Combining Features and Intensity for Wide-Baseline Non-Rigid Surface Registration

Jim Braux-Zin[1,2]
jim.braux-zin@cea.fr
Romain Dupont[1]
romain.dupont@cea.fr
Adrien Bartoli[2]
adrien.bartoli@gmail.com

[1] CEA, LIST
91 191 Gif-sur-Yvette
France
[2] ISIT, Université d'Auvergne
63 000 Clermont-Ferrand
France

Abstract

Non-rigid surface registration consists in estimating the deformation of a known flat surface, usually by fitting a warp such as a Thin-Plate Spline or a Free-Form Deformation. Common techniques are split in two categories: (1) pixel-based surface tracking where important deformations can be estimated over a video sequence as long as the frame to frame steps are small and (2) wide-baseline feature-based registration, able to directly handle large deformations at the cost of reduced accuracy. We introduce a new direct data term robustly merging feature and pixel-based costs in a pyramidal variational approach. By using a robust estimator we achieve an implicit optimal filtering of features and automatic balancing between the two terms.

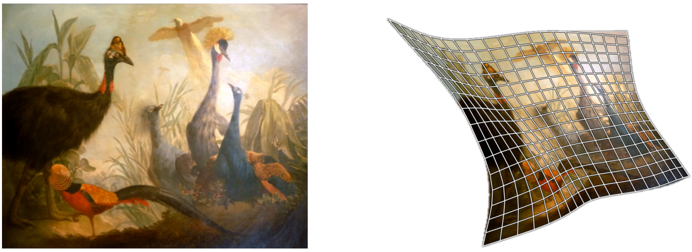


Figure 1: Example of surface registration with our method

1 Introduction

Non-rigid surface registration consists in estimating the deformation between two images of a known surface, usually by fitting a warp such as a Thin-Plate Spline or a Free-Form Deformation. It is a fundamental Computer Vision tool with applications such as augmented reality, non-rigid 3D reconstruction, or deformation analysis. Common techniques are split in two categories: detection and tracking. Pixel-based tracking [5, 6] tracks deformations over a video sequence through variational optimization and can handle large deformations as long as the frame to frame steps are small. Feature-based surface detection [9, 10, 16] directly estimates a potentially important deformation from an image and a source template using feature matches [2, 8]. Feature-based techniques tend to be more robust but less accurate than pixel-based ones since they discard much of the available information. Moreover, the warp fitting process needs an outlier filtering step using a robust estimator [9, 16] or a local smoothness prior [10] which is tricky to tune right in order to remove all outliers while keeping all good matches. There has been surprisingly few attemps to combine the strengths of the two approaches. Pizarro *et al.* [10] propose to initialize the pixel-based algorithm [6] with a feature-based warp but they still depend much on the quality of the former. In the more general field of optical flow estimation the integration of features into a variational estimation has been proven successful by Brox *et al.* [3]. However they rely on custom descriptors only suited for small-baseline displacements. We propose here an extension of this approach dedicated to non-rigid surface registration presented from the features filtering point of view. We propose to upgrade the pixel-based method [6] — already quite robust thanks to the handling of occlusions and the use of a coarse-to-fine optimization — with the introduction of a new feature-based term in the cost function. Thus we make this method viable for wide-baseline registration. Our framework is given in Section 2, starting from the models used (2.1), then the introduction of our feature-based cost (2.2) and at last the bending-energy and pixel-based priors we consider for match filtering (2.3). Section 3 includes qualitative and quantitative results with comparisons to state of the art feature-based techniques.

2 Proposed Framework

2.1 Models

Inputs. Our goal is to estimate the deformation between a flat template $\mathcal{I}_0$ and an image $\mathcal{I}$ of the deformed surface. Ω denotes the set of pixels of $\mathcal{I}_0$ and can be seen as continuous using for instance linear interpolation of the pixel values. Accompanying the image pair, a set $\mathcal{F}$ of point feature matches (which may contain erroneous matches) has been computed beforehand.

