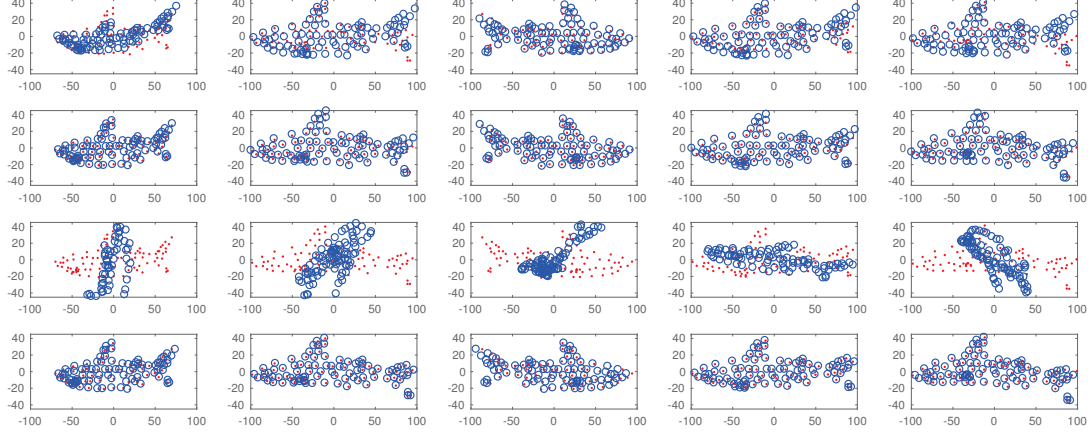
Figure 3. Reprojections (blue) to random frames against ground truth projections (red) with the Shark dataset; $K = 1$. (1st row) Pseudoinverse Method Dai *et al.* [10]; (2nd row) Block Matrix Method Dai *et al.* [10]; (3rd row) Kong&Lucey's Priorless Compressible Method [21]; (4th row) proposed ISA.
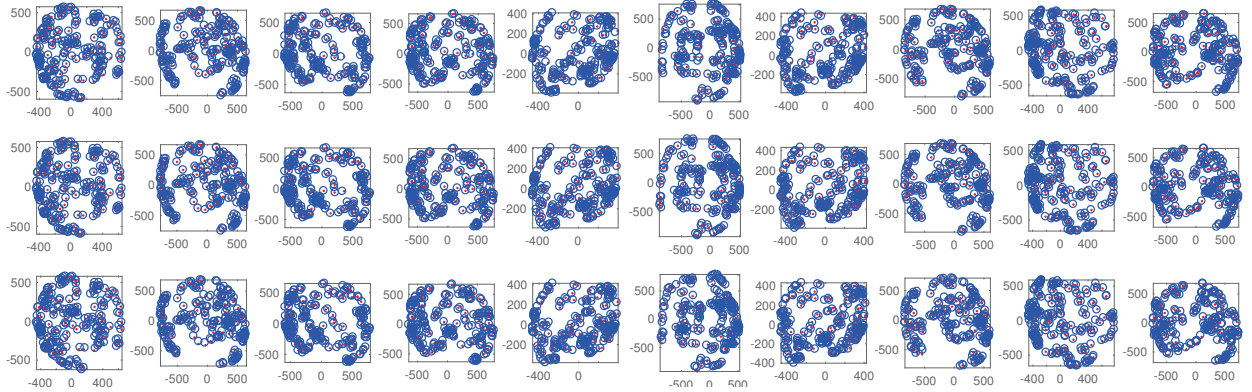

Figure 4. Reprojections (blue) to random frames against ground truth projections (red) with the Balloon deflation dataset; $K = 5$. (1st row) Pseudoinverse Method Dai *et al.* [10]; (2nd row) Block Matrix Method Dai *et al.* [10]; (3rd row) Proposed ISA.

