Uncalibrated Non-rigid Factorisation with Automatic Shape Basis Selection

Sami S. Brandt[1,2], Pekka Koskenkorva[2], Juho Kannala[2], and Anders Heyden[3,1]
{sbrandt,pkoskenk,jkannala}@ee.oulu.fi; heyden@maths.lth.se
[1]Malmö University, Applied Mathematics Group, Östra Varvsgatan 11A, 20506 Malmö, Sweden
[2]University of Oulu, Machine Vision Group, P.O. Box 4500, 90014 University of Oulu, Finland
[3]Lund University, Centre for Mathematical Sciences, Box 118, SE-22100 Lund, Sweden

Abstract

We propose an extension to the non-rigid factorisation method to solve the affine structure and motion of a deformable object, where the shape basis is selected automatically. In contrast to earlier approaches, we assume a general uncalibrated, affine camera model whereas most of the previous approaches assume a special case such as an orthographic, weak-perspective or paraperspective camera model. In general, there is a global affine ambiguity for the shape bases. It turns out that a natural way of selecting the shape bases is to pick up the bases that are statistically as independent as possible. The independent bases can be found by independent subspace analysis (ISA) which leads to the minimisation of mutual information between the basis shapes. After selecting the shape basis by ISA, only the in-the-subspace affine ambiguities remain from the general affine ambiguity. To solve the remaining unknowns of the general affine transformation, we propose an iterative method that recovers the block structure of the factored motion matrix. The experiments are provided with synthetic structure and real face expression data in 2D and 3D, which show promising results.

1. Introduction

The structure-from-motion problem of non-rigid objects has recently received lot of attention in the computer vision community. The rigid affine SFM problem can be solved by rigid factorisation [16] using a rank constraint of the measurement matrix and the singular value decomposition. The factorisation approach can be generalised for non-rigid scenes [3] by modelling the structure deformations as a linear combinations of rigid basis shapes. However, in the non-rigid case, there are additional constraints for the measurement matrix due to its block structure. Many solutions to recover the block structure have been proposed of which the earliest are [18, 2]. A usual tactic is to compute the initial estimate by some means after which to impose the block form by bundle adjustment [8]. Under certain assumptions, even closed-form solution can be found [19]. The use of the rotation constraints however requires the use of some more restricting camera model such as weak-perspective projection [19]. Many authors introduce priors to well-pose the non-rigid structure-from-motion problem [17, 1, 14, 7, 5] and handle missing data by a statistical framework. As far as our method is concerned, the missing data can be similarly handled as in [1, 14], or by the EM algorithm [7].

To our knowledge, there is however no completely satisfactory general solution available that would suit for an uncalibrated affine camera model while minimising the reprojection error in the non-rigid factorisation problem. This holds also for the determination of the number of the shape bases and their affine ambiguity, although Xiao's basis constraints is one way to fix it in the weak perspective case. To determine the number of the shape bases, Torresani suggests in [18] an iterative method to increase the number of shape bases until sufficiently low reprojection error is achieved. This approach is developed further in [1] as a coarse-to-fine ordering of the deformation modes is proposed. Nevertheless, a method that satisfyingly solves these problems under the general uncalibrated affine camera model, to our knowledge, has not been proposed.

The key point of this paper is the observation that the non-rigid factorisation problem is essentially *blind source separation* problem. Independent component analysis (ICA) provides a solution of the blind source separation problem by defining the goal as finding such linear combinations of data so that the individual components are statistically as independent as possible [11]. The approach can be seen as mutual information minimisation or non-linear factor analysis, which characterise the underlying structure factors of the data. ICA has turned out to be a powerful statistical method and has established its position in the sig-

nal processing community. The multidimensional extension of ICA is known as independent subspace analysis (ISA) [10], where, in contrast to ICA, independence is assumed between subspaces instead of one-dimensional signals.

As Theis [15] points out, no general ISA solution is available yet and for now ISA problem has to be solved in every case separately. The pioneering work [10] solves the ISA problem in the context of phase- and shift invariant features in image data. The same authors have recently developed a fixed-point algorithm for ISA [12], which is however prone to convergence to local minima. Specific ISA solutions are additionally available for face recognition and separation of mixed audio sources [4, 6].

The non-rigid factorisation problem can be seen as an ISA problem since the shape bases are typically multidimensional, 3D shapes. Given an ISA solution, an affinity in each component (subspace) as well as permutations of the components of the same dimension give another ISA solution [15]. In our context this means that after we have selected the shape basis by ISA, only the in-the-subspace affine ambiguities and between-the-subspace permutation ambiquity remain from the general affine ambiguity. In practise, the ISA problem can be solved with the help of ICA by first finding independent directions in the signal space and then finding the dependent subsets of signals that form the ICA components. We follow this approach by first solving the ICA components by FastICA [9] algorithm and then search for the dependent subspaces by finding the permutations that reveal the block structure of the measurement matrix. We additionally propose how the in-the-subspace affine ambiguities can be solved that remain from the general affine ambiguity after applying ISA.