Deformation model. We use a two-dimensional Free-Form Deformation model based on cubic B-splines to parametrize the deformation. We use the formulation from [7] that we briefly recall here. We define a $n_x \times n_y$ regular grid of control points $\mathbf{s}_{i,j}$ with a step δ, and the displacements $\mathbf{u}_{i,j}$ associated to control points. The warp is defined as a linear combination of the 16 neighboring control points:

$$\mathcal{W}(\mathbf{q};\mathbf{u}) = \sum_{k=0}^{3}\sum_{l=0}^{3} \mathcal{B}_k(v)\mathcal{B}_l(w)(\mathbf{s}_{i+k,j+l} + \mathbf{u}_{i+k,j+l}) \tag{1}$$

where $\mathbf{q} = (q_x, q_y)^T$, $i = \text{floor}(q_x/n_x) - 1$, $j = \text{floor}(q_y/n_y) - 1$, $v = q_x/n_x - \text{floor}(q_x/n_x)$, $w = q_y/n_y - \text{floor}(q_y/n_y)$ and the $\mathcal{B}_k$ are the basis functions of the B-spline:

$$\mathcal{B}_0(v) = (1-v)^3/6 \qquad \mathcal{B}_1(v) = (3v^3 - 6v^2 + 4)/6$$
$$\mathcal{B}_2(v) = (-3v^3 + 3v^2 + 3v + 1)/6 \qquad \mathcal{B}_3(v) = u^3/6$$

This model has the desirable property of producing a twice continuously differentiable warp. Moreover, the control points only have a local influence which makes the warp estimation computationally cheap in many cases.

We define the discrete finite difference partial derivatives in a direction $\mathbf{d}$:

$$\begin{gathered} D_{\mathbf{d}}(\mathbf{q},\mathbf{u}) = \frac{\mathcal{W}(\mathbf{q}+\delta\mathbf{d},\mathbf{u})-\mathcal{W}(\mathbf{q}-\delta\mathbf{d},\mathbf{u})}{2\delta} \\ D_{\mathbf{d}}^{(l)}(\mathbf{q},\mathbf{u}) = \frac{\mathcal{W}(\mathbf{q},\mathbf{u})-\mathcal{W}(\mathbf{q}-\delta\mathbf{d},\mathbf{u})}{\delta} \qquad D_{\mathbf{d}}^{(r)}(\mathbf{q},\mathbf{u}) = \frac{\mathcal{W}(\mathbf{q}+\delta\mathbf{d},\mathbf{u})-\mathcal{W}(\mathbf{q},\mathbf{u})}{\delta} \end{gathered} \tag{2}$$

Cost model. We adopt the following model for the cost function of non-rigid registration:

$$\varepsilon(\mathbf{u},\mathcal{F},\mathcal{I},\mathcal{I}_0) = \varepsilon_f(\mathbf{u},\mathcal{F}) + \varepsilon_{\text{prior}}(\mathbf{u},\mathcal{I},\mathcal{I}_0) \tag{3}$$

where ε_f is our feature-based cost function given in Section 2.2 and $\varepsilon_{\text{prior}}$ is an energy encoding other constraints (smoothness of the warp, self-occlusion handler and pixel-wise brightness constancy) to filter out false matches and regularize the output.

2.2 Feature-Based Cost

We call $\mathcal{F}$ the set of matches composed of two point features $\mathbf{f}_0 \in \mathcal{I}_0$ and $\mathbf{f} \in \mathcal{I}$. As in [3], the feature-based cost is naturally a function of the distance between the warp displacement $\mathbf{u}$ and the displacement $\mathbf{f}-\mathbf{f}_0$ induced by the matches. We have

$$\varepsilon_f(\mathbf{u},\mathcal{F}) = \lambda_f \sum_{(\mathbf{f}_0,\mathbf{f})\in\mathcal{F}} \iint_{\Omega} \omega(\mathbf{q},\mathbf{f}_0)\Psi_\sigma\left(\|\mathbf{u}-(\mathbf{f}-\mathbf{f}_0)\|_2\right) \mathrm{d}\mathbf{q} \tag{4}$$

The bilinear influence fonction ω has a radius of one pixel and is normalized to make the overall influence of features independent from their density. If $(d_x,d_y)^\mathsf{T} = \text{abs}(\mathbf{q}-\mathbf{f}_0)$, it is defined as:

$$\hat{\omega}(\mathbf{q},\mathbf{f}_0) = (1-d_x)(1-d_y) \text{ if } d_x < 1 \text{ and } d_y < 1, 0 \text{ otherwise} \tag{5}$$

$$\omega(\mathbf{q},\mathbf{f}_0) = \frac{\hat{\omega}(\mathbf{q},\mathbf{f}_0)}{\sum_{(\mathbf{g}_0,\mathbf{g})\in\mathcal{F}} \hat{\omega}(\mathbf{q},\mathbf{g}_0)} \tag{6}$$

When several features are associated to the same control points, the contribution of the feature-based cost can be seen as a weighted vote of each feature.

