Direct Estimation of Non-Rigid Registrations with Image-Based Self-Occlusion Reasoning

V. Gay-Bellile$^{1,2}$, A. Bartoli$^1$, P. Sayd$^2$

$^1$ LASMEA, Clermont-Ferrand, France
$^2$ LCEI, CEA LIST, Saclay, France

Vincent.Gay-Bellile@univ-bpclermont.fr

Abstract

The registration problem for images of a deforming surface has been well studied. External occlusions are usually well-handled. In 2D image-based registration, self-occlusions are more challenging. Consequently, the surface is usually assumed to be only slightly self-occluding.

This paper is about image-based non-rigid registration with self-occlusion reasoning. A specific framework explicitly modeling self-occlusions is proposed. It is combined with an intensity-based, i.e. direct, data term for registration. Self-occlusions are detected as shrinking areas in the 2D warp.

Experimental results on several challenging datasets show that our approach successfully registers images with self-occlusions while effectively detecting the occluded regions.

1. Introduction

Registering monocular images of a deforming surface is important for tasks such as video augmentation by texture editing, non-rigid Structure-from-Motion and deformation capture. This is a difficult problem since the appearance of imaged surfaces varies due to several phenomena such as camera pose, surface deformation, lighting, motion blur and occlusions. The latter causes difficult issues in 2D image registration. They can be classified into external occlusions and self-occlusions. External occlusions are caused by an object entering the space between the camera and the surface of interest. Self-occlusions are due to the surface being bent in such a way that a part of it occludes another one. Deformation estimation in the presence of self-occlusions could be formulated in 3D, see e.g. [7]. However, recovering a 3D surface, its deformations and the camera pose from a monocular video sequence is intrinsically ill-posed. While prior information can be used to disambiguate the problem, see e.g. [3], it is common to avoid a full 3D model by using image-based deformation models, e.g. [1, 5, 8].

Previous work on 2D image registration, e.g. [1, 8, 11, 9, 13], usually deals with external and self-occlusions within an outlier rejection framework, e.g. the X84 rule, based on the data term. These methods handle external occlusions and very limited amounts of self-occlusion, as figure 1 illustrates.

Figure 1. Classical methods fail on extreme self-occlusions. Example of image registration using a classical outlier rejection algorithm (see §5 for details). Left: success on an external occlusion. Right: failure on an extreme self-occlusion.

Self-occlusions thus have to be dealt with in a different manner. We propose a specific framework for non-rigid registration in spite of self-occlusions. The basic idea is to consider self-occluded pixels not as outliers, but as points at which the unknown warp is injective. This is implemented with two components: The warp is constrained to shrink rather than to ‘fold’ and self-occlusions are detected as the warp shrinkage areas. This allows accurate deformation estimation in the presence of self-occlusions. Experimental results demonstrating our framework are reported.
Roadmap. Previous work is reviewed in §2. Self-occlusion detection is explained in §3. Direct registration with self-occlusion reasoning is proposed in §4. Experimental results on real data are reported in §5. Finally, we give our conclusions and discuss future work in §6.

Notation. The images to be registered are written $I_i$ with $i = 1, \ldots, n$. The template, e.g. the occlusion-free region of interest in the first image, is denoted $I_0$. The warp is written $W$. It depends on a parameter vector $u_i$ for image $I_i$ and maps a points $q$ in the template to the corresponding point $q_i$ in image $i$ as: $q_i = W(q; u_i)$.

We drop the image index $i$ for clarity reasons in most of the paper. We write $\mathcal{R}$ the set of pixels of interest. We respectively denote $E_c(d; q; u)$, $E_l(d; q; u)$ and $E_r(d; q; u)$ the central, left and right derivatives of the warp along direction $d \in S^1$, e.g. $E_c(d; q; u) = \frac{W(q+d; u) - W(q-d; u)}{2\epsilon}$ ($S^1$ is the unit circle). We denote $v^x$ and $v^y$ the $x$-component and the $y$-component of a two-dimensional vector $v$.

2. Previous Work

Registering images of deformable surfaces has received a growing attention over the past few years. However, the self-occlusion issue is usually not explicitly tackled.

In [11], a feature-based non-rigid surface detection algorithm is proposed. The robustness of this approach to self-occlusions is not described clearly but experimental results show that only small self-occlusions are handled. Dealing with extreme self-occlusions such as those of figures 5 is very challenging. The number of features might not be large enough, in particular in the neighborhood of the self-occlusion boundary, to recover the correct warp. Direct non-rigid image registration generally yields more accurate deformation estimates. Classical methods, e.g. [1, 8] do not take self-occlusions into account. Some works explored this issue by using specific patterns. For example, Lin et al. [9] proposed a very robust algorithm for tracking a Near Regular Texture (NRT). A visibility map is used for dealing with external and self-occlusions. It is based on geometric and appearance properties of NRTs.

