DEFORMABLE MOTION 3D RECONSTRUCTION BY UNION OF REGULARIZED SUBSPACES

Antonio Agudo *Francesc Moreno-Noguer*

Institut de Robòtica i Informàtica Industrial, CSIC-UPC, 08028, Barcelona, Spain

ABSTRACT

This paper presents an approach to jointly retrieve camera pose, time-varying 3D shape, and automatic clustering based on motion primitives, from incomplete 2D trajectories in a monocular video. We introduce the concept of order-varying temporal regularization in order to exploit video data, that can be indistinctly applied to the 3D shape evolution as well as to the similarities between images. This results in a union of regularized subspaces which effectively encodes the 3D shape deformation. All parameters are learned via augmented Lagrange multipliers, in a unified and unsupervised manner that does not assume any training data at all. Experimental validation is reported on human motion from sparse to dense shapes, providing more robust and accurate solutions than state-of-the-art approaches in terms of 3D reconstruction, while also obtaining motion grouping results.

***Index Terms*—** Non-Rigid Structure from Motion, Order-Varying Regularization, Union of Regularized Subspaces

1. INTRODUCTION

In the last decade, many efforts have been done towards developing 3D perception algorithms. Initially, rigidity constraints were exploited to acquire robust and accurate 3D geometries from monocular video, posing a well-posed problem without accounting for any extra sensor [1]. These formulations were later extended to the non-rigid domain, where many different 3D shape configurations may yield very similar 2D observations. Addressing this ambiguous scenario requires incorporating more sophisticated priors than those utilized in the rigid case, able to constraining the solution space. In the literature, this problem is known as Non-Rigid Structure from Motion (NRSfM), and consists in retrieving shape and motion from 2D point trajectories in a RGB video without the need for a pre-trained model.

The most standard priors for NRSfM enforce a low-rank constraint over the entire shape [2, 3], the 3D point trajectories [4, 5, 6] or the force patterns that induce the deformations [7]. These approaches, though, consider only one single low-rank modality, and are prone to fail in situations that need a larger level of expressiveness. To solve this limitation, recent approaches have been formulated in terms of a union of subspaces, but without properly encoding the physical priors on an image sequence. In these cases, it is worth noting that most frames are similar to their neighbors in video data. In this paper, we propose to exploit this sequential nature, by incorporating two neighbor-penalty terms into a union-of-subspaces model: one to enforce consecutive temporal similarities, and another to impose smooth deformations over time.

2. RELATED WORK

Estimating time-varying 3D shape while retrieving camera motion from solely the observation of 2D point tracks in a RGB video is a severely under-constrained problem that demands more sophisticated prior knowledge. The most popular approach to address the inherent ambiguity of the NRSfM problem consists in assuming the 3D shape to lie in a low-rank subspace. In this context, factorization-based approaches were presented by using shape [3, 8, 9], trajectory [4, 5, 6], shape-trajectory [10, 11], and force [7] models, where the dimensionality of the subspace was assumed to be known. More recently, other formulations have imposed a low-rank constraint by directly minimizing the rank of a matrix representing the 3D shape. To do this, these formulations rely on the data lie in a single [12, 13], in a union of temporal [14] or spatio-temporal [15] subspaces, or by means of multiple unions of them [16]. On top of these models, additional spatial [2] or temporal [17, 18, 19, 20] smoothness constraints have also been considered, even obtaining real-time solutions [21, 22]. However, the sequential nature in video data has not been properly exploited in previous formulations, considering both deformations and similarities. In this work, we account for temporal consistency using a novel formulation based on a union of regularized subspaces. This constraint allows establishing a double temporal regularization, that can be imposed by using a unique order-varying smoothness matrix. We first penalize deviations on consecutive motion similarities by incorporating a temporal Laplacian regularization term, and then, our model is complemented by means of a temporal regularization over the 3D shape.