M-estimator for implicit outliers filtering. When dealing with unfiltered features including erroneous matches, a robust estimator is needed to deal with outliers. Brox *et al.* [3] use the Huber approximation of the L^1 norm, while Pilet *et al.* use a non convex estimator which was shown to be efficient at filtering the matches when reducing its radius at each iteration in an annealing manner. We use the classical Geman McClure estimator

$$\Psi_\sigma(x) = \frac{x^2}{\sigma + x^2} \tag{7}$$

which has a strong filtering power without any explicit thresholding (see Figure 2). The integration into the optimization is straightforward using iteratively reweighted least squares with weights: $w(x) = \frac{1}{x}\frac{\mathrm{d}\Psi(x)}{\mathrm{d}x} = \frac{2\sigma}{(\sigma+x^2)^2}$.

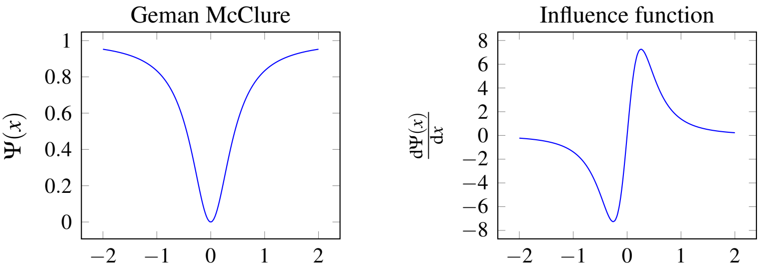


Figure 2: Geman McClure M-estimator with $\sigma = 0.2$

Pyramidal scheme for automatic balancing. The use of coarse-to-fine optimization exhibits automatic balancing of the feature-based cost. Indeed, as shown in [3], the features are quasi-dense at low resolution, constraining strongly the updates of the warp, while at high resolution, only a few control points are affected and the regularization limits their influence. Moreover, the pyramidal approach allows one to use a constant radius for our robust M-estimator during the whole process with each upsampling step naturally increasing its selectivity.

Discussion. The cost function (4) looks similar to the ones proposed in [9] and [3] but there are some key differences that explain our better results (see Section 3). Contrary to them we do not rely on any confidence measure of the matches, often unreliable and much less important than the spatial coherency implied by the robust estimator as shown in [9]. This allows us to use any current and future feature matcher without changes in our algorithm. Moreover by using all the available information — features, intensity, suitable deformation model and self-occlusions handling — we have the unique combination of a robust non-convex estimator with a rich prior dedicated to non-rigid deformations.

2.3 Priors

We propose a combination of three complementary energies as the prior cost introduced in (3): a global regularizer, a term handling self-occlusions and a pixel-based cost that we explain in the following paragraphs:

$$\varepsilon_{\text{prior}}(\mathbf{u}, \mathcal{I}, \mathcal{I}_0) = \varepsilon_b(\mathbf{u}) + \varepsilon_s(\mathbf{u}) + \varepsilon_d(\mathbf{u}, \mathcal{I}, \mathcal{I}_0) \tag{8}$$

Bending energy. For smooth deformations of deformable surfaces without crumpling, the bending energy is a well suited constraint [1]. It penalizes the variations of the second order derivatives and is defined as:

$$\varepsilon_b(\mathbf{u}) = \lambda_b \iint_\Omega \left(\frac{\partial \mathcal{W}(\mathbf{q};\mathbf{u})^2}{\partial^2 q_x} \right) + \left(\frac{\partial \mathcal{W}(\mathbf{q};\mathbf{u})^2}{\partial q_x \partial q_y} \right) + \left(\frac{\partial \mathcal{W}(\mathbf{q};\mathbf{u})^2}{\partial^2 q_y} \right) \mathrm{d}\mathbf{q} \tag{9}$$

