HAL
open science

Non-rigid Shape Matching Using Geometry and Photometry

Nicolas Thorstensen, Renaud Keriven

To cite this version:

Nicolas Thorstensen, Renaud Keriven. Non-rigid Shape Matching Using Geometry and Photometry. ACCV 2009 9th Asian conference on Computer Vision, Sep 2009, Xi'an, China. pp.644-654. hal-00711070

HAL Id: hal-00711070

https://enpc.hal.science/hal-00711070

Submitted on 2 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Non-rigid Shape matching using Geometry and Photometry

Nicolas Thorstensen[1] and Renaud Keriven[1]

Universite Paris-Est, Ecole des Ponts ParisTech, IMAGINE
{thorstensen,keriven}@imagine.enpc.fr
http://imagine.enpc.fr

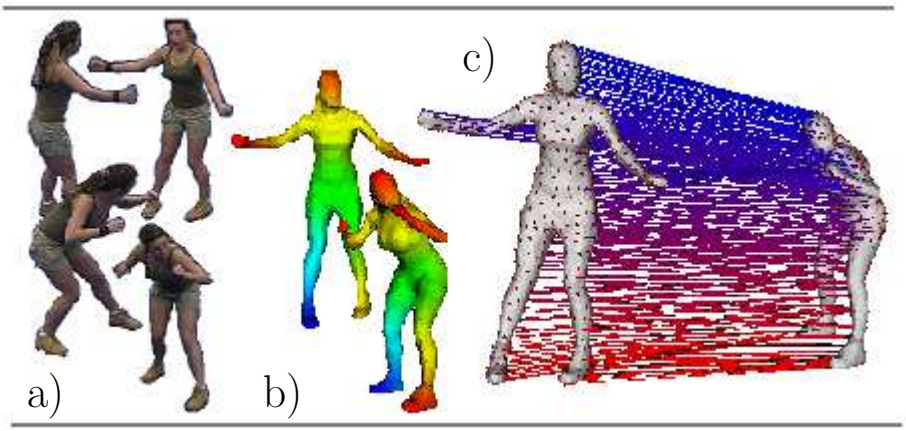

Fig. 1. a) Multiple views of a person in two different poses; b) Mesh based representation of the reconstructed object and heat map of the geodesic distance to the lower left foot; c) Correspondence result after optimization

Abstract. In this paper, we tackle the problem of finding correspondences between three-dimensional reconstructions of a deformable surface at different time steps. We suppose that (i) the mechanical underlying model imposes time-constant geodesic distances between points on the surface; and that (ii) images of the real surface are available. This is for instance the case in spatio-temporal shape from videos (e.g. multi-view stereo, visual hulls, etc.) when the surface is supposed approximatively unstretchable. These assumptions allow to exploit both geometry and photometry. In particular we propose an energy based formulation of the problem, extending the work of Bronstein *et al.* [1]. On the one hand, we show that photometry (i) improves accuracy in case of locally elastic deformations or noisy surfaces and (ii) allows to still find the right solution when [1] fails because of ambiguities (e.g. symmetries). On the other hand, using geometry makes it possible to match shapes that have

undergone large motion, which is not possible with usual photometric methods. Numerical experiments prove the efficiency of our method on synthetic and real data.

1 Introduction

Most of the objects observed in the real world are non-rigid. This makes them particularly important for computer vision. What makes non-rigid shapes challenging is that their associated deformations exhibit a potentially infinite number of degrees of freedom. As a consequence they are hard to analyze. One typical example is the three-dimensional (3D) reconstruction of a person in a multiple cameras environment. More generally, matching 3D-reconstructed shapes have numerous applications, among which are reconstruction-based animation, motion analysis, shape recognition, physical phenomena analysis, etc.

A subject observed at different times usually results in meshes in different poses. In such a situation, image-based matching algorithms [2, 3] tend to fail, as the motion in between the two poses is too large. Assuming the object surface is approximately unstretchable, Bronstein *et al.* [1] designed a matching framework where geodesic distances between automatically chosen key points are preserved. Yet, such a criterion fails in case of ambiguities like symmetries. It also yields inaccuracy when the unstretchability hypothesis is violated.

