PIECEWISE BÉZIER SPACE: RECOVERING 3D DYNAMIC MOTION FROM VIDEO

ANTONIO AGUDO

INSTITUT DE ROBÒTICA I INFORMÀTICA INDUSTRIAL, CSIC-UPC, BARCELONA, SPAIN

2488

Institut de Robòtica i Informàtica Industrial

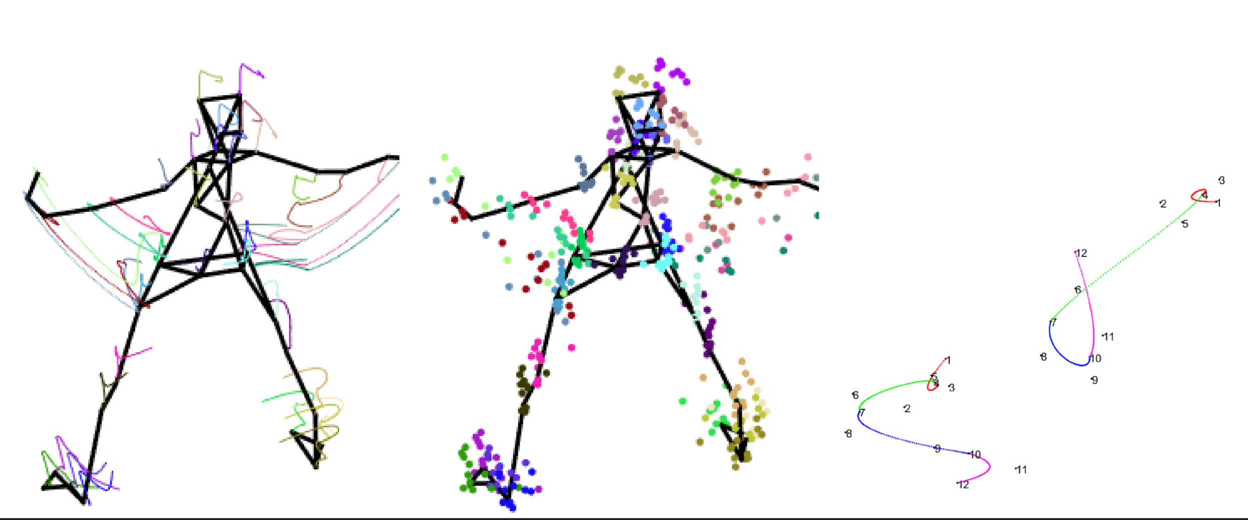
CSIC Consejo Superior de Investigaciones Científicas

UPC

PROBLEM STATEMENT

- Given an uncalibrated *monocular video* where a piece-varying non-rigid shape can be observed.
- We want to jointly recover the *motion* of the camera, the *3D non-rigid shape model* and the *temporal grouping into deformations*.
- We present a **piecewise Bézier Space** to model non-rigid motions with **physical meaning**, that can automatically enforce C^0, C^1 and C^2 continuities. No training data are needed.

Trajectories — Control points — Pieces



PIECEWISE BÉZIER CURVES

- A **Bézier curve** $\mathbf{p}(s)$ is defined by a *linear combination* of K Bernstein basis polynomials of degree $K-1$ restricted to the continuous interval $s=[0,\ldots,1]$ as:

$$\mathbf{p}(s)=\sum_{k=1}^{K}\binom{K-1}{k-1}s^{k-1}(1-s)^{K-k}\mathbf{c}_k=\sum_{k=1}^{K}b_k(s)\mathbf{c}_k$$

with $b_k(s)$ and $\mathbf{c}_k=[c_{xk},c_{yk},c_{zk}]^\top$ are the k-th Bernstein basis and 3D control point, respectively. $\binom{K-1}{k-1}$ is a binomial coefficient.

- A global curve is represented by low-order $K=\{3,4\}$ *piecewise Bézier curves.*
- Quadratic (qua.) and cubic (cub.) curves are used as a function of control points K, defining the number of pieces (P) and transitions (T).
- C^0*-continuity* is guaranteed by construction as a unique control point acts as the last and first in two pieces.

