

Multi-image Interpolation based on Graph-Cuts and Symmetric Optic Flow

Christian Linz, Christian Lipski, Marcus A. Magnor

Computer Graphics Lab, TU Braunschweig, Germany
{linz, lipski, magnor}@cg.cs.tu-bs.de

Abstract

Multi-image interpolation in space and time has recently received considerable attention. Typically, the interpolated image is synthesized by adaptively blending several forward-warped images. Blending itself is a low-pass filtering operation: the interpolated images are prone to blurring, even if correspondences are perfect. Furthermore, ghosting artifacts appear as soon as the underlying correspondence fields are imperfect. We address both issues and propose a multi-image interpolation algorithm that avoids blending. Instead, we cast multi-image interpolation as a labeling problem and decide for each pixel in the synthesized view from which input image to sample. Combined with a symmetrical long-range optical flow formulation for correspondence field estimation, our approach yields crisp interpolated images without ghosting artifacts.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation/Viewing algorithms; I.3.7 [COMPUTER GRAPHICS]: Three-Dimensional Graphics and Realism/Animation; I.3.m [COMPUTER GRAPHICS]: Miscellaneous/Image- and Video-based Rendering; I.4.8 [IMAGE PROCESSING AND COMPUTER VISION]: Scene Analysis/Time-varying imagery.

1. Introduction

The synthesis of in-between images from different view-points and/or time instants is experiencing a renaissance. Mahajan et al. recently proposed a high-quality interpolation technique for two images that is based on finding an optimal path for a pixel transitioning from one image to the other [MHM*09]. The strength of this approach is that the path framework allows each pixel to transition to the other image somewhere along the path, whenever a good correspondence is found. Further on, each pixel in the interpolated view is sampled from exactly one source image, thus avoiding ghosting or blurring artifacts. A major drawback of this approach is that the path idea can only be applied to two images; a direct extension to multi-image interpolation without resorting to intermediate interpolated images is not feasible.

Stich et al. recently introduced a perception motivated spatio-temporal image interpolation [SLAM08, SLW*10]. His approach is based on adaptively blending four forward-warped images. While this approach delivers high-quality interpolation results, it can suffer from ghosting and blurring

artifacts as soon as the underlying correspondence fields are imperfect.

In this paper, we combine the strengths of both approaches and propose a direct multi-image interpolation that avoids blending several images. Instead, in our approach we decide for each pixel in an interpolated sequence from which of the source images to sample best. Inspired by Mahajan et al. [MHM*09], we also reconstruct the interpolated image sequence using Poisson reconstruction after having filled spatio-temporal holes using an extended image inpainting technique. Combined with an extended long-range correspondence estimation with occlusion reasoning, our approach yields high-quality image interpolation results without ghosting.

Our contributions are thus

1. the formulation of multi-image interpolation as a labeling problem,
2. and the presentation of a powerful long-range correspondence estimation technique.

The remainder of the paper is structured as follows: we

discuss related work in Sect. 2 before we introduce our correspondence estimation technique in Sect. 3. We then describe our multi-image interpolation algorithm in Sect. 4 and evaluate its results on several scenes in Sect. 5. We conclude with a brief discussion, Sect. 6

2. Related work

Correspondence estimation. Dense correspondence estimation for pairs of images is a well-researched field in computer vision. Since an in-depth discussion of all approaches is out of the scope of this paper, we refer the reader to [BRS*07] for a recent survey. Most of these methods are based on a differential approximation of the image brightness constancy assumption and are hence only valid for small motions. Larger motions, which are common in image interpolation, are usually estimated by employing a multi-resolution framework. However, object motion which is larger than the object itself cannot be estimated since the object will vanish in the image pyramid before the displacement is small enough to be estimated. To account for this, Brox et al. [BBM09] recently incorporated descriptor matching in a variational framework to guide optical flow estimation for larger motions. As descriptors, they propose to use region descriptors of a hierarchical segmentation of the image, similar to the SIFT descriptor [Low04]. Their approach combines the power of descriptor matching with the regularization properties of a variational approach.

A different approach has been taken by Steinbruecker et al. [SPC09,STD09]. Starting at a standard variational formulation and making use of techniques known from quadratic relaxation [ZPB07], they arrive at a formulation with a point-wise data term and a convex smoothness term which are coupled via an additional flow field. For both data and smoothness term, a globally optimal solution can be found. The solution for the data term can simply be computed by a complete search, alleviating the need for coarse-to-fine warping strategies. One further appealing property of this approach is that any point-wise error measure can be integrated into the data term.

Occlusion handling is an important aspect of optical flow computation since no sensible correspondences can be found for occluded regions. Only few algorithms handle occlusion directly. Alvarez et al. [ADPS07] enforce symmetrical flow fields in a variational framework. They do not explicitly handle flow in occluded regions, but only impose a higher penalty on them. In contrast to this, Ince et al. [IK08] simultaneously estimate optical flow and occluded regions in a variational framework. Optical flow in occluded regions is inpainted from neighboring visible regions using image-driven anisotropic diffusion.