Self-occlusion handling. Self-occlusions appear when a deformable surface folds such as part of it is hidden by another of its parts. They can be handled gracefully by noting that the derivative of the warp vanishes in one direction [6]. Using the finite difference approximation (2) this translates to:

$$\mathbf{q} \text{ self-occluded} \quad \Leftrightarrow \quad \exists \mathbf{d} \mid D_{\mathbf{d}}(\mathbf{q};\mathbf{u}) = 0 \tag{10}$$

It has been shown [6] that the smallest partial derivative σ_0 is linked to the Jacobian $\mathbf{J}$ by $\sigma_0 = \min_{\|\mathbf{d}\|=1} \mathbf{d}^\top \mathbf{J}^\top \mathbf{J} \mathbf{d}$ and after spectral decomposition of $\mathbf{O} = \mathbf{J}^\top \mathbf{J}$:

$$\sigma_0 = \frac{1}{2}\left(\mathbf{O}_{11} + \mathbf{O}_{12} - \sqrt{(\mathbf{O}_{11} - \mathbf{O}_{22})^2 + 4\mathbf{O}_{12}^2}\right) \tag{11}$$

A smooth step function $\mathcal{S}(x,k,r) = \frac{1}{1+\exp(-k(x-r))}$ translates σ_0 to a self-occlusion probability:

$$\mathcal{P}_{\text{SO}} = 1 - \mathcal{S}(\sigma_0, 40, 0.1) \tag{12}$$

In order for the criterion (10) to hold, the warp must be prevented from folding in self-occluded areas, which is achieved through the addition of a dedicated *shrinker* term:

$$\begin{aligned} \varepsilon_s(\mathbf{u}) &= \lambda_s \sum_{\mathbf{q}\in\Omega} \sum_{\mathbf{d}\in\mathcal{D}} \sum_{c\in\{x,y\}} \gamma\left\{ \left(D_{\mathbf{d}}^{(l)}(\mathbf{q};\mathbf{u})\right)_c \left(D_{\mathbf{d}}^{(r)}(\mathbf{q};\mathbf{u})\right)_c \right\} \\ \gamma(x) &= 0 \text{ if } x \geq 0 \text{ and } x^2 \text{ otherwise} \end{aligned} \tag{13}$$

where $\mathcal{D}$ is a discretized set of directions and the function γ penalizes points where the right and left derivatives have opposite signs (see [6] for details).

Pixel-based data term. The data term used is based on the brightness constancy assumption: corresponding pixels in the two images are assumed to have similar intensities. The resulting data cost is the sum of the squared differences:

$$\varepsilon_d(\mathbf{u}, \mathcal{I}, \mathcal{I}_0) = \lambda_d \sum_{\mathbf{q}\in\Omega} (1 - \mathcal{P}_{\text{SO}})\big(\mathcal{I}_0(\mathbf{q}) - \mathcal{I}(\mathcal{W}(\mathbf{q};\mathbf{u}))\big)^2 \tag{14}$$

The direct data term is not to be trusted in self-occluded areas, so we multiply it by $1 - \mathcal{P}_{\text{SO}}$.

This model fails in the presence of illumination changes and several solutions have been proposed to address this issue such as structure-texture decomposition [17], Light-Invariant [11] or CENSUS [12, 18] transforms. Our feature-based data term is independent of the direct term used so we restrict ourselves to the simple model (14) for clarity.

Discussion. It is unusual to consider a pixel-wise error as a prior but we made this conceptual choice to fit our approach into the model (3) and allow an easier comparison with methods based on feature filtering. Moreover, in the wide-baseline setting the feature-based data term can usually converge without the pixel-wise prior but the inverse is not true, which proves that the latter is weaker.

