2D-3D REGISTRATION OF DEFORMABLE SHAPES WITH MANIFOLD PROJECTION

Alessio Del Bue, Marko Stosic, Marija Dodig[†], *João Xavier*[*]

Instituto Superior Técnico
Institute for Systems and Robotics
Av. Rovisco Pais 1 – 1049-001 Lisboa – Portugal

[†]Universidade de Lisboa
CELC
1649-003 Lisboa, Portugal

ABSTRACT

We present an algorithm able to register a known 3D deformable model to a set of 2D matched points extracted from a single image. Unlike previous approaches, the problem is solved simultaneously for both the rigid and non-rigid parameters of the model. The key advantage of our approach is the projection of an initial affine estimation of the motion parameters into the *motion manifold* corresponding to the exact parametrization of the problem. This projection is formulated as the minimization of the distance between the affine solution and the surface of the manifold. Such optimization results in a quadratically constrained quadratic minimization problem that can be efficiently solved with standard optimization tools. Synthetic and real tests demonstrate the effectiveness of the approach.

***Index Terms*—** Image Registration, Deformable Modelling, Quadratic Optimization, Manifold Projection.

1. INTRODUCTION

The problem of registering deformable shapes from images has stirred growing interests from the research community. Such interest rises from the practical necessity of automatically aligning shapes which vary their topology over time. This is rather important in a medical context where accurate image registration is a fundamental step for any clinical analysis. In a more theoretical sense, the complexity of this task has created an increasing attention over this problem. The loss of rigidity constraints renders the 2D-3D registration either ill-conditioned or extremely non-linear. Ill-conditioning led to the introduction of extrinsic and intrinsic priors in respect to the object pose and its deformation parameters respectively. On the other hand, non-linearities resulted in methods with an iterative nature which are prone to reach local solution given the extension of the parameter space. Other approaches proposed a simplification of the problem as a two-stage procedure [1]. First, a rigid registration of the shape is followed by an optimization of the deformable parameters. However, this procedure is likely to be biased especially in the presence of deformations which are strongly directional and asymmetrical [2].

This paper deals with the case of aligning a given 3D deformable model to a set of 2D points taken from an image. The novelty here introduced consists in finding a solution for the model parameters which exactly lie in the so-called *motion manifold* [3] of the parameter space. In such way, the estimation over this space reflects exactly the structure of the problem leading to a simultaneous estimation of the rigid and deformable motion parameters. Instead of solving the problem directly on the surface space, we instead advocate a procedure which project an initial affine solution onto the manifold surface. This not only results in an efficient algorithm but also it avoids the locality of an iterative algorithm minimizing over the manifold. In particular, we assume the camera matrices being orthographic which is a reasonable approximation when the shape profile is small compared to the distance from the camera. Such cameras lie on the $V_{2,3}$ Stiefel manifold[1]. Projecting into the space of deformable parameters given this constraint results in a non-convex problem which we show that can be efficiently solved by semidefinite relaxation. To the authors knowledge, the only approach trying to solve simultaneously for all the motion parameters is the work of Xiao et al. [2]. However their method requires several 2D observations of the non-rigid shape and a prior over the set of basis shapes modelling the deformation. The work of GU and Kanade [4] can be also related to our work with the difference that it computes a solution for both the matching and registration problems while we restrict ourself exclusively to a 2D-3D registration of a deformable model (i.e. in our case the assignments are already given). Still, when registering the matched 2D points to the 3D shape, they decouple the 2D rotation and deformation parameters solving then for the registration and matching using an iterative EM procedure.

[*]This work was supported by Fundação para a Ciência e a Tecnologia (ISR/IST pluriannual funding) through the POS_Conhecimento Program (include FEDER funds) and grant SIPM-PTDC/EEA-ACR/73749/2006, grant MODI-PTDC/EEA-ACR/72201/2006, the project ISFL-1-1431 and scholarship SFRH/BPD/26607/2006

[1]The Stiefel manifold $V_{k,m}$ may be viewed as the collection of all $m \times k$ matrices whose columns form an orthonormal set. More precisely, the (real) Stiefel manifold $V_{k,m}$ is the collection of all ordered sets of k orthonormal vectors in Euclidean space $\mathbb{R}^m$.