The paper is organised as follows. In Section 2, we review the standard non-rigid model. Independent component analysis is presented and applied to the factorisation problem in Section 3. In Section 4, we show how the non-rigid factorisation can be solved as an ISA problem. Experiments are presented in Section 5 and conclusions are in Section 6.

2. Standard 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. The modelled 2D projection $\hat{\mathbf{m}}_j^i$ for a 3D point $\mathbf{x}_j$, corresponding to the measurement $\mathbf{m}_j^i$, is then

$$\hat{\mathbf{m}}_j^i = \mathbf{M}^i \mathbf{x}_j + \mathbf{t}^i = \mathbf{M}^i \left(\sum_k \alpha_k^i \mathbf{b}_{kj} \right) + \mathbf{t}^i, \qquad (1)$$

where $\mathbf{M}^i$ is 2×3 projection matrix to the image i, $\mathbf{t}^i$ is the corresponding translation vector, and the coefficients α_k^i and the shape basis points $\mathbf{b}_{kj}$ form the linear combination, which is supposed to represent the non-rigid deformation. Assuming Gaussian noise, the maximum likelihood solution with respect to the unknown 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 function

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

or equivalently

$$\min \|\mathbf{W} - \hat{\mathbf{W}}\|_{\text{fro}}^2, \qquad (3)$$

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

$$\mathbf{W} \simeq \underbrace{\begin{pmatrix} \alpha_1^1 \mathbf{M}^1 & \alpha_2^1 \mathbf{M}^1 & \cdots & \alpha_K^1 \mathbf{M}^1 \\ \alpha_1^2 \mathbf{M}^2 & \alpha_2^2 \mathbf{M}^2 & \cdots & \alpha_K^2 \mathbf{M}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^I \mathbf{M}^I & \alpha_2^I \mathbf{M}^I & \cdots & \alpha_K^I \mathbf{M}^I \end{pmatrix}}_{\triangleq \mathbf{M}} \underbrace{\begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \vdots \\ \mathbf{B}_K \end{pmatrix}}_{\triangleq \mathbf{B}}, \qquad (4)$$

where $\mathbf{B}_k = (\mathbf{b}_{k1}\ \mathbf{b}_{k2}\ \cdots\ \mathbf{b}_{kJ})$. From (4) it follows that the noise free measurement matrix has the rank constraint $R \triangleq \operatorname{rank} \hat{\mathbf{W}} \leq 3K$. The matrix that minimises (3) with the rank constraint can be obtained from the singular value decomposition of $\mathbf{W} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}}$ by truncating the smallest singular values, keeping the R largest, and removing the corresponding singular vectors, giving

$$\hat{\mathbf{W}} = \underbrace{\left(\frac{1}{\sqrt{J}} \tilde{\mathbf{U}} \tilde{\mathbf{S}} \right)}_{\triangleq \tilde{\mathbf{M}}} \underbrace{\left(\sqrt{J} \tilde{\mathbf{V}}^{\mathrm{T}} \right)}_{\triangleq \tilde{\mathbf{B}}} = \underbrace{\tilde{\mathbf{M}} \mathbf{A}}_{\triangleq \hat{\mathbf{M}}} \underbrace{\mathbf{A}^{-1} \tilde{\mathbf{B}}}_{\triangleq \hat{\mathbf{B}}} = \hat{\mathbf{M}} \hat{\mathbf{B}}, \quad (5)$$

where $\mathbf{A}$ is an unknown affine transformation. The singular value decomposition does not generally force the block structure for $\tilde{\mathbf{M}}$ which $\mathbf{M}$ has in (4). To find the estimates for the uncalibrated, affine non-rigid structure $\hat{\mathbf{B}}$ and motion matrix $\hat{\mathbf{M}}$, we need to search for the best affine transformation $\mathbf{A}$ that recovers the block structure of the matrix as well as possible; the recovery of $\mathbf{A}$ will be studied in Section 3.

3. Shape Bases via ICA

3.1. Why ICA?

The independent component analysis is suitable for problems of blind source separation where one seeks to recover independent source signals and the mixing coefficients after examining only mixed signals. For instance, with two source signals $s_1(n)$, $s_2(n)$, $n = 1, 2, \ldots$, and observed signals

$$x_1(n) = a_{11} s_1(n) + a_{12} s_2(n), \qquad (6)$$

$$x_2(n) = a_{21} s_1(n) + a_{22} s_2(n). \qquad (7)$$

In ICA, it is sufficient to assume that the source signals are *statistically independent*, to recover the mixing coefficients and the source signals from the observed signals.

Solving the affine transformation in (5) can be seen as the blind signal separation problem above. The shape basis is a linear combination of the shape bases, where an individual signal is interpreted to be a row vector of $\hat{\mathbf{B}}$. In other words,

$$\tilde{\mathbf{B}} = \mathbf{A}\hat{\mathbf{B}}, \tag{8}$$

i.e., $\mathbf{A}$ has the role of the the mixing matrix in the ICA model. In non-rigid SFM problem, statistical independence of the shape basis vectors is not an unrealistic assumption while it does not even need to hold exactly. For instance, facial expressions, can be assumed to contain statistically independent factors, the base expressions. If the components are not strictly statistically independent, we will just obtain the *projection pursuit* directions instead.