3 Experimental Results

In this section, we demonstrate the validity of the joint optimization of the feature-based and direct data terms. We use the same parameters for all the experiments: 6 pyramid levels, a grid step $\delta = 5$px, $\lambda_d = 1$, $\lambda_f = 800$, $\lambda_b = 5000$, $\lambda_s = 100000$, $\sigma = 0.2$. Our Matlab implementation is based on the publicly available code of [6] for the Gauss-Newton optimization of a direct data term. We compare our method with the public implementations of state of the art feature filtering methods FBDSD [10] based on local smoothness assumption and RANSAC [16] based on plane fitting with a robust estimator. The general optical flow methods such as LDOF [3] are not designed for strong deformations and do not produce results worth mentioning on the following experiments. The feature matches are obtained with the OpenCV implementations of the FAST [13] and SIFT [8] detectors, and the SIFT descriptors, with default parameters. A cross-check filtering step eliminates ambiguous matches. Qualitative results Figures 3 and 4 show the accuracy and robustness of our method demonstrated in the next sections through in-depth quantitative evaluations.

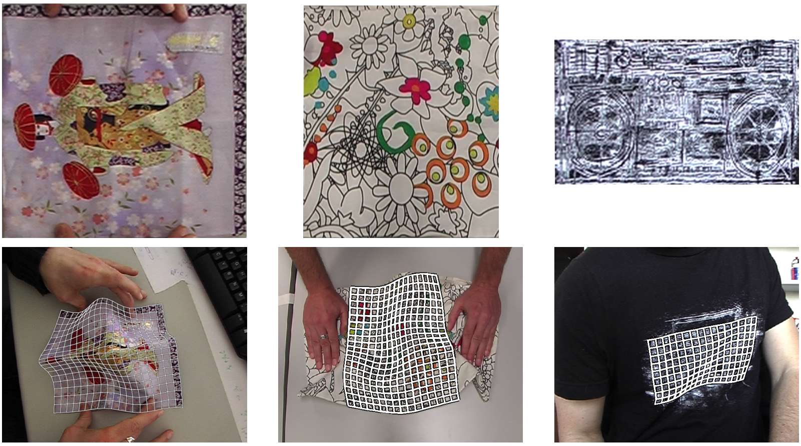


Figure 3: Qualitative results on the EPFL dataset [14, 15]. First row: templates, second row: deformed surface with estimated deformation.

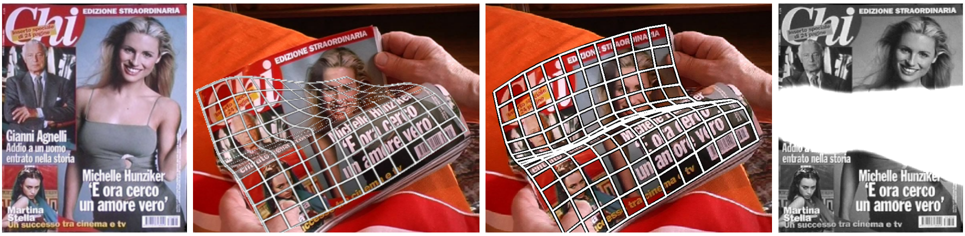


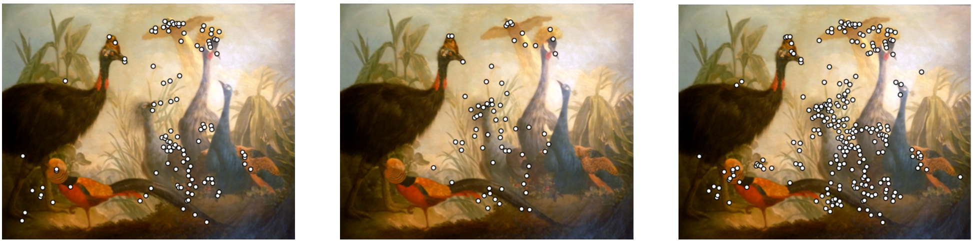
Figure 4: Result on a challenging case from the ETHZ toys dataset [4]. From left to right: template, estimated deformation with FBDSD [10], estimated deformation and estimated self-occlusions (in white) with our method.

3.1 A Case Against Feature-Only Filtering

In this first experiment, we use a synthetic deformation showed in Figure 1 which provides ground truth. We generate 331 feature matches. We run the two filtering approaches on those matches, and refine the output with a pass of the pixel-based method [6]. The average warp errors are listed in Table 1. To explain the results, we separate the matches into 111 inliers within a 2 pixels of the ground-truth warp, and 220 outliers and compare them with the filtered features. Figure 5 shows the output of the feature filtering approaches. For the warp fitting process, it is crucial that the features are as uniformly distributed on the template as possible. FBDSD seems to be overly selective with all the remaining features concentrated in the middle of the template. RANSAC is on the contrary too permissive.