Image interpolation. Image interpolation also has many antecedents. A generally applicable, feature-based method for interpolating between two different images is presented

by Beier and Neely [BN92]. Chen and Williams show how general image interpolation can be used for view interpolation [CW93]. For improved rendering performance, McMillan and Bishop propose a planar-to-planar, forward mapped image warping algorithm [MB95]. Mark et al. adapt the method to achieve high frame rates for post-rendering [MMB97], while Zhang et al. apply feature-based morphing to light fields [ZWGS02]. Lee et al. extended the feature-based method presented by Beier and Neely [BN92] to more than two images [LWS98]. Stich et al. [SLAM08, SLW*08, SLW*10] recently proposed an algorithm for perceptually plausible image interpolation in space as well as time. This approach is the basis for the Virtual Video Camera System by Lipski et al. [LLB*10], suitably extended to more than two images. Recently, Mahajan et al. presented a path-based interpolation for image pairs that operates in the gradient domain and prevents ghosting/blurring and many occlusion artifacts visible in morphing-based methods [MHM*09]. Unfortunately, the path idea does not easily generalize to more than two input images without resorting to intermediate interpolations.

In our approach, we combine the idea of transitioning instead of blending of Mahajan et al. with the image interpolation of Stich et al. Instead of estimating paths, we resort to robust long-range symmetric correspondence estimation based on SIFT features. Exploiting symmetry in the correspondence estimation, occlusion can be easily detected. Inspired by the image fusion approach proposed by Agarwala [ADA*04], we transition from one image source to another by formulating the transition as a labeling problem and optimizing it via graph-cuts.

3. Correspondence Estimation

Our correspondence estimation algorithm is based on the approach presented by Steinbruecker et al. [SPC09]. This approach separates the data-term, i.e. the brightness constancy assumption $I_1 - I_2(\mathbf{x} + \mathbf{w}_{1,2}) \approx 0$, and the smoothness-term, i.e. $\nabla \mathbf{w}_{1,2} \approx \vec{0}$, that are the basis for the estimation of the correspondence map $\mathbf{w}_{1,2}$. It hence allows for the integration of arbitrary data-terms, especially data-terms that are not differentiable. Steinbruecker et al. already show how to integrate patch-based data-terms into this framework [STD09]. The key idea of this approach is based on the work of Zach et al. [ZPB07] where instead of direct minimization of the total-variation L_1 formulation

$$\min_{\mathbf{w}_{1,2}} \int_{\Omega} \alpha |I_1 - I_2(\mathbf{x} + \mathbf{w}_{1,2})| + |\nabla \mathbf{w}_{1,2}| d\mathbf{x}, \quad (1)$$

an auxiliary variable $\tilde{\mathbf{w}}_{1,2}$ is introduced and the problem

$$\min_{\mathbf{w}_{1,2}, \tilde{\mathbf{w}}_{1,2}} \int_{\Omega} \alpha |I_1 - I_2(\mathbf{x} + \mathbf{w}_{1,2})| + \frac{2}{\theta} \|\mathbf{w}_{1,2} - \tilde{\mathbf{w}}_{1,2}\|_2^2 + |\nabla \tilde{\mathbf{w}}_{1,2}| d\mathbf{x} \quad (2)$$

is considered. For small θ the solution of the original problem and the auxiliary problem are the same, but the latter

problem permits an elegant and fast solution: Equation (2) is solved iteratively for $\tilde{\mathbf{w}}_{1,2}$ keeping $\mathbf{w}_{1,2}$ fixed, and for $\mathbf{w}_{1,2}$ keeping $\tilde{\mathbf{w}}_{1,2}$ fixed, see [ZPB07] for details. Depending only on $\mathbf{w}_{1,2}$ and no longer on $\nabla \mathbf{w}_{1,2}$ the latter problem can be solved point-wise by considering

$$\tilde{E}(\mathbf{x}) = \alpha |I_1 - I_2(\mathbf{x} + \mathbf{w}_{1,2})| + \frac{2}{\theta} \|\mathbf{w}_{1,2} - \tilde{\mathbf{w}}_{1,2}\|_2^2. \quad (3)$$

Since Eq. (3) can be solved point-wise, essentially any non-linear data-term can be used and optimized globally by an exhaustive search.