2. PROBLEM STATEMENT

A 2D point j is defined in non-homogeneous coordinates as a vector $\mathbf{w}_j = (u_j \ \ v_j)^T$ where u_j and v_j are the horizontal and vertical image coordinates respectively. A set of p points lying over a deformable body can be then expressed in matrix form as:

$$\mathtt{W}_{2\times p} = \left[\begin{array}{ccc} \mathbf{w}_1 & \dots & \mathbf{w}_p \end{array} \right]. \tag{1}$$

We now introduce the image projection and deformable models describing the space of variations of the image coordinates. A deformable 3D shape $\mathtt{X}_{3\times p}$ at a certain time instance is projected by an orthographic camera $\mathtt{R}_{2\times 3}$ as:

$$\mathtt{W} = \mathtt{RX} + \mathtt{T} \tag{2}$$

where $\mathtt{R}$ lies onto the Stiefel manifold (i.e. $\mathtt{RR}^T = \mathtt{I}_{2\times 2}$) which from now it will be called a *Stiefel matrix* and $\mathtt{T}$ is a $2 \times p$ matrix such that $\mathtt{T} = \mathbf{t}\mathbf{1}^T$. The vector $\mathbf{t}$ is the image centroid of the object coordinates stored in $\mathtt{W}$ and $\mathbf{1}$ a vector of p ones.

2.1. THE NON-RIGID MODEL

A deformable shape varies its topology at each time instance. Such variations can be modelled as a linear approximation using a set of k basis shapes which expresses the modes of variation of the shape. Such models have been already used with 2D points (Active Shapes Models [5]), morphable models for face analysis and animation [6] and non-rigid structure from motion [7]. In such framework, the 3D structure $\mathtt{X}$ is given as:

$$\mathtt{X} = \sum_{d=1}^{k} l_d \mathtt{B}_d, \qquad \mathtt{B}_d \in \mathbb{R}^{3\times p}, \ \ l_d \in \mathbb{R} \tag{3}$$

where $\mathtt{B}_d$ are the given basis shapes and l_d the configuration weights which linear combination gives the deformation at each time instance. A more compact form can be obtained substituting the previous equation in eq. (2) giving:

$$\mathtt{W} = \left[\begin{array}{ccc} l_1\mathtt{R} & \dots & l_k\mathtt{R} \end{array} \right] \left[\begin{array}{c} \mathtt{B}_1 \\ \vdots \\ \mathtt{B}_k \end{array} \right] + \mathtt{T} = \mathtt{MS} + \mathtt{T} \tag{4}$$

where $\mathtt{M}$ is a $2 \times 3k$ matrix and $\mathtt{S}$ a $3k \times p$ matrix. It is convenient to define here the homogeneous formulation of $\mathtt{S}$ as $\bar{\mathtt{S}} = [\mathtt{S}^T \mathbf{1}]^T$. To notice the particular structure of $\mathtt{M}$ which entails strong non-linearities given by the repetitive structure of the *Stiefel matrices*. The space of these matrices is the actual *motion manifold* where our solution has to lie.

2.2. THE NON-RIGID COST FUNCTION

The registration problem now can be written as an optimization of the following cost function:

$$\min_{\mathtt{R}, l_1 \dots l_k, \mathbf{t}} \left\{ \|\mathtt{W} - (\mathtt{MS} + \mathtt{T})\|_F^2 \right\} \tag{5}$$

given the constraints $\mathtt{RR}^T = \mathtt{I}_{2\times 2}$. Optimizing (5) results in the minimization of a non-linear cost function with non-linear constraints. However in the next section we re-formulate the problem in such a way it is solvable via quadratic optimization. To notice that in this work we consider the assignment problem between 3D model and 2D observations already given.

3. 2D-3D REGISTRATION WITH PROJECTIONS

Our algorithm can be summarized as it follows:

Algorithm 1 2D-3D registration of deformable shapes

Require: A set of 2D points $\mathtt{W}$ and the assignment to the deformable model $\mathtt{S}$.

Ensure: The motion matrix $\mathtt{M}$ as in eq. (4).

1: Compute an initial affine estimate $\bar{\mathtt{M}}$ as $\bar{\mathtt{M}} = \mathtt{W}\bar{\mathtt{S}}^T \left(\bar{\mathtt{S}}\bar{\mathtt{S}}^T\right)^{-1}$.

2: Extract the translation $\mathbf{t}$ as $\bar{\mathtt{M}} = \left[\begin{array}{c|c} \tilde{\mathtt{M}} & \mathbf{t} \end{array} \right]$.

3: Project $\tilde{\mathtt{M}}$ onto the deformable *motion manifold* by solving

$$\min_{\mathtt{R}, l_1 \dots l_k} \left\{ \left\| \tilde{\mathtt{M}} - \left[\begin{array}{ccc} l_1\mathtt{R} & \dots & l_k\mathtt{R} \end{array} \right] \right\|_F^2 \right\} \tag{6}$$

Now the problem is to find the best projection given $\mathtt{R}$ lying onto a Stiefel manifold as given in eq. (6). This problem was previously solved using a two-step procedure (i.e. first estimating $\mathtt{R}$ and then l_i [1]) or by considering some priors on the parameter space [2]. Differently, we will show that is possible to reformulate the problem in Step 3 in such way we can project to the correct *motion manifold* of the parameter space.