We believe this experiment demonstrates that separating the feature filtering and warp fitting steps is misguided because each optimization is based on incomplete knowledge and cannot compensate the weaknesses of the other. On the other hand, our joint optimization produces near-optimal results from the same data.

With this 320 × 400 template, the processing time of the unoptimized Matlab implementation is approximately 20 seconds. This is longer than the feature-based techniques but almost the same as the method [6] since our feature-based term adds only little overhead.



(a) 111 inliers (< 2px)

(b) FBDSD [10] filtering

43 false positives
86 false negatives

(c) RANSAC [16] filtering

143 false positives
2 false negatives

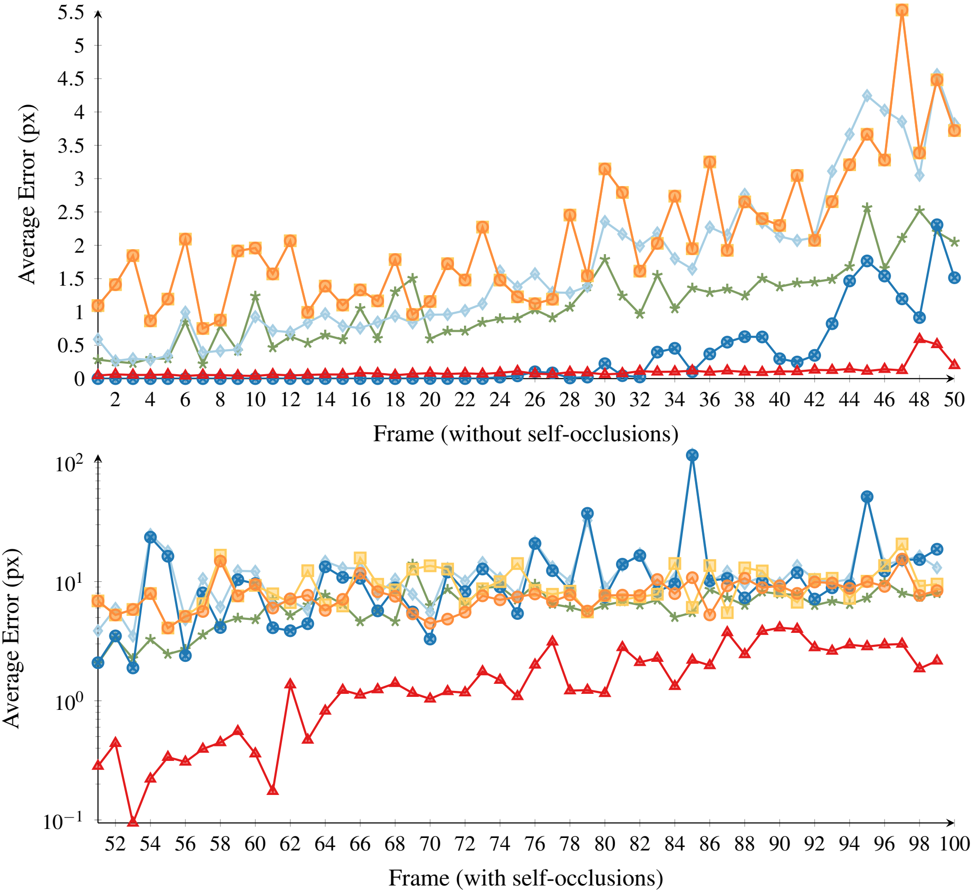
Figure 5: The feature-only filtering strategies either miss some true matches or leave false matches. See text Section 3.1.

Method	FBDSD	FBSD+P	RANSAC	RANSAC+P	Inliers	Inliers+P	Ours
Average error (px)	12.234	6.37	7.91	5.07	6.92	4.73	1.35

Table 1: Comparison of feature-based approaches on the image pair Figure 1. The "+P" variants are refined warps with pixelic information using [6]. "Inliers" designates the warp fitted to the 2px inliers (see text).