3.2. Mutual Information Minimisation

As the criterion for statistical independence, ICA can be defined as the minimisation of the mutual information of N scalar random variables $X_1, X_2, \ldots, X_N$, or

$$I\{X_1, X_2, \ldots, X_N\} = \sum_{i=1}^{N} H\{X_i\} - H\{\mathbf{X}\}, \tag{9}$$

where H denotes the differential entropy defined as

$$H\{\mathbf{X}\} = -\int p(\mathbf{x}) \log p(\mathbf{x}) \mathrm{d}\mathbf{x}, \tag{10}$$

where $\mathbf{x}$ denotes the realisation of the random vector $\mathbf{X}$. The differential entropy has a connection to coding theory and, loosely speaking, (9) describes how much the compression code length of the vector $\mathbf{X}$ increases if all its elements are coded separately instead of being coded together. Mutual information is zero if and only if the random variables are statistically independent. Minimisation of the mutual information can thus be taken as a natural definition for ICA.

Usual preprocessing, data centring and whitening makes the implementation of the ICA easier. Assuming that the signals are centred, uncorrelated, and with unity variance, it can be shown that mutual information takes the form

$$I\{X_1, X_2, \ldots, X_N\} = C - \sum_i \mathcal{N}\{X_i\}, \tag{11}$$

where C is a constant and $\mathcal{N}$ denotes the negentropy defined as

$$\mathcal{N}\{\mathbf{X}\} = H\{\mathbf{X}_\mathrm{G}\} - H\{\mathbf{X}\}, \tag{12}$$

where $\mathbf{X}_\mathrm{G}$ is the Gaussian random vector with the same covariance as $\mathbf{X}$, thus the minimisation of the mutual information is equivalent to the problem of finding the maximally non-Gaussian directions from the centred and whitened data.

3.3. Centring and Whitening

Proposition 3.1 *Factorisation by SVD produces a shape basis that is centred and whitened, that is, the columns of the intermediate basis $\tilde{\mathbf{B}}$ have zero mean and they are uncorrelated, with the identity matrix as the covariance matrix.*

Proof. We first prove that the columns of $\mathbf{B}$ have zero mean. Using the fact that the measurement matrix has been mean corrected, it follows that

$$\mathbf{0} = \frac{1}{J}\mathbf{W}\mathbf{1} = \frac{1}{J}\tilde{\mathbf{U}}\tilde{\mathbf{S}}\tilde{\mathbf{V}}^\mathrm{T}\mathbf{1} = \tilde{\mathbf{M}}\left(\frac{1}{J}\tilde{\mathbf{B}}\mathbf{1}\right) = \tilde{\mathbf{M}}\bar{\mathbf{c}}, \tag{13}$$

where $\bar{\mathbf{c}}$ denotes the mean of $\tilde{\mathbf{B}}$ over the column dimension. Hence, $\bar{\mathbf{c}} \in \mathrm{Ker}\{\tilde{\mathbf{M}}\}$. However, the rank of $\tilde{\mathbf{M}}$ is R, thus $\mathrm{Ker}\{\tilde{\mathbf{M}}\} = \mathbf{0}$ implies $\bar{\mathbf{c}} = \mathbf{0}$.

We then show that the columns of $\tilde{\mathbf{B}}$ are uncorrelated. Let $\mathbf{c}_j$, $j = 1, \ldots, J$, denote the column vectors and $\mathbf{r}_r$, $r = 1, \ldots, R$, the row vectors of $\tilde{\mathbf{B}}$. The sample covariance matrix of the columns is

$$\hat{\mathbf{C}} = \frac{1}{J}\sum_j (\mathbf{c}_j - \bar{\mathbf{c}})(\mathbf{c}_j - \bar{\mathbf{c}})^\mathrm{T} = \frac{1}{J}\sum_j \mathbf{c}_j\mathbf{c}_j^\mathrm{T}. \tag{14}$$

On the other hand, the rows of $\tilde{\mathbf{B}}$ are orthogonal so that

$$\mathbf{r}_r^\mathrm{T}\mathbf{r}_q = \begin{cases} J & \text{if } r = q, \\ 0 & \text{otherwise,} \end{cases} \tag{15}$$

so we may write

$$\begin{aligned} \frac{1}{J}\sum_j \mathbf{c}_j\mathbf{c}_j^\mathrm{T} &= \frac{1}{J}\begin{pmatrix} \mathbf{r}_1^\mathrm{T}\mathbf{r}_1 & \mathbf{r}_1^\mathrm{T}\mathbf{r}_2 & \cdots & \mathbf{r}_1^\mathrm{T}\mathbf{r}_R \\ \mathbf{r}_2^\mathrm{T}\mathbf{r}_1 & \mathbf{r}_2^\mathrm{T}\mathbf{r}_2 & \cdots & \mathbf{r}_2^\mathrm{T}\mathbf{r}_R \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{r}_R^\mathrm{T}\mathbf{r}_1 & \mathbf{r}_R^\mathrm{T}\mathbf{r}_2 & \cdots & \mathbf{r}_R^\mathrm{T}\mathbf{r}_R \end{pmatrix} \\ &= \mathbf{I}, \end{aligned} \tag{16}$$

and the claim follows.