3.1. MOTION MANIFOLD PROJECTION

The projection is carried out on each $2 \times 3k$ sub-matrix $\tilde{\mathtt{M}}$ as defined in Algorithm 1 and it corresponds to solve the minimization problem as stated in eq. (6). Our aim is to project the affine solution obtained previously to a motion subspace which exactly respect the given constraints (i.e $\mathtt{R}_i$ be a 2×3 *Stiefel matrix*). This is equivalent to minimizing separately all the 2×3 blocks of $\tilde{\mathtt{M}}$ giving:

$$\min_{\mathtt{R}} \sum_{d=1}^{k} \min_{l_d} \left\|\tilde{\mathtt{M}}_d - l_d\mathtt{R}\right\|_F^2 \tag{7}$$

which is equivalent to:

$$\min_{\mathtt{R}} \sum_{d=1}^{k} \min_{l_d} \left\{ \left\|\tilde{\mathtt{M}}_d\right\|^2 + l_d^2 \left\|\mathtt{R}\right\|^2 - 2 l_d \operatorname{Tr}[\tilde{\mathtt{M}}_d^T \mathtt{R}] \right\}. \tag{8}$$

We can then reformulate the problem by computing the minimum first for l_d given $\mathtt{R}$. This resolves in computing the minimum of the quadratic function in l_d given by $f(l_d) = a\, l_d^2 - 2\, b\, l_d + c$. Such minimum is found in $l_d = b/a$ giving:

$$l_d = \frac{\mathrm{Tr}[\tilde{\mathtt{M}}_d^T \mathtt{R}]}{\|\mathtt{R}\|^2} = \frac{1}{2}\,\mathrm{Tr}[\tilde{\mathtt{M}}_d^T \mathtt{R}]$$

putting this value back in eq. (7) and following with the simplification, the minimization can be written as:

$$\min_{\mathbf{r}=vec(\mathtt{R}^T)} \quad \mathbf{r}^T \left[-\sum_{d=1}^{k} \mathbf{m}_d \mathbf{m}_d^T \right] \mathbf{r} = \min_{\mathbf{r}=vec(\mathtt{R}^T)} \quad \mathbf{r}^T\, \mathtt{E}\, \mathbf{r} \quad (9)$$

where $\mathbf{r} = vec(\mathtt{R}^T)$ and $\mathbf{m}_d = vec(\tilde{\mathtt{M}}_d^T)$. The matrix $\mathtt{R}$ must satisfy the constraint of being a *Stiefel matrix* (i.e. $\mathtt{R}\mathtt{R}^T = \mathtt{I}_{2\times 2}$). This quadratic minimization problem presents non-convex constraints given by $\mathtt{R}$, however, it is possible to obtain a tight convex relaxation. We rewrite (9) as

$$\min_{\mathbf{r}=vec(\mathtt{R}^T)} \mathrm{Tr}(\mathtt{E}\mathbf{r}\mathbf{r}^T) = \min_{\mathtt{X}\in S} \mathrm{Tr}(\mathtt{E}\mathtt{X}), \qquad (10)$$

where S is the set of all real symmetric 6×6 matrices $\mathtt{X} = \begin{bmatrix} \mathtt{A} & \mathtt{B} \\ \mathtt{B}^T & \mathtt{C} \end{bmatrix}$, with $\mathtt{A} \in \mathbb{R}^{3\times 3}$, satisfying

$$\mathtt{X} \succcurlyeq 0, \qquad (11)$$

$$\mathrm{Tr}(\mathtt{A}) = \mathrm{Tr}(\mathtt{C}) = 1, \quad \mathrm{Tr}(\mathtt{B}) = 0, \qquad (12)$$

$$\mathrm{rank}\, \mathtt{X} = 1. \qquad (13)$$

This quadratic problem, has a nonconvex constraint ($\mathrm{rank}\, \mathtt{X} = 1$). So, we do a convex relaxation of this problem. Since the cost function is linear we have

$$\min_{\mathtt{X}\in S} \mathrm{Tr}(\mathtt{E}\mathtt{X}) = \min_{\mathtt{X}\in \mathrm{co}(S)} \mathrm{Tr}(\mathtt{E}\mathtt{X}), \qquad (14)$$

where $\mathrm{co}(S)$ is the convex hull of the set S.