3.1. SIFT data term

We exploit this desirable property and additionally integrate the SIFT descriptor [Low04] into the data term as proposed by [LYT*08]. Our data term now reads

$$\begin{aligned} \tilde{E}(\mathbf{x}) = & \alpha (|I_1 - I_2(\mathbf{x} + \mathbf{w}_{1,2})| + \|S_1 - S_2(\mathbf{x} + \mathbf{w}_{1,2})\|_2) \\ & + \frac{2}{\theta} \|\mathbf{w}_{1,2} - \tilde{\mathbf{w}}_{1,2}\|_2^2, \end{aligned} \quad (4)$$

with S_1 and S_2 denoting the dense SIFT image for I_1 and I_2 , respectively. By incorporating the SIFT descriptor into the data term, we gain increased robustness against illumination changes, incorporating a small neighborhood into the data term.

3.2. Edge data term

Following the work of Stich et al. [SLW*10], maintaining edge correspondences is important for high-quality image interpolation. We thus further integrate the edge-matching approach into Eq. (4), yielding

$$\begin{aligned} \tilde{E}(\mathbf{x}) = & \alpha (|I_1 - I_2(\mathbf{x} + \mathbf{w}_{1,2})| + \|S_1 - S_2(\mathbf{x} + \mathbf{w}_{1,2})\|_2) \\ & + \beta f(E_1) \|E_1 - \mathbf{w}_{1,2}\|_2 \\ & + \frac{2}{\theta} \|\mathbf{w}_{1,2} - \tilde{\mathbf{w}}_{1,2}\|_2^2, \end{aligned} \quad (5)$$

where E_1 is a sparse correspondence prior derived from matched edge pixels and $f(x) = 1$ iff E_1 has a valid entry and $f(x) = 0$ otherwise.

3.3. Symmetry data term

Symmetry is another important aspect for high-quality image interpolation. The input images are warped towards each other and the pixel values are blended; mismatching pixel values will lead to visible artifacts. Enforcing a symmetry constraint already in the computation of the flow fields is thus the basis for high-quality interpolation without cross-fading artifacts. We hence further add a symmetry term to Eq. (5), resulting in

$$\begin{aligned} \tilde{E}(\mathbf{x}) = & \alpha (|I_1 - I_2(\mathbf{x} + \mathbf{w}_{1,2})| + \|S_1 - S_2(\mathbf{x} + \mathbf{w}_{1,2})\|_2) \\ & + \beta f(E_1) \|E_1 - \mathbf{w}_{1,2}\|_2 \\ & + \gamma (1 - g(\|\mathbf{w}_{1,2} + \mathbf{w}_{2,1}(\mathbf{x} + \mathbf{w}_{1,2})\|_2)) \end{aligned}$$

$$+ \frac{2}{\theta} \|\mathbf{w}_{1,2} - \tilde{\mathbf{w}}_{1,2}\|_2^2, \quad (6)$$

where $g(x) = 1/(1 + kx^2)$, $k = 0.25$, is a weighting function used by Ince et al. [IK08] and originally proposed by [PM90]. This weighting function penalizes small flow deviations and leaves large deviations untouched, allowing other data terms to take control when no symmetry can be established. Regions where sensible correspondences can't be established are potentially occluded.

Eq. (6) is optimized by a full search in an $n \times n$ window making use of the compute power of recent GPUs. n denotes the maximal flow length in pixels in x- and y-direction. The optimization is started with $\theta = 2 \cdot n$ and is run for 10 iterations with θ decreasing linearly to 0.01. After each iteration, the result is median filtered in a 5×5 neighborhood and 10 smoothing iterations are applied. A good starting point for the remaining parameters is given by $\alpha = 8.0$, $\beta = 4.0$ and $\gamma = 1.0$

3.4. Occlusion detection

Since we enforce symmetric flow in the optimization, we can now use the geometric mismatch of forward flow $\mathbf{w}_{1,2}$ and backward flow $\mathbf{w}_{2,1}$ to detect occluded regions by considering $m(\mathbf{w}_{1,2}, \mathbf{w}_{2,1}) = \|\mathbf{w}_{1,2} + \mathbf{w}_{2,1}(\mathbf{x} + \mathbf{w}_{1,2})\|_2^2$. Thresholding the geometric mismatch m gives a binary occlusion map O , regions where flow values are not symmetric and hence unreliable. We repair the flow values in those regions by transferring the idea of geodesic matting [BS09] to motion inpainting. To this end, we make the assumption that the occluded region belongs either to foreground or background and its affiliation can be determined by color. As a first step, we identify all occluded blobs $\{b_i\}_{1 \dots N}$ in the binary occlusion map O ; large blobs are split perpendicular to their major axis. The splitting threshold is usually set to 100 pixels. For every blob b_i , we then examine a five pixel wide boundary around the blob and determine two boundary clusters \mathcal{F}_i and \mathcal{B}_i by clustering unoccluded flow values using k-means, Fig. 2. For each cluster, we estimate a color probability density function (PDF) in Lab space. In order to identify the splitting boundary in b_i , we then compute the weighted geodesic distance $d(x)$ for foreground and background PDF for each pixel $x \in \Omega_{b_i}$, accordingly, see [BS09] for details. To inpaint flow values into blob b_i , we simply copy the median flow vector from the assigned border class.