□

3.4. Recovery of the ICA Shape Bases

Since the intermediate basis is centred and white, we immediately have the following result.

Proposition 3.2 *The affine transform $\tilde{\mathbf{A}}$ that recovers the normalised independent components $\tilde{\mathbf{B}}_\mathrm{ICA}$ from $\tilde{\mathbf{B}}$, such that $\tilde{\mathbf{B}} = \tilde{\mathbf{A}}\tilde{\mathbf{B}}_\mathrm{ICA}$, is an orthogonal transform.*

Proof. By taking the expectation over both sides in (16), and denoting the columns of $\tilde{\mathbf{B}}_\mathrm{ICA}$ by $\mathbf{s}_j$, it follows that

$$\begin{aligned} \mathbf{I} &= \mathrm{E}\left\{\frac{1}{J}\sum_{j=1}^{J}\mathbf{c}_j\mathbf{c}_j^\mathrm{T}\right\} = \mathrm{E}\left\{\frac{1}{J}\sum_{j=1}^{J}\tilde{\mathbf{A}}\mathbf{s}_j\mathbf{s}_j^\mathrm{T}\tilde{\mathbf{A}}^\mathrm{T}\right\} \\ &= \tilde{\mathbf{A}}\left(\frac{1}{J}\sum_{j=1}^{J}\mathrm{E}\left\{\mathbf{s}_j\mathbf{s}_j^\mathrm{T}\right\}\right)\tilde{\mathbf{A}}^\mathrm{T} = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^\mathrm{T}. \end{aligned} \tag{17}$$

□

In practise, $\tilde{\mathbf{A}}$ is computed from $\tilde{\mathbf{B}}$ using the FastICA algorithm [9], which is a numerical implementation to minimise (11).

4. Signal Sorting and Subspace Affinity

Even with ICA, there are two ambiguities that cannot be recovered: (1) the relative variances or energies of the independent components, and (2) their order. If one acknowledges that the projection pursuit directions form the shape bases, we would only have a permutation and relative basis vector scaling ambiguities left in the affine transform. More generally, however, the ISA model allows any dependence structure between basis vectors in an independent subspace, so we have a permutation and in-the-subspace affine ambiguities that need to be determined. We will do this by searching for the nearest matrix with the block structure shown in (4).

The mixing matrix $\mathbf{A}$ will thus have the form

$$\mathbf{A} = \tilde{\mathbf{A}}\mathbf{P}\mathbf{D}, \tag{18}$$

where $\mathbf{P}$ is a permutation matrix and $\mathbf{D}$ is a block diagonal matrix containing the in-the-subspace affinities $\mathbf{D}_k$, $k = 1, 2, \ldots, K$ of the independent subspaces in $\hat{\mathbf{B}}$, thus,

$$\hat{\mathbf{M}} = \tilde{\mathbf{M}}\tilde{\mathbf{A}}\mathbf{P}\mathbf{D} \tag{19}$$

and

$$\hat{\mathbf{B}} = \mathbf{D}^{-1}\mathbf{P}^{-1}\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}. \tag{20}$$

The remaining problem is to find $\mathbf{P}$ and $\mathbf{D}$ so that the block matrix form of (4) is recovered as well as possible.

The best block matrix form is searched in the least squares sense, i.e., we compute

$$\min_{\substack{\alpha_k^i, \mathbf{M}^i, \mathbf{D}_k, \mathbf{P} \\ i=1,2,\ldots,I \\ k=1,2,\ldots,K}} \|\mathbf{M} - \tilde{\mathbf{M}}\tilde{\mathbf{A}}\mathbf{P}\mathbf{D}\|_{\mathrm{fro}}^2, \tag{21}$$

subject to $\|\mathbf{M}^i\|_{\mathrm{fro}}^2 = 1$, $i = 1, 2, \ldots, I$, $\|\mathbf{D}_k\|_{\mathrm{fro}}^2 = 1$, $k = 1, 2, \ldots, K$. The minimisation could be performed directly by bundle adjustment but we additionally propose an alternative, straightforward iteration method based on the following two conditional solutions.

Proposition 4.1 *The estimates for α_k^i and $\mathbf{M}^i$, $i = 1, 2, \ldots, I$, $k = 1, 2, \ldots, K$, that minimise (21), conditioned to an estimate for the in-the-subspace affinities $\mathbf{D}$, and the permutation matrix $\mathbf{P}$, can be computed by SVD.*