Here, we approximate the convex hull $\mathrm{co}(S)$ by the set of real symmetric 6×6 matrices $\mathtt{X}$ that satisfy

$$\mathtt{X} \succcurlyeq 0, \qquad (15)$$

$$\mathrm{Tr}(\mathtt{A}) = \mathrm{Tr}(\mathtt{C}) = 1, \quad \mathrm{Tr}(\mathtt{B}) = 0, \qquad (16)$$

$$\begin{bmatrix} \mathtt{I}_{3\times 3} - \mathtt{A} - \mathtt{C} & \mathbf{w} \\ \mathbf{w}^T & 1 \end{bmatrix} \succcurlyeq 0, \qquad (17)$$

with $\mathbf{w}$ given by

$$\mathbf{w} = \begin{bmatrix} b_{23} - b_{32} \\ b_{31} - b_{13} \\ b_{12} - b_{21} \end{bmatrix} \qquad (18)$$

where $\mathtt{B} = [b_{ij}]$. Moreover, this set is defined only by linear matrix inequalities (LMI).

Hence, we have that our problem (9) is equivalent to finding the minimum of a linear function ($\mathrm{Tr}(\mathtt{E}\mathtt{X})$) on a convex set ($\mathrm{co}(S)$), which is given only by LMI (15)-(17). By using SeDuMi [8], we quickly obtain the optimal matrix $\mathtt{X}$ for (14) with $\mathrm{rank}\, \mathtt{X} = 1$. By factorizing $\mathtt{X} = \mathbf{r}\mathbf{r}^T$, we obtain the optimal *Stiefel matrix* as $\mathtt{R}^T = vec^{-1}(\mathbf{r})$. For more details and proofs about the tight convex relaxation proposed the reader can check [9]. The computed *Stiefel matrix* $\mathtt{R}$ is then used to recover the weights l_d from eq. (8), obtaining a non-rigid motion matrix that satisfies the metric constraints. This allows to solve for the motion parameters as described in Algorithm 1.

	Noise %					
	0	10	20	30	40	50
$\mathtt{W}_{err}$	0	0.00332	0.00334	0.00336	0.00334	0.00335
$\mathtt{S}_{err}$	0	0.100	0.114	0.114	0.114	0.115

Table 1. The table shows the rms error for the 2D error $\mathtt{W}_{err}$ and the 3D reconstruction error $\mathtt{S}_{err}$ computed on 3500 random trials for each level of gaussian noise. The 2D image points are always constrained in a square with a side 2 units long. Instead, the 3D error $\mathtt{S}_{err}$ is always shown as a percentages relative to the overall 3D shape size.

4. EXPERIMENTS

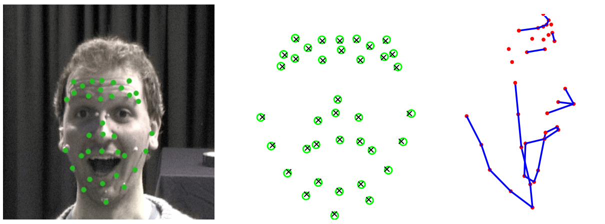
We evaluate our algorithm with synthetic experiments to test the converge of the algorithm and to evaluate its resilience to noise. Real tests will demonstrate the algorithm capabilities in dealing with realistic deformations.

4.1. SYNTHETIC TEST

A 3D model $\mathtt{S}$ is obtained by randomly generating each basis shapes $\mathtt{B}_1$, ..., $\mathtt{B}_k$. The orthographic camera matrices are obtained by generating random rotations and then by truncating the last row of the matrix. The configuration weights l_d are generated more carefully since we require that $\mathtt{B}_1$ is a mean 3D shape of the deforming object (as it happens naturally). Thus we constraints the l_d such that $\|l_1\mathtt{B}_1\| > \sum_{d=2}^{k} \|l_d\mathtt{B}_d\|$. Then we add a random translation $\mathbf{t}$ to the projected points and we form $\mathtt{W}$. Finally, a $\mathtt{N}_{2\times p}$ Gaussian noise matrix is added to the measurement with a noise strength represented as $\|\mathtt{N}\| \,/\, \|\mathtt{W}\|$. The experiments show a peculiar resilence to noise leading to rather similar results at each level. The results with no noise always converged to the optimum.