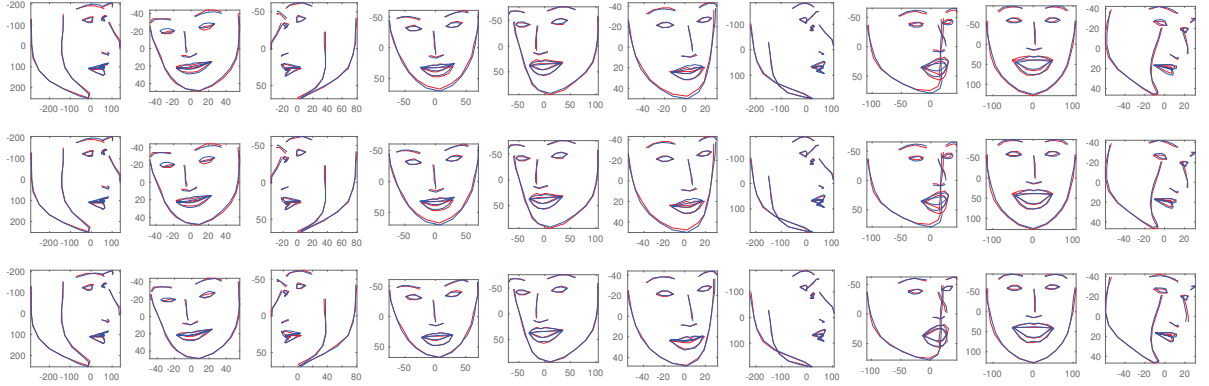

Figure 5. Reprojections (blue) to random frames against ground truth projections (red) with the LS3D-W face dataset; $K = 9$. (1st row) Pseudoinverse Method Dai *et al.* [10]; (2nd row) Block Matrix Method Dai *et al.* [10]; (3rd row) Proposed ISA.

5. CONCLUSIONS

In this paper we proposed a generalisation for non-rigid structure-from-motion. In contrast to the earlier belief that the recovery of shape basis would be ambiguous without prior information, we have shown that only assuming statistical independence between the 3D basis shapes yields
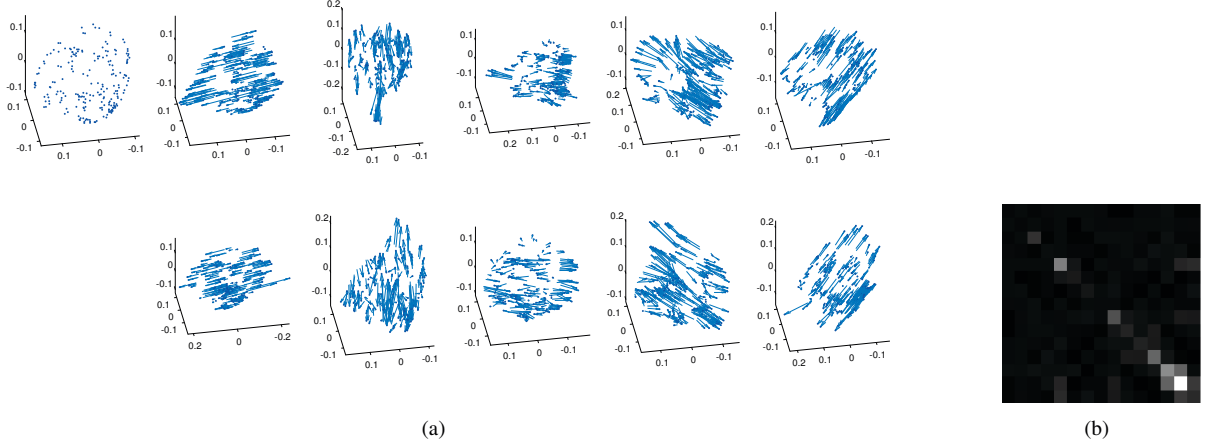
(a)

(b)

