Non-Rigid Optical Flow with Laplacian Cotangent Mesh Constraints

Anonymous CVPR submission

Paper ID 2090

Abstract

In this paper we present a novel non-rigid optical flow algorithm for dense image correspondence and non-rigid image registration. The algorithm introduces a smoothness term based on laplacian cotangent mesh deformation. Such deformation approaches are popular in graphics research, particularly for preserving local detail. In our work we introduce the concept for optical flow smoothness with the similar motivation of imposing local detail preservation. The algorithm applies a mesh to the image with a resolution up to one vertex per pixel. This uses edge angles defined by the mesh faces to ensure sensible local deformations between image pairs . The formulation is contained wholly inside the optical flow optimisation, and can be applied in a straightforward manner to a wide range of image tracking and registration problems. We evaluate our approach on several data sets against several popular optical flow algorithms. We show our approach to provide improved local detail preservation when registering image sequences containing non-rigidly deforming objects.

1. Introduction

Many important problems in computer vision require the tracking of non-rigid objects, calculation of dense image correspondences and the registration of image sequences containing highly non-rigid deformation. Existing algorithms to achieve this include model based tracking [10], dense patch identification and matching [8], group-wise image registration [5], space-time tracking [15] and optical flow [13, 4, 7].

Optical flow is an attractive formulation since it provides a dense displacement field between image pairs. In most standard approaches, assumptions regarding brightness constancy between images and smoothness in motion between neighboring pixels are adopted [4, 13]. Sun et al [14] adopt a different approach and introduce a method to overcome these constraints by learning a probabilistic model for flow estimation. However, their approach requires training pre-calculated flow ground truths, which are difficult to obtain.

In the general optical flow model, it is common to adopt a data term consisting of color and gradient constraints (e.g. Brox and Malik [4]) and an additional smoothness term. However, most previous optical flow formulations only consider global smoothness and ignore formulations that preserve local image detail.

Many optical flow techniques concentrate on problems where the scene movement is largely rigid in nature. However, there are many problem cases where we would like to calculate flow given highly non-rigid global and local image displacements over long image sequences. One recent problem highlighting this particular case is the alignment of 3D dynamic facial sequences containing highly non-rigid deformations [6, 3]. The problem requires non-rigidly aligning a set of images to a reference - e.g. a neutral facial expression. Each image - referred to as a UV map [1] is accompanied by a corresponding 3D mesh, and each mesh has a difference vertex topology. Once the UV maps are registered to a reference image (e.g. a neutral expression), vertex correspondence can be imposed. The technique is popular in 3D Morphable Model construction [6, 2].

Beeler et al [1]] and Bradley et al [3] adopt a slightly different approach to mesh correspondence. In their solutions, image displacement is calculated from camera views and then used to deform a reference mesh from an initial frame through a 3D sequence. The optical flow provides guides for adjusting pixel positions, and the mesh reduces artefacts by imposing a constraint to precent faces on the mesh from becoming inverted or flipped.

Mesh and image deformation research in graphics is an active area of research [9]. Such techniques provide flexible methods to invoke deformation while preserving some desired properties such as local geometric detail. As such, it is interesting to also consider such solutions as smoothness constraints in optical flow calculation, and this forms the central basis of our presented work.

[1] UV refers to the XY location of a pixel in the image. UV map is the graphical term for the *texture* for a 3D model. Each UV location maps to a 3D vertex on a corresponding mesh

1.1. Contributions

In this paper we present a novel non-rigid optical flow algorithm for dense image correspondence and non-rigid image registration. The algorithm introduces a smoothness term based on laplacian cotangent mesh deformation. Such deformation approaches are popular in graphics research, particularly for preserving local detail. In our work we introduce the concept for optical flow smoothness with the similar motivation of imposing local detail preservation. The algorithm applies a mesh to the image with a resolution up to one vertex per pixel. This uses edge angles defined by the mesh faces to ensure sensible local deformations between image pairs . The formulation is contained wholly inside the optical flow optimisation, and can be applied in a straightforward manner to a wide range of image tracking and registration problems. We evaluate our approach on several data sets against several popular optical flow algorithms. We show our approach to provide improved local detail preservation when registering image sequences containing non-rigidly deforming objects.

Our paper is organized as follows. In the section 2 we outline our strategy for calculating optical flow displacements between image pairs. In Section 3 we then describe how we adopt this algorithm in a coarse-to-fine framework. Section 4 presents an evaluation of our approach on 12 sequences of non-rigidly deforming objects. Section 5 then draws conclusions and discusses future directions.