K	P	T	Type	K	P	T	Type
3	1	0	1 qua.	10	3	2	3 cub.
4	1	0	1 cub.	11	4	3	2 qua. + 2 cub.
5	2	1	2 qua.	12	4	3	1 qua. + 3 cub.
6	2	1	1 qua. + 1 cub.	13	4	3	4 cub.
7	2	1	2 cub.	14	5	4	2 qua. + 3 cub.
8	3	2	2 qua. + 1 cub.	15	5	4	1 qua. + 4 cub.
9	3	2	1 qua. + 2 cub.	16	5	4	5 cub.

3D DYNAMIC SHAPE, GROUPING AND MOTION FROM 2D POINT TRACKS

- Under orthography, a time-deforming 3D scenario S can be projected as $\mathbf{W}$ such that:

$$\underbrace{\begin{bmatrix}\mathbf{w}_1^1 & \ldots & \mathbf{w}_N^1\\ \vdots & \ddots & \vdots\\ \mathbf{w}_1^F & \ldots & \mathbf{w}_N^F\end{bmatrix}}_{\mathbf{W}}=\underbrace{\begin{bmatrix}\mathbf{R}^1 & & \\ & \ddots & \\ & & \mathbf{R}^F\end{bmatrix}}_{\mathbf{G}}\underbrace{\begin{bmatrix}\mathbf{s}_1^1 & \ldots & \mathbf{s}_N^1\\ \vdots & \ddots & \vdots\\ \mathbf{s}_1^F & \ldots & \mathbf{s}_N^F\end{bmatrix}}_{\mathbf{S}}$$

- Piecewise Bézier curves are used for modelling the evolution of every point coordinate over time.
- Shape $\mathbf{S}=\mathbf{BC}$, with $\mathbf{B}\in\mathbb{R}^{3F\times 3K}$ a known matrix with predefined piecewise basis, and $\mathbf{C}\in\mathbb{R}^{3K\times N}$ a matrix of unknown control points as:

$$\mathbf{B}=\begin{bmatrix}\mathbf{I}_3\otimes(\mathbf{b}^1(s))^\top\\ \vdots\\ \mathbf{I}_3\otimes(\mathbf{b}^F(s))^\top\end{bmatrix},\quad \mathbf{C}=[\boldsymbol{\kappa}^1\ \ldots\ \boldsymbol{\kappa}^N]$$

- N $3K$-dimensional vectors as $\boldsymbol{\kappa}^n=[c^n_{x1},\ldots,c^n_{xK},c^n_{y1},\ldots,c^n_{yK},c^n_{z1},\ldots,c^n_{zK}]^\top$, and F K-dimensional vectors $\mathbf{b}^f(s)=[b_1(s),\ldots,b_K(s)]^\top$.
- C^1 and C^2 continuities can be enforced by the constraint $(\mathbf{I}_3\otimes\mathbf{M})\mathbf{C}=(\mathbf{I}_3\otimes\mathbf{N})\mathbf{C}$, with $\mathbf{M}$ and $\mathbf{N}$ known matrices.
- Considering temporal smoothness priors, a union of temporal piecewise subspaces, and orthogonality, we propose to solve the optimization problem:

$$\begin{aligned}\arg\min_{\boldsymbol{\Phi}}\ & \|(\mathbf{V}\otimes\mathbf{1}_2)\odot(\mathbf{W}-\bar{\mathbf{W}})\|_F^2+\beta\|\mathbf{W}\|_*\\ &+\alpha(\|\mathbf{S}-\mathbf{BC}\|_F+\|(\mathbf{I}_3\otimes(\mathbf{M}-\mathbf{N}))\mathbf{C}\|_F)\\ &+\gamma(\|\hat{\mathbf{S}}\|_*+\|\mathbf{T}\|_*)+\lambda\|\mathbf{E}\|_{2,1}+\zeta g(\mathbf{R}^f)\\ \text{subject to}\ & \mathbf{W}=\mathbf{GS},\ \ \mathbf{S}=(\mathbf{I}_3\otimes\hat{\mathbf{S}}^\top)\mathbf{A}\\ & \hat{\mathbf{S}}=\hat{\mathbf{S}}\mathbf{T}+\mathbf{E},\ \ \hat{\mathbf{S}}\mathbf{F}=\mathbf{0},\ \ \mathbf{R}^f\mathbf{R}^{f\top}=\mathbf{I}_2\end{aligned}$$

- We present a two-step strategy to minimize the cost function: 1) completing missing entries, 2) estimating camera rotation, clustering and 3D reconstruction.
- Augmented Lagrange multipliers are considered.