This work has been partially supported by the Spanish Ministry of Science and Innovation under project HuMoUR TIN2017-90086-R; and by the Spanish State Research Agency through the María de Maeztu Seal of Excellence to IRI MDM-2016-0656.

3. NON-RIGID STRUCTURE FROM MOTION

Let us consider a set of P 3D points viewed along F image frames. We denote by $\mathbf{x}_p^f = [x_p^f, y_p^f, z_p^f]^\top$ the 3D location of the p-th point at frame f, and by $\tilde{\mathbf{w}}_p^f = [u_p^f, v_p^f]^\top$ its 2D projection in the image plane. Under orthography, the camera translation can be computed as the mean of the observations, such that $\mathbf{t}^f = \sum_i \tilde{\mathbf{w}}_p^f / P$, and subtracting it in each frame we can obtain zero-mean measurements as $\mathbf{w}_p^f = \tilde{\mathbf{w}}_p^f - \mathbf{t}^f$. Considering all views and points, we can define the mapping 3D-to-2D point coordinates as:

$$\underbrace{\begin{bmatrix} \mathbf{w}_1^1 & \dots & \mathbf{w}_P^1 \\ \vdots & \ddots & \vdots \\ \mathbf{w}_1^F & \dots & \mathbf{w}_P^F \end{bmatrix}}_{\mathbf{W}} = \underbrace{\begin{bmatrix} \mathbf{R}^1 & & \\ & \ddots & \\ & & \mathbf{R}^F \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} \mathbf{x}_1^1 & \dots & \mathbf{x}_P^1 \\ \vdots & \ddots & \vdots \\ \mathbf{x}_1^F & \dots & \mathbf{x}_P^F \end{bmatrix}}_{\mathbf{X}},$$

where $\mathbf{W}$ is a $2F \times P$ matrix storing the 2D trajectories arranged in columns, $\mathbf{G}$ is a $2F \times 3F$ block diagonal matrix, made of the F truncated 2×3 camera rotations $\mathbf{R}^f$, and $\mathbf{X}$ is a $3F \times P$ matrix with the 3D locations of the points along the sequence, also arranged in columns. The NRSfM problem consists in retrieving the deformable shape $\mathbf{X}$ and camera motion $\mathbf{G}$ matrices from 2D trajectories $\mathbf{W}$ in a RGB video.

For later computations, we also include another interpretation of the shape matrix, the matrix $\mathbf{Y}$, that re-arranges the entries of $\mathbf{X}$ into a new $3P \times F$ matrix. These two shape interpretations can be mapped one onto the other by means of $\mathbf{X} = (\mathbf{I}_3 \otimes \mathbf{Y}^\top)\mathbf{A}$ and $\mathbf{Y} = (\mathbf{X}^\top \otimes \mathbf{I}_3)\mathbf{B}$, where $\otimes$ is the Kronecker product operator, $\mathbf{I}_3$ the identity matrix, and $\mathbf{A}$ and $\mathbf{B}$ are binary matrices of size $9N \times N$ and $9F \times F$, respectively.

4. ORDER-VARYING TEMPORAL REGULARIZATION

In this paper, we propose several temporal regularization priors to encode the sequential relationships in video data. We will use this type of priors to enforce a temporal smoothing of some model parameters, such as the time-varying 3D location of an observed object in the monocular video. To this end, we draw inspiration in the theory of finite differences, writing a smoothness constraint in terms of a finite number of values along a temporal direction. For instance, we can impose a first-order approximation of the type $\mathbf{x}_p^f \approx \mathbf{x}_p^{f+1}$ to enforce smooth deformations, i.e., the 3D location of the p-th point in two neighboring frames does not change much. In a similar manner, we could impose higher-order approximations to extend the influence of the neighborhood, obtaining different levels of regularization.