2. Laplacian Mesh Based Optical Flow

In this section we present our mesh based optical flow approach. In our formulation we consider a pair of consecutive frames in an image sequence. We denote the current frame by $I_i(\mathbf{X})$ and its successor by $I_{i+1}(\mathbf{X})$, where $\mathbf{X} = (x, y)^T$ is a pixel location in the image domain Ω. We define the optical flow displacement between $I_i(\mathbf{X})$ and $I_{i+1}(\mathbf{X})$ as $\mathbf{w}_i(\mathbf{X}) = (\mathbf{u}(\mathbf{X}), \mathbf{v}(\mathbf{X}))^T$.

In our proposed optical flow approach we use the following general formulation:

$$\mathbf{E}(\mathbf{w}) = \sum_{\Omega} (\mathbf{E}_{Data}(\mathbf{w}) + \lambda \mathbf{E}_{Global}(\mathbf{w}) + \epsilon \mathbf{E}_{Lap}(\mathbf{w})) \quad (1)$$

where $\mathbf{E}_{Data}(\mathbf{w})$ denotes a data term that contains both color and gradient constancy assumptions (see section 2.1) on pixel values between $I_i(\mathbf{X})$ and $I_{i+1}(\mathbf{X})$. This first term is similar to those derived in Brox and Malik [4] and our motivation for including these terms is also the same.

We also introduce two smoothness terms into our formulation. Similar to Brox and Malik [4] and Horn and Schunk [13], our first term $\mathbf{E}_{Global}(\mathbf{w})$ controls global flow smoothness. Our second term represents the core contribution of our work, i.e. a Laplacian Cotangent mesh constraint $\mathbf{E}_{Lap}(\mathbf{w})$). We now outline all three terms in detail.

2.1. Data Term Definition

Following the standard optical flow assumption regarding brightness constancy, we assume that the color value of a pixel is not changed by its displacement through an image sequence. In addition, we also assume a gradient constancy assumption. This is added to provide additional stability where the first assumption (brightness constancy) is violated by changes in illumination. Our data term encoding these assumptions is therefore formulated as:

$$\mathbf{E}_{Data}(\mathbf{w}) = \sum_{\Omega} \Psi(I_{i+1}(\mathbf{X}+\mathbf{w}) - I_i(\mathbf{X}))^2 + \ldots \quad (2)$$
$$\Psi(\nabla I_{i+1}(\mathbf{X}+\mathbf{w}) - \nabla I_i(\mathbf{X}))^2$$

We use the increasing concave function $\Psi(s^2) = \sqrt{s^2 + \epsilon^2}$ with $\epsilon = 0.001$ similar to Brox and Malik [4] to solve this formation. This results in allowing $L1$ minimization. The remaining term $\nabla = (\delta_x, \delta_y)^T$ is a spatial gradient.

2.2. Global Smoothness Constraint

Our first smoothness term is a dense pixel based regularizer that penalizes global variation. The objective is to produce a globally smooth optical flow field, and it is defined as:

$$\mathbf{E}_{Global}(\mathbf{w}) = \sum_{\Omega} \Psi(|\nabla \mathbf{u}|^2 + |\nabla \mathbf{v}|^2) \quad (3)$$

We again use the robust function Ψ to solve this as in the data term.

2.3. Laplacian Mesh Smoothness Constraint

We now introduce our proposed Laplacian cotangent mesh constraint. The aim of this constraint is to account for non-rigid motion in scene deformation. Our term is inspired by mesh deformation research in graphics, and we introduce its use in optical flow estimation for this first time in our work.

Although non-rigid motion is highly nonlinear, the movement of pixels in such deformations still often has strong correlations in local regions. To represent this, we propose a novel sparse smoothness constraint based on a Laplacian framework and differential representation.

We assume that the frame is initially covered by a triangular mesh denoted by $M = (V, E, F)$. We define n as the number of vertices, V as the set of vertex coordinates, E as the set of edges, and F as the set of faces. The location of each vertex is represented using absolute cartesian coordinates. The k-th vertex is denoted by $\mathbf{v}_k \epsilon V$. In this paper we set the distance between adjacent vertices to be 10 pixels as

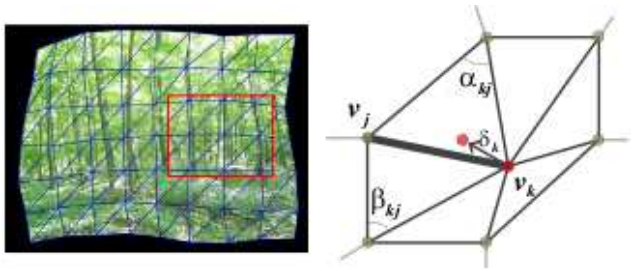

Figure 1. 1-ring neighbourhood of vertex.

we found this to give the most stable tracking results in our experiments (see Section 4).