EXPERIMENTAL RESULTS

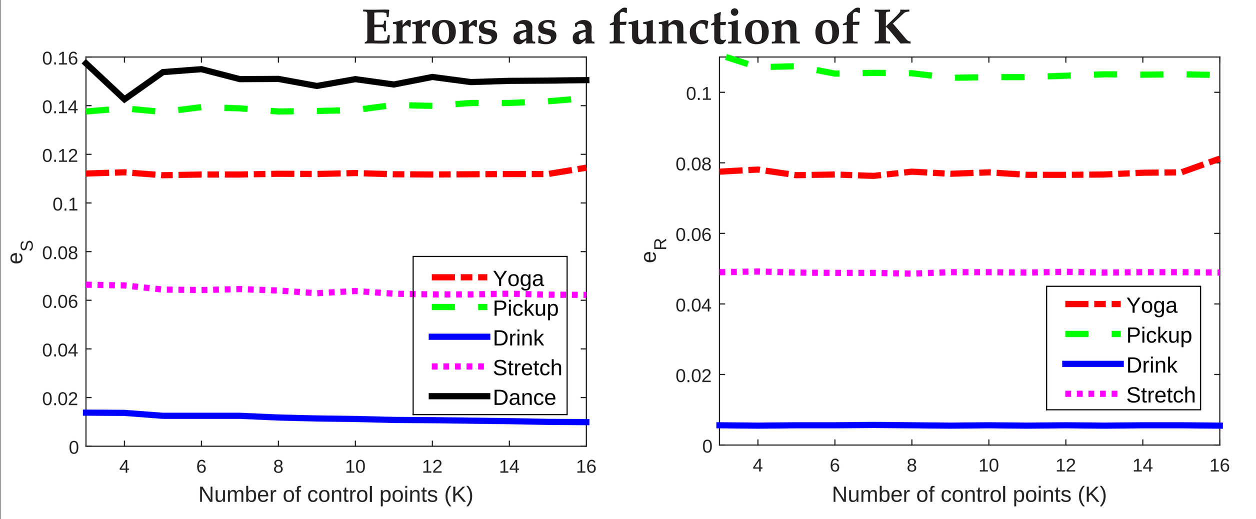
Quantitative Evaluation: Human Motion Capture Sequences

Data \ Met.	EM-PPCA [?]		MP [?]		PTA [?]		CSF [?]		KSTA [?]	
	e_R	$e_S(K)$	e_R	$e_S(K)$	e_R	$e_S(K)$	e_R	$e_S(K)$	e_R	$e_S(K)$
Noise-free observations										
Drink	.186	.261(7)	.330	.357(12)	.006	.025(13)	.006	.022(6)	.006	.020(12)
Stretch	.749	.458(7)	.832	.900(8)	.055	.109(12)	.049	.071(8)	.049	.064(11)
Yoga	.688	.445(8)	.854	.786(2)	.106	.163(11)	.102	.147(7)	.102	.148(7)
Pick-up	.417	.423(14)	.249	.429(5)	.155	.237(12)	.155	.230(6)	.155	.233(6)
Dance	–	.339(4)	–	.271(5)	–	.296(5)	–	.271(2)	–	.249(4)
Average error:		.385		.549		.166		.148		.143
Relative error:		4.16		5.93		1.79		1.60		1.54
Noisy observations										
Drink	.231	.250(7)	.329	.517(12)	.043	.045(13)	.043	.044(6)	.043	.042(12)
Stretch	.819	.886(7)	.872	.975(8)	.091	.144(12)	.091	.121(8)	.091	.166(11)
Yoga	.700	.507(8)	.858	.791(2)	.124	.174(11)	.125	.168(7)	.125	.172(7)
Pick-up	.499	.807(14)	.250	.407(5)	.148	.228(12)	.148	.224(6)	.148	.222(6)
Dance	–	.336(4)	–	.282(5)	–	.299(5)	–	.266(2)	–	.248(4)
Average error:		.557		.594		.178		.165		.170
Relative error:		4.64		4.95		1.48		1.37		1.42