3.5. User interaction

Despite the expressiveness of the SIFT descriptor, there are cases where the automatic correspondence estimation does not produce correct results. To still guarantee high quality interpolation results, we resort to user interaction in difficult cases. User correspondences $\tilde{\mathbf{w}}_{1,2}$ are specified in an interactive tool using brushes and are readily integrated into Eq. 6,

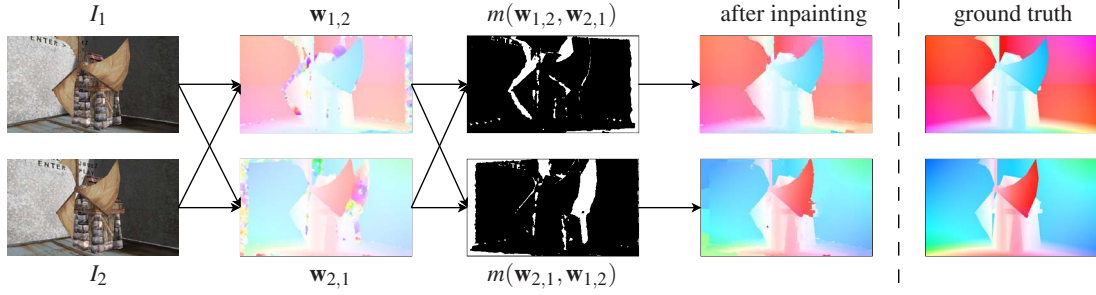


Figure 1: We first estimate correspondences by optimizing Eq. (6). We then detect occluded regions, compute color statistics along the boundary of the occluded regions and inpaint the flow values based on color similarity. The last column shows a color coding of the ground truth flow fields for visual comparison.



Figure 2: Flow repair: For each occluded blob, we determine two boundary classes \mathcal{F}_i (green) and \mathcal{B}_i (blue) by clustering the flow values along the boundary of the occluded blob b_i (red). For each boundary class, color statistics are computed in Lab space and each pixel of the occluded blob is inpainted with flow values from the most similar boundary.

resulting in

$$\begin{aligned} \tilde{E}(\mathbf{x}) = & \alpha (|I_1 - I_2(\mathbf{x} + \mathbf{w}_{1,2})| + \|S_1 - S_2(\mathbf{x} + \mathbf{w}_{1,2})\|_2) \\ & + \beta f(E_1) \|E_1 - \mathbf{w}_{1,2}\|_2 \\ & + \gamma (1 - g(\|\mathbf{w}_{1,2} + \mathbf{w}_{2,1}(\mathbf{x} + \mathbf{w}_{1,2})\|_2)) \\ & + \delta f(\hat{\mathbf{w}}_{1,2}) \|\hat{\mathbf{w}}_{1,2} - \mathbf{w}_{1,2}\|_2 \\ & + \frac{2}{\theta} \|\mathbf{w}_{1,2} - \tilde{\mathbf{w}}_{1,2}\|_2^2. \end{aligned}$$

4. Multi-image interpolation

Stich et al. [SLAM08] proposed a GPU-based image interpolation algorithm that is founded on mesh-based forward warping and adaptive blending of images. In this approach, occlusion and disocclusion are handled by depth heuristics and a connectedness measure, respectively. This algorithm naturally extends to more than two images; the virtual view

I_v can be synthesized as

$$I_v = \sum_{i=1}^n \mu_i \tilde{I}_i,$$

with

$$\tilde{I}_i \left(\mathbf{x} + \sum_{j=1, \dots, n, j \neq i} \mu_j \mathbf{w}_{ij}(\mathbf{x}) \right) = I_i(\mathbf{x}) \quad (7)$$

and μ_i denoting a barycentric weighting scheme in $n - 1$ dimensional space, see [LLB*10] for a detailed explanation of the construction of the space and the derivation of the weights μ_i . While this approach produces good results if the correspondence fields \mathbf{w}_{ij} match up exactly, the blending actually produces a low-pass filtered image for imprecise correspondence fields. Further, the adaptive blending weights derived from the connectedness measure often result in streaking artifacts in disoccluded regions, cf. Fig. 4(a).

Our proposed approach also relies on forward warped images, i.e. we also warp each input image to the desired position in by applying Eq. (7). In contrast to Stich et al., our approach cuts the underlying warping mesh open in disoccluded regions by measuring the triangle stretch. Instead of blending the forward warped images to get the interpolated image, we formulate the interpolation as a labeling problem which is explained in the following subsection.

