Adaptive Metric Registration of 3D Models to Non-rigid Image Trajectories

Alessio Del Bue*

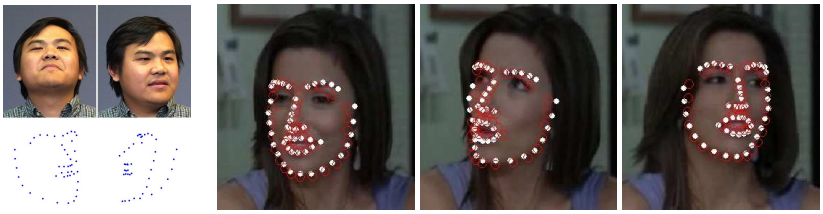
Istituto Italiano di Tecnologia
Via Morego 30, 16163 Genova, Italy
alessio.delbue@iit.it

Abstract. This paper addresses the problem of registering a 3D model, represented as a cloud of points lying over a surface, to a set of 2D deforming image trajectories in the image plane. The proposed approach can adapt to a scenario where the 3D model to register is not an exact description of the measured image data. This results in finding the best 2D–3D registration, given the complexity of having both 2D deforming data and a coarse description of the image observations. The method acts in two distinct phases. First, an affine step computes a factorization for both the 2D image data and the 3D model using a joint subspace decomposition. This initial solution is then upgraded by finding the best projection to the image plane complying with the metric constraints given by a scaled orthographic camera. Both steps are computed efficiently in closed-form with the additional feature of being robust to degenerate motions which may possibly affect the 2D image data (i.e. lack of relevant rigid motion). Moreover, we present an extension of the approach for the case of missing image data. Synthetic and real experiments show the robustness of the method in registration tasks such as pose estimation of a talking face using a single 3D model.

1 Introduction

The analysis of non-rigid motion has great relevance in many life science and engineering tasks. This need arises from the observation that most of the natural shapes are constantly modifying their topology. Such variations may appear smooth and tiny as in the bending of the arm muscles or drastic and violent, as in the reactions taking place at the molecular level. Such degrees of variation have consequently brought new challenges in the Structure from Motion (SfM) [3, 12, 2, 1] and image registration fields [4, 8, 14]. The problem is made more difficult because the assumption of rigidity is now broken and the classical metric constraints used in rigid SfM [11] are weakened if not irremediably lost. Here specifically, we study the problem of registering a 3D model to a set of 2D trajectories extracted from an image sequence. Our challenge is represented by the fact that the 3D model to register may not be an exact description of the 2D motion shown in the image sequence as exemplified in Figure 1 in a face analysis domain. The aim is to provide a new set of tools which adapt to the new information provided by the image sequence. This problem occurs more often thanks to the rapid advance-

* This work was partially supported by FCT, under ISR/IST plurianual funding (POSC program, FEDER) and grant MODI-PTDC/EEA-ACR/72201/2006. Thanks to J. Peyras and J. Xiao for providing the image sequences and tracks.



(a) Rigid sequence (b) Non-rigid 2D image sequence

Fig. 1. The figure shows an example of our problem. In the top row of figure (a), a 3D shape can be recovered from a rigid image sequence with standard SfM algorithms. The model in the bottom row has now to be registered to a new non-rigid image sequence (b) with 2D trajectories extracted from a subject with different somatic traits. We seek the best registration given both 2D and 3D data which satisfy the metric constraints of the shapes. White dots represent the 2D image data and the red circles ○ our algorithm result.

ments of the modern sensor technologies. Nowadays, it is a more likely occurrence to have available measurements coming from different devices. However, temporally evolving data is mainly restricted to 2D observation (e.g. video from cameras, MRI and ultrasound images) while full 3D information is captured at sparser time instances (e.g. scans given by CT and range sensors). For this reason, a robust 2D-3D registration of data coming from different sources is more often required. Moreover such registration has to adapt to the given observed image motion, since it is likely that the given 3D surface may not be an exact representation of the evolving shape.

This paper proposes a novel registration procedure that adapts the given 3D shape to the 2D data. In order to solve the problem, a general two-step formulation is introduced. First, a compact low-rank description is extracted from both the 2D measurements and the 3D rigid shape. This first decomposition is up to a generic affine transformation. Then, this solution is corrected by finding the best transformation that complies with the metric constraints given the image motion and the shape to register. To the authors knowledge, the closest work to the proposed algorithm is the one by Xiao et al. [14] where the scope of the authors was not only restricted to registration but also to the inference of a full deformable model. Their closed form solution however makes use of the assumption that there exists a set of independent basis shapes and results may vary if this choice is not accurate as noted in [2, 12]. Full 3D reconstruction is out of the scope of this paper since our main aim is to find the most appropriate rigid motion describing the non-rigid image trajectories without any assumptions about the model underlying the deformations.

1.1 Contributions and paper organization

We first introduce the mathematical framework and a standard solution for the 2D–3D registration problem with rigid models. Such an algorithm however cannot cope properly when the registration is done with inaccurate 3D observations such as the one shown in Figure 1(b). The proposed method instead performs an affine registration procedure which is derived from the work of Del Bue [5]. The first contribution is a

new set of metric constraints which jointly force the projection constraints for the 2D data and 3D data. This problem is then formulated by finding a corrective transform which enforces the given constraints. This optimization can be solved either in closed form with Least Squares (LS) or by defining the problem with a cost function which is minimised using convex optimization. In this way, we consequently not only perform a registration but also the reconstruction of a new rigid shape or deformable model which adapts automatically to the image measurement and 3D shape geometric constraints. This solution is particularly robust to degenerate 2D motion given this new set of metric constraints. The second contribution is an iterative extension of the proposed approach which deals with the likely event of missing data in the 2D image trajectory.