Figure 6. (a) Affine ISA Shape Basis on the Balloon deflation dataset with $K = 5$. (left column) Rigid affine 3D Shape $\mathbf{B}_0$; (the other columns) estimated non-rigid ISA shape basis component on both sides around the rigid shape. The arrows illustrate the drift of the points from the mean shape positions. The basis shapes are the components $\mathbf{B} = \mathbf{B}_0 \pm \alpha_k \hat{\mathbf{B}}_k$, where α_k is a positive scalar. (b) The $3K \times 3K$ mode covariance matrix $\mathbf{C}$ demonstrating the 3×3 block diagonal structure.
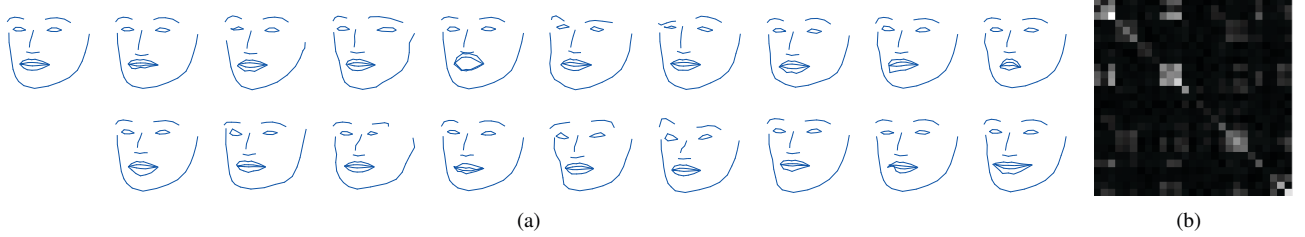

(a)

(b)

Figure 7. (a) Affine ISA shape basis for the LS3D-W data set with $K = 9$. (Left column) rigid affine 3D Shape $\mathbf{B}_0$; (the other columns) the 9 estimated 3D ISA basis shapes $\mathbf{B} = \mathbf{B}_0 \pm \alpha_k \hat{\mathbf{B}}_k$, where α_k is a positive scalar. (b) The $3K \times 3K$ mode covariance matrix $\mathbf{C}$ demonstrating the 3×3 block diagonal structure.

an uncalibrated, affine shape basis and affine, non-rigid structure and motion estimates. In analogy to the theory about rigid structure-from-motion, estimating an affine reconstruction instead of an Euclidean one yields a simpler solution for the non-rigid structure-from-motion problem, and independent subspace analysis serves as a natural tool to resolve the basis ambiguity. Our experiments showed that the approach is suitable for large data sets and it facilitates modelling and analysis of non-rigid structures in an uncalibrated setting. The approach hence opens the way for solving the non-rigid structure-from-motion problem for stationary motion sequences. In future, we are going to extend our methodology to handle missing data and to cope with more versatile statistical dependencies between the shape bases.

A. IRLS FOR BLOCK-FORM RECOVERY

Problem: Find $\alpha_k^i \in \mathbb{R}$ and $\mathbf{D}_k \in \mathbb{R}_{3 \times 3}$ such that

$$\sum_{i,k} \|\mathbf{M}_k^i \mathbf{D}_k - \alpha_k^i \mathbf{M}^i\|_{\mathrm{Fro}}^2 \longrightarrow \min, \qquad (15)$$

subject to $\|\mathbf{D}_k\|_{\mathrm{Fro}} = 1$, $k = 1, 2, \dots, K$, where $\mathbf{M}_k^i, \mathbf{M}^i \in \mathbb{R}_{2 \times 3}$.