4.1. Graph-cut based interpolation

Inspired by the interactive digital photomontage by Agarwala et al. [ADA*04], we now show how we avoid blending several images at each pixel. Our approach is based on solving an optimization problem that decides for each pixel in the virtual view I_v from which of the n source images best to take the color information. To this end, we formulate the view synthesis as a labeling problem in a 3D MRF framework incorporating temporal coherence. Note that from now on we consider the entire interpolated sequence and perform all computations on this spatio-temporal volume. The goal is

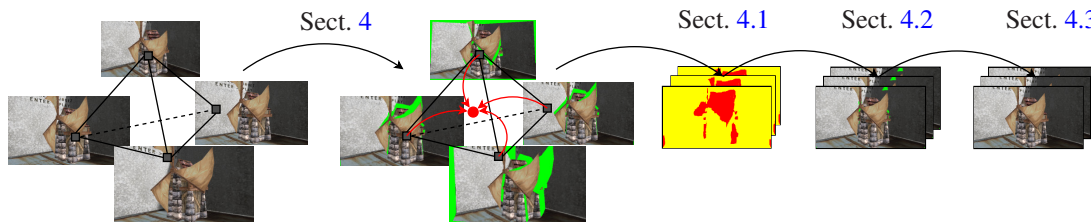


Figure 3: Schematic overview of the proposed interpolation. First, the input images are forward-warped to each target position of the interpolated sequence. The interpolated sequence is then constructed as the solution of a spatio-temporal labeling problem. Holes in the interpolated sequence are inpainted and the result is finally reconstructed in the gradient domain.

to assign to each pixel $p \in I_v$ a label $L(p)$ indicating which of the source images $\tilde{I}_{L(i)}$ the pixel should be taken from. In particular, we optimize the following energy:

$$E(L) = \sum_{p \in I_v} E_D(p, L(p)) + \lambda \sum_{p, q \in \mathcal{N}} E_S(p, q, L(p), L(q)), \quad (8)$$

where E_D measures the quality of the current labeling, E_S controls the smoothness of the labeling and p, q are neighboring pixels in a 6-connected[†] neighborhood $\mathcal{N} \subset \{I_v\}$.

Our data cost function

$$E_D(p, L(p)) = 4.0 \cdot P_{dis}(\Delta_p) \cdot e^{1 - \mu_{L(p)}},$$

favors pixel that receive a low disocclusion penalty $P_{dis}(\Delta_p)$. We compute the disocclusion penalty based on the area of the associated triangle Δ_p in the underlying warp mesh as

$$P_{dis}(\Delta_p) = \begin{cases} 0 & \text{if } \Delta_p \leq 0.5 \\ e^{1.25(\Delta_p - 0.5)^2} - 1 & \text{else.} \end{cases}$$

We further assume that images with a high barycentric weight $\mu_{L(p)}$ only have a low distortion.

The smoothness term

$$E_S(p, q, L(p), L(q)) = X \cdot Y$$

is composed of

$$X = \left(\|\tilde{I}_{L(p)}(p) - \tilde{I}_{L(q)}(p)\|_2 + \|\tilde{I}_{L(p)}(q) - \tilde{I}_{L(q)}(q)\|_2 \right)$$

and

$$Y = 2.0 - (\nabla_{p,q} \tilde{I}_{L(p)} + \nabla_{p,q} \tilde{I}_{L(q)}),$$

see Bhat et al. [BZS*07]. It prefers cuts through regions of homogeneous colors or along prominent edge structures. When considered from a perception point of view, this is desirable since (1) cuts in homogeneous regions will most likely go unnoticed and (2) cuts along prominent edges will keep structural information intact.

[†] Our neighborhood consists of 4 spatial and 2 temporal neighbors.

We find a labeling that is the approximate global minimum of Eq. (8) using the alpha-expansion algorithm proposed by Boykov et al. [BVZ01]. For our test scenes, we set $\lambda = 4$.

4.2. Spatio-temporal inpainting

The forward warping approach with mesh cutting discussed in Sect. 4 introduces holes in each source image where pixels are disoccluded. Eventually, some of those areas are invisible in all source images, such that our algorithm does not find a sensible labeling for those areas and hence no color value, cf. Fig. 4(b). Those areas have to be filled with perceptually plausible color values in a temporally consistent manner. To this end, we adapt the inpainting method presented by Telea [Tel04] to three dimensions and inpaint the spatio-temporal holes in the interpolated sequence. In our implementation, we favor inpainting along the temporal direction, over inpainting along the spatial dimensions by giving the temporal dimension a higher weight. This is justified as follows: invisible regions potentially occur at occlusion edges; inpainting along the spatial dimensions would lead to diffusing wrong color information over those occlusion boundaries. When inpainting along the temporal dimension, we exploit that the invisible region becomes visible at some point earlier or later in the sequence and we are hence able to diffuse color information in a perceptually plausible way, Fig. 4(c). We apply a temporal weighting factor of 0.9 and a spatial weighting factor of 0.1 in all our test scenes.