To enforce a temporal smoothness constraint over all object points along the sequence, we define a $F \times F$ matrix $\mathbf{L}_o$, where the subindex denotes the order of the approximation. This is a highly sparse matrix, especially for low-order approximations. For instance, when considering the first terms of approximation, $\mathbf{L}_1$ can be modeled by a lower bidiagonal matrix, $\mathbf{L}_2$ by a tridiagonal one and so on. It is worth pointing out that our matrix can degenerate also into $\mathbf{L}_0 \equiv \mathbf{I}_F$ when no relations appear in the data. To better illustrate the structure of these matrices, we next explicitly write some examples of the first-order forward, the second-order central, and the fourth-order central –with boundary conditions– difference approximations:

$$\mathbf{L}_1 = \begin{bmatrix} -1 & & & & \\ 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{bmatrix}, \mathbf{L}_2 = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & \ddots & \ddots & \ddots \\ & & & -1 & 1 \end{bmatrix}, \tag{1}$$

$$\mathbf{L}_4 = \begin{bmatrix} 0 & 1 & -1 & & & & \\ 1 & -16 & 16 & -1 & & & \\ -1 & 16 & -30 & 16 & -1 & & \\ & -1 & 16 & -30 & 16 & -1 & \\ & & -1 & 16 & -30 & 16 & -1 \\ & & & \ddots & \ddots & \ddots & \vdots \\ & & & & -1 & 1 & 0 \end{bmatrix}. \tag{2}$$

5. 3D RECONSTRUCTION AND MOTION GROUPING FROM 2D TRAJECTORIES

After these preliminary concepts, we next formulate the general problem and then propose an efficient and unified optimization strategy to solve it, which does not need training data.

5.1. Problem Formulation

For simultaneously solving the NRSfM problem (time-varying shape $\mathbf{X}$ and motion $\mathbf{G}$) and clustering the motion into action primitives, we propose encoding the shape deformation by means of a union of regularized temporal subspaces in combination with temporal smoothness priors. To enforce the union of temporal subspaces, we consider an $F \times F$ affinity matrix $\mathbf{F}$, such that $\mathbf{Y} = \mathbf{YF} + \mathbf{N}$, where $\mathbf{N}$ represents a $3P \times F$ residual noise. We assume the matrices $\mathbf{Y}$ and $\mathbf{F}$ to be low-rank and, therefore, they can be computed by minimizing their rank. Since this is a non-convex NP-hard problem we used the nuclear norm instead, which is its convex relaxation [23, 24]. Additionally, we consider extra penalty terms to exploit the temporal closeness in time video data, by means of the matrix $\mathbf{L}_o$. To this end, we add a temporal Laplacian regularization [25] function $p(\mathbf{F})$ to incorporate the temporal information in the affinity matrix. Additionally, we include a temporal regularization over the time-varying 3D shape matrix, by enforcing the constraint $\mathbf{YL}_o$.

We denote by $\mathbf{\Omega} \equiv \{\mathbf{W}, \mathbf{G}, \mathbf{F}, \mathbf{X}, \mathbf{Y}, \mathbf{N}\}$ the set of all model parameters that needs to be estimated. Our input data consist of partial 2D point tracks in a RGB video $\bar{\mathbf{W}}$, and the corresponding visibility matrix $\mathbf{V} \in \mathbb{R}^{F \times P}$, with $\{1, 0\}$ entries indicating whether a point in a specific frame is visible or not. Considering orthonormality constraints on camera rotations, our problem can be written as:

$$\begin{aligned}\arg\min_{\mathbf{\Omega}} \quad & \| (\mathbf{V} \otimes \mathbf{1}_2) \odot \left(\mathbf{W} - \bar{\mathbf{W}}\right) \|_F^2 + \beta\|\mathbf{W}\|_* + \gamma\|\mathbf{Y}\|_* \\ & + \gamma\|\mathbf{F}\|_* + \phi\, p(\mathbf{F}) + \lambda\|\mathbf{N}\|_1 \qquad (3)\end{aligned}$$