4.2. REAL TEST

We tested our algorithm in a face registration scenario. In the first test, we used a PCA model of a face computed from 3D ground truth data obtained from a VICON motion capture system. The subject in this learning stage was performing different facial expressions and slight head rotations. The PCA model consisted of $k = 5$ basis shapes and 37 points (i.e. $\mathtt{S}_{15\times 37}$). We then selected an image (Figure 1a) from a video of the same subject and tried to register the points after



a) b) c)

Fig. 1. Image a) shows the selected points in the given image frame. Image b) shows the selected point positions (green circles) and the estimated 2D position (black crosses). Image c) shows the registered 3D shape to the image using the given PCA model of the subject.

a manual assignment of the correspondences computed from the image. Figure 1b shows the results for the alignment in the image plane and Figure 1c the given 3D shape obtained from the estimated configuration weights as stated in eq. (3).

A second test[2] was made using a face model $\mathtt{S}_{12\times 74}$ obtained from an image sequence where a subject was tracked using an Active Appearance Model (AAM) tracker [10]. A non-rigid structure from motion algorithm [11] was then used to extract the basis shapes from the 2D observations. Such model was then used to register the 2D points of a different subject as shown in Fig. 2(b). In this case the model was not exactly matching the target anatomical features, however it was still possible to compute sensible depth estimate after running the registration algorithm. In such sense a sensible registration may be obtained even if the model does no perfectly match the given 2D data. To conclude, the whole algorithm was implemented in MATLAB and it requires few seconds in order to perform the minimization since it is independent by the number of points and basis shapes in $\mathtt{S}$.

a) b) c)

Fig. 2. The first 2 figures in image a) show two samples of the 75 frames long sequence used to generate the 3D model. The last image of a) shows the sample used for the registration. Image b) shows the selected point positions (green circles) and the estimated 2D position (black crosses). Image c) shows the registered 3D shape to the image.

5. CONCLUSION

We have presented a method to register 3D deformable models expressed as a combination of basis shapes to a set of 2D points. This algorithm solves the problem by considering all the rigid and non-rigid parameters involved in the registration at once. In such case, the proposed optimal projection to the *motion manifolds* of non-rigid shapes has shown promising results. To notice that such projections can be successfully used for the task of reconstructing 3D deformable models from a set of images as shown in [12]. As future work we are planning to tackle the problem of the point assignment given the mathematical insights provided by this work and to extend the optimization to different types of camera models.

[2] Thanks to J. Xiao for providing the sequences and the 2D image tracks

6. REFERENCES

[1] T. Cootes, “An Introduction to Active Shape Models,” *Image Processing and Analysis*, pp. 223–248, 2000.

[2] J. Xiao, B. Georgescu, X. Zhou, D. Comaniciu, and T. Kanade, “Simultaneous Registration and Modeling of Deformable Shapes,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition, New York, NY*, 2006, pp. 2429–2436.

[3] M. Marques and J. P. Costeira, “Estimating 3D shape from degenerate sequences with missing data,” *Computer Vision and Image Understanding*, 2008.

[4] L. Gu and T. Kanade, “A Generative Shape Regularization Model for Robust Face Alignment,” in *Proceedings of the 10th European Conference on Computer Vision*, 2008, vol. 1, pp. 413–426.

[5] T. F. Cootes and C. J. Taylor, “Active shape models – smart snakes,” in *Proc. British Machine Vision Conference*, 1992, pp. 266–275.

[6] V. Blanz and T. Vetter, “A morphable model for the synthesis of 3d faces,” in *SIGGRAPH ’99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, 1999, pp. 187–194.

[7] C. Bregler, A. Hertzmann, and H. Biermann, “Recovering non-rigid 3D shape from image streams,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head, South Carolina*, June 2000, pp. 690–696.

[8] J.F. Sturm, “Using SeDuMi 1.02, A Matlab toolbox for optimization over symmetric cones,” *Optimization Methods and Software*, vol. 11, no. 1, pp. 625–653, 1999.

[9] Marija Dodig, Marko Stošić, and João Xavier, “On minimizing a quadratic function on stiefel manifolds,” Tech. Rep., Instituto de Sistemas e Robotica, 2009, Available at www.isr.ist.utl.pt/˜jxavier/ctech.pdf.

[10] J. Xiao, S. Baker, I. Matthews, , and T. Kanade, “Real-time combined 2d+3d active appearance models,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition, New York, NY*, June 2004, vol. 2, pp. 535 – 542.

[11] A. Del Bue, F. Smeraldi, and L. Agapito, “Non-rigid structure from motion using ranklet–based tracking and non-linear optimization,” *Image and Vision Computing*, vol. 25, no. 3, pp. 297–310, March 2007.

[12] M. Paladini, A. Del Bue, M. Stošić, M. Dodig, J. Xavier, and L. Agapito, “Factorization for non-rigid and articulated structure using metric projections,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Miami, Florida, USA, June 2009.