Proof. Let $\mathbf{N} = \tilde{\mathbf{M}}\tilde{\mathbf{A}}\mathbf{P}\mathbf{D}$, and let $\mathbf{N}_k^i$ denote the 2×3 blocks of $\mathbf{N}$, so the objective function takes the form

$$\begin{aligned} E &= \|\mathbf{M} - \underbrace{\tilde{\mathbf{M}}\tilde{\mathbf{A}}\mathbf{P}\mathbf{D}}_{\triangleq \mathbf{N}}\|_{\mathrm{fro}}^2 = \sum_{i,k} \|\alpha_k^i \mathbf{M}^i - \mathbf{N}_k^i\|_{\mathrm{fro}}^2 \\ &= \sum_{i,k} \|\alpha_k^i \mathrm{vec}(\mathbf{M}^i) - \mathrm{vec}(\mathbf{N}_k^i)\|^2 \\ &= \sum_i \Big\| \begin{pmatrix} \alpha_1^i \mathrm{vec}(\mathbf{M}^i) & \alpha_2^i \mathrm{vec}(\mathbf{M}^i) & \cdots & \alpha_K^i \mathrm{vec}(\mathbf{M}^i) \end{pmatrix} \\ &\quad - \underbrace{\begin{pmatrix} \mathrm{vec}(\mathbf{N}_1^i) & \mathrm{vec}(\mathbf{N}_2^i) & \cdots & \mathrm{vec}(\mathbf{N}_K^i) \end{pmatrix}}_{\triangleq \mathcal{N}^i} \Big\|_{\mathrm{fro}}^2 \end{aligned} \tag{22}$$

The estimates for the unknowns in each image i are thus obtained by computing SVD for the matrix $\mathcal{N}^i$ and truncating all the singular values and the corresponding singular vectors except the largest so that $\mathcal{N}^i \approx s^i \mathbf{u}^i {\mathbf{v}^i}^{\mathrm{T}}$. Denoting $\alpha^i = (\alpha_1^i, \alpha_2^i, \ldots, \alpha_K^i)$, we may write

$$\hat{\mathbf{M}}^i = \mathrm{mat}(\mathbf{u}^i), \quad \hat{\alpha}^i = s^i \mathbf{v}^i, \quad i = 1, 2, \ldots, I. \tag{23}$$

□

Proposition 4.2 *There is a one parameter family of solutions for the in-the-subspace affinities $\mathbf{D}$ that minimise the error (21) given estimates for α_k^i, $\mathbf{M}^i$, $i = 1, 2, \ldots, I$, $k = 1, 2, \ldots, K$, and the permutation matrix $\mathbf{P}$.*

Proof. Let $\mathbf{L} = \tilde{\mathbf{M}}\tilde{\mathbf{A}}\mathbf{P}$ so that we minimise

$$\begin{aligned} E &= \|\mathbf{M} - \mathbf{L}\mathbf{D}\|_{\mathrm{fro}}^2 = \sum_{i,k} \|\alpha_k^i \mathbf{M}^i - \mathbf{L}_k^i \mathbf{D}_k\|_{\mathrm{fro}}^2 \\ &= \sum_{i,k} \Bigg\| \underbrace{\alpha_k^i \mathrm{vec}(\mathbf{M}^i)}_{\triangleq \mathbf{a}_k^i} - \underbrace{\begin{pmatrix} \mathbf{L}_k^i & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_k^i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{L}_k^i \end{pmatrix}}_{\triangleq \mathcal{L}_k^i} \underbrace{\mathrm{vec}(\mathbf{D}_k)}_{\mathbf{d}_k} \Bigg\|^2 \\ &= \sum_k \underbrace{\left(\sum_i {\mathbf{a}_k^i}^{\mathrm{T}} \mathbf{a}_k^i \right)}_{\triangleq a_k^2} - 2 \underbrace{\left(\sum_i {\mathbf{a}_k^i}^{\mathrm{T}} \mathcal{L}_k^i \right)}_{\triangleq \mathbf{c}_k^{\mathrm{T}}} \mathbf{d}_k + \\ &\quad + {\mathbf{d}_k}^{\mathrm{T}} \underbrace{\left(\sum_i {\mathcal{L}_k^i}^{\mathrm{T}} \mathcal{L}_k^i \right)}_{\triangleq \mathcal{L}_k} \mathbf{d}_k \end{aligned} \tag{24}$$

Taking into account the constraint on the Frobenius norm of the matrices $\mathbf{D}_k$, $k = 1, 2, \ldots, K$, the Lagrangian takes the form

$$L = \sum_k a_k^2 - 2\mathbf{c}_k^{\mathrm{T}} \mathbf{d}_k + {\mathbf{d}_k}^{\mathrm{T}} \mathcal{L}_k \mathbf{d}_k + \lambda_k \left(\|\mathbf{d}_k\|^2 - 1 \right). \tag{25}$$

Algorithm 1 ISA/ICA-based non-rigid factorisation

1. Select a permutation matrix $\mathbf{P}$ that has not yet been selected.
2. Set the in-the-subspace affinities $\hat{\mathbf{D}}_k = \mathbf{I}$, $k = 1, 2, \ldots, K$.
3. Compute the estimates for α_k^i, $\mathbf{M}^i$, conditioned to an estimate for the in-the-subspace affinities $\hat{\mathbf{D}}_k$, $i = 1, 2, \ldots, I$, $k = 1, 2, \ldots, K$.
4. Compute the estimates for $\mathbf{D}_k$, $k = 2, 3, \ldots, K$, given the previous estimates $\hat{\alpha}_{k'}^i$, $\hat{\mathbf{M}}^i$, $i = 1, 2, \ldots, I$, $k' = 1, 2, \ldots, K$.
5. Iterate between steps 3-4 until convergence. Store the estimates and the obtained reprojection error estimate.
6. If there are permutations left, go to step 1. Otherwise, select the permutation and the estimates that minimise the reprojection error.