$$\begin{aligned}\text{subject to} \quad & \mathbf{W} = \mathbf{G}\mathbf{X} \\ & \textstyle\sum_{f=1}^{F} \mathbf{R}^f {\mathbf{R}^f}^\top = F\mathbf{I}_2 \\ & \mathbf{Y} = \mathbf{Y}\mathbf{F} + \mathbf{N} \\ & (\mathbf{I}_3 \otimes \mathbf{Y}^\top)\mathbf{A} = \mathbf{X} \\ & \mathbf{Y}\mathbf{L}_o = \mathbf{0}\end{aligned}$$

where $\odot$ denotes a Hadamard product, $\mathbf{1}$ is a vector of ones, $\|\cdot\|_*$ and $\|\cdot\|_1$ represent the nuclear norm and the l_1-norm, respectively, and $\|\cdot\|_F$ indicates the Frobenius norm. $\{\beta, \gamma, \phi, \lambda\}$ are penalty weight coefficients. All these values were determined empirically using a validation sequence, and kept constant for the rest of all experiments.

The overall problem in Eq. (3) is non-convex, and it can be approximated by a three-step strategy. First of all, we solve a matrix-completion problem to compute full measurements $\mathbf{W}$ from incomplete data $\bar{\mathbf{W}}$, enforcing the measurement matrix to be low rank. After that, we retrieve the camera motion matrix $\mathbf{G}$. Both previous steps are performed as described in [15]. Finally, we jointly estimate shape $\mathbf{X}$ (or its alternative interpretation $\mathbf{Y}$) along with the affinity matrix $\mathbf{F}$, solving the subproblem:

$$\arg\min_{\mathbf{F},\mathbf{Y},\mathbf{X},\mathbf{N}} \quad \gamma(\|\mathbf{Y}\|_* + \|\mathbf{F}\|_*) + \phi\, p(\mathbf{F}) + \lambda\|\mathbf{N}\|_1 \qquad (4)$$

$$\begin{aligned}\text{subject to} \quad & \mathbf{W} = \mathbf{G}\mathbf{X} \\ & \mathbf{Y} = \mathbf{Y}\mathbf{F} + \mathbf{N} \\ & (\mathbf{I}_3 \otimes \mathbf{Y}^\top)\mathbf{A} = \mathbf{X} \\ & \mathbf{Y}\mathbf{L}_o = \mathbf{0}\end{aligned}$$

5.2. Simultaneous 3D Shape and Grouping

To solve the objective function in Eq. (4), we devise an optimization algorithm based on Augmented Lagrange Multipliers (ALM), being the equivalent Lagrangian form of Eq. (4):

$$\begin{aligned}\arg\min_{\mathbf{J},\mathbf{F},\mathbf{X},\mathbf{Y},\mathbf{N}} \quad & \gamma(\|\mathbf{Y}\|_* + \|\mathbf{J}\|_*) + \phi\,\mathrm{tr}(\mathbf{F}\mathbf{L}_o\mathbf{F}^\top) + \lambda\|\mathbf{N}\|_1 \qquad (5) \\ & + \mathrm{tr}(\mathbf{M}_1^\top(\mathbf{W} - \mathbf{G}\mathbf{X})) + \frac{\alpha}{2}\|\mathbf{W} - \mathbf{G}\mathbf{X}\|_F^2 \\ & + \mathrm{tr}(\mathbf{M}_2^\top(\mathbf{Y} - \mathbf{Y}\mathbf{F} - \mathbf{N})) + \frac{\alpha}{2}\|\mathbf{Y} - \mathbf{Y}\mathbf{F} - \mathbf{N}\|_F^2 \\ & + \mathrm{tr}(\mathbf{M}_3^\top((\mathbf{I}_3 \otimes \mathbf{Y}^\top)\mathbf{A} - \mathbf{X})) + \mathrm{tr}(\mathbf{M}_4^\top\mathbf{Y}\mathbf{L}_o) \\ & + \frac{\alpha}{2}\|(\mathbf{I}_3 \otimes \mathbf{Y}^\top)\mathbf{A} - \mathbf{X}\|_F^2 + \frac{\alpha}{2}\|\mathbf{Y}\mathbf{L}_o\|_F^2 \\ & + \mathrm{tr}(\mathbf{M}_5^\top(\mathbf{F} - \mathbf{J})) + \frac{\alpha}{2}\|\mathbf{F} - \mathbf{J}\|_F^2\end{aligned}$$