A method for retexturing non-rigid objects from a single viewpoint is proposed in [13]. This approach is based on texture and shading information. Deformable surface tracking is required to obtain texture coordinate before retexturing. The most impressive results are obtained with specific patterns. They allow to obtain very accurate texture coordinates even if some areas are slightly occluded.

A classical technique such as the z-buffer allows to predict the occluding contour when a 3D model is used. Ilic et al. [7] uses an implicit representation of 3D shapes. Occluding contours are computed as the solution of an ordinary differential equation. This allows one to recover the shape of a deforming object in monocular videos. This method can be applied to a restricted set of self-occlusions only: those for which the boundaries can not be a priori segmented defeat this method, see e.g. the one shown on figure 5.

To summarize, dealing with self-occlusions in purely 2D image-based registration is a very challenging problem that has not yet received a commonly agreed solution in the community.

3. Self-occlusion Detection Framework

3.1. Overview

The basic idea is to consider self-occluded pixels as points at which the unknown warp shrinks. Note that a natural warp behavior such as folding does not allow detection in 2D. We thus use the following three basic components. First, the warp is constrained not to fold onto itself but to shrink along the self-occlusion boundaries. Second, benefiting from this property, the warp is used to detect the self-occluded pixels. Third, these pixels are ignored in the data term in the error function. This is in contrast with classical methods which use the image data, i.e. the pixel intensities, for rejecting self-occluded parts as outliers.

In other words, the warp is constrained to be one-to-one in visible and externally occluded parts and many-to-one in self-occluded regions. The proposed approach, based on the shrinking property to detect the self-occluded pixels, outputs a binary self-occlusion map $H(q; u)$ with $q \in \mathcal{R}$ and $H(q; u) = 0$ if pixel $q$ is occluded by the surface and $H(q; u) = 1$, otherwise.

3.2. Preventing Foldings, Enforcing Shrinkage

A regularized warp naturally folds onto himself in case of extreme self-occlusions. Warp shrinkage is enforced by penalizing loops and folds, via the penalty term $E_{\text{sh}}$. This behavior is required by the self-occlusion detection modules described in §3.3. Folds make the warp many-to-many in visible parts. These configurations are characterized by a variation in the sign of the partial derivatives of the warp along some direction. We note that diffeomorphic warps are proposed in [5]. They enforce one-to-one correspondences by preventing warp folds. Diffeomorphic warps are unadapted in our context since we model self-occluded areas by a many-to-one warp.

Our penalty is built via the following function:

$$\gamma(r) = \begin{cases} 0 & \text{if } r \ge 0 \\ r^2 & \text{otherwise.} \end{cases}$$

It is applied to the element-wise product between the left and the right derivatives of the warp evaluated at the points in $\mathcal{R}$ and integrated along direction $d \in S^1$. It allows to penalize those points for which the right and left derivatives...
have opposite signs. It enforces the warp not to fold but rather to shrink along the self-occluded area. The shrinking constraint is given by:
\[
\mathcal{E}_{\text{sh}}(\mathbf{u}) = \sum_{\mathbf{q} \in \mathbb{R}} \int_{\mathbf{d} \in \mathbb{S}^1} \sum_{c \in \{x,y\}} \gamma((\mathbf{E}_c^d(\mathbf{d}, \mathbf{q}, \mathbf{u}))E_c^d(\mathbf{d}, \mathbf{q}, \mathbf{u}))d\mathbf{d}.
\]
In practice, the integral is discretized on a set of 8 directions.

3.3. Detecting Self-Occlusions

One consequence of the shrinking property is that for any self-occluded pixel \( \mathbf{q} \), there exists at least one direction \( \mathbf{d} \in \mathbb{S}^1 \) such that the partial derivatives of the warp at \( \mathbf{q} \) in direction \( \mathbf{d} \) vanish. We thus define the self-occlusion map \( \mathcal{H}(\mathbf{q}; \mathbf{u}) \) as:
\[
\mathcal{H}(\mathbf{q}; \mathbf{u}) = \begin{cases} 0 & \exists \mathbf{d} \in \mathbb{S}^1 \mid \|\mathbf{E}_c(\mathbf{d}; \mathbf{q}; \mathbf{u})\| < r \\ 1 & \text{otherwise.} \end{cases}
\]