The paper is structured as follows. Section 2 introduces the problem and a first initial solution. Section 3 presents the new approach when the 3D shape needs adaptation to deal with the variations in the 2D data. In the case of missing data, Section 4 provides an iterative solution to the problem. Section 5 shows synthetic and real data while Section 6 points out the possible improvements and direction for future work.

2 Rigid and non-rigid 2D–3D registration

2.1 Rigid registration with an exact 3D model

Consider first the problem of registering a single rigid shape to a set of 2D image trajectories. The 2D image measurements are stored in a single matrix $\mathtt{W}$ of size $2F \times P$ with the following structure:

$$\mathtt{W} = \begin{bmatrix} \mathbf{w}_{11} & \dots & \mathbf{w}_{1P} \\ \vdots & \ddots & \vdots \\ \mathbf{w}_{F1} & \dots & \mathbf{w}_{FP} \end{bmatrix} = \begin{bmatrix} \mathtt{W}_1 \\ \vdots \\ \mathtt{W}_F \end{bmatrix}, \tag{1}$$

where F and P are the total number of frames and the number of points respectively. The 2-vector $\mathbf{w}_{ij} = (u_{ij}\, v_{ij})^T$ stores the image coordinates at each frame i and point j. Given a known rigid shape $\mathtt{B}$ of size $3 \times P$ our aim is then to compute the best projection that aligns the 3D shape to the 2D data. In this work there are two main assumptions. The assignments between the image trajectories in $\mathtt{W}$ and the 3D points in $\mathtt{B}$ are given and that, initially, $\mathtt{W}$ does not contain missing data. However, this last assumption will be relaxed later in this paper.

The image projection model considered here is a scaled orthographic model denoted as a 2×3 matrix $\mathtt{M}_i$ such that $\mathtt{M}_i = c_i \mathtt{R}_i$ with the orthogonality constraints given by $\mathtt{R}_i \mathtt{R}_i^T = \mathtt{I}_2$. The 2D–3D registration problem can be then re-stated as the optimization of the following cost function:

$$\min_{\mathtt{R}_i \mathtt{R}_i^T = \mathtt{I}_2} \quad \left\| \mathtt{W} - \mathtt{M}\mathtt{B} \right\|^2 \tag{2}$$

where $\mathtt{M}$ is the matrix obtained by stacking all the sub-blocks $\mathtt{M}_i$ for each frame as:

$$\mathtt{M} = \begin{bmatrix} \mathtt{M}_1 \\ \vdots \\ \mathtt{M}_F \end{bmatrix}. \tag{3}$$

A solution to this problem satisfying the exact orthographic constraints can be obtained in two steps. First, by finding an affine Maximum Likelihood (ML) solution using the pseudoinverse of $\mathtt{B}$ giving $\tilde{\mathtt{M}} = \mathtt{W}\mathtt{B}^T(\mathtt{B}\mathtt{B}^T)^{-1}$ and then forcing the orthogonality constraints in $\tilde{\mathtt{M}}$. This final step is not performed globally for the collection of the 2×3 sub-blocks $\tilde{\mathtt{M}}_i$ as done in the Tomasi-Kanade factorization [11]. Instead, the affine block is projected into the closest scaled orthographic camera matrix $c_i\mathtt{R}_i$ as presented by Marques and Costeira [7] in a 3D reconstruction context. Such projection is given by:

$$\mathtt{R}_i = \mathtt{U}\mathtt{V}^T \quad \text{and} \quad c_i = (\sigma_1 + \sigma_2)\backslash 2 \tag{4}$$

where $\tilde{\mathtt{M}}_i = \mathtt{U}\mathtt{D}\mathtt{V}^T$ is the SVD of the affine motion matrix and σ_d for $d = 1, 2$ are the singular values stored in $\mathtt{D}$. Such projection is preferred to the global LS solution which may not exactly comply with the scaled orthographic camera matrix constraints. Differently, eq. (4) always gives a matrix $\mathtt{R}_i$ that complies with the given constraints as pointed out in [7, Appendix B].

Note that the solution obtained in step 2 of Algorithm 1 is optimal with the assumption of isotropic and zero-mean Gaussian noise affecting the measurements in $\mathtt{W}$. Such assumption is generally valid when accurate 2D measurements are obtained from the image tracks of a rigid object. However, when trajectories are extracted from shapes with consistent directional deformations, such assumption is violated as it was noticed by Xiao et al. [14] in a medical context.

Algorithm 1 Rigid registration with image projections

Require: The 2D image data $\mathtt{W}$ and the 3D shape $\mathtt{B}$.

Ensure: A metric 2D–3D registration of the shape to the image measurements.

1: Compute the image centroid of $\mathbf{t} = \frac{1}{P}\mathtt{W}\mathbf{1}_P$ and register the data as $\bar{\mathtt{W}} = \mathtt{W} - \mathbf{t}\mathbf{1}_P^T$
2: Estimate the affine motion $\tilde{\mathtt{M}}$ as $\tilde{\mathtt{M}} = \bar{\mathtt{W}}\mathtt{B}^T(\mathtt{B}\mathtt{B}^T)^{-1}$.
3: Project each 2×3 sub-block $\tilde{\mathtt{M}}_i$ to the closest scaled orthographic matrix using eq. (4).