where the matrix $\mathbf{J}$ is a dual variable. In addition, we also introduce the Lagrange multipliers: $\mathbf{M}_1 \in \mathbb{R}^{2F \times N}$, $\{\mathbf{M}_2, \mathbf{M}_4\} \in \mathbb{R}^{3N \times F}$, $\mathbf{M}_3 \in \mathbb{R}^{3F \times N}$, and $\mathbf{M}_5 \in \mathbb{R}^{F \times F}$, and $\alpha > 0$ is a penalty weight to improve convergence. $\mathrm{tr}(\cdot)$ denotes the trace of a matrix.

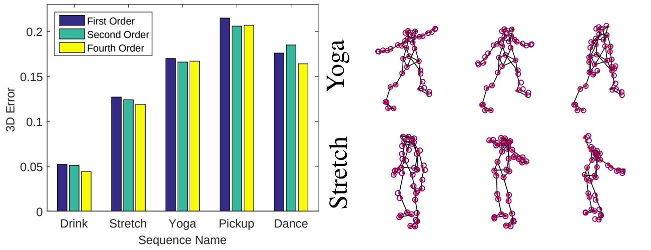

Fig. 1. **Temporal-regularization evaluation on human motion capture video sequences with noisy observations. Left:** 3D reconstruction error e_S per video sequence. For every case, we evaluate first-, second-, and fourth-order approximations. **Right:** Some sample frames of the *Yoga* and *Stretch* sequences. Red dots correspond to the shape estimated with our approach, and purple circles are the ground truth. Best viewed in color.

The problem in Eq. (5) can be efficiently resolved by estimating every model parameter separately and in closed form, while keeping fixed the rest of model parameters. Particularly, to solve the nuclear-norm problems, we apply a singular value thresholding minimization [26] with a shrinkage operator $S_{\frac{\gamma}{\alpha}}(x) = \max(0, x - \frac{\gamma}{\alpha})$. For the l_1-norm minimization problem, we use the element-wise shrinkage operator $S_{\frac{\lambda}{\alpha}}(x) = \max(0, x - \frac{\lambda}{\alpha})$ [27].

6. EXPERIMENTAL EVALUATION

We now present our experimental evaluation on several human motion videos, considering articulated and continuous motion, face and full body configurations, and scenarios with missing data. First of all, we evaluate our algorithm on the articulated human motion dataset introduced in [4], which includes five motion activities. As it is common in the literature [6, 11, 12], we will report the normalized mean 3D error e_S, and the mean rotation error e_R. For further details, we refer the reader to these papers. Additionally, we also provide the object grouping error e_G as defined in [15], after applying spectral clustering [28] over the estimated matrix $\mathbf{F}$.

We first evaluate our approach by considering different alternatives to enforce the temporal regularization by the matrix $\mathbf{L}_o$. Specifically, we test over first-, second- and fourth-order approximations. As shown in Fig. 1-left, when using higher-order approximation to enforce the temporal regularization, the estimated 3D reconstructions are in general more accurate. We also compare our URS (Union of Regularized Subspaces) algorithm with $\mathbf{L}_4$ against seven state-of-the-art methods: EM-PPCA [3], MP [8], PTA [4], CSF [11], KSTA [10], BMM [12], and PPTA [6]. In contrast to the rest of approaches, our method does not require tuning the subspace rank R, which had to be done for every competing approach and experiment, considering the basis rank that produced the lowest e_S. We consider two situations: noise-free observations, and 2D trajectories artificially