4.3. Label-based view synthesis

The labeling found by our optimization completely defines how to construct the virtual view I_v : one simply samples each interpolated pixel from the appropriate source image. We follow the approach of Mahajan et al. [MHM*09] and sample the image in the gradient domain. The interpolated image is then reconstructed by solving the 3D Poisson equation, i.e. we solve

$$\nabla^2 I_v = \nabla \cdot G,$$

where $G(\mathbf{x}) = (G_x, G_y, G_t)$ denote the gradients of the virtual view in the x and y direction and along the path through

the spatio-temporal volume. The Laplacian operator is computed as $\nabla^2 I_v = \frac{\partial^2 I_v}{\partial x^2} + \frac{\partial^2 I_v}{\partial y^2} + \frac{\partial^2 I_v}{\partial t^2}$ and the divergence of the gradient field is computed as $\nabla \cdot G = (\frac{\partial G_x}{\partial x} + \frac{\partial G_y}{\partial y} + \frac{\partial G_t}{\partial t})$. We take the first and last frame of the interpolation path as well as the one-pixel boundary of each intermediate frame as boundary conditions for the Poisson reconstruction.

5. Results and Evaluation

To evaluate the correctness of the proposed multi-image interpolation approach, we use a synthetic scene with ground-truth flow fields. We compare the proposed approach to the multi-image interpolation method proposed by Stich et al. [SLAM08], Fig. 4. Each image is interpolated from four input images. The blending-based interpolation technique suffers from streaking artifacts in disoccluded and totally invisible regions, despite the use of ground-truth correspondence fields. Our approach clearly produces less artifacts and is visually more pleasing. We next evaluate our interpolation approach on a high-speed camera sequence from the Middlebury optical flow benchmark. This scene features fast motion of small objects and represents a difficult test case for common optical flow approaches. We evaluate our interpolation algorithm on flow fields computed using the original algorithm of Steinbruecker et al., i.e., computed by solving Eq. (3), and on flow fields computed using our proposed extension including SIFT, edge and symmetry data terms, cf. Eq. (6). Without our proposed extension, the interpolation suffers from artifacts such as a distorted ball and visible seams running over the girl's skirt. Compared to the approach of Mahajan et al., we obtain visually comparable results with our full approach, Fig. 5. In addition to Mahajan et al., our approach naturally extends to more than two images without resorting to intermediate interpolations, thus avoiding additional image resampling that potentially leads to loss in quality. We next compare our full approach to the multi-scale approach recently proposed by Stich et al. [SLW*10], Fig. 6. Again, our approach produces sharper images without ghosting artifacts. For all scenes, we refer the reader to the accompanying video for better assessment of image quality.

5.1. Limitations

The inpainting approach used to fill occluded regions with sensible flow information relies on the assumption that the scene consists of two layers and that the occluded region is in direct adjacency to the region it actually belongs to. We further make the assumption that the layers can be clearly distinguished by their color statistics. If one of those assumptions is violated, as is the case for parts of the roof in the windmill sequence, Fig. 7, wrong flow information will be filled in. Incorporating high-level scene segmentation into the inpainting step might be an interesting direction for future work.

Currently, the spatio-temporal inpainting approach is limited to regions with only little texture. In highly textured regions, our inpainting will not be able to fill in matching details. In the future, we plan to examine texture synthesis algorithms for this task.

A further drawback of our interpolation approach is that it is not real-time capable. Most computation time is spent in solving the labeling problem. For a resolution of 960×540 pixels, the interpolation of a single frame takes between 10 and 20 seconds, depending on the quality of the initial labeling.

6. Conclusion

We have presented an algorithm for ghosting-free multi-image interpolation. Our approach is based on a novel formulation of image interpolation as a labeling problem. Combined with a symmetric long-range correspondence estimation technique based on SIFT features and edge matching, our interpolation technique yields high-quality results superior to state of the art.

Future work will focus on finding better ways to repair correspondence fields in occluded regions. One possible approach would be to integrate a high-level scene segmentation into this process.

References

- [ADA*04] AGARWALA A., DONTCHEVA M., AGRAWALA M., DRUCKER S., COLBURN A., CURLESS B., SALESIN D., COHEN M.: Interactive digital photomontage. *ACM Trans. Graph.* 23, 3 (2004), 294–302. 2, 4
- [ADPS07] ALVAREZ L., DERICHE R., PAPADOPOULOS T., SANCHEZ J.: Symmetrical dense optical flow estimation with occlusions detection. *IJCV* 75, 3 (2007), 371–385. 2
- [BBM09] BROX T., BREGLER C., MALIK J.: Large Displacement Optical Flow. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2009), pp. 41–48. 2
- [BN92] BEIER T., NEELY S.: Feature-based image metamorphosis. *Computer Graphics (Proc. of SIGGRAPH'93)* 26, 2 (1992), 35–42. 2
- [BRS*07] BAKER S., ROTH S., SCHARSTEIN D., BLACK M., LEWIS J., SZELISKI R.: A Database and Evaluation Methodology for Optical Flow. In *IEEE International Conference on Computer Vision (ICCV)* (2007), pp. 1–8. 2
- [BS09] BAI X., SAPIRO G.: Geodesic matting: A framework for fast interactive image and video segmentation and matting. *International Journal of Computer Vision* 82 (2009), 113–132. 3, 4