\( \|\mathbf{E}_c(\mathbf{d}; \mathbf{q}; \mathbf{u})\| = 0 \) implies that points \( \mathcal{W}(\mathbf{q} + c\mathbf{d}; \mathbf{u}) \) and \( \mathcal{W}(\mathbf{q} - c\mathbf{d}; \mathbf{u}) \) are identical. We fix \( r \) slightly over 0, e.g. 0.1, in order to tolerate noisy warps. In practice, the exhaustive search of directions \( \mathbf{d} \) is replaced by a minimization problem:
\[
\sigma_0 = \min_{\mathbf{d} \in \mathbb{S}^1} \|\mathbf{E}_c(\mathbf{d}; \mathbf{q}; \mathbf{u})\|^2.
\]
A pixel \( \mathbf{q} \) is labeled as self-occluded by comparing \( \sigma_0 \) with the threshold \( r \). This minimization problem has a closed-form solution. Let \( \mathcal{J} \) be the Jacobian matrix of \( \mathcal{W} \) evaluated at \( (\mathbf{q}; \mathbf{u}) \), i.e. \( \mathcal{J} = \left( \begin{array}{cc} \frac{\partial \mathcal{W}_x}{\partial x} & \frac{\partial \mathcal{W}_x}{\partial y} \\ \frac{\partial \mathcal{W}_y}{\partial x} & \frac{\partial \mathcal{W}_y}{\partial y} \end{array} \right) \). We have \( \mathbf{E}_c(\mathbf{d}; \mathbf{q}; \mathbf{u}) = \mathcal{J}\mathbf{d} \), and thus:
\[
\sigma_0 = \min_{\mathbf{d} \in \mathbb{S}^1} \|\mathbf{E}_c(\mathbf{d}; \mathbf{q}; \mathbf{u})\|^2 = \min_{\mathbf{d} \in \mathbb{S}^1} \mathbf{q}^T \mathcal{J}^T \mathcal{J} \mathbf{d}.
\]
Spectral decomposition of the symmetric matrix \( \mathcal{O} = \mathcal{J}^T \mathcal{J} \) gives:
\[
\sigma_0 = \frac{1}{2} \left( \mathcal{O}_{1,1} + \mathcal{O}_{2,2} - \sqrt{2} \right).
\]

4.4. The Cost Function

Direct non-rigid registration algorithms usually use as discrepancy function \( \mathcal{E}_{\text{data}} \) the two-norm of the difference \( \mathcal{D} \) between the template and the current image, warped toward the template, i.e. \( \mathcal{D}(\mathbf{q}) = \mathcal{I}_0(\mathbf{q}) - \mathcal{I}(\mathcal{W}(\mathbf{q}; \mathbf{u})) \), giving:
\[
\mathcal{E}_{\text{data}}(\mathbf{u}) = \sum_{\mathbf{q} \in \mathbb{R}} \mathcal{D}^2(\mathbf{q}; \mathbf{u}).
\]

More sophisticated data terms [10, 12] robust to noise or lighting variations can also be used.

For purely twodimensional registration, as is the case in this paper, smoothness constraints are used. These soft constraints can be implicitly incorporated in a parameterized warp as e.g. in Thin-Plate Spline warps. We use a simple regularization term \( \mathcal{E}_{\text{reg}} \) that is added to the error function:
\[
\sum_{\mathbf{q} \in \mathbb{R}} \mathcal{D}^2(\mathbf{q}; \mathbf{u}) + \lambda_{\text{reg}} \mathcal{E}_{\text{reg}}(\mathbf{u}).
\]

Denote \( \mathcal{U} \) the displacement field reshaped on the mesh grid:
\( \mathbf{u} = \text{vect}(\mathcal{U}) \). The regularization penalty on the displacement field \( \mathcal{U} \) is a discrete approximation to the bending energy:
\[
\mathcal{E}_{\text{reg}}(\mathbf{u}) = \int_{\mathbb{R}} \int_{\mathbb{R}} \left( \frac{\partial \mathcal{U}}{\partial x} \right)^2 + 2 \left( \frac{\partial \mathcal{U}}{\partial x\partial y} \right)^2 + \left( \frac{\partial \mathcal{U}}{\partial y} \right)^2 \, dx \, dy.
\]

Other choices are possible. The bending energy empirically appears to be well suited for the case of smooth surfaces.

Dealing with occlusions is usually done through a robust estimator in the data term:
\[
\sum_{\mathbf{q} \in \mathbb{R}} \rho (\mathcal{D}^2(\mathbf{q})) + \lambda_{\text{reg}} \mathcal{E}_{\text{reg}}(\mathbf{u}).
\]

As said above, it is not sufficient for self-occlusions. We rather weight each term by the self-occlusion map \( \mathcal{H} \) described in §3. Simultaneously estimating the self-occlusion map and the displacement field is complicated and subject to many local minima. A two-step minimization scheme is used. First, the parameter vector is updated by using the current estimate of the self-occlusion map. Second, the latter is updated with equation (1). The associated error function is given by:
\[
\sum_{\mathbf{q} \in \mathbb{R}} \mathcal{H}(\mathbf{q}; \mathbf{u}) \mathcal{D}^2(\mathbf{q}; \mathbf{u}) + \lambda_{\text{reg}} \mathcal{E}_{\text{reg}}(\mathbf{u}),
\]
where \( \mathbf{u} \) is the current parameter estimate.