Data \ Met.	EM-PPCA [3]		MP [8]		PTA [4]		CSF [11]		KSTA [10]		BMM [12]		PPTA [6]		URS (Ours)		
	e_R	$e_S(R)$	e_R	$e_S(R)$	e_R	$e_S(R)$	e_R	$e_S(R)$	e_R	$e_S(R)$	e_R	$e_S(R)$	e_R	$e_S(R)$	e_R	e_S	$e_G[\%]$
Drink	.186	.261(7)	.330	.357(12)	.006	.025(13)	.006	.022(6)	.006	.020(12)	.007	.027(12)	.006	.011(30)	.006	**.009**	0.8(2)
Stretch	.749	.458(7)	.832	.900(8)	.055	.109(12)	.049	.071(8)	.049	.064(11)	.068	.103(11)	.058	.084(11)	.058	**.061**	4.1(3)
Yoga	.688	.445(8)	.854	.786(2)	.106	.163(11)	.102	.147(7)	.102	.148(7)	.088	**.115(10)**	.106	.158(11)	.106	.143	0.3(2)
Pick-up	.417	.423(14)	.249	.429(5)	.155	.237(12)	.155	.230(6)	.155	.233(6)	.121	**.173(12)**	.154	.235(12)	.154	.221	3.7(3)
Dance	–	.339(4)	–	.271(5)	–	.296(5)	–	.271(2)	–	.249(4)	–	.188(10)	–	.229(4)	–	**.165**	–
Average error:		.385		.549		.166		.148		.143		.121		.143		**.119**	
Relative error:		3.23		4.61		1.39		1.24		1.20		1.02		1.20		1.00	
Drink	.231	.250(7)	.329	.517(12)	.043	.045(13)	.043	.044(6)	.043	.042(12)	.044	.056(12)	.042	**.038(30)**	.042	.044	3.6(2)
Stretch	.819	.886(7)	.872	.975(8)	.091	.144(12)	.091	.121(8)	.091	.166(11)	.098	.183(11)	.091	.123(11)	.091	**.119**	8.4(3)
Yoga	.700	.507(8)	.858	.791(2)	.124	.174(11)	.125	.168(7)	.125	.172(7)	.136	.195(10)	.124	.174(11)	.125	**.167**	0.0(2)
Pick-up	.499	.807(14)	.250	.407(5)	.148	.228(12)	.148	.224(6)	.148	.222(6)	.141	.212(12)	.148	.228(12)	.148	**.207**	3.1(3)
Dance	–	.336(4)	–	.282(5)	–	.299(5)	–	.266(2)	–	.248(4)	–	.236(10)	–	.222(4)	–	**.164**	–
Average error:		.557		.594		.178		.165		.170		.176		.157		**.140**	
Relative error:		3.97		4.24		1.27		1.18		1.21		1.26		1.12		1.00	

Table 1. **Quantitative comparison on human-motion sequences.** We include rotation e_R and reconstruction e_S errors for competing techniques: EM-PPCA [3], MP [8], PTA [4], CSF [11], KSTA [10] and BMM [12], and PPTA [6]; and for our URS approach. For each solution, we also provide in parentheses the rank R of the linear subspace that produced the lowest e_S error. Relative error is always represented with respect to URS reconstruction. For ours, we also include grouping error $e_G[\%]$, and the number of motion groups in parentheses. The symbol "$-$" denotes that ground truth data is not available. **Top:** Noise-free observations. **Bottom:** Noisy observations.