2.2 Registration bias with inexact models

Deformation directionality is less noticeable when non-rigid motion is nearly isotropic to the shape centroid or with strong symmetries. Figure 2 shows a case when a 2D image of a cylinder is bending and the actual registration given Algorithm 1 with a rigid 3D shape from the ground truth at rest. As expected, a consistent bias in the 2D–3D registration appears when the shape is bending towards the direction of maximal variation. In such cases, a rigid registration of a single $\mathtt{B}$ is unfit since it cannot deal with the deformations. When the data is non-rigid we have at each frame that:

$$\mathtt{W}_i = c_i\mathtt{R}_i\mathtt{X}_i \quad \text{with} \quad \mathtt{X}_i \in \Re^{3\times P} \tag{5}$$

where $\mathtt{X}_i$ represents the metric time-varying shape. For the whole set of 3D shapes, the most popular representation used is to parameterize $\mathtt{X}_i$ as a set of linear basis shapes [3] giving $\mathtt{X}_i = \sum_d l_{id}\mathtt{S}_d$. These linear bases are usually sufficient to represent a generic set of deformations however they may require a high number of basis shapes when dealing with non-linear deformations as to the bending cylinder in Figure 2.

Frame 1 Frame 120

Fig. 2. Black dots • represent the 2D measurements, red circles ○ a half-cylinder 3D shape registered by Algorithm 1 and blue crosses × the results by the proposed Algorithm 2. The image data show the cylinder starting from a rest position in Frame 1 where the registration is perfect. The cylinder is bending at the last 3 semi-circles and the registration at the maximum deformation is strongly biased toward the deformation direction.

3 Adaptive registration using joint subspaces

Algorithm 1 may perform well when $\mathtt{B}$ represents a single instance of the deformations appearing in the image sequence. However such a case is unlikely in many registration scenarios and a method which encompasses some degree of adaptation may strongly reduce the registration error. In the following, the given surface $\mathtt{B}$ is not a current observation of the 2D image trajectories (i.e. $\mathtt{X}_i \neq \mathtt{B}$ for $i = 1 \dots F$). This will consequently affects the estimated motion parameters in Algorithm 1 giving an additional bias from the unfitness of $\mathtt{B}$. In order to reduce this effect we propose a different approach which first finds an affine joint subspace belonging to the set $\{\mathtt{W}, \mathtt{B}\}$ and then computes the best solution to registration given the joint metric constraints.

3.1 Affine joint subspace computation

The main idea here is to join the information contained in $\mathtt{B}$ with the available measurements in $\mathtt{W}$ in order to extract an affine fit which is dependent on both components. In order to do so, we follow the strategy used in [5] for a 3D reconstruction scenario. A *Generalised Singular Value Decomposition* (GSVD) is used to compute a joint row space between the image data and the model to register. In such a way, we decompose both matrices with GSVD as:

$$\begin{aligned} \mathtt{W} &= \mathtt{U}\,\mathtt{D}_U\,\mathtt{X}^T \\ \mathtt{B} &= \mathtt{V}\,\mathtt{D}_V\,\mathtt{X}^T \end{aligned} \tag{6}$$

where $\mathtt{X}^T$ is a $P \times P$ matrix which spans the common row space of $\{\mathtt{W}, \mathtt{B}\}$, $\mathtt{U}$ is a $2F \times 2F$ matrix with orthonormal columns ($\mathtt{U}^T\mathtt{U} = \mathtt{I}$) and $\mathtt{V}$ is a 3×3 matrix such that $\mathtt{V}^T\mathtt{V} = \mathtt{I}$. The diagonal value matrices $\mathtt{D}_U$ and $\mathtt{D}_V$ of size $2F \times P$ and $3 \times P$ respectively are given by:

$$\mathtt{D}_U = \begin{bmatrix} \Sigma_U & \mathtt{0} \\ \mathtt{0} & \mathtt{I} \end{bmatrix} \text{ and } \mathtt{D}_V = \begin{bmatrix} \Sigma_V & \mathtt{0} \end{bmatrix}. \tag{7}$$

The diagonal matrices $\Sigma_U = diag(\sigma_1, \ldots, \sigma_3)$ and $\Sigma_V = diag(\mu_1, \ldots, \mu_3)$ of size 3×3 are constrained such that $\Sigma_U^2 + \Sigma_V^2 = \mathtt{I}$ and the diagonal entries ordered as:

$$0 \leq \sigma_1 \leq \ldots \leq \sigma_3 \leq 1 \text{ and } 1 \geq \mu_1 \geq \ldots \geq \mu_3 > 0.$$

In order to guarantee a well-conditioned decomposition a single scaling of the data is performed imposing that $\|\mathtt{W}\|^2 = \|\mathtt{B}\|^2$ [6]. Given the initial factorisation with GSVD, it is possible after some matrix operations [5] to arrange the different factors as:

$$\begin{aligned} \mathtt{W} &= \tilde{\mathtt{M}}_{2f \times t} \tilde{\mathtt{S}}_{t \times p} = [\mathtt{M}_J | \, \mathtt{M}_I] \begin{bmatrix} \mathtt{B}_J \\ \mathtt{B}_I \end{bmatrix} \\ \mathtt{B} &= \mathtt{N}_{3\times 3} \mathtt{B}_J \end{aligned} \tag{8}$$

where the J subscript refers to the components obtained from the joint space of $\mathtt{B}$ and $\mathtt{W}$ while the I refers to the remaining ones. The dimensionality of the joint row space $\mathtt{B}_J$ depends directly on the dimension of the model to register. Thus, in the case of rigid registration, the matrix $\mathtt{B}_J$ has size $3 \times P$ and the $r = (t-3)$ dimension of $\mathtt{B}_I$ depends on the rank of the independent components. Such value can be estimated by inspecting the singular values of the remaining 2D data and choosing a r which contains most of the energy. Notice that this parameter is not important for the proposed approach since it relies only on the joint components $\mathtt{M}_J$ and $\mathtt{B}_J$.