Differentiating L with respect to $\mathbf{d}_k$, and setting the result to zero yields

$$\mathbf{d}_k = \left(\mathcal{L}_k + \lambda_k \mathbf{I}\right)^{-1} \mathbf{c}_k, \quad k = 1, 2, \ldots, K, \tag{26}$$

which is a one parameter family of solutions in λ_k. The solution satisfying the unity norm for $\mathbf{d}_k$ is unique and can be found by one-dimensional search, where the corresponding objective function is strictly increasing in λ_k.

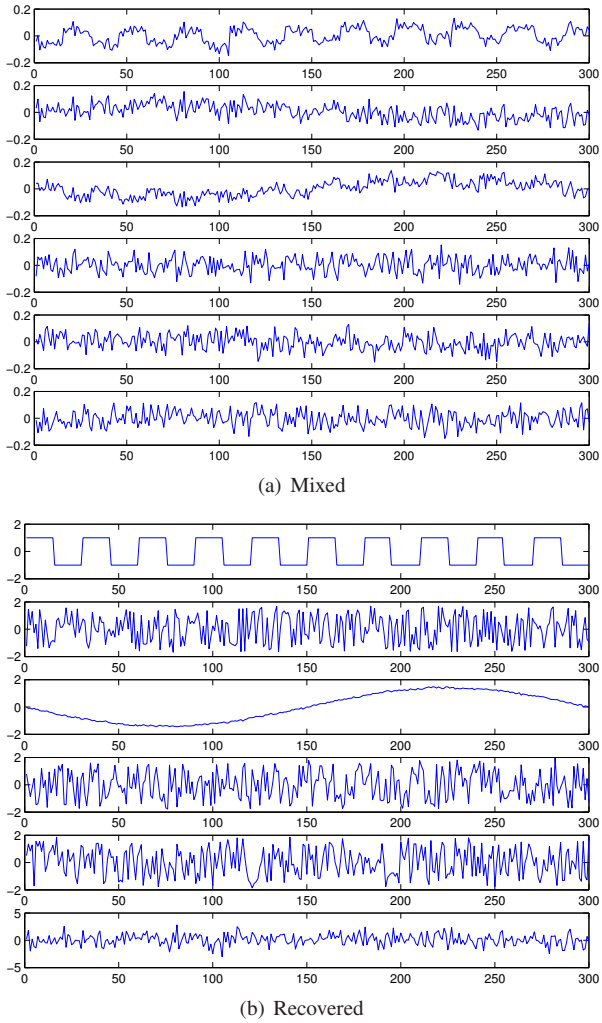
□

Our iterative method for minimising (21) is thus constructed by searching over all the possible permutations $\mathbf{P}$ and, given a permutation searching for the best block matrix form and in-the-subspace affinity, iterating the results of the Proposition 4.1 and 4.2. To avoid a singular solution we additionally fix, without a loss of generality, the first in-the-subspace affinity $\mathbf{D}_1$ to the identity matrix. The algorithm is summarised in Algorithm 1.

5. Experiments

We first show a simple proof of the principle by creating a synthetic object which contains two 3D point sets as shape bases each of containing 300 3D points. The first basis was generated as

$$\begin{cases} x_1 = x(n), & x(n) \sim \text{Uniform}(-0.5, 0.5) \\ y_1 = y(n), & y(n) \sim \text{Uniform}(-0.5, 0.5) \\ z_1 = z(n), & z(n) \sim \text{Uniform}(-0.5, 0.5) \end{cases} \tag{27}$$



(a) Mixed

(b) Recovered

Figure 1: Synthetic 3D basis shapes shown as one-dimensional signals.

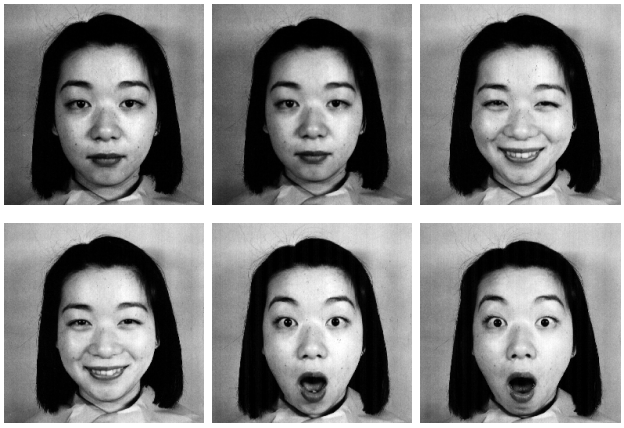


Figure 2: Training images: two neutral, two smiling and two surprised expressions.

and the second as

$$\begin{cases} x_2 = \sin \frac{n}{600\pi} \\ y_2 = \text{sign}\left(\sin \frac{n}{60\pi}\right) \\ z_2 = z(n), \quad z(n) \sim \text{N}(0,1) \end{cases} . \tag{28}$$

We modelled 6 projection views by generating six orthographic cameras pointing towards random directions. The mixing coefficients $\alpha = \{\alpha_k^i\}$ for the basis shapes were set as

$$\alpha_0 = \begin{pmatrix} 1 & 0.55 & 0.1 & 1 & 0.55 & 0.1 \\ 0.87 & 0.91 & 0.16 & -0.5 & -0.97 & -0.31 \end{pmatrix}.$$

and the measurement matrix was built using the model (4).