Considering a small mesh region, each vertex v_k has a 1-ring neighbourhood, denoted $\mathcal{A}_{ring}$, as shown in Figure 1. The relationship between v_k and v_j on the i-th frame can be represented as follows:

$$\delta_k = \frac{1}{\mid \mathcal{A}^k_{Voronoi} \mid} \sum_{j \epsilon N(k)} \frac{1}{2}(cot\alpha_{kj} + cot\beta_{kj})(\mathbf{v}_k - \mathbf{v}_j) \quad (4)$$

where $N(i) = \{j \mid (k,j)\epsilon E\}$, α_{kj} and β_{kj} are the two angles opposite the edge $\mathbf{e}_{kj} = \mathbf{v}_k - \mathbf{v}_j$. $\mid \mathcal{A}_{Voronoi} \mid$ is the Voronoi region area of $\mathbf{v}_k$ which provides error bounds and robustness against different dimensions of $\mathcal{A}_{ring}$. The $\mid \mathcal{A}_{Voronoi} \mid$ for the $\mathcal{A}_{ring}$ at $\mathbf{v}_k$ can be calculated using

$$\mid \mathcal{A}^k_{Voronoi} \mid = \frac{1}{8} \sum_{j \epsilon N(k)} (cot\alpha_{kj} + cot\beta_{kj}) \mid \mathbf{v}_k - \mathbf{v}_j \mid^2 \quad (5)$$

In a planar region, the vector δ_k represents the motion trend of $\mathbf{v}_k$ towards a new position. We assume that the flow motion $\mathbf{w}^k$ and $\mathbf{w}^j$ will behave in a similar way to the vertex displacements at positions $\mathbf{v}_k$ and $\mathbf{v}_j$. We define the displacement as $\delta = (\delta_x, \delta_y)^T$, where δ_x is movement in the horizontal direction and δ_y movement in the vertical direction. We therefore describe displacement in terms of flow vectors as follows:

$$\delta = \sum_{k \epsilon V} (\frac{1}{\mid \mathcal{A}_{Voronoi} \mid} \sum_{j \epsilon N(k)} \frac{1}{2}(cot\alpha_{kj} + cot\beta_{kj}) \ldots (\mathbf{w}^k - \mathbf{w}^j)) \quad (6)$$

Based on this formulation, we define out Laplacian smoothness constraint as:

$$E_{Lap}(\mathbf{w}) = \sum_n \Psi(\mid \nabla\delta_x \mid^2 + \mid \nabla\delta_y \mid^2) \quad (7)$$

In our formulation, the mesh constraint helps preserve local detail by minimising the angle differences in local neighborhoods. The main reason for this is that by constraining local edge angles – embedded in the cotangent formulation – the term penalizes local displacements which may cause overlap with other pixels. Our experiments in Section 4 have shown that this constraint leads to the better preservation of image detail given non-rigid motion changes.

Input: two images and a triangle mesh
1. n-levels Gaussian pyramids are constructed for both the images and the mesh. Parameters $l = 0$ (pyramid level) and $\mathbf{w} = 0$ (empty flow field) are set 2. The flow field is propagated to level $l + 1$ 3. The energy function (1) is minimized 4. If $l \neq n - 1$ then $l = l + 1$ and go to step 2.
Output: optical flow field

Table 1. The overall framework of our method.

3. Optical Flow Framework

In the previous section we outlined our optical flow strategy given a pair of frames. In this section we describe our overall framework in relation to a coarse-to-fine optimization strategy.

Table 1 gives an overview of our framework. In order handle large-displacement flow between images, we use a coarse-to-fine framework. Our algorithm takes two frames and a triangle mesh as input. In this paper our mesh is overlayed over the entire image surface. Each vertex in the mesh has 6 adjacent neighbors in its 1-ring surroundings. Every vertex has a distance of 10 pixels from its horizontal and vertical adjacent neighbors. As previously mentioned, we found this value to provide the most satisfactory results in our evaluation (see Section 4).