3.2 Joint metric upgrade

The next step is to find a corrective transform for both the affine subspaces $\mathtt{M}_J$ and $\mathtt{N}$ which complies with the metric constraints of the 3D shape to register and the 2D image trajectories. This results in computing a 3×3 transformation matrix $\mathtt{Q}$ which enforces the metric constraints such that $\mathtt{M}_J\mathtt{Q} = \mathtt{M}$ and $\mathtt{N}\mathtt{Q} = \mathtt{Z}$ where $\mathtt{Z}$ is a rotation matrix with $\mathtt{Z}\mathtt{Z}^T = \mathtt{I}_3$. The following problem is non-linear given the joint set of orthogonality constraints. However, a closed form solution can be computed if we consider the quadratic form $\mathtt{H} = \mathtt{Q}\mathtt{Q}^T$ and forming the orthogonality constraints as:

$$\begin{gathered} \mathbf{m}_{ui}^T \mathtt{H} \mathbf{m}_{ui} - \mathbf{m}_{vi}^T \mathtt{H} \mathbf{m}_{vi} = 0 \\ \mathbf{m}_{ui}^T \mathtt{H} \mathbf{m}_{vi} = 0 \\ \mathtt{N}\mathtt{H}\mathtt{N}^T = \mathtt{I}_3 \end{gathered}$$

where $\mathbf{m}_{ui}$ and $\mathbf{m}_{vi}$ refer to the motion components of the horizontal and vertical image coordinates respectively such that:

$$\mathtt{M}_{Ji} = \begin{bmatrix} \mathbf{m}_{ui}^T \\ \mathbf{m}_{vi}^T \end{bmatrix} \text{ where } \mathtt{M}_J = \begin{bmatrix} \mathtt{M}_{J1} \\ \vdots \\ \mathtt{M}_{JF} \end{bmatrix}. \tag{9}$$

As follows $\mathtt{H}$ is a symmetric matrix which can be computed with LS for the six unique parameters by rearranging eq. (9). Then, if $\mathtt{H}$ is positive semidefinite, the matrix $\mathtt{Q}$ is

given by $\mathtt{H} \xrightarrow{eig} \mathtt{Q} = \mathtt{U}\sqrt{\Delta}$ with $\mathtt{U}$ and Δ being the eigenvectors and eigenvalues respectively. On the contrary, if the matrix is not positive semidefinite, we estimate the closest $\mathtt{Q}$ by defining:

$$\mathtt{F} = \begin{bmatrix} \mathtt{M}_1\tilde{\mathtt{Q}} \\ \vdots \\ \mathtt{M}_F\tilde{\mathtt{Q}} \end{bmatrix} \quad \text{and} \quad \mathtt{G} = \begin{bmatrix} ((\mathtt{M}_1\tilde{\mathtt{Q}})^T)^\dagger \\ \vdots \\ ((\mathtt{M}_F\tilde{\mathtt{Q}})^T)^\dagger \end{bmatrix} \tag{10}$$

where $\tilde{\mathtt{Q}}$ is a SVD approximation of $\mathtt{Q}$ using the estimated $\mathtt{H}$ (i.e $\tilde{\mathtt{Q}} = \mathtt{UD}$ if $\mathtt{H} = \mathtt{UDV}^T$). Then the closest $\mathtt{Q}$ given the metric constraints is computed as $\mathtt{Q} = \tilde{\mathtt{Q}}\sqrt{\mathtt{F}\backslash\mathtt{G}}$ where $\backslash$ denotes the left matrix division.

Alternatively to this solution, we obtained more accurate results by solving the problem using Semi-Definite Programming (SDP). In this case we can compute explicitly $\mathtt{H}$ such that $\mathtt{H} \succeq \mathtt{0}$. First we define the cost function by separating the joint motion matrix $\mathtt{M}_J$ in its horizontal and vertical image components such that:

$$\mathtt{M}_{Ju} = \begin{bmatrix} \mathbf{m}_{u1}^T \\ \vdots \\ \mathbf{m}_{uf}^T \end{bmatrix} \quad \text{and} \quad \mathtt{M}_{Jv} = \begin{bmatrix} \mathbf{m}_{v1}^T \\ \vdots \\ \mathbf{m}_{vf}^T \end{bmatrix} \tag{11}$$

The problem is then re-formulated as the minimization of the following cost function:

$$\begin{aligned} \min_{\mathtt{H}} \quad & \Big\{ \big\|\mathrm{diag}(\mathtt{M}_u\mathtt{H}\mathtt{M}_v^T)\big\| + \big\|\mathrm{diag}(\mathtt{M}_u\mathtt{H}\mathtt{M}_u^T - \mathtt{M}_v\mathtt{H}\mathtt{M}_v^T)\big\| \\ & \qquad + \big\|\mathtt{N}\mathtt{H}\mathtt{N}^T - \mathtt{I}_3\big\| \Big\} \end{aligned} \tag{12}$$

such that

$$\begin{aligned} \mathtt{H} &\succeq \mathtt{0} \\ \mathbf{m}_{u1}^T\mathtt{H}\mathbf{m}_{u1}^T &= d \end{aligned}$$

where the last constraint $\mathbf{m}_{u1}^T\mathtt{H}\mathbf{m}_{u1}^T = d$ imposes an arbitrary value over the first frame to avoid the zero solution. This problem can be solved efficiently with current SDP toolboxes such as SeDuMi [10] since optimization is run over a small 3×3 matrix independently from the size of $\mathtt{W}$ and $\mathtt{B}$.