1.1 Our contributions

Therefore we propose to take the best of both worlds and design an extension of the work of [1] by adding a photometric term to their energy. We obtain a robust multi-resolution 3D surface mapping procedure that combines photometric and geometric information(c.f. Figure 1). We experiment it for non-rigid surface correspondence between two surfaces observed a different time steps in a multiple cameras environment and demonstrate its superiority.

The paper is organized as follows. Section 2 reviews previous work for non-rigid shape matching. Section 3 states the problem formulation and Section 4 explores the algorithmic implementation. Finally, numerical experiments on real data are reported in Section 5 and Section 6 concludes.

2 Related Work

The correspondence problem is one of the fundamental challenges in computer vision. Might it be in the context of optical flow, calibration or surface registration. For rigid surface and point cloud registration, Iterative Closest Point (ICP) and its variants [4] are the standard algorithms. Operating on a purely geometric level, they rely on approximated differential quantities, e.g. curvature, or more robust surface descriptors [5]. Recently, several algorithms also address the problem of non rigid surface registration. They can be mainly divided into two categories: geometric and photometric. Whereas geometric methods assume the geometry known, photometric methods estimate structure and motion.

2.1 Geometric methods

A non-rigid counterpart to ICP was introduced in [6]. Further in [7], the authors propose a non rigid registration method by piecewise rigid registration and local thin plate spline alignment of range scan images. Elad *et al.* [8] take an intrinsic point of view of the matching problem. By isometrically embedding two surfaces into a common Euclidean space the problem is reduced to the simpler of rigid matching between surfaces. Following this idea, Bronstein *et al.* [9] take it one step further and solve the correspondence problem by embedding the surface directly into another. The major drawback of this approach is the difficulty to handle ambiguities like shape symmetries. As the method solely relies on the distance function of the surface, the algorithm tends to wrongly match pairs of symmetric points (see Figure 2).

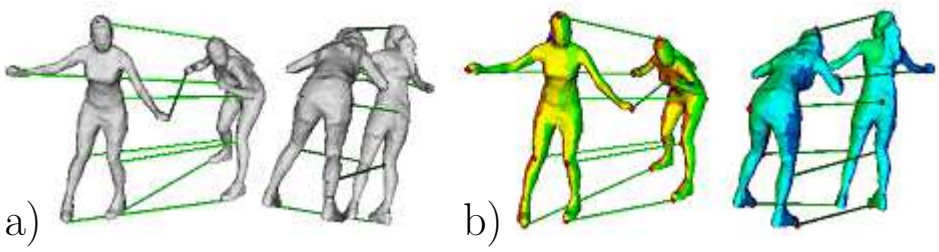

Fig. 2. a) Initialization using solely geometric information is sensible to symmetry in shapes. b) Color coded labeling of the meshes (see text) and our photometric plus geometric initialization, where ambiguities have been solved.

2.2 Photometric methods

On the photometric side, non rigid registration is commonly known under the name of scene flow. Vedula *et al.* [2] were the first to introduce the concept of scene flow which is the 3D extension of 2D optical flow. It is the simultaneous reconstruction of structure and motion between time frames. Further work on scene flow was undertaken by other teams [10, 11, 3]. As pointed out by Starck *et al.* [12], such approaches are limited to small displacements.

2.3 Where geometry meets photometry

Lastly, one can take advantages of both approaches. Several recent papers address this problem. In [13], Starck *et al.* provide a method to register shapes from silhouette. The method embeds the two surfaces into a common spherical domain. Using the analytic expression of geodesic distances on the sphere allows to minimize the geometric distance between correspondences and distance between the associated color feature. This work is the most related to ours in

the sense that they perform an isometric mapping into a common metric space and use geodesic distances as regularization for the appearance minimization. In order to be robust with respect to topological changes, hey then developed a matching algorithm based on a Markov random field optimization [12]. A much more heuristic approach is used by Ahmed *et al.* [14]. A set of SIFT features is computed which are then used for initialization of a refinement model. Lastly, the work of Varanasi *et al.* [15] makes also use of SIFT. The 3D position of the interest points are used to compute a sparse estimate of the motion field. By applying a diffusion operator on the sparse motion field, they recover a dense motion field.

Most of the algorithms presented so far rely on heuristics to put two temporal meshes into correspondence. In this work we want to consider a more rigorous model and propose a variational refinement method using ideas from variational stereo [16] and geometric embedding methods [8, 9]. In this way, we recover temporal correspondences between meshes at different resolution and independently of the combinatorial configuration. Further, our method can handle local elastic deformations such that the correspondence is consistent with the observed images.