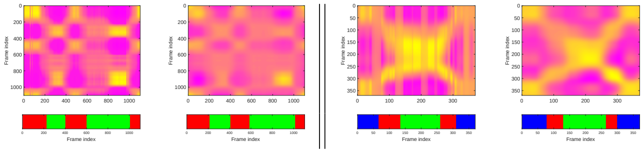

Fig. 2. **Affinity matrices.** In both cases, we represent our estimated matrix $\mathbf{F}$ over noisy observations, and the corresponding ground truth. We also represent the group bar. **Left:** *Drink* sequence. **Right:** *Stretch* sequence.

corrupted by zero-mean Gaussian noise with standard deviation $\sigma_{noise} = 0.01\rho$, with ρ being the maximum distance of an image point to the centroid of all the points. Table 1 summarizes the 3D reconstruction errors for all methods, datasets, and situations. Note that our approach consistently outperforms state-of-the-art in terms of 3D reconstruction, especially for noisy observations, reducing the 3D error of other approaches by large margins between 12% and 424%. Some examples of our 3D reconstructions for the *Yoga* and *Stretch* datasets are shown in Fig. 1-right. In Fig. 2, we show a qualitative comparison between our affinity estimation and the ground truth, together with corresponding grouping bars. We observe the clustering we obtain is very accurate.

We also validate the robustness of our algorithm to occlusions, by processing an American-sign-language sequence, where a man is moving the head while talking and hand gesturing [7]. Figure 3-top shows some frames and our 3D reconstruction, as well as the affinity and groups estimation. Finally, we also test our method to dense data. To this end, we process a back sequence with 20,561 2D trajectories taken from [13]. In Fig. 3-bottom is displayed the 3D reconstruction for some images, along with the estimated similarities and groups. Despite being only qualitative, the reconstruction results seem very accurate.

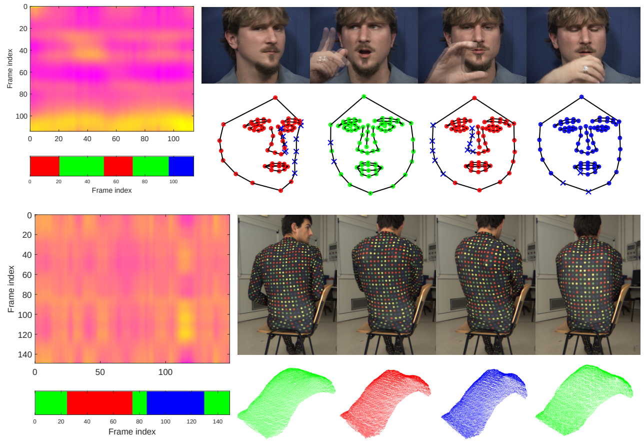

Fig. 3. **Face and Back sequences.** In both cases, we represent the same information. **Left:** Motion affinity matrix we recover, and the corresponding group bar. **Right:** Images and a general view of the reconstructed shape. Every color corresponds to a motion group. Blue crosses are missing points.

7. CONCLUSION

We have presented a novel formulation to solve the NRSfM problem in a unified and unsupervised manner. For this purpose, we have proposed a union of regularized subspaces that enforces both temporally consistent 3D reconstructions and grouping samples into motion primitives. An energy-based formulation is designed to encode the problem, that is solved using augmented Lagrange multipliers. We show that besides providing correct motion grouping, our method produces more accurate solutions than the rest of competing approaches to recover human motion, can cope with missing entries and handles dense data. In the future, we aim at using this research to solve the problem in a sequential fashion, retrieving the model parameters as the data arrives.

8. REFERENCES

[1] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski, "Building Rome in a day," in *ICCV*, 2009.

[2] M. Lee, J. Cho, C. H. Choi, and S. Oh, "Procrustean normal distribution for non-rigid structure from motion," in *CVPR*, 2013.

[3] L. Torresani, A. Hertzmann, and C. Bregler, "Non-rigid structure-from-motion: estimating shape and motion with hierarchical priors," *TPAMI*, vol. 30, no. 5, pp. 878–892, 2008.

[4] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade, "Non-rigid structure from motion in trajectory space," in *NIPS*, 2008.

[5] H. S. Park, T. Shiratori, I. Matthews, and Y. Sheikh, "3D reconstruction of a moving point from a series of 2D projections," in *ECCV*, 2010.