3.3 Registration algorithm and discussions

The full approach is finally summarized in Algorithm 2. The idea at the basis of this procedure is to obtain the best possible registration even if the 3D shape to register is not an exact description of the image data. In this sense, given the first initial 3D shape $\mathtt{B}$, we search for a common representation of the set $\{\mathtt{W}, \mathtt{B}\}$ using GSVD. This representation is then used to find the best metric solution given a joint set of metric constraints. This not only solves for the registration, but also compute a new metric shape $\hat{\mathtt{B}}$ given the contribution of both data.

Enforcing the metric constraints for both the 2D measurements and the 3D shape give robustness to degenerate motion in $\mathtt{W}$. This happens often in non-rigid motion analysis whenever a non-rigid shape is not performing enough rigid motion compared to the

Algorithm 2 Rigid registration using a joint subspace

Require: The 2D image data $\mathtt{W}$ and the 3D shape $\mathtt{B}$.

Ensure: A metric 2D–3D registration of the shape to the non-rigid image measurements.

1: Compute the image centroid of $\mathbf{t} = \frac{1}{P}\mathtt{W}\mathbf{1}_P$ and register the data as $\bar{\mathtt{W}} = \mathtt{W} - \mathbf{t}\mathbf{1}_P^T$

2: Estimate the joint affine motions $\mathtt{M}_J$ and $\mathtt{N}$ together with the joint shape $\mathtt{B}_J$ as in Section 3.1.

3: Given the affine solution, compute the best metric motion and shape as shown in section 3.2 such that:

$$\mathtt{W}_B = \mathtt{M}_J\mathtt{Q}\ \mathtt{Q}^{-1}\mathtt{S}_J = \hat{\mathtt{M}}\ \hat{\mathtt{B}} \tag{13}$$

$$\mathtt{B} = \ \mathtt{N}\mathtt{Q}\ \mathtt{Q}^{-1}\mathtt{S}_J = \hat{\mathtt{Z}}\ \hat{\mathtt{B}} \tag{14}$$

4: Project each 2×3 sub-block $\hat{\mathtt{M}}_i$ to the closest scaled orthographic matrix using eq. (4).

variations given by the deformations. In such cases, obtaining a reliable estimation of the depth of the shape is rather complex since, without rotation, it is very ambiguous to compute reliable estimates.

4 Registration with missing data

If the 2D image trajectories are interrupted due to occlusions or tracking failures, we have to additionally solve for the missing entries in $\mathtt{W}$. In such a task, the cost function to optimise is the following:

$$\min_{\mathtt{R}_i\mathtt{R}_i^T = \mathtt{I}_2} \ \left\| \mathtt{D} \odot (\mathtt{W} - \mathtt{MB}) \right\|^2 \tag{15}$$

where $\mathtt{D}$ is a $2F \times P$ mask matrix with either 1 if the 2D point is present or 0 if it is missing. Given the missing entries, it is not possible to solve for the cost function in closed form. Thus we revert to an iterative approach. Provided an initialisation of the missing entries, the approach first computes an affine solution with GSVD for $\mathtt{M}$ given $\mathtt{S}$. After a projection to the correct orthographic camera matrices, missing entries in $\mathtt{W}$ are filled given the 3D shape estimated with the joint subspaces provided by the GSVD. The algorithm stops when the updated values have minimal variations from one iteration to the other. Regarding the initialisation, best results were achieved by filling the missing entries at each trajectory with the mean value computed from the known trajectory points in $\mathtt{W}$. Note that in this case we have also to estimate the shape 2D centroid $\mathbf{t}$ at each iteration of the algorithm since it depends on the estimated missing data. The algorithm is resumed in the table for Algorithm 3.

5 Experiments

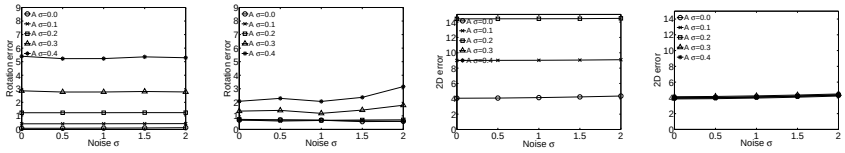
5.1 Synthetic data

The algorithm performance are evaluated with the following synthetic experimental setup. The 2D data is created from a randomly generated cloud of 20 points $\mathtt{S}_{mean}$

Algorithm 3 Rigid registration with missing data

Require: An initialisation for the 2D image data $\mathtt{W}$ and the 3D shape $\mathtt{B}$.
Ensure: A metric 2D–3D registration of the shape to the non-rigid image measurements.
1: Compute the image centroid from the current estimate of $\mathtt{W}$ as $\mathbf{t} = \frac{1}{P}\mathtt{W}\mathbf{1}_P$.
2: Given $\bar{\mathtt{W}} = \mathtt{W} - \mathbf{t}\mathbf{1}_P^T$ and $\mathtt{B}$, estimate the joint affine motions $\mathtt{M}_J$ and $\mathtt{N}$ together with the joint shape $\mathtt{B}_J$ as in Section 3.1.
3: Given the affine solution, compute the best metric motion and shape with Algorithm 2.
4: Project each 2×3 sub-block $\hat{\mathtt{M}}_i$ to the closest scaled orthographic matrix using eq. (2).
5: Given the metric solution $\hat{\mathtt{M}}$ and $\hat{\mathtt{B}}$, input the missing entries as $\bar{\mathtt{W}} = \hat{\mathtt{M}}\ \hat{\mathtt{B}}$.
6: Iterate until the update on the 2D missing data points is less then a given threshold.