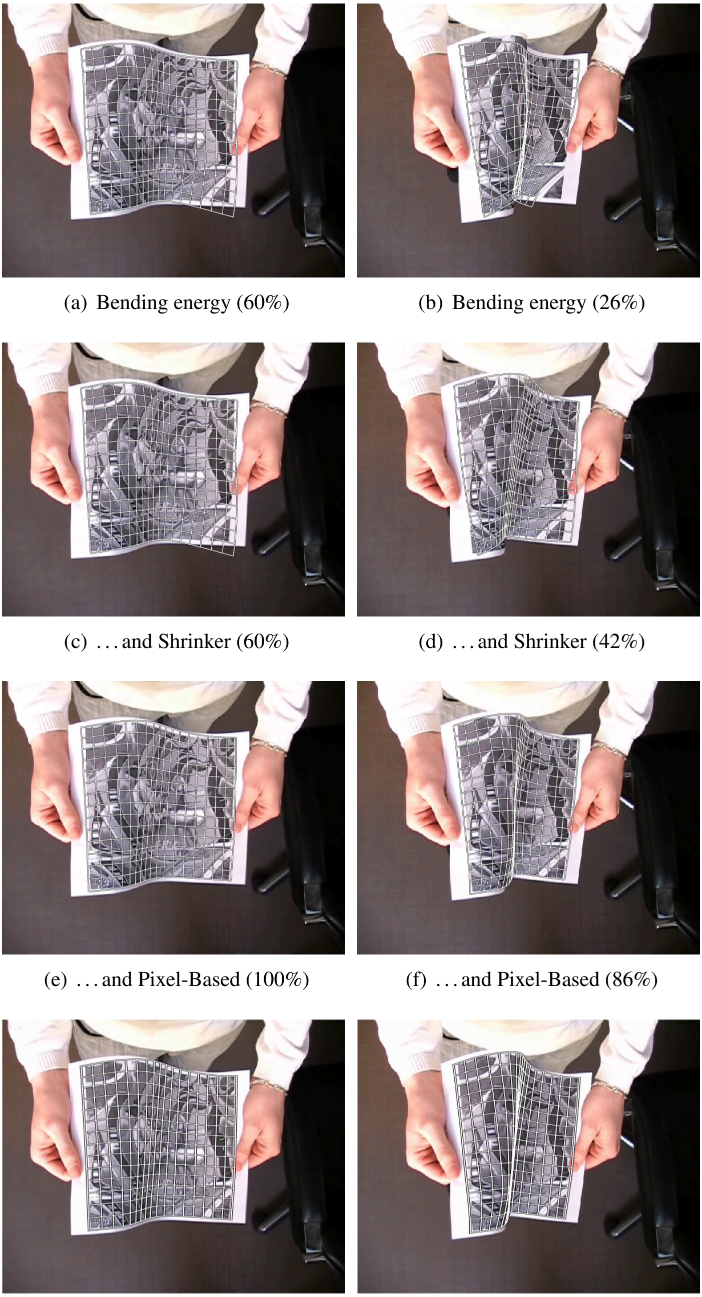
3.2 Quantitative Evaluation on a Real Sequence

After this synthetic experiment, we observe the behaviors of the considered algorithms on a real sequence. We use the sequence accompanying the implementation of [6]. The tracking-based results [6] are accurate enough compared to feature-based methods to serve as a ground truth. Over the 100 first frames of the sequence we compare the feature-based approaches using only the template, the current frame and SIFT matches. We also use this sequence to evaluate the different priors presented Section 2.3 and run our method with only the bending energy regularization (9) which is almost equivalent to [9], then adding the shrinker term (13) for better behavior in the presence of self-occlusions, and at last adding the direct data term (14) to exploit all the available information.



Method	average error w/o SO	with SO
—✶— RANSAC [16]	1.07	6.22
—◆— FBDSD [10]	1.59	14.06
—⊗— FBDSD+P [10]	0.29	12.74
—■— Bending energy regularization (9)	2.02	9.14
—●— . . . and shrinker term (13)	2.02	7.74
—▲— . . . and pixel-based data term (14)	0.10	1.60

Figure 6: Evolution of the Free-Form Deformation control points average error relative to the tracking based method [6]. Our method is evaluated with the bending energy prior —■—, adding the shrinker term —●— for better robustness to self-occlusions, and the pixel-based data term for better accuracy —▲—. For reference we include the results of the methods [16] (RANSAC) —✶— and [10] (FBDSD) feature-based warps —◆— and refined with pixellic data term (FBDSD+P) —⊗—. Best viewed in color.