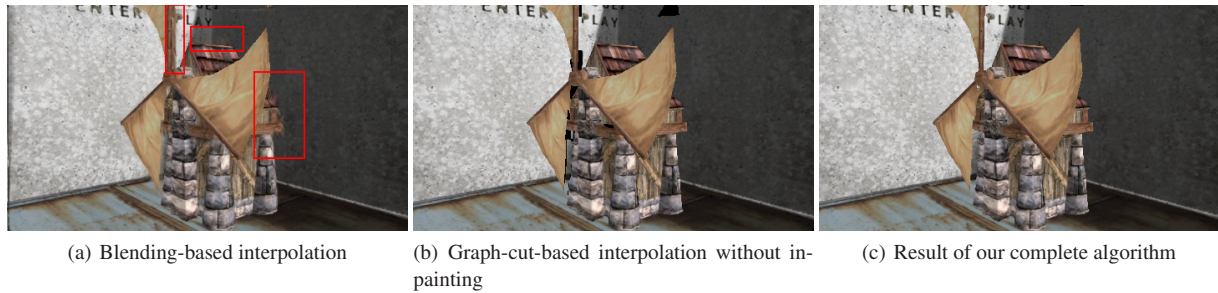


Figure 4: We evaluate the correctness of our algorithm on a synthetic scene with ground-truth correspondence fields. Each image is interpolated from four input images. Our approach is superior to the blending-based approach, especially in disoccluded regions. In those regions, the blending-based approach suffers from annoying streaks, despite the use of ground truth flow fields.



Figure 5: Results on the backyard sequence from the Middlebury benchmark, interpolated from two images. Left: result of the Moving Gradients approach by Mahajan et al. [MHM*09], middle: our proposed interpolation without inclusion of SIFT and symmetry in the correspondence estimation, and right: our full approach. The full approach yields comparable results to Mahajan et al. Note that the left image has been extracted from video; the low quality might be due to low resolution.

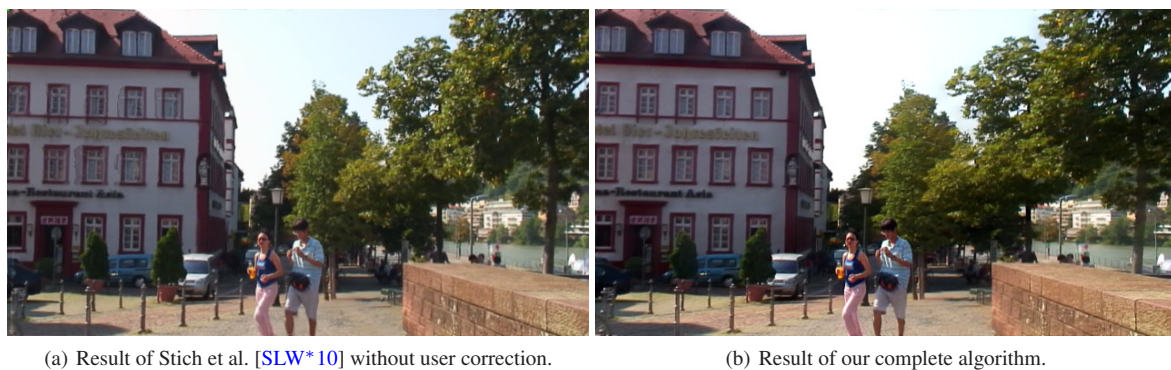
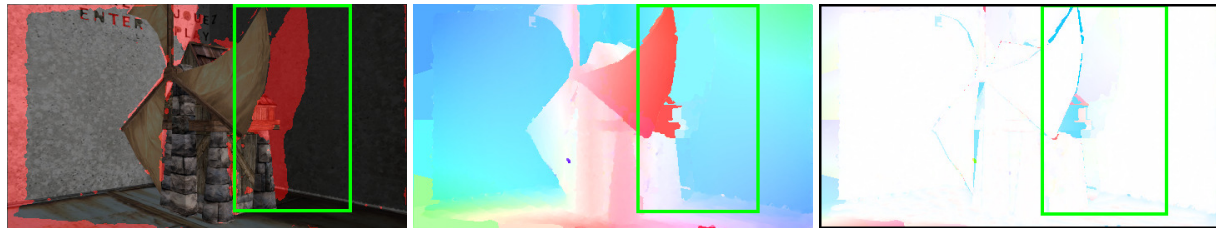


Figure 6: Results on the Heidelberg stereo sequence, interpolated from three images. Left: result of the approach of Stich et al. [SLW*10] using their edge-based correspondence estimation. No user corrections have been applied. The image suffers from ghosting noticeable around the windows, and appears blurry overall. Right: our algorithm yields crisp images without ghosting.