[6] A. Agudo and F. Moreno-Noguer, "A scalable, efficient, and accurate solution to non-rigid structure from motion," *CVIU*, vol. 167, no. 2, pp. 121–133, 2018.

[7] A. Agudo and F. Moreno-Noguer, "Learning shape, motion and elastic models in force space," in *ICCV*, 2015.

[8] M. Paladini, A. Del Bue, M. Stosic, M. Dodig, J. Xavier, and L. Agapito, "Factorization for non-rigid and articulated structure using metric projections," in *CVPR*, 2009.

[9] V. Golyanik and D. Stricker, "Dense batch non-rigid structure from motion in a second," in *WACV*, 2017.

[10] P. F. U. Gotardo and A. M. Martinez, "Kernel non-rigid structure from motion," in *ICCV*, 2011.

[11] P. F. U. Gotardo and A. M. Martinez, "Computing smooth time-trajectories for camera and deformable shape in structure from motion with occlusion," *TPAMI*, vol. 33, no. 10, pp. 2051–2065, 2011.

[12] Y. Dai, H. Li, and M. He, "A simple prior-free method for non-rigid structure from motion factorization," in *CVPR*, 2012.

[13] R. Garg, A. Roussos, and L. Agapito, "Dense variational reconstruction of non-rigid surfaces from monocular video," in *CVPR*, 2013.

[14] Y. Zhu, D. Huang, F. de la Torre, and S. Lucey, "Complex non-rigid motion 3D reconstruction by union of subspaces," in *CVPR*, 2014.

[15] A. Agudo and F. Moreno-Noguer, "DUST: Dual union of spatio-temporal subspaces for monocular multiple object 3D reconstruction," in *CVPR*, 2017.

[16] A. Agudo, M. Pijoan, and F. Moreno-Noguer, "Image collection pop-up: 3D reconstruction and clustering of rigid and non-rigid categories," in *CVPR*, 2018.

[17] A. Bartoli, V. Gay-Bellile, U. Castellani, J. Peyras, S. Olsen, and P. Sayd, "Coarse-to-fine low-rank structure-from-motion," in *CVPR*, 2008.

[18] A. Agudo, J. M. M. Montiel, L. Agapito, and B. Calvo, "Modal space: A physics-based model for sequential estimation of time-varying shape from monocular video," *JMIV*, vol. 57, no. 1, pp. 75–98, 2017.

[19] A. Agudo and F. Moreno-Noguer, "Combining local-physical and global-statistical models for sequential deformable shape from motion," *IJCV*, vol. 122, no. 2, pp. 371–387, 2017.

[20] M. Lee, C. H. Choi, and S. Oh, "A procrustean Markov process for non-rigid structure recovery," in *CVPR*, 2014.

[21] A. Agudo, B. Calvo, and J. M. M. Montiel, "3D reconstruction of non-rigid surfaces in real-time using wedge elements," in *ECCVW*, 2012, pp. 113–122.

[22] A. Agudo, F. Moreno-Noguer, B. Calvo, and J. M. M. Montiel, "Real-time 3D reconstruction of non-rigid shapes from single moving camera," *CVIU*, vol. 153, no. 12, pp. 37–54, 2016.

[23] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717, 2008.

[24] Y. Chen, H. Xu, C. Caramanis, and S. Sanghavi, "Robust matrix completion with corrupted columns," in *ICML*, 2011.

[25] Z. Zhang and K. Zhao, "Low-rank matrix approximation with manifold regularization," *TPAMI*, vol. 35, no. 7, pp. 1717–1729, 2013.

[26] J.F. Cai, E. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM JO*, vol. 20, no. 4, pp. 1956–1982, 2010.

[27] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *UIUC Technical Report UILU-ENG-09-2215*, 2009.

[28] W. Y. Chen, Y. Song, H. Bai, C.J. Lin, and E. Chang, "Parallel spectral clustering in distributed systems," *TPAMI*, vol. 33, no. 3, pp. 568–586, 2010.