3 Variational matching

Let S^1 and S^2 be the two surfaces to match. Each of them is observed by a certain number of cameras. Although not required, we simplify notations and suppose that the number and positions of the cameras are constant, so that both surfaces are observed by n cameras defined by their respective projections $\Pi_1, \dots \Pi_n$. We denote by I_i^k the ith image of surface k.

Following [9], we use a Lagrangian point of view where a set of m correspondences are constrained to move on the surface such that they minimize a given energy. We denote (P_i^1, P_i^2) such a correspondence where $P_i^k \in S^k$ and $\Theta = \{(P_i^1, P_i^2), 1 \leq i \leq m\}$ the set of all correspondences. m is a fixed integer that can be estimated during initialization (see Section 4.2). Our energy is a function of parameter Θ that writes:

$$E_{tot}(\Theta) = \alpha E_{geom}(\Theta) + \beta E_{photo}(\Theta) \tag{1}$$

The first term E_{geom} is the geometric part, taken from the work of Bronstein *et al.* [1] while E_{photo} is our photometric attachment. As usual, α and β are positive constants that control the relative weights among these terms. Minimizing energy (1) with respect to Θ will position the correspondences on the mesh such that their projections in the images minimize a photometric dissimilarity measure while geodesic distances on the surfaces are respected.

3.1 Geometry

Bronstein *et al.* [9] propose to embed near-isometric surfaces one into another by minimizing the following energy:

$$E_{geom}(\Theta) = \sum_{i>j} (d_{S^1}(P_i^1, P_j^1) - d_{S^2}(P_i^2, P_j^2))^2. \tag{2}$$

where d_{S^k} is the geodesic distance on surface S^k. Again, this energy suffers mainly from two weaknesses: (i) symmetries yield ambiguities and (ii) if the object undergoes locally elastic deformations between shape S^1 and shape S^2, geodesic distances are not exactly preserved. Note that this is also the case when the surface are 3D reconstructions, since they are unavoidably noisy.

3.2 Photometry

For our image matching term E_{photo}, we chose the normalized cross-correlation to measure similarity between corresponding points. Its simplicity, robustness in the presence of varying lighting conditions and differentiability make it a common choice in variational methods. Each surface point is generally seen from several cameras and one might be tempted to correlate multiple pairs of images. However, in our experiments, the number of cameras is relatively small. Thus, using information from only one pair of camera for each surface point reveals to be enough.

As a first step, for each surface S^k, we associate to each point $M \in S^k$ an optimal image $I^k_{l^k(M)}$. Choice of labels l^k might be guided by different criteria. Here, we compute partitions of the surfaces following [17]. This method assigns smoothly each point to a label corresponding to the camera from which it is best viewed. Using graph-cut optimization, the labeling is obtained by minimizing a weighted sum of two terms which realizes a good trade off between resolution and color continuity, while respecting occlusions (c.f. Figure 2b).

Let $l^k(i)$ be a short notation for $l^k(P_i^k)$, our photometric energy then writes:

$$E_{photo}(\Theta) = \sum_{i=1}^{m} g[NCC(I^1_{l^1(i)} \circ \Pi_{l^1(i)}, I^2_{l^2(i)} \circ \Pi_{l^2(i)})(P_i^1, P_i^2)] \tag{3}$$

where g is a positive decreasing function and $NCC(f^1, f^2)(M^1, M^2)$ denote the normalized cross-correlation of functions f^1 and f^2 between two related neighborhoods of points M^1 and M^2.

Following the stereovision work of Keriven *et al.* [16], we approximate locally the surfaces by their tangent planes at points M^k. In their case, only one surface is considered and M_1 and M_2 are the same point, with the same tangent plane, thus the same neighborhood. Their correlation boils down to correlating image regions related by a homography. In our case, we suppose that the tangent plane to S^1 at point M^1 and the tangent plane to S^2 at M^2 are related by a given two-dimensional isometry $\mathcal{I}_{M^1,M^2}$ sending M^1 to M^2. Under this assumption,

neighborhoods on the respective tangent planes are related and the correlation $NCC(f^1, f^2)(M^1, M^2)$ is correctly defined. Moreover, it (and its derivatives) remains easy to compute since corresponding image regions are still related by a homography. Introducing the isometry $\mathcal{I}_{M^1,M^2}$ are each point pair (M^1, M^2)

