Multiview 3D Warps
Supplemental Material

Alessio Del Bue
Istituto Italiano di Tecnologia (IIT)
alessio.delbue@iit.it

Adrien Bartoli
ALCoV - ISIT, Université Clermont 1
adrien.bartoli@gmail.com

Abstract

The supplemental material supports the paper with a more detailed explanation of the optimisation procedure used for the estimation of the Multiview 3D Warps. It also includes a summary of the videos included in the submission with additional experiments on real and synthetic sequences. This material will be possibly made available online after the reviewing process and it will be included in an extended version of this paper submission that is in preparation.

1. Additional details on the optimisation of Equation (10), Section 4.3

As pointed out in Sec. 4.3 of the submitted paper, the inference of the multiview 3D warps implies the 3D localisation of the time-varying control points contained in the $l \times 3f$ matrix $\mathtt{P}$ and the orthographic camera matrices stored in the $3f \times 2f$ matrix $\mathtt{M}$. The problem results in the factorisation of the $l \times 2f$ *warp projection matrix* $\mathtt{G}$ such that:

$$\min_{\mathtt{M},\mathtt{P}} = \|\mathtt{G} - \mathtt{P}\mathtt{M}\|^2 \quad \text{subject to} \quad \mathtt{M}_i^\top \mathtt{M}_i = \mathtt{I}_2, \qquad (1)$$

where $\mathtt{P}$ and $\mathtt{M}$ are defined as:

$$\mathtt{P} = \begin{bmatrix} \mathtt{P}_1 & \dots & \mathtt{P}_f \end{bmatrix} \quad \text{and} \quad \mathtt{M} = \begin{bmatrix} \mathtt{M}_1 & \dots & \mathtt{0} \\ \vdots & \ddots & \vdots \\ \mathtt{0} & \dots & \mathtt{M}_f \end{bmatrix}. \qquad (2)$$

This problem given the time-varying motion of the control points in 3D is ill-posed – a set of 3D points configurations will generate the same 2D projection onto the image plane. In order to select a correct solution among many, we enforce one of the priors listed in Sec. 3.2. The chosen statistical prior states that there exist a set of linear *warp deformation basis* that can model the time-varying motion of the control points. The 3D position at time instance i of the $l \times 3$ control points $\mathtt{P}_i$ is defined as:

$$\mathtt{P}_i = \sum_{d=1}^{D} r_{id}\mathtt{B}_d \quad \text{with} \quad \mathtt{B}_d \in \Re^{l\times 3}, r_{id} \in \Re \qquad (3)$$

where $\mathtt{B}_d$ for $d = 1 \dots D$ represents a set of D linear modes of deformation that are linearly combined by the weights r_{id} to obtain a given configuration of the control points. Given Eqs. (3), (2) and (1) we have that at a frame i the final cost function to optimise is the following:

$$\min_{\mathtt{M}_i,\mathtt{B}_d,r_{id}} \left\| \mathtt{G}_i - \sum_{d=1}^{D} r_{id}\mathtt{B}_d\mathtt{M}_i \right\|^2 \quad \text{subject to} \quad \mathtt{M}_i^\top \mathtt{M}_i = \mathtt{I}_2 \quad (4)$$

where $\mathtt{G}_i$ is the $l \times 2$ matrix containing the the projected control points into the image (i.e.$\mathtt{G} = \begin{bmatrix} \mathtt{G}_1 & \dots & \mathtt{G}_f \end{bmatrix}$). This optimisation problem is well-known in the Non-rigid Structure from Motion (NRSfM) literature [6, 4, 3]. Notice that, even if the optimisation problem is similar, the purpose of the estimation is different. The Multiview 3D Warps estimate a dense deformation field while classical NRSfM estimates the location of a sparse set of 3D points. Efficient solvers are available which can simultaneously optimise $\mathtt{M}_i$, $\mathtt{B}_d$ and r_{id} given the non-linear constraints present in $\mathtt{M}_i$. In our implementation we adopt the general purpose BALM solver [3] even if other implementations may be used if other priors are needed. A further point, the original BALM implementations was modified in order to fix the first basis $\mathtt{B}_1$ to the rest configuration of the control points $\mathtt{P}_0$. This assures consistency with the first warp initialisation stage in Section 4.1 where the parameters of the warp were estimated using the rest configuration $\mathtt{P}_0$. In order to initialise the iterative algorithm we used the pre-estimated orthographic camera matrices used in the warp placement stage.

2. Video sequence summary

The video sequence downloadable at http://goo.gl/22t30 presents the results for

the experiments presented in the paper and it adds more experimental evidence by showing the method applied to more real and synthetic sequences[1]. The video data (images and 2D tracks) presented in the supplemental material and in the experimental section of the paper come from several works on 3D Reconstruction, 2D registration and monocular 3D tracking. We have chosen such different sources for our experiments in order to asses that the proposed multiview 3D warps can deal with the data handled by different techniques in the state of the art.