Data \ Met.	BMM [?]		PPTA [?]		URS [?]			TRUS [?]			(Ours)		
	e_R	$e_S(K)$	e_R	$e_S(K)$	e_R	e_S	$e_G[\%]$	e_R	e_S	$e_G[\%]$	e_R	e_S	$e_G[\%]$
Noise-free observations													
Drink	.007	.027(12)	.006	.011(30)	.006	**.009**	0.8(2)	.006	**.009**	0.6(2)	.005	**.009**	0.6(2)
Stretch	.068	.103(11)	.058	.084(11)	.058	.061	4.1(3)	.058	**.060**	4.1(3)	.048	.062	4.3(3)
Yoga	.088	.115(10)	.106	.158(11)	.106	.143	0.3(2)	.091	.133	0.2(2)	.076	**.111**	0.1(2)
Pick-up	.121	.173(12)	.154	.235(12)	.154	.221	3.7(3)	.147	.209	3.0(3)	.104	**.138**	1.4(3)
Dance	–	.188(10)	–	.229(4)	–	.165	–	–	.150	–	–	**.143**	–
Average error:		.121		.143		.119			.112			**.092**	
Relative error:		1.31		1.54		1.28			1.21			**1.00**	
Noisy observations													
Drink	.044	.056(12)	.042	.038(30)	.042	.044	3.6(2)	.036	**.034**	1.4(2)	.037	.036	1.3(2)
Stretch	.098	.183(11)	.091	.123(11)	.091	**.119**	8.4(3)	.091	**.119**	5.1(3)	.091	.120	4.9(3)
Yoga	.136	.195(10)	.124	.174(11)	.125	.167	0.0(2)	.112	**.162**	0.2(2)	.115	.164	0.2(2)
Pick-up	.141	.212(12)	.148	.228(12)	.148	.207	3.1(3)	.147	.205	2.5(3)	.103	**.136**	1.2(3)
Dance	–	.236(10)	–	.222(4)	–	.164	–	–	.157	–	–	**.146**	–
Average error:		.176		.157		.140			.135			**.120**	
Relative error:		1.47		1.31		1.17			1.12			**1.00**	

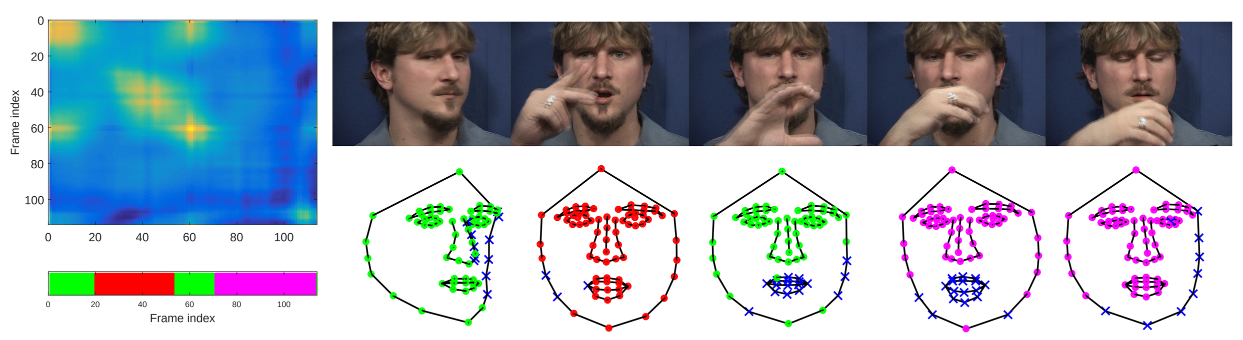
EM-PPCA (Expectation-Maximization Probabilistic Principal Component Analysis), MP (Metric Projections), PTA (Point Trajectory Approach), CSF (Column Subspace Fitting), KSTA (Kernel Shape Trajectory Approach), BMM (Block Matrix Approach), PPTA (Probabilistic Point Trajectory Approach), URS (Union of Regularized Subspaces), TRUS (Temporal Regularized Union of Subspaces)