In order to minimize the energy function (1), we follow the coarse-to-fine strategy outlined in [4]. We construct an *n*-level image pyramid for both input images, and perform the same step for the mesh. We use a down sampling factor 0.75 and perform 12 warping steps on each pyramid level using bicubic interpolation. We apply the global smoothness constraint E_{Global} to each level of the pyramid setting λ to 0.85. The laplacian constraint E_{Lap} is only applied on the bottom two (highest resolution) levels of the pyramid, setting ϵ to a value of 0.9. This is because the mesh is defined on the highest level image, and since the vertex number is fixed, down-sampling the mesh can result in more vertices than pixels. This pyramid is equivalent to the Euler-Lagrange equations and we solve these using Conjugate gradients in a similar manner to Brox and Malik [4].

4. Results

In this section we describe experiments evaluation our Laplacian based optical flow algorithm against existing approaches. We use three data sets consisting of 12 image sequences obtained from Salzmann et al [11] and Garg et al [7]. Data from Salzmann et al consists of 6 video sequences between 68 and 271 frames, and 2 motion capture sequences of 54 and 103 frames respectively. The motion capture sequences contain non-rigid objects with reflective markers captured using a Vicon optical motion capture system. In order to use this data, we added a synthetic texture to the 3D marker positions and rendered the deforming image sequence in 2D. Data from Garg et al consists of 4 sequences each 60 frames long, also containing non-rigidly deforming objects.

Optical flow evaluation on non-rigid sequences is an area which has received little attention, unlike evaluation on rigid scenes using e.g. the Middlebury dataset [12]. We therefore derive what we believe to be a fair comparison of the performance of our method using a data set of non-rigid movement compiled from several recent papers investigating non-rigid motion. In order to evaluate the performance of our algorithm we use a concatenative alignment strategy overviewed in Figure 2. We calculate the forward flow between pairs of image frames, and then use this flow to align an image back to the first frame in the sequence. Images from frame I_i are therefore aligned to the first frame by concatenating the consecutive flow fields. We perform this for each sequence using our Laplacian flow and also the optical flow algorithms of Brox and Malik [4] and Horn and Schunk [13]. After alignment, we then calculate the Mutual Information (MI) between the first frame in the image sequence and the aligned image.

Figures 3, 4, 5 and 6 give visual comparisons of the alignment methods. Figure 7 shows the mutual information between a registered image and the image in the first frame of the sequence. The accompanying video also shows registration examples for all of the 12 video sequences.

Figure 7 clearly shows improved mutual information for our method between registered images and the first image in the sequence. This is supported by the visual comparisons in Figures 3, 4, 5 and 6. The results show improved detail preservation in all sequences, as well as improved boundary preservation particularly in the more highly deforming scenes (e.g. Figure 3). Figures 4 and 6 show reasonable global registration for both our method and Brox and Malik. However, a close examination shows improved detail preservation in our approach. This is also supported (and is visually clearer) by results in our supporting video.

As would be expected, mutual information reduces for frames further from the reference. This is due to accumulating optical flow error as well as occlusions that appear as the sequence progresses. One rationale for the improvement in detail preservation for our proposed approach is that our Laplacian smoothness term is providing a more robust constraint on the movement of local pixel regions. In Brox and Malik [4] and Horn and Shunk [13] the smoothness term is global, and does not directly consider local pixel deformations.

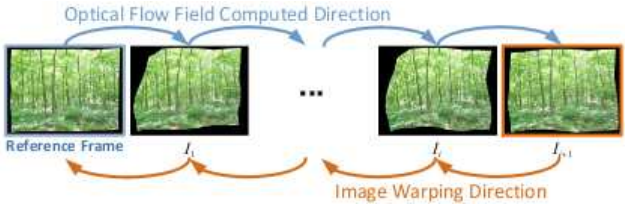

Figure 2. For evaluation the optical flow field is computed for each adjacent pair of frames. A frame is then warped back to the reference by concatenating flow fields.

5. Conclusions

In this paper we have presented a novel optical flow formulation that imposes a Laplacian mesh based smoothness constraint. The additional term focuses on preserving local image deformations and performs particularly well on image sequences containing non-rigidly deforming objects. We have compared our approach to several popular optical flow formulations and demonstrated its ability to better preserve local image detail.

In our evaluations we have identified some remaining issues relating to improving flow accuracy over long periods. This forms our direction for future work. The problem is partially due to occlusions forming in the sequences. This leaves large scope for investigating the application of additional robustness terms in flow formulations, or some other feature based stability measure for long sequences.