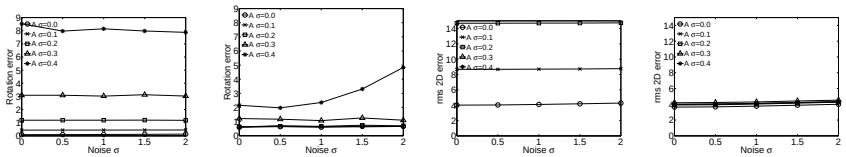
sampled inside a sphere of radius one. Deformations were constructed with a set of K random linear basis $\mathtt{S}_1 \ldots \mathtt{S}_K$. Each time-varying shape $\mathtt{X}_i$ was computed by the linear combination of random linear weights giving $\mathtt{X}_i = \mathtt{S}_{mean} + \sum_{d=1}^{K} l_{id}\mathtt{S}_d$. In order to control the deformation intensity, the *Deformation Power ratio* (DPr) is defined as: $DPr = ||f\mathtt{S}_{mean}|| \setminus ||\sum_{i=1}^{f}\sum_{d=1}^{K} l_{id}\mathtt{S}_d||$. Finally, 50 random orthographic camera matrices $\mathtt{R}_i$ and translation $\mathbf{t}_i$ are used to form the 2D measurements onto the image plane. The generation of the shape to register is made by selecting an initial random $\mathtt{X}_i = \mathtt{B}$. Then, in order to simulate distortion in $\mathtt{B}$, random affine transformation $\mathtt{A}$ are applied to $\mathtt{B}$ such that: $\tilde{\mathtt{B}} = \mathtt{AB}$. In more detail, this distortion was computed as $\mathtt{A} = \mathtt{I}_3 + \aleph$ where $\aleph$ was a 3×3 matrix of Gaussian noise with variance $\sigma_\aleph$. To conclude, zero-mean Gaussian noise with variance of $\sigma_{\mathtt{W}}$ image pixel was added to the measurements stored in $\mathtt{W}_{100\times 20}$. The 2D data was finally scaled in order to fit into a 320×240 image frame. In the following tests the root mean squared (rms) error was always used to compute the 2D registration error in pixels per point and the rotation misalignment in degrees given the known ground truth.



Algorithm 1 Algorithm 2 Algorithm 1 Algorithm 2

Fig. 3. Results for a synthetic sequence with $DPr = 0.45$. The figures show the result for the rotation error in degrees and the rms 2D error in image pixel per point.

Figure 3 shows a test result obtained by fixing $DPr = 0.15$ and after running 200 trials for each configuration of noise and affine distortion $\mathtt{A}$ (i.e. 25 configurations in total). The results show that both algorithms are relatively robust to the added image Gaussian noise however a difference is noticeable when evaluating the 2D and rotation error at increasing distortions rates for the 3D shape $\mathtt{B}$. An important fact to keep in mind when evaluating the 2D errors is that we are evaluating a registration of a rigid shape to



Algorithm 1 Algorithm 2 Algorithm 1 Algorithm 2

Fig. 4. Results for a synthetic sequence with $DPr = 0.45$. The figures show the result for the rotation error in degrees and the rms 2D error in image pixel per point.

a non-rigid sequence. Thus there is always a constant residual when plotting the error (bottom plots in Figure 3). In contrast, here we put more emphasis on the worsening of the error with the increase of the affine distortion $\mathtt{A}$. In such case, Algorithm 2 is rather robust for both 2D and rotation error due to the distortions $\aleph$ until the last level of noise where the algorithm starts to perform worse. Algorithm 1 reports a very high 2D error up to 18 pixels rms for the stronger distortion (out of the plot scale). This is expected as the shape is fixed. More interesting is the plot showing the rotational error, indicating slightly better results for tiny distortions in respect to Algorithm 2 but then diverging again up to 5 degrees for higher distortions. Figure 4 shows analogous behaviors for both algorithms but in the case of stronger deformations in the image measurements ($DPr = 0.45$). Algorithm 2 shows decreased the performance as expected but still maintains reasonable values. Differently, Algorithm 1 reaches a misalignment up to 9 degrees.