Solution:

Let $\mathbf{d}_k = \mathrm{vec}(\mathbf{D}_k)$, $\mathbf{m}^i = \mathrm{vec}(\mathbf{M}^i)$. Now, for $i = 1, 2, \dots, I$, $k = 1, 2, \dots, K$,

$$\|\mathbf{M}_k^i \mathbf{D}_k - \alpha_k^i \mathbf{M}^i\|_{\mathrm{Fro}}^2$$

$$= \left\| \underbrace{\begin{pmatrix} \mathbf{M}_k^i & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_k^i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{M}_k^i \end{pmatrix}}_{\triangleq \mathbf{N}_k^i} \mathbf{d}_k - \alpha_k^i \mathbf{m}^i \right\|_2^2 \quad (16)$$

$$= \left\| \begin{pmatrix} \mathbf{N}_k^i & -\mathbf{m}^i \end{pmatrix} \begin{pmatrix} \mathbf{d}_k \\ \alpha_k^i \end{pmatrix} \right\|_2^2.$$

By collecting all the coefficients $\mathbf{N}_k^i, \mathbf{m}^i$ into a $6IK \times (9 + I)K$ matrix $\mathbf{N}$ the problem (15) is equivalent to the constrained least squares problem

$$\|\mathbf{N}(\mathbf{d}_1, \dots, \mathbf{d}_K, \alpha_1^1, \dots, \alpha_K^1, \alpha_1^2, \dots, \alpha_K^2, \dots, \alpha_K^I)\|_2$$
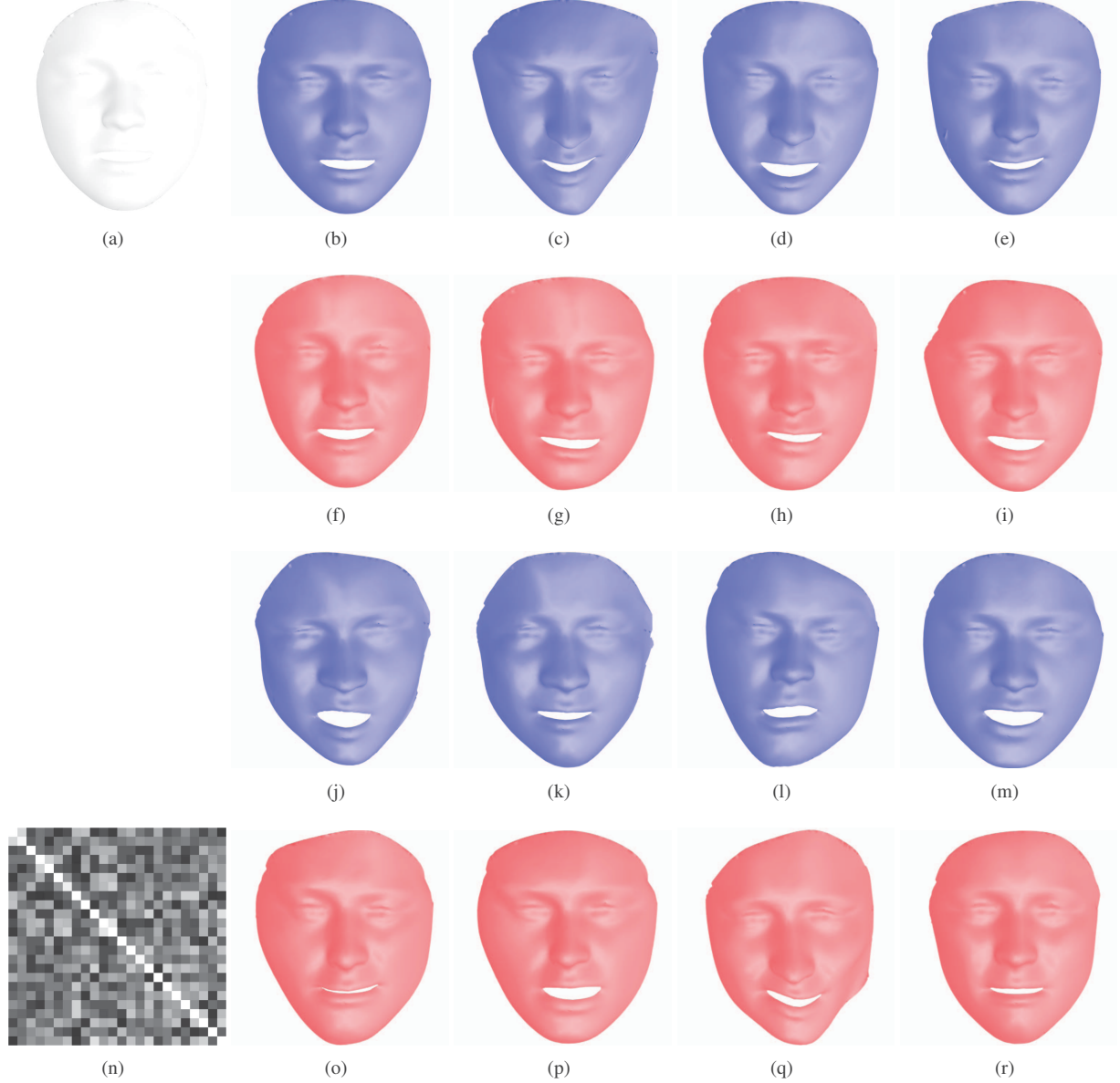$$\longrightarrow \min, \qquad (17)$$