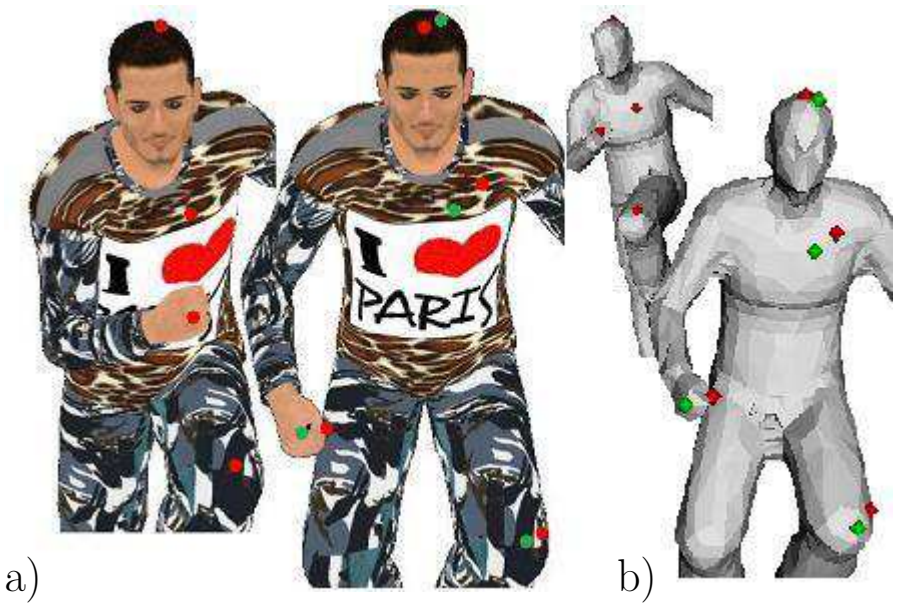

Fig. 3. a) On the left, the front image of the first mesh, red dots being the projections of some key points. On the right, the front image of the second mesh, showing the reprojections of the corresponding points, in red obtained with the method in [9] (initialized with our method), in green obtained with our method. b) Shows the same correspondences directly on the meshes with the same color code. The inaccuracies of [9], here due to local elastic deformations, are corrected by our photometric + geometric criterion.

might be thought as problematic since one would require to match the surfaces to know it, yielding a chicken and egg problem. Practically, this is not the case. As usual, we will minimize the energy by mean of a gradient descent starting from a coarse initialization (see Section 4). This approximate solution reveals to be sufficient to obtain a robust $\mathcal{I}_{M^1,M^2}$. We proceed the following way: (i) each correspondence point P_i^k defines a geodesic distance map $d_{S^k}(P_i^k, .)$ on S^k; (ii) the gradients of these distance maps at a given point M^k defines local directions that should correspond from one surface to the other if M^1 corresponds to M^2; (iii) as a consequence, the best[1] isometry from the tangent plane at M^1 to the one at M^2 that sends both M^1 to M^2, and the distance gradients directions at M^1 to the ones at M^2, is a good estimate of $\mathcal{I}_{M^1,M^2}$. Please note that computing

[1] in the least squares sense

distance maps is no extra cost since it is part of E_{geom}. Note also that the isometries $\mathcal{I}$ are actually needed only for the pairs (P_i^1, P_i^2) and that they will be refined during the gradient descent iterations as the pairs moves.

4 Optimization

4.1 Discretization

We suppose that the surfaces are both discretized as collections of triangles. Following [9], points P_i^k are taken as barycenters of triangle vertices. Θ consists in a choice of triangles and corresponding barycentric coordinates. The geodesic distances between all vertices of the mesh are computed using the Fast Marching algorithm for triangular surfaces [18, 19]. Geodesic distance is then interpolated like in [9] (note that the problem is not only to interpolate the distance at some barycenter, but also the distance **to** some barycenter).

For the photometric part of the energy, discretization is not a particular issue: the labeling method [17] is designed for triangle meshes, and we use a standard normal interpolation method to estimate the tangent planes.

Minimizing the energy with respect to both the P_i^1's and the P_i^2's is obviously not well posed. Although different cases might happen, in our experiments we have no further constraint on the choice of the points to be matched. Thus, we fix points P_i^1 to their initial position (see Section 4.2) and minimize the energy with respect to the positions of points P_i^2.