References

[1] T. Beeler, F. Hahn, D. Bradley, B. Bickel, P. A. Beardsley, C. Gotsman, R. W. Sumner, and M. H. Gross. High-quality passive facial performance capture using anchor frames. *ACM Transactions on Graphics (proceedings of ACM SIGGRAPH)*, 30(4):75, 2011. 1

[2] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. *ACM Transactions on Graphics (proceedings of ACM SIGGRAPH)*, pages 187–194, 1999. 1

[3] D. Bradley, W. Heidrich, T. Popa, and A. Sheffer. High resolution passive facial performance capture. *ACM Transactions on Graphics (proceedings of ACM SIGGRAPH)*, pages 41:1–41:10, 2010. 1

[4] T. Brox and J. Malik. Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33:500–513, 2011. 1, 2, 3, 4

[5] T. Cootes, S. Marsland, C. Twining, K. Smith, and C. Taylor. Groupwise diffeomorphic non-rigid registration for automatic model building. In *Proceedings of the European Conference on Computer Vision*, volume 3024 of *Lecture Notes in Computer Science*, pages 316–327. Springer, 2004. 1

[6] D. Cosker, E. Krumhuber, and A. Hilton. A facs valid 3d dynamic action unit database with applications to 3d dynamic morphable facial modelling. In *13th International Conference on Computer Vision (ICCV)*, November 2011. 1

[7] R. Garg, A. Roussos, and L. Agapito. Robust trajectory-space tv-l1 optical flow for non-rigid sequences. In *In Proc. of EMMCVPR*, pages 300–314, 2011. 1, 4

[8] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski. Non-rigid dense correspondence with applications for image enhancement. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2011)*, 30(4):70:1–70:9, 2011. 1

[9] A. Jacobson, I. Baran, J. Popović, and O. Sorkine. Bounded biharmonic weights for real-time deformation. *ACM Transactions on Graphics (proceedings of ACM SIGGRAPH)*, 30(4):78:1–78:8, 2011. 1

[10] I. Matthews and S. Baker. Active appearance models revisited. *Int. J. Comput. Vision*, 60(2):135–164, 2004. 1

[11] M. Salzmann, R. Hartley, and P. Fua. Convex optimization for deformable surface 3-d tracking. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference onn*, 0:1–8, 2007. 4

[12] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vision*, 47:7–42, April 2002. 4

[13] B. G. Schunck and B. K. P. Horn. Determining optical flow. pages 144–156, 1981. 1, 2, 4

[14] D. Sun, J. P. Lewis, and M. Black. Learning optical flow. In *In Proc. ECCV*, pages 83–97, 2008. 1

[15] L. Torresani and C. Bregler. Groupwise diffeomorphic non-rigid registration for automatic model building. In *Proceedings of the European Conference on Computer Vision*, volume 3024 of *Lecture Notes in Computer Science*, pages 316–327. Springer, 2002. 1