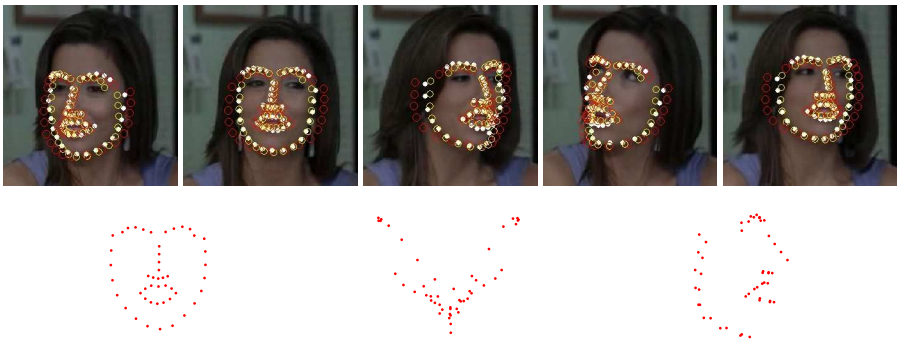


Fig. 5. Real sequence 2D–3D registration with a 3D shape as in Figure 1(b). In the top row, white dots show the 2D tracks extracted from the sequence. Red circles ○ shows the registration with Algorithm 1. Yellow circles ○ show the registration with Algorithm 2 which achieves better reprojection error especially in the eyeborow, mouth and temple areas. Bottom row shows frontal, top and side view of the joint 3D shape $\hat{\mathtt{B}}$ obtained from the registration algorithm.

5.2 Real data

The scenario here considered is the registration of a 3D face model to a set of 2D image trajectories obtained from an AAM tracker [13]. Notice that the 3D model that represents our $\mathtt{B}$ of size 3×48 was computed from a subject with different somatic traits as it was shown in Figure 1(a). The model building of $\mathtt{B}$ was performed using 2D points obtained from nearly rigid motion of the subject followed by a rigid 3D reconstruction using factorization [11]. The target 2D sequence came from a different video footage as presented in Figure 5. Results for both registration algorithms are shown in Figure 5 with the reprojected image tracks. Algorithm 2 shows its properties of adaptation by registering and computing a joint shape closer to the new subject traits. This can be noticed especially in the different eyebrow shape compared with the registration of the original $\mathtt{B}$ obtained by Algorithm 1. Finally for this test, bottom row of Figure 5 shows three views of the reconstructed joint 3D shape $\hat{\mathtt{B}}$ which qualitatively describe well the 3D shape of the subject.

A further test presents the performance of the algorithms in the case of a degenerate talking face sequence. This test is especially aimed to show the relevance of the joint metric constraints in this type of image sequences. We used the same rigid shape as the previous example and plotted the registration over the image sequence in Figure 6. Again the subjects presented different physical traits from the reference 3D model $\mathtt{B}$. Figure 6(a) shows a side and top view of the joint shape $\hat{\mathtt{B}}$ computed with the joint metric constraints as in Section 3.2. Figure 6(b) instead presents the same computation omitting the cost term $\left\|\mathtt{NHN}^T - \mathtt{I}_3\right\|$ in eq. (12). The resulting 3D shape is geometrically distorted and it is not representing the correct metric characteristics.

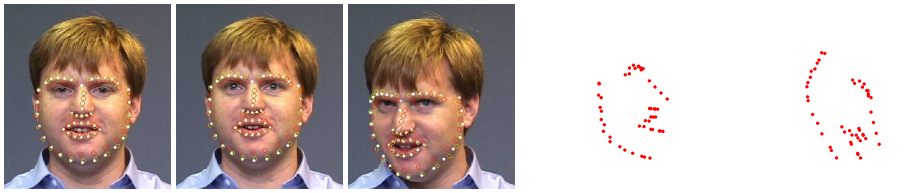


Image sequence with registration (a) Joint metric (b) 2D metric

Fig. 6. The figure shows the registration results for Algorithm 1 (red circles ○) and Algorithm 2 (yellow circles ○). The top three figures show the image sequence of a subject talking and performing minimal rigid motion. Registration is made with the 3D shape as shown in Figure 1(b). Bottom line shows first (a) two views of the shape $\hat{\mathtt{B}}$ extracted using the joint metric constraints and figure (b) the distorted shape obtained from the metric constraints of the 2D data alone.

A final experiment shows the algorithm behavior on the IMM database [9] which contains a set of 240 manually annotated face images. The dataset is divided in 6 different poses for 40 subjects. Among those six poses, 2 of them are showing non-rigid motion. Each face is manually annotated with 58 points as shown in Figure 7. A global mean 3D shape is reconstructed from all the subjects by running a rigid Tomasi and Kanade [11] factorization on the first, third and fourth pose of each subject. These

frames were showing predominant rigid motion thus they were appropriate for the task. Figure 7 shows as well three views of the 3D rigid reconstruction.

This mean shape was then registered to every image in the database using Algorithm 1 and Algorithm 2 as presented in the paper. Note that in this case we have 40 sequences for each subject composed by six frames. Figure 8 shows the results on 2 subjects. White dots show the 2D tracks manually extracted from each short sequence. Red circles ○ shows the registration with Algorithm 1. Yellow circles ○ show the registration with Algorithm 2. Again the proposed algorithm shows its adaptation capabilities when dealing with a large set of people with different somatic traits.

Fig. 7. a) A subject from the database. The white dots represent the manually annotated 2D points. b) Front, top and side views of the mean 3D shape reconstructed from the database.

Fig. 8. Four selected frames from subject ♯22 and ♯35 in the IMM database.

5.3 Evaluation with missing data

The performance of Algorithm 3 was initially tested with synthetic data as showed in the previous setup. Given the same amount of points and image frames, The affine distortion was fixed to $\sigma_{\aleph} = .20$ and $DPr = 0.25$. The evaluation included 25 tests for each configuration of missing data and noise level (225 tests in total) Experiments were made with increasing percentages of missing data and showed a robustness of the

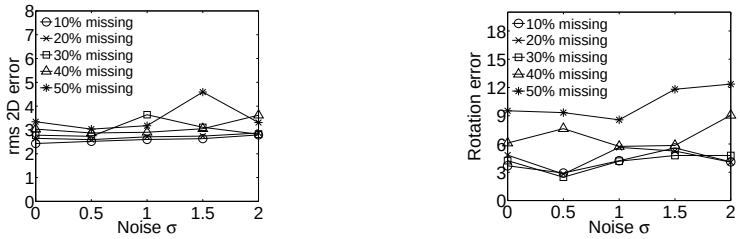


Fig. 9. Results for a synthetic sequence with $DPr = 0.25$ using Algorithm 3 and randomly generated noise and missing data.