(a) Source image overlaid with detected occluded regions. (b) Repaired flow field. Repair fails in the marked region. (c) Difference to ground truth flow field.

Figure 7: Flow repair fails if new entities are discovered or color statistics are not distinctive enough. In this case, the roof of the windmill is a new entity which violates all of our assumptions.

- [BVZ01] BOYKOV Y., VEKSLER O., ZABIH R.: Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* 23, 11 (2001), 1222–1239. 5
- [BZS*07] BHAT P., ZITNICK C. L., SNAVELY N., AGARWALA A., AGRAWALA M., CURLESS B., COHEN M., KANG S. B.: Using photographs to enhance videos of a static scene. In *Proc. of EGSR* (June 2007), Eurographics, pp. 327–338. 5
- [CW93] CHEN S. E., WILLIAMS L.: View interpolation for image synthesis. In *Proc. of ACM SIGGRAPH'93* (New York, 1993), ACM Press/ACM SIGGRAPH, pp. 279–288. 2
- [IK08] INCE S., KONRAD J.: Occlusion-Aware Optical Flow Estimation. *IEEE Trans. Image Processing* 17, 8 (2008), 1443–1451. 2, 3
- [LLB*10] LIPSKI C., LINZ C., BERGER K., SELLENT A., MAGNOR M.: Virtual Video Camera: Image-based Navigation Through Space and Time. *Computer Graphics Forum* (2010), to appear. 2, 4
- [Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* 60, 2 (2004), 91–110. 2, 3
- [LWS98] LEE S., WOLBERG G., SHIN S.: Polymorph: morphing among multiple images. *IEEE Computer Graphics and Applications* (1998), 58–71. 2
- [LYT*08] LIU C., YUEN J., TORRALBA A., SIVIC J., FREEMAN W. T.: Sift flow: Dense correspondence across different scenes. In *ECCV '08* (Berlin, Heidelberg, 2008), Springer-Verlag, pp. 28–42. 3
- [MB95] MCMILLAN L., BISHOP G.: Plenoptic Modeling. In *Proc. of ACM SIGGRAPH'95* (New York, 1995), ACM Press/ACM SIGGRAPH, pp. 39–46. 2
- [MHM*09] MAHAJAN D., HUANG F., MATUSIK W., RAMAMOORTHY R., BELHUMEUR P.: Moving Gradients: A Path-Based Method for Plausible Image Interpolation. *ACM Transactions on Graphics* 28, 3 (2009), 42:1–42:11. 1, 2, 5, 6
- [MMB97] MARK W., MCMILLAN L., BISHOP G.: Post-Rendering 3D Warping. In *Proc. of Symposium on Interactive 3D Graphics* (1997), pp. 7–16. 2
- [PM90] PERONA P., MALIK J.: Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 12, 7 (1990), 629–639. 3
- [SLAM08] STICH T., LINZ C., ALBUQUERQUE G., MAGNOR M.: View and Time Interpolation in Image Space. *Computer Graphics Forum (Proc. of PG'08)* 27, 7 (2008), 1781–1787. 1, 2, 4, 7
- [SLW*08] STICH T., LINZ C., WALLRAVEN C., CUNNINGHAM D., MAGNOR M.: Perception-motivated Interpolation of Image Sequences. In *Proc. of ACM APGV'08* (Los Angeles, USA, 2008), ACM Press, pp. 97–106. 2
- [SLW*10] STICH T., LINZ C., WALLRAVEN C., CUNNINGHAM D., MAGNOR M.: Perception-motivated Interpolation of Image Sequences. *ACM Transactions on Applied Perception (TAP)* (2010), to appear. 1, 2, 3, 6, 7
- [SPC09] STEINBRUECKER F., POCK T., CREMERS D.: Large Displacement Optical Flow Computation without Warping. In *IEEE International Conference on Computer Vision (ICCV)* (September 2009), pp. 1609–1614. 2
- [STD09] STEINBRUECKER F., T.POCK, D.CREMERS: Advanced Data Terms for Variational Optic Flow Estimation. In *Vision, Modeling, and Visualization (VMV)* (November 2009), pp. 155–162. 2
- [Tel04] TELEA A.: An image inpainting technique based on the fast marching method. *Journal of Graphical Tools Vol.9, No.1* (2004), 25–36. 5
- [ZPB07] ZACH C., POCK T., BISCHOF H.: A duality based approach for realtime TV- L^1 optical flow. In *Pattern Recognition* (2007), vol. 4713, pp. 214–223. 2
- [ZWGS02] ZHANG Z., WANG L., GUO B., SHUM H.-Y.: Feature-based light field morphing. *ACM Trans. on Graphics* 21, 3 (2002), 457–464. 2