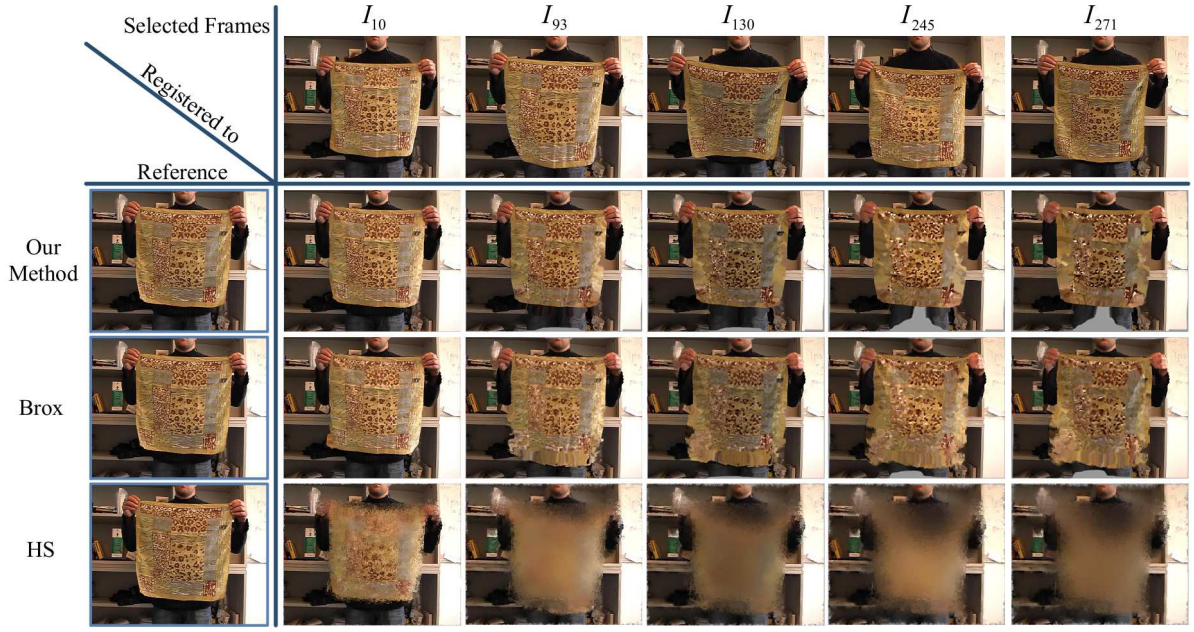

Figure 3. Registration results on Salzmann et al's *EPFLcloth* sequence.

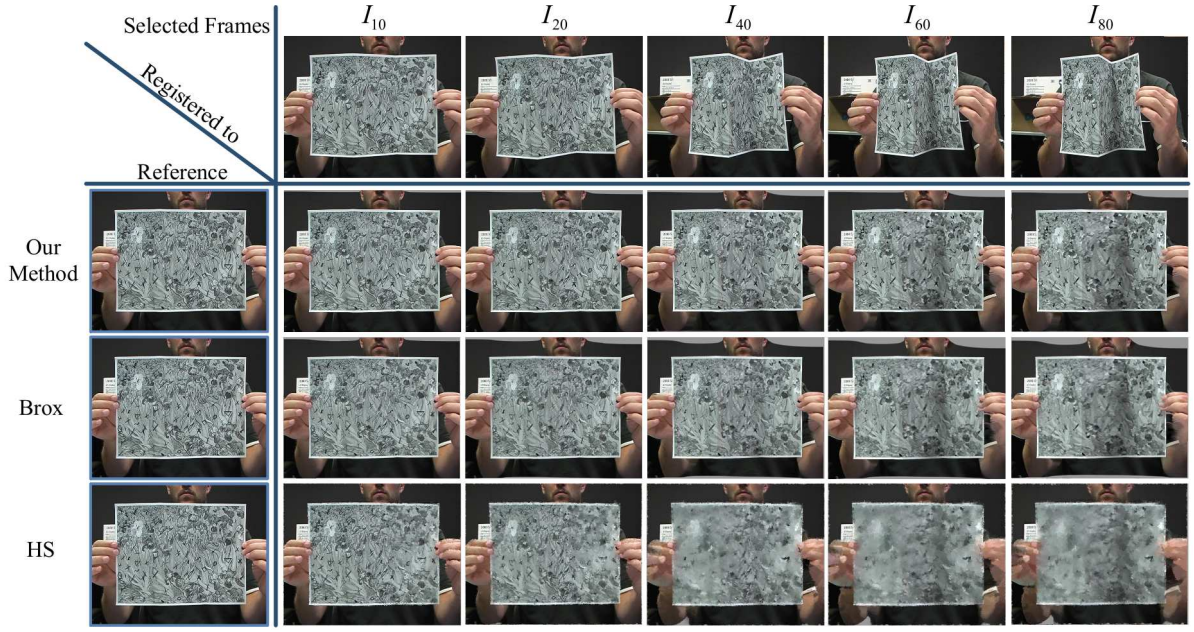

Figure 4. Registration results on Salzmann et al's *PaperCrease* sequence.