As written above, we use a classical gradient descent. Properly minimizing it is not trivial because the problem is non-convex. In order to cope with local minima, we apply a multi resolution strategy, considering the problem at several scales. Once a solution is found at a coarse scale, it is used to initialize the problem at a finer scale. Our problem has two scalable dimensions. The first one is the number of correspondences and the second is the scale of the images. This leads to a two step multi resolution scheme. Starting with a small number of correspondences, we iteratively increase the number of points by interpolating the solution from the coarser level to the next finer level. This scheme is adapted from [20]. Then, at each level, we perform a gradient descent in a multi scale manner using a Gaussian pyramid of the images.

4.2 Initialization

We first have to initialize the correspondences. Copying [20], we take advantage of the geodesic distance maps and use the farthest point sampling (FPS) strategy [21] to get geometry-based feature points on the surfaces. For near isometric surfaces we can expect the sampling to be almost identically distributed on both surfaces S^1 and S^2 [9]. Taking photometry into account to avoid geometric ambiguities, we then reject points that have an autocorrelation score below a given threshold, thus corresponding to non textured regions. As in [20], points are then associated using branch and bound optimization [22], yielding m initial

pairs. Here, to the initial geodesic distance based criterion, we add a photometric one in order to get rid of geometric ambiguities. Because no correlation is possible (at this stage, tangent planes cannot yet be related by isometries), we use SIFT descriptor based similarity, being thus invariant to scale and orientation. The results of the initialization of the correspondences can be viewed in Figure 2. Note how geometric ambiguities are solved.

4.3 Gradient descent

Optimization is performed at all scales until convergence is reached, i.e. norm of the gradient is below a given threshold. The expressions of the gradients of the geometric and the photometric parts of our energy can be found in [20] and [16] respectively. Remember that Θ consists in these coordinate but also in the choice of the triangles to which the barycenters are related. As in [9], the gradients are computed for a fixed such choice. However, when a point P_i^k gets out of its related triangle, we force it to stop at the reached edge and assign it to the triangle at the "other side" of this edge. Doing this way, points travel gently from one triangle to another if needed.

5 Results

In order to validate the proposed method, we run several experiments on real and synthetic data. First, we test it on a synthetic dataset. In a second experiment, we validate our algorithm on real images.

5.1 Validation

Our first experiment focuses on the validation of our energy by testing the algorithm on a synthetic dataset (8 cameras) and comparing to the result of [9]. This experiment aims at justifying the photometric part of the energy. The parameter α is set to 1.0 whereas β equals 8.0. We take 12 pairs of correspondences and a 3 level Gaussian pyramid. The advantages of our initialization having already be demonstrated on Figure 2, we rather launch the original method proposed by Bronstein *et al.* [9] with our initialization. The red dots in the left image of Figure 3a are some of the P_i^1 projected on the front image of S^1. The red dots in the right image of Figure 3a are the projections of the corresponding P_i^2 obtained after running the optimization of [9]. The green dots correspond to the result obtained with our combined photometric-geometric optimization. One can clearly see, the green dots are consistent with the initial sampling in the left image although the zone around the knee and shoulder exhibit elastic deformation. Whereas the red dots in the right image ignore the image signal and are pushed away by the local elastic deformations. Figure 3b shows the same points on the meshes.