To solve the inverse problem, we decomposed the measurement matrix on the basis of (5) and computed the independent components by FastICA for the rows of the shape basis matrix $\hat{\mathbf{B}}$, shown in Fig. 1(b). It can be seen that the original signals were succesfully recovered. We see also the well known fact that ICA cannot recover the order or the sign of the signals.

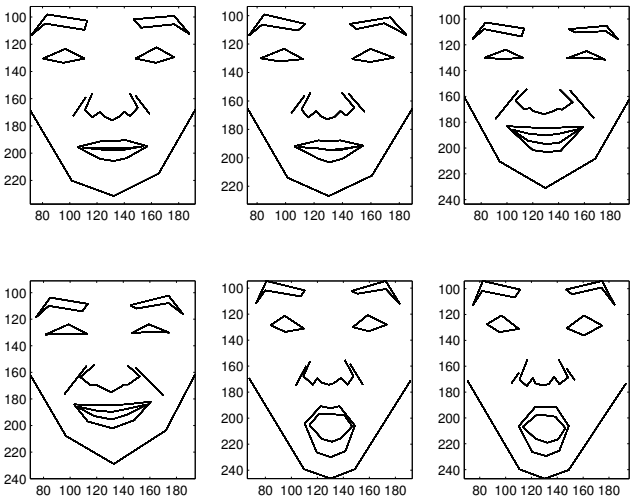


Figure 3: Manually extracted training shapes.

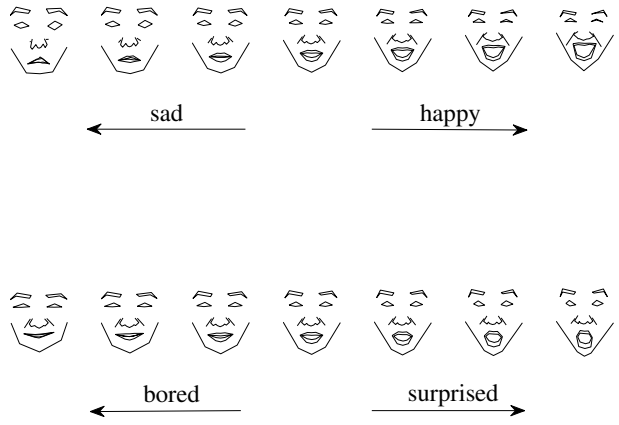


Figure 4: The two estimated ICA basis shapes. The mean expression over the training shapes is shown in the middle and the ICA component added with an increasing (decreasing) weight to the right (left).

Another experiment was carried out with real face expression data. We picked up six training images (Fig. 2) from the Japanese Female Facial Expression (JAFFE) database [13]. From each image, 54 feature points were manually exctracted to form training shapes (Fig. 3) for our algorithm. In this example, we assumed a 2×2 affine model for the transformation from the shape basis to the image, as the 3D information was very weak in the training data. From the measurement matrix, the independent shape components were then extracted by the ISA method given in Algorithm 1, assuming the model with two shape bases. The recovered shape bases are illustrated in Fig. 4. The first basis shape describes sad-happy mood of the female, whereas the second mode represents bored-surprised emotion. Note that the method is able to generalise the semantic sad-happy and bored-surprised modes on the basis of the training data, even though there is no unhappy example in the training data.

In the third experiment, we tested our algorithm with the real 3D data; the facial motion capture data in [17]. The data contains 3D position measurements of 40 markers attached to the subject's face where the subject performed a range of facial expressions and dialogue while rotating his head horizontally. The sequence consisted of 316 frames in total. The 3D data was projected to images by orthogonal projection where the camera was kept in a fixed position. From the projection data, the camera projection matrices as well as the shape bases were recovered using the Algorithm 1 with the assumption of two shape bases ($K = 2$). For comparison, we computed the result with the EM-LDS method [17]. The results are shown in Table 1 where it can be seen that the relative error $\|\mathbf{W} - \hat{\mathbf{W}}\|_{\text{fro}} / \|\mathbf{W}\|_{\text{fro}}$ is smaller than with the EM-LDS method. As another proof of the principle, we illustrate the reprojection of the recovered 3D points on to images in Figure 5 and show the recovered shape bases in Figure 6. It is intuitively pleasing to see that the emotions sad–angry and neutral-mouth-shut–neutral-mouth-open are recovered as the statistically independent shape bases by ISA.