Finally, the shrinking penalty is added to the cost function. The global energy to be minimized is thus:
\[ E(u) = \sum_{q \in R} H(q; \tilde{u})D^2(q; u) + \lambda_{\text{reg}} \mathcal{E}_{\text{reg}}(u) + \lambda_{\text{sh}} \sum_{q \in R} \sum_{d \in F} \sum_{c \in \{x, y\}} \gamma(\mathcal{E}_c(d, q, u)) \mathcal{E}_c(d, q, u). \]  

The global error function (5) is minimized using the Gauss-Newton algorithm. Note that the sparse structure of the Jacobian matrix, see figure 2, is taken into account in the minimization. A hierarchical approach [2] is used. It is mainly required at the disocclusion stage. The coarse-to-fine refinement step consists to propagate the displacement field and the self-occlusion map \( \mathcal{H} \) using pyramid expansion operations [4]. The latter has to be binarized. All the non-zero labels are fixed at one. This over-estimates the self-occlusion map, preventing misalignment at the boundaries.

We note that warp shrinking is the natural behavior of the warp when pixels vanish due to foreshortening under perspective projection. In these cases, the detection module is still valid. The shrinking constraint does not however activates since the warp does not fold. The proposed method naturally deals with these configurations.

The Jacobian and Hessian (Gauss-Newton approximation) matrices have a very sparse structure. These example from the paper sequence of figure 4.

5. Experimental results

We tested our approach on several videos with different kinds of surfaces (paper, fabric). A 2-level pyramid is used. Various experiments show that the hierarchical framework is necessary to recover the warp at disocclusion but that additional levels do not significantly improve the registration. The image registration algorithm described in §4 takes few seconds per frame with our unoptimized Matlab code. A regular grid has been defined for visualization purposes. The latter is less dense than the one used to guide the warp.

The one-to-one constraint requirement. As said above, loops and folds are non-admissible warp configurations. In the absence of the shrinking penalty, they naturally appear when the surface is extremely self-occluded, as shown on figure 3. The self-occlusion detection is wrong since the warp derivative is not null for the whole self-occluded area. This defeats the registration process. Adding the shrinking constraint to the error function forces the warp not to fold but rather to shrink. Self-occlusion detection is then successful.

The paper sequence. Figure 4 shows registration results on the first paper sequence. Registration of visible parts is accurate while the warp shrinks well in self-occluded regions. Recovering the true warp at disocclusion is done without misalignment.

Comparison with state-of-the-art. We compare our approach with a classical method for robust direct registration. It consists to minimize equation (3) with a hierarchical Gauss-Newton algorithm. Results are shown on figure 5. The Huber function [6] is employed as the robust kernel. Results in figure 1 are obtained with this procedure. Our approach successfully deals with self-occlusions while robust methods fail since they do not constrain the warp to shrink in self-occluded regions and yield inaccurate registration of...
visible parts when the surface is extremely self-occluded. It is unlikely that they manage to keep track of the surface at the disocclusion stage. We note that a re-initialization procedure such as non-rigid surface detection [11] might be used when the registration is lost. The proposed approach allows to recover the warp at disocclusion without misalignment. A re-initialization procedure is not required unless the surface is totally occluded, contrarily to classical robust approaches.

**Self-occluded boundaries.** Self-occluded surface boundaries are particular cases. Classical approaches and the one we propose are both valid i.e. non-visible pixels are correctly rejected and visible ones are correctly registered. However, with our self-occlusion approach, the occlusion boundary is tracked as shown on figure 6. Minimizing equation (4) enforces the warp to shrink along the self-occlusion boundary. The self-occlusion boundary is lost with the classical approaches.

**Other kinds of surfaces.** Experiments on a fabric, figure 7, show that the specific framework we propose deal with self-occlusions for many kinds of surfaces.

### 6. Conclusion

We addressed the important and seldom explored issue of self-occlusions in non-rigid registration. A specific framework is proposed. The main idea is to constrain the warp to shrink in self-occluded regions while detecting them based on this property. Experimental results on real videos show that our approach successfully deals with extreme self-occlusions while classical robust methods fail. Future work will concentrate on improving the accuracy along the boundary of extreme self-occlusions.

### References


Figure 5. Registration results on the second paper sequence. Top: the input images $I_i$. Middle: erroneous registration results with a classical outlier rejection approach. Bottom: successful registration results with the proposed method.

Figure 7. Registration results on the BMVC bag sequence. Top: the template $I_0$ with detected self-occlusions shown in white. Bottom: a grid illustrating the warp.