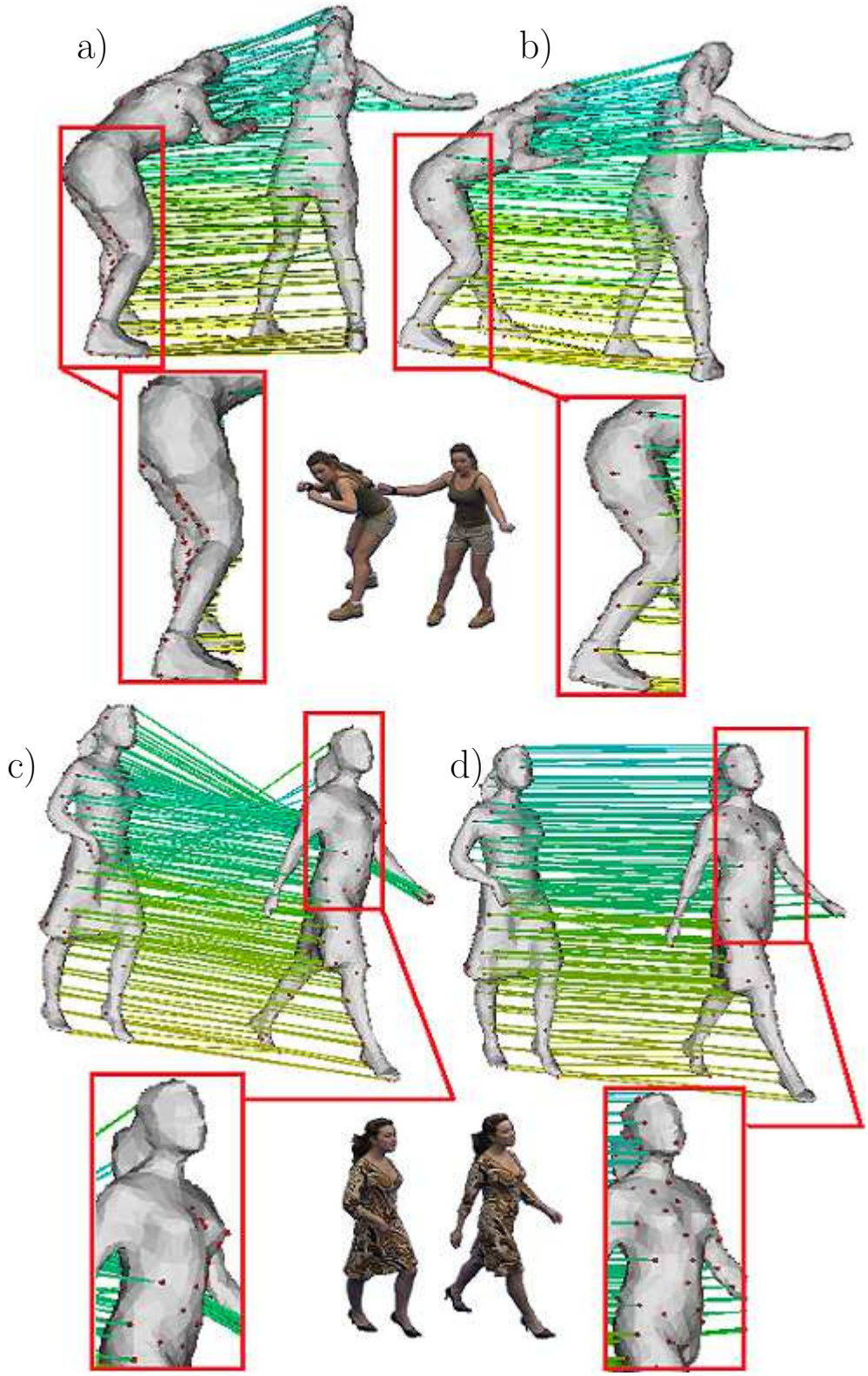

Fig. 4. Results on a real-worl dataset[23]. a) and c) show the results obtained using the method in [9]. b) and d) are obtained using our method.

5.2 Real data

In order to see how our method performs, we run several experiments on image data courtesy of J. Starck[2] [23], again with 8 cameras. In this experiment the number of correspondences is 150 and the number of image levels is set to 3. α and β are set to 0.9 and 1.5 respectively. The results are depicted in Figure 4. Notice how the method of [9] fails to solve the matching problem. Local elastic deformations are observed in both cases (Figure 4a and 4c) and wrong matches occur because of symmetry. Nevertheless, our method can handle the symmetries and local elastic deformations as can be noticed the around the hair and the back in Figure 4b and in the zones located on the skirt and the hair in Figure 4d.

6 Conclusion

We have described a variational framework for the correspondence problem of non-rigid surfaces coming from multi-view reconstruction or from any setup that provides images of the scene. By using photometric and geometric information, our method improves the one by Bronstein *et al.* [9] and allows to cope with elastic stretches and symmetries of the shape. Like in the initial work, and contrary to usual photometric methods, we are not limited to small deformations. Numerical experiments show the success of our method. Our future work includes extending our framework to surfaces with varying topology. A first step in this direction could be the use of more robust embedding such as the one proposed in [24]. Another improvement is the use of second order information in the gradient descent in order to perform Newton steps. Finally, we notice that regularization is important in low-textured image regions. Therefore we currently investigate the effect of automatically setting the control weights as proposed in [25] in another context.