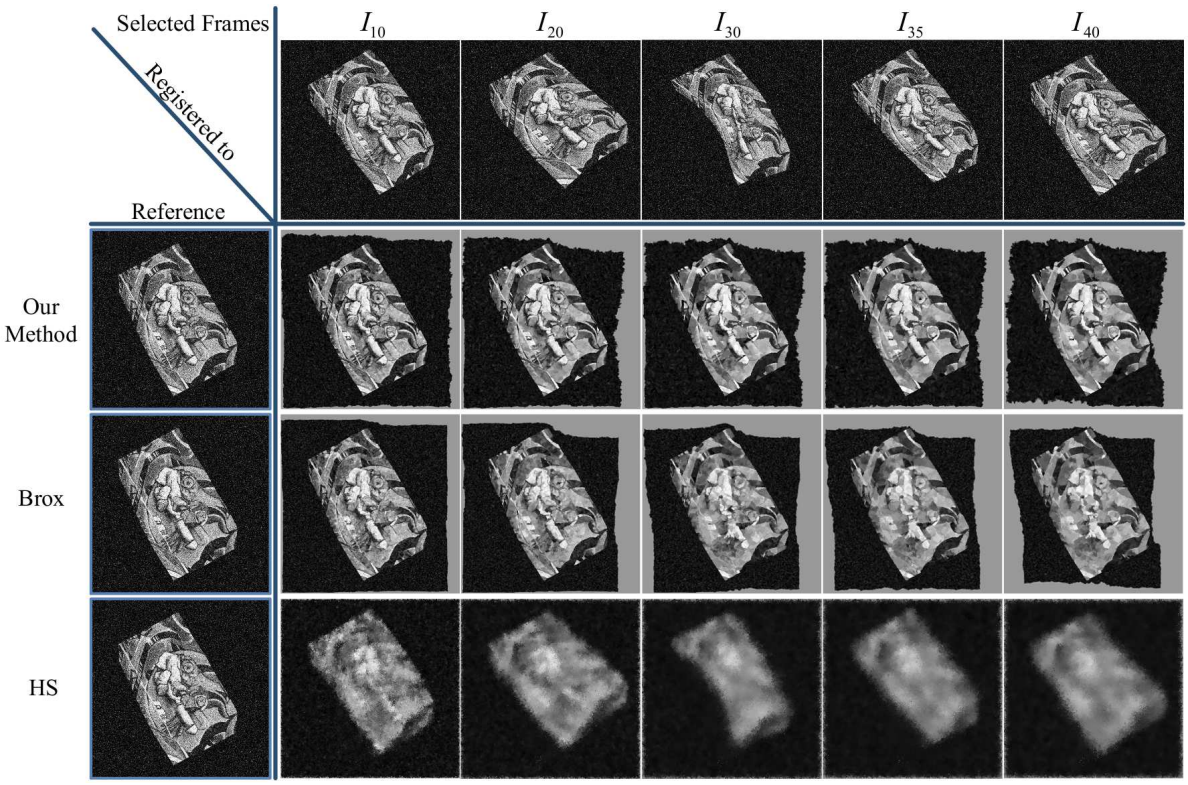

Figure 5. Registration results on Garg et al's *GN* sequence.

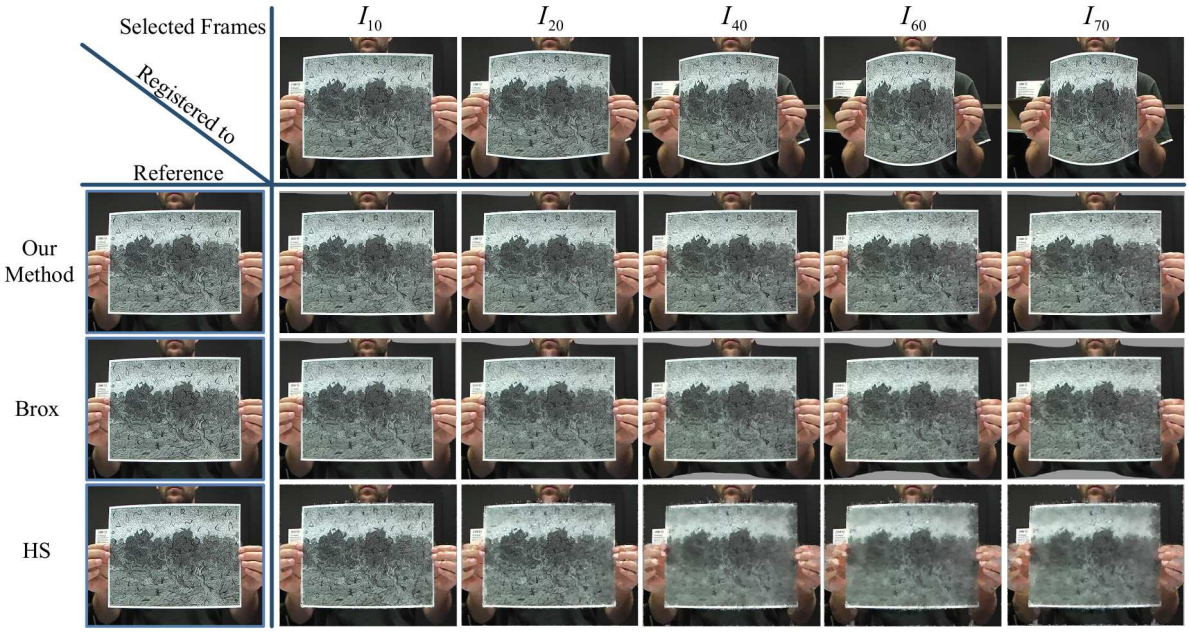

Figure 6. Registration results on Salzmann et al's *PaperBend* sequence.