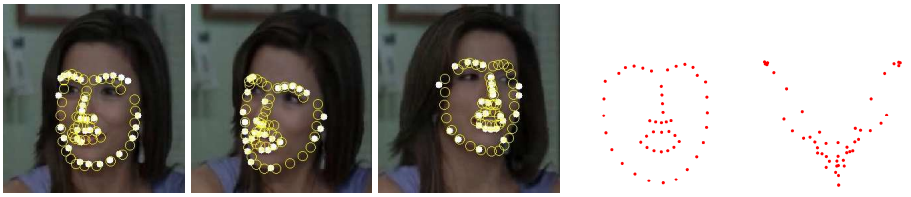


Image sequence with registration Joint 3D shape

Fig. 10. The figures on the top show the image sequence together with the registration given by Algorithm 3 with 30% of missing data. White dots show the available 2D data while the yellow circles ○ represent the estimated registration. The three images on the bottom present front, top and side view of the joint 3D shape.

approach until 40% ratio as shown in Figure 9. The maximum number of iterations was fixed to 50 and a stop criteria was fixed at 10^{-6} on the update of the reprojection error of the missing 2D points. Note here that, even if the reprojection error is minimised for the case of 50% missing data, the error in degrees is around 10 units thus we can consider the registration compromised. For higher levels of missing data, the algorithm fails to obtain a reliable registration and thus results are not presented in the plots.

The real test shown in Figure 10 presents the results on the sequence in Figure 5 where occlusions were randomly created up to a 30% ratio. The algorithm converged after 74 iterations with a threshold on the 2D points update of 10^{-6}. The registration quality is barely degraded still showing a reasonable estimate of face side and frontal profiles. We realised that most of the misalignement were present when the shape was turning on the side. It is possible to notice that now there is less symmetry in the reconstructed 3D shape with a wider gap in the side view corresponding to points lying on the upper jaw. Still most of the depth of the shape was estimated reliably.

6 Conclusions

This paper presented a new approach to the 2D–3D registration problem in the case of non-rigid 2D image trajectories and a shape represented as a set of 3D points. The method is designed for the case when the shape is not an exact description of the 2D

trajectories and it can deal with degeneracies in the 2D motion. This solution is targeted for the face analysis and medical registration scenario where often single 3D observations have to be fit to a set of 2D trajectories. The formulation, given the joint subspace may also give some intuition on how to solve the greatest crux of these methods; the matching problem between the 3D shape and the 2D image points. This will represent the starting point for future investigations together with the application of the proposed joint metric constraints to the tracking and non-rigid 3D reconstruction domains.

References

1. Bartoli, A., Gay-Bellile, V., Castellani, U., Peyras, J., Olsen, S., Sayd, P.: Coarse-to-Fine Low-Rank Structure-from-Motion. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska. pp. 1–8 (2008)
2. Brand, M.: A direct method for 3D factorization of nonrigid motion observed in 2D. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, San Diego, California. pp. 122–128 (2005)
3. Bregler, C., Hertzmann, A., Biermann, H.: Recovering non-rigid 3D shape from image streams. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head, South Carolina. pp. 690–696 (June 2000)
4. Cootes, T.F., Taylor, C.J.: Active shape models. In: Proc. British Machine Vision Conference. pp. 265–275 (1992)
5. Del Bue, A.: A factorization approach to structure from motion with shape priors. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska. pp. 1–8 (2008)
6. Hansen, P.: Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion. Society for Industrial Mathematics (1998)
7. Marques, M., Costeira, J.P.: Estimating 3D shape from degenerate sequences with missing data. Computer Vision and Image Understanding (2008)
8. Shen, D., Davatzikos, C.: An adaptive-focus deformable model using statistical and geometricinformation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8), 906–913 (2000)
9. Stegmann, M.B., Ersbøll, B.K., Larsen, R.: FAME – a flexible appearance modelling environment. IEEE Trans. on Medical Imaging 22(10), 1319–1331 (2003)
10. Sturm, J.: Using SeDuMi 1.02, A Matlab toolbox for optimization over symmetric cones. Optimization Methods and Software 11(1), 625–653 (1999)
11. Tomasi, C., Kanade, T.: Shape and motion from image streams under orthography: A factorization approach. International Journal of Computer Vision 9(2), 137–154 (1992)
12. Torresani, L., Hertzmann, A., Bregler., C.: Non-rigid structure-from-motion: Estimating shape and motion with hierarchical priors. IEEE Transactions on Pattern Analysis and Machine Intelligence pp. 878–892 (2008)
13. Xiao, J., Baker, S., Matthews, I., Kanade, T.: Real-time combined 2d+3d active appearance models. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, Washington D.C. vol. 2, pp. 535 – 542 (June 2004)
14. Xiao, J., Georgescu, B., Zhou, X., Comaniciu, D., Kanade, T.: Simultaneous Registration and Modeling of Deformable Shapes. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, New York, NY. pp. 2429–2436 (2006)