Acknowledgement We would like to thank A.M. and M.M. Bronstein for their kind help and for providing their code.

References

1. Bronstein, A., Bronstein, M., Kimmel, R.: Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. Proc. National Academy of Sciences (PNAS) **103**(5) (January 2006) 1168–1172
2. Vedula, S.and Baker, S., Rander, P., Collins, R., Kanade, T.: Three-dimensional scene flow (2005)
3. Pons, J.P., Keriven, R., Faugeras, O.: Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. The International Journal of Computer Vision **72**(2) (Apr 2007) 179–193

[2] `http://personal.ee.surrey.ac.uk/Personal/J.Starck/`

4. Gelfand, N., Mitra, N.J., Guibas, L.J., Pottmann, H.: Robust global registration. In: In M. Desbrun and H. Pottmann, editors, Eurographics Association, ISBN 3-905673-24-X. (2005) 197–206
5. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (2002) 509–522
6. Amberg, B., Romdhani, S., Vetter, T.: Optimal step nonrigid icp algorithms for surface registration. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2007)
7. Brown, B., Rusinkiewicz, S.: Global non-rigid alignment of 3-D scans. ACM Transactions on Graphics (Proc. SIGGRAPH) **26**(3) (August 2007)
8. Elad, A., Kimmel, R.: On bending invariant signatures for surfaces. In: IEEE Trans. on Pattern Analysis and Machine Intell. Volume 25 of LNCS. (October 2003) 12851295
9. Bronstein, A., Bronstein, M., Kimmel, R.: Efficient computation of isometry-invariant distances between surfaces. In: SIAM J. Sci. Comput. Volume 28. (2006)
10. Furukawa, Y., Ponce, J.: Dense 3d motion capture from synchronized video streams. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (June 2008)
11. Huguet, F., Devernay, F.: A variational method for scene flow estimation from stereo sequences. In: ICCV. (2007) 1–7
12. Starck, J., Hilton, A.: Correspondence labelling for wide-timeframe free-form surface matching. IEEE International Conference on Computer Vision (ICCV) (2007)
13. Starck, J., Hilton, A.: Spherical matching for temporal correspondence of non-rigid surfaces. In: ICCV. (2005) 1387–1394
14. Ahmed, N., Theobalt, C., Roessl, C., Thrun, S., Seidel: Dense correspondence finding for parametrization-free animation reconstruction from video. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (June 2008)
15. Varanasi, K., Zaharescu, A., Boyer, E., Horaud, R.: Temporal surface tracking using mesh evolution. In: Proceedings of the Tenth European Conference on Computer Vision. Volume Part II of LNCS. (October 2008) 30–43
16. Keriven, R., Faugeras, O.: Complete dense stereovision using level set methods. In: 5th European Conference on Computer Vision. (1998)
17. Allène, C., Pons, J.P., Keriven, R.: Seamless image-based texture atlases using multi-band blending. In: 19th International Conference on Pattern Recognition, Tampa, US (Dec 2008)
18. Sethian, J.A.: Fast marching methods. SIAM Review **41** (1999) 199–235
19. Kimmel, R., Sethian, J.: Computing geodesic paths on manifolds (1998)
20. Bronstein, M., Bronstein, A., Kimmel, R.: Efficient computation of isometry-invariant distances between surfaces. In: Techn. Report CIS-2006-02, Dept. of Computer Science, Technion, Israel. (2006)
21. Moenning, C., Dodgson, N.A.: Fast marching farthest point sampling. In: EuroGraphics. (Sept. 2003)
22. Dasgupta, S., Papadimitriou, C., Vazirani, U.: Algorithms. McGrawHill (2007)
23. Starck, J., Hilton, A.: Surface capture for performance based animation. IEEE Computer Graphics and Applications **27(3)** (2007) 21–31
24. Bronstein, A., Bronstein, M., Kimmel, R., Mahmoudi, M., Sapiro, G.: A gromov-hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching. International Journal of Computer Vision (IJCV) (submitted)

25. Vu, H., Keriven, R., Labatut, P., Pons, J.: Towards high-resolution large-scale multi-view stereo. Conference on Computer Vision and Pattern Recognition(CVPR) (Jun 2009)