The submitted video in the supplemental material is divided in different parts, we describe them briefly in the next sections.

Cushion sequence

Figure 1. Cushion sequence results. Starting from the left image we have the 2D original image tracks with missing data, the reprojected dense mesh in green and the control points location in red.

We show results for the cushion sequence originally presented in [3]. The video sequence has 50 frames with missing data up to 40%. The three video shots present first the tracked 90 points and the amount of noise affecting the 2D trajectories. Missing data are shown with a red circles. The second part shows the projected green mesh after the shape augmentation stage. Finally, we show the control points in red that warp the deformations at each frame.

Paper sequence

The sequence of 210 frames show the set of 140 points tracked from a bending paper while the camera is rotating around the sheet. The sequence was originally presented in [1]. Again, after learning the 3D warp, we augment the estimated mean shape with a dense mesh which is then reprojected (in green) into the image frame. The last video shot shows the cloning process where a red mesh was replicated from the original one. In this sequence we added a stronger rotation in respect to the example presented in the paper ($\alpha = 30$, $\beta = 10$, and $\gamma = 70$ degrees).

[1]The codec used is the Windows Media Codec running for every Windows Machine and most video decoders. If playing the video is problematic with your installed player, we suggest the use of the VLC media player available for most of the Operating Systems (http://www.videolan.org/vlc/).

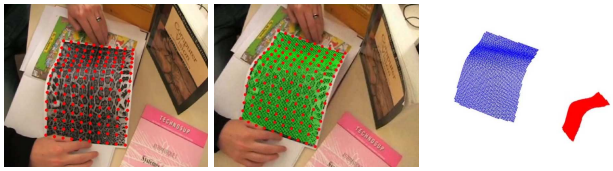

Figure 2. Paper sequence results. Starting from the left image we have the 2D original image tracks in red, the reprojected dense mesh in green and the cloning of the original mesh in blue into a new resized and scaled mesh in red. Second row, The re-texturing of the ICCV 2011 logo onto the original deformable surface.

Figure 3. Paper sequence logo re-texturing.

For the same sequence, we also show the re-texturing of a logo onto the deformable surface learned using the Multiview 3D warps. Even if the object shape is planar the re-textured logo follow the image deformations appearing in the sequence.

Extra material

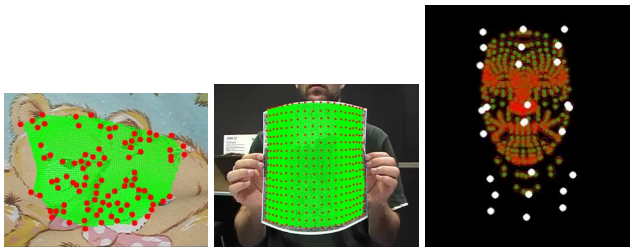

Figure 4. Extra sequences presented in the video. Starting from the left image we present a blanket deforming, another bending paper sequence and a synthetic example showing the control points motion.

Some extra sequences that are not presented in the paper show two further real experiments and a synthetic one. **Bears sequence** [4]: the video shows a 105 frames long sequence with 94 tracked points. Notice that there are parts of the images without dense sampling of the surface. However the green grid bends accordingly to the deformation appear-

ing in the video. **Paper sequence** [5]: this sequence shows another bending motion and the warping of the projected mesh. It is 70 frames long and the number of tracked points amount to 340. **Synthetic face sequence** [2]: This face is completely synthetic and it was created using CG tools. It is interesting to show the control points motion related to the deformations. Especially with some specific deformations (eye closing), the 3D control points are moving accordingly to model the 2D displacement. Also notice that the control points rotate and follow the shape rigid motion (a rotation on the z-axis).

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 Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska*, pages 1–8, 2008. 2

[2] A. Del Bue and L. Agapito. Stereo non-rigid factorization. *International Journal of Computer Vision*, 66(2):193–207, February 2006. 3

[3] A. Del Bue, J. Xavier, L. Agapito, and M. Paladini. Bilinear factorization via augmented lagrange multipliers. In *European Conf. on Computer Vision*, 2010. 1, 2

[4] S. Olsen and A. Bartoli. Implicit non-rigid structure-from-motion with priors. *Journal of Mathematical Imaging and Vision*, 31(2):233–244, 2008. 1, 2

[5] M. Salzmann, R. Hartley, and P. Fua. Convex optimization for deformable surface 3-d tracking. In *Proc. International Conference on Computer Vision*, pages 1–8, 2007. 3

[6] L. Torresani, A. Hertzmann, and C. Bregler. Non-rigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pages 878–892, 2008. 1