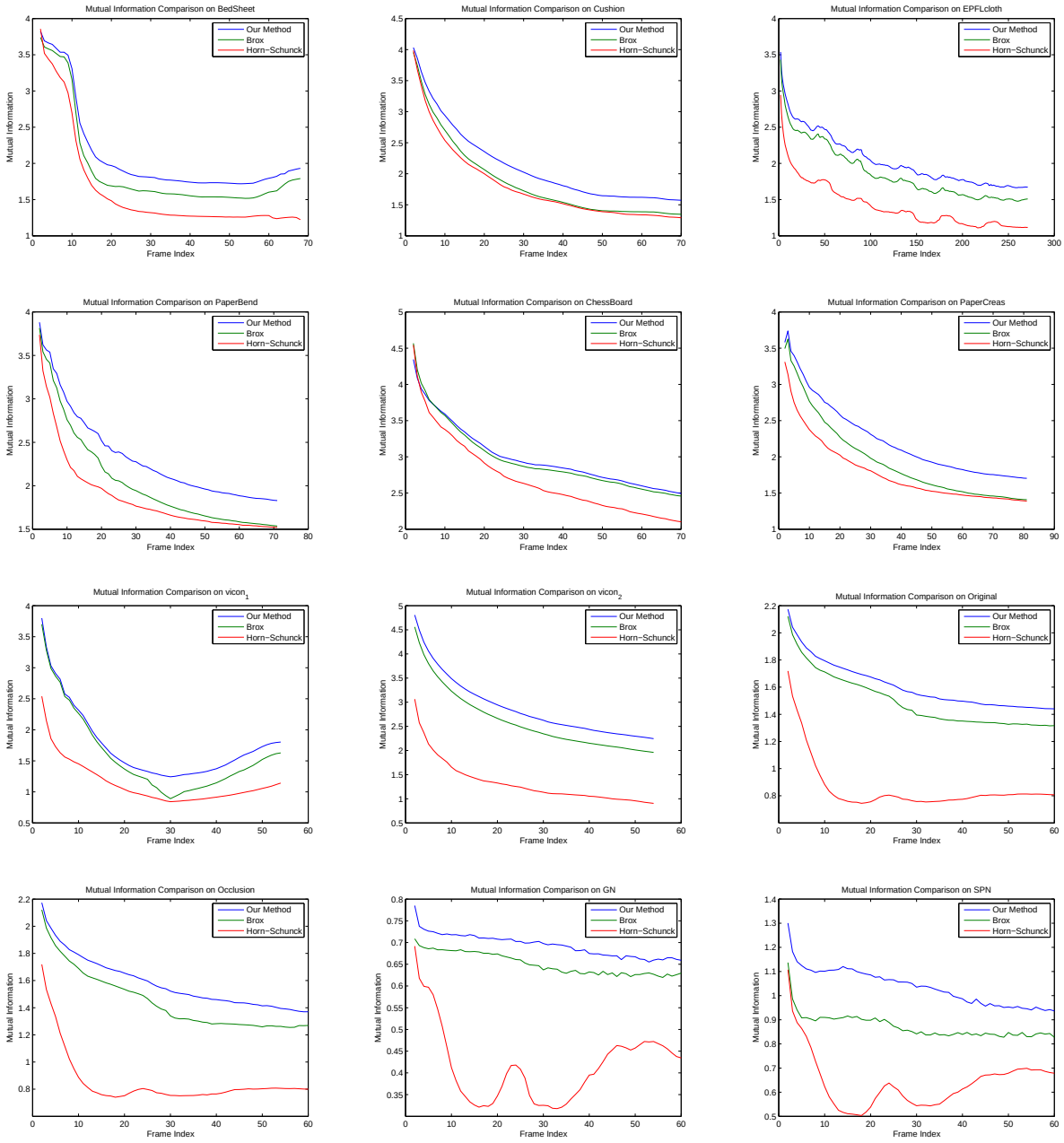

Figure 7. Mutual Information comparisons on each non-rigid image sequence for multiple algorithms: our approach, Brox et al, and Horn and Schunk.