(a) Bending energy (60%)

(b) Bending energy (26%)

(c) . . . and Shrinker (60%)

(d) . . . and Shrinker (42%)

(e) . . . and Pixel-Based (100%)

(f) . . . and Pixel-Based (86%)

(g) Tracking-based [6]

(h) Tracking-based [6]

Figure 7: Comparison of our results using SIFT matches (198 matches on the left, 155 matches on the right) with the tracking based method [6]. The percentage is the amount of control points within a 2 pixels radius of the tracking-based results.

Quantitative results are plotted Figure 6 and show that our approach is clearly superior except for very small deformations where it is on par with current state of the art. Sample output on two frames displayed Figure 7 stresses the importance of the priors. The shrinker has no effect in the absence of self-occlusions but can bring a great improvement otherwise. The pixel-based data term is essential for correct fitting on the warp near the boundaries of the surface, where feature density is usually low.

Conclusion

We introduced a new model of non-rigid surface registration to jointly optimize feature and pixel-based costs through a pyramidal variational scheme. We demonstrated the viability of the approach and showed results where we clearly outperformed other state of the art methods, quantitatively and qualitatively on synthetic and real data. All results were obtained with the same parameter set which proves that our contribution is robust. Future works involve incorporating better features and more robust pixel-based data terms. A real-time implementation on a compiled language is also envisioned for augmented-reality applications.

Acknowledgement. This research has received funding from the EU's FP7 through the ERC research grant 307483 FLEXABLE.

References

[1] A. Bartoli, M. Perriollat, and S. Chambon. Generalized Thin-Plate Spline warps. *IJCV*, 2010.

[2] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded Up Robust Features. *ECCV*, 2006.

[3] T. Brox and J. Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *PAMI*, 2011.

[4] V. Ferrari, T. Tuytelaars, and L. Gool. Simultaneous object recognition and segmentation by image exploration. In *ECCV*. 2004.

[5] R. Garg, A. Roussos, and L. Agapito. A variational approach to video registration with subspace constraints. *IJCV*, 2013.

[6] V. Gay-Bellile, A. Bartoli, and P. Sayd. Direct estimation of nonrigid registrations with image-based self-occlusion reasoning. *PAMI*, 2010.

[7] S. Lee, G. Wolberg, K. Chwa, and S. Shin. Image metamorphosis with scattered feature constraints. *Visualization and Computer Graphics*, 1996.

[8] D. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *IJCV*, 2004.

[9] J. Pilet, V. Lepetit, and P. Fua. Fast Non-Rigid Surface Detection, Registration and Realistic Augmentation. *IJCV*, 2008.

[10] D. Pizarro and A. Bartoli. Feature-based deformable surface detection with self-occlusion reasoning. *IJCV*, 2012.

[11] D. Pizarro, J. Peyras, and A. Bartoli. Light-invariant fitting of active appearance models. In *CVPR*, 2008.

[12] R. Ranftl, S. Gehrig, T. Pock, and H. Bischof. Pushing the limits of stereo using variational stereo estimation. In *Intelligent Vehicles Symposium (IV)*, 2012.

[13] E. Rosten, R. Porter, and T. Drummond. Faster and better: A machine learning approach to corner detection. *PAMI*, 2010.

[14] M. Salzmann, R. Hartley, and P. Fua. Convex optimization for deformable surface 3-d tracking. In *ICCV*, 2007.

[15] Mathieu Salzmann and Pascal Fua. Reconstructing sharply folding surfaces: A convex formulation. In *CVPR*, 2009.

[16] Q. Tran, T. Chin, G. Carneiro, M. Brown, and D. Suter. In defence of RANSAC for outlier rejection in deformable registration. In *ECCV*, 2012.

[17] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers. An improved algorithm for TV-L1 optical flow. In *Statistical and Geometrical Approaches to Visual Motion Analysis*. 2009.

[18] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *ECCV*, 1994.