Figure 8. Reconstructions of the $K = 8$ basis shapes computed from the Binghamton, BU3DFE [27] dataset. (a) The mean rigid shape $\mathbf{B}_0$; (b–r, except n) the estimated 3D ISA basis shapes $\mathbf{B} = \mathbf{B}_0 \pm \alpha_k \hat{\mathbf{B}}_k$, shown in red and blue, respectively. (n) Mode covariance matrix.

subject to $\|\mathbf{d}_k\|_2 = 1$, $k = 1, 2, \ldots, K$. The estimate can be found by the iteratively reweighted least squares by first assuming that $\alpha_k^{1,(n)} = 1/K$ for $n = 0$, $k = 1, \ldots, K$ and finding the solution of the reduced system

$$\|\mathbf{N}_{\backslash \alpha^1}(\mathbf{d}_1, \ldots, \mathbf{d}_K, \alpha_1^2, \ldots, \alpha_K^2, \ldots, \alpha_K^I) - \mathbf{c}^{(n)}\|_2 \longrightarrow \min, \tag{18}$$

where $\mathbf{N}_{\backslash \alpha^1}$ is constructed from $\mathbf{N}$ by dropping the columns corresponding to α_k^1, for all k, and $\mathbf{c}^{(n)} = -\mathbf{N}_{\alpha^1}(\alpha_1^{1,(n)}, \ldots, \alpha_K^{1,(n)})$. The estimate for $\alpha_k^{1,(n+1)}$ is computed as $\alpha_k^{1,(n+1)} \leftarrow \alpha_k^{1,(n)}/\|\mathbf{d}_k^{(n)}\|$, and the compu-

tation is iterated until convergence. This reweighting follows from the weighting $w_k^i = \|\mathbf{d}_k^{(n)}\|^{-2}$ in the iterated reweighted least squares (IRLS) scheme seeking to adjust the mixing weights in the first view that results in the unity Frobenius norms for the estimate of $\mathbf{D}_k$. The IRLS solution typically converges in only a few iterations, so the computational overhead is negligible.

References

[1] H. Aanæs and F. Kahl. Estimation of deformable structure and motion. In *Workshop on Vision and Mod-*

elling of Dynamic Scenes (ECCVW), 2002. 1

[2] I. Akhter, Y. Sheikh, and S. Khan. In defense of orthonormality constraints for nonrigid structure from motion. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1534–1541, June 2009. 2

[3] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Nonrigid structure from motion in trajectory space. In *Advances in Neural Information Processing Systems (NIPS)*, pages 41–48, 2009. 2

[4] M. Awiszus, H. Ackermann, and B. Rosenhahn. Learning disentangled representations via independent subspaces. In *Proc. Third International Workshop on Robust Subspace Learning and Applications in Computer Vision, in conjunction to ICCV 2019*, October 2019. 2