Table 1: Face motion capture data (316 frames) with EM-LDS algorithm [17] and our ISA-algorithm as $K = 2$.

method	relative 2D error
EM-LDS	0.0164
ISA	0.0110

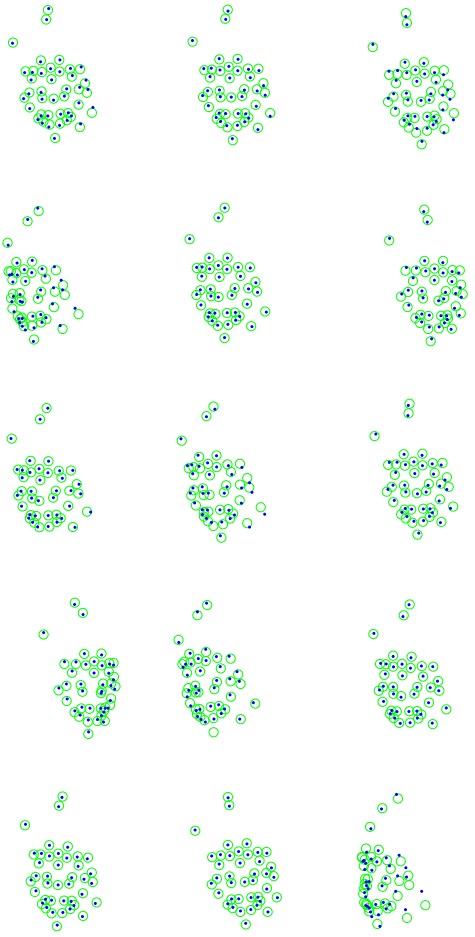


Figure 5: 2D reprojections of the face motion capture data [17], with $K = 2$, in 15 frames picked uniformly from the sequence. Our algorithm was given 2D tracks as inputs. Ground-truth features are shown as green circles; reprojections are blue dots.

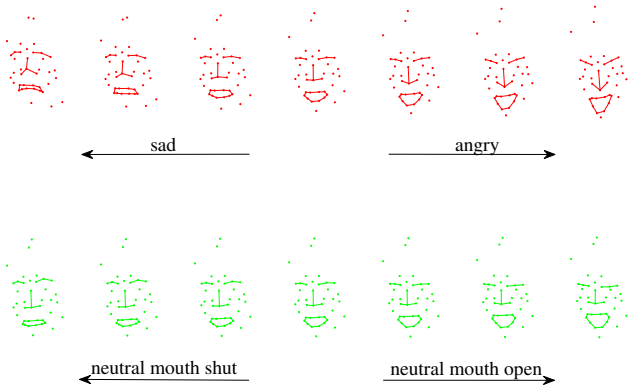


Figure 6: The three-dimensional shape modes in the face motion capture data [17] reprojected into a randomly picked frame number 51. The first mode represents the sad–angry emotion where as the second a neutral mouth-shut–mouth-open expression.

6. Conclusions

In this paper, we have approached the uncalibrated non-rigid structure-from-motion problem as a blind source separation problem. As common approach for non-rigid affine SFM is to model shapes as a linear combination of rigid basis shapes, the challenging problem has been in finding the basis shapes, consistent with the non-rigid factorisation model, that provide a descriptive model for the non-rigid data. Our solution, based on independent subspace analysis, searches for the shape bases that are statistically as independent as possible. The ISA solution reduces the global affine ambiguity of the standard non-rigid factorisation into in-the-subspace affine ambiguities, for which we proposed an iterative solution to find the solution that preserves the block structure of the measurement matrix as well as possible. In contrast to the exisiting algorithms for non-rigid factorisation, our approach is general, uncalibrated, i.e., the solution is given up to an unknown 3D affinity. The solution can be updated to Euclidean by applying metric constrains as it is conventional in the stratified approach for SFM. Our results are promising while showing that meaningful statistically independent shape basis can be found e.g. in facial expression data both which explain the non-rigid data. In the future, we intend to investigate other ISA algorithms that might be more effective in the non-rigid factorisation problem.

References

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

[2] M. Brand. Morphable 3D models from video. In *In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 456–463, 2001.

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

[4] M. A. Casey. Separation of mixed audio sources by independent subspace analysis. Mitsubishi Electric Research Laboratory, 2001.

[5] A. Del Bue. A factorization approach to structure from motion with shape priors. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[6] H. K. Ekenel and B. Sankur. Feature selection in the independent component subspace for face recognition. *Pattern Recognition Letters*, 25:1377–1388, 2004.

[7] A. Gruber and Y. Weiss. Factorization with uncertainty and missing data: exploiting temporal coherence. In *in NIPS*, 2003.

[8] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2003.

[9] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.

[10] A. Hyvärinen and P. Hoyer. Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces. *Neural Computation*, 12(7):1705–1720, 2000.

[11] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent component analysis*. Adaptive and learning systems for signal processing, communications and control. John Wiley and Sons, Inc., 2001.

[12] A. Hyvärinen and U. Köster. FastISA: A fast fixed-point algorithm for independent subspace analysis. In *ESANN*, 2006.

[13] M. J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets. In *International Conference on Automatic Face and Gesture Recognition*, pages 200–205, 1998.

[14] S. I. Olsen and A. Bartoli. Implicit non-rigid structure-from-motion with priors. *Journal of Mathematical Imaging and Vision*, 31(2-3):233–244, 2008.

[15] F. J. Theis. Towards a general independent subspace analysis. In *Neural Information Processing Systems*, 2006.

[16] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137–154, 1992.

[17] 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*, 30(5), 2008.

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

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