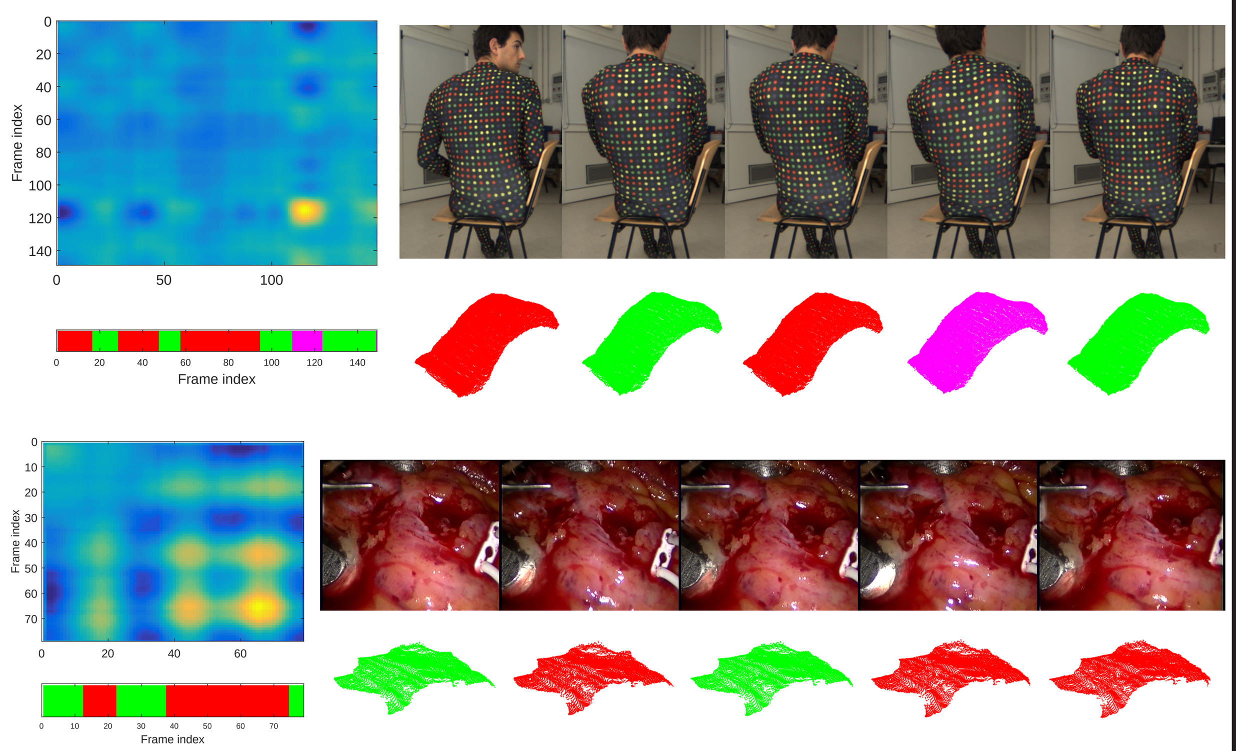
Errors as a function of K



- **Rotation and reconstruction errors:**
  Error reduction is consistent as K increases.
  Error always remains within reasonable bounds.
- **Accuracy and Generality:**
  Motion, 3D shape and grouping estimations.
  Accurate solutions for both noise-free and noisy observations.
  Efficient approach in a commodity laptop.

Qualitative Evaluation: Sparse Face, and Dense Back and Heart





- **Sparse data:**
  Robust to self-occlusions and lack of visibility.
  Detecting three type of deformations.
- **Dense data:**
  Videos from 20,561 to 68,295 points.
  Physically possible solutions.
  A wide variety of scenarios.

This work has been partially supported by the Spanish Ministry of Science and Innovation under project HuMoUR TIN2017-90086-R, by the ERA-Net Chistera project IPALM PCI2019-103386, and the María de Maeztu Seal of Excellence to IRI MDM-2016-0656.