[5] A. Bartoli, V. Gay-Bellile, U. Castellani, J. Peyras, S. Olsen, and P. Sayd. Coarse-to-fine low-rank structure-from-motion. In *2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, June 2008. 1

[6] M. Brand. Morphable 3d models from video. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2001. 1, 2

[7] S. S. Brandt, P. Koskenkorva, J. Kannala, and A. Heyden. Uncalibrated non-rigid factorisation with automatic shape basis selection. In *IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 352–359, Sept 2009. 2

[8] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3d shape from image streams. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 690–696, June 2000. 1, 2

[9] A. Bulat and G. Tzimiropoulos. How far are we from solving the 2d & 3d face alignment problem? (and a dataset of 230,000 3d facial landmarks). In *International Conference on Computer Vision (ICCV)*, 2017. 5

[10] Y. Dai, H. Li, and M. He. A simple prior-free method for non-rigid structure-from-motion factorization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2012. 2, 4, 5, 6

[11] A. Del Bue and L. Agapito. Non-rigid 3d shape recovery using stereo factorization. In *Asian Conference on Computer Vision (ACCV)*, 2004. 1

[12] A. Del Bue, X. Llad, and L. Agapito. Non-rigid metric shape and motion recovery from uncalibrated images using priors. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1191–1198, June 2006. 1

[13] A. Del Bue, J. Xavier, L. Agapito, and M. Paladini. Bilinear modeling via augmented lagrange multipliers. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, pages 1496 –1508, August 2012. 1, 2

[14] V. Golyanik, B. Taetz, G. Reis, and D. Stricker. Extended coherent point drift algorithm with correspondence priors and optimal subsampling. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2016. 5

[15] P. F. U. Gotardo and A. M. Martinez. Computing smooth time trajectories for camera and deformable shape in structure from motion with occlusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, pages 2051–2065, Oct 2011. 2

[16] R. Hartley and R. Vidal. Prespective nonrigid shape and motion recovery. In *Proc. Eccv*, 2008. 1

[17] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, Inc., 2002. 2, 4

[18] A. Hyvärinen and U. Köster. Fastisa: A fast fixed-point algorithm for independent subspace analysis. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, 2006. 2, 3

[19] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, July 1997. 2, 3

[20] S. H. N. Jensen, A. Del Bue, M. E. B. Doest, and H. Aanæs. A benchmark and evaluation of non-rigid structure from motion. *arXiv preprint arXiv:1801.08388*, 2018. 5

[21] C. Kong and S. Lucey. Prior-less compressible structure from motion. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4123–4131, June 2016. 2, 4, 5, 6

[22] M. Paladini, A. Del Bue, J. Xavier, L. Agapito, M. Stosic, and M. Dodig. Optimal metric projections for deformable and articulated structure-from-motion. *International Journal of Computer Vision (IJCV)*, 96:252–276, 2012. 1

[23] C. Tomasi and T. Kanade. Shape and motion form image streams under orthography: A factorization approach. *International Journal of Computer Vision (IJCV)*, 9(2):137–154, November 1992. 3

[24] L. Torresani, A. Hertzmann, and C. Bregler. Non-rigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 30(5):878–892, May 2008. 1, 2, 4

[25] L. Torresani, D. B. Yang, E. J. Alexander, and C. Bregler. Tracking and modeling non-rigid objects with rank constraints. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume I, pages 493–500, 2001. 1

[26] J. Xiao, J. Chai, and T. Kanade. A closed-form solution to non-rigid shape and motion recovery. *International Journal of Computer Vision (IJCV)*, 67(2):233–246, 2006. 1, 2

[27] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato. A 3d facial expression database for facial behavior research. In *7th International Conference on Automatic Face and Gesture Recognition*, pages 211–216, April 2006. 5, 8