Pacific Graphics 2008
T. Igarashi, N. Max, and F. Sillion
(Guest Editors)

Volume 27 (2008), Number 7

View and Time Interpolation in Image Space

Timo Stich, Christian Linz, Georgia Albuquerque, Marcus Magnor

Computer Graphics Lab, TU Braunschweig , Germany

Abstract

The ability to interpolate between images taken at different time and viewpoints directly in image space opens up new possiblities. The goal of our work is to create plausible in-between images in real time without the need for an intermediate 3D reconstruction. This enables us to also interpolate between images recorded with uncalibrated and unsynchronized cameras. In our approach we use a novel discontiniuity preserving image deformation model to robustly estimate dense correspondences based on local homographies. Once correspondences have been computed we are able to render plausible in-between images in real time while properly handling occlusions. We discuss the relation of our approach to human motion perception and other image interpolation techniques.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Bitmap and framebuffer operations

1. Introduction

In-between images of different viewpoints can be rendered by deforming recorded images according to epipolar constraints (e.g. [CW93, SD96, ZKU*04]). However, the problem of estimating the correspondences without additional camera calibrations and time synchronization is ill posed in the general case. If also time interpolation is considered, a deformable 3D model of the scene must be obtained to restrict the search for appropriate correspondences. Still, some constraints independent of additional information can be enforced to support the computation of suitable solutions. We propose a novel discontinuity preserving deformation model that captures such relaxed restrictions of general 3D motions. With our model, occlusions are handled correctly and we can interpolate directly in image space without the need to reconstruct the 3D surfaces, motion and camera parameters. The benefit of this approach is that it becomes possible to robustly estimate correspondence fields that can be used to interpolate in time and view from images recorded with unsynchronized, uncalibrated cameras and surfaces that are hard to reconstruct in 3D such as flames and hair.

2. Related Work

Image morphing, denotes interpolation between images depicting different objects from user-defined correspondences. Algorithms like [BN92], are often used in the movie industries to create visual effects. Other warping techniques have been discussed by Wolberg [Wol98], including the popular thin-plate spline interpolation which is based on point correspondences. A computationally more complex method based on point and line features was proposed by Schaefer et al. [SMW06]. In general, image morphing methods are solely based on user specified features and are thus work intensive when interpolating image sequences. Additionally, when motion discontinuities need to be taken care of, an pre-segmentation of the image into layers is necessary.

Optical flow refers to the flow field created by the spatiotemporal trajectories of image patches during an image sequence. Since the pioneering work on local and global optical flow reconstruction by Lucas and Kanade [LK81] and Horn and Schunck [HS81], respectively, a multitude of computational approaches have been devised [BSL*07] . These flow fields computed with optical flow algorithms can also be used to interpolate images. The comparison in Section 6 shows that our approach yields better results in the case of time and view interpolation.

Image-based rendering (IBR) methods achieve highly realistic rendering results using a collection of timely synchronized calibrated photographs. While some IBR methods rely solely on the number of images to minimize aliasing

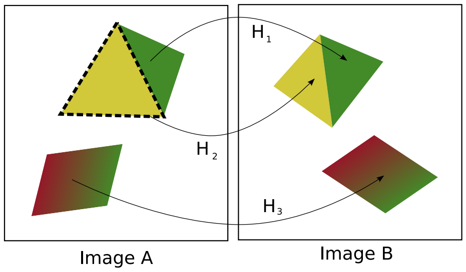

Figure 1: *Correspondences for views of a dynamic 3D scene consisting of planar surfaces can be described in image space by homographies. We define a translet as the pair of an image segment of a 3D plane and a corresponding homography. For example a translet of image A is the outlined image segment showing the bright face of the pyramid and the corresponding homography H_2 which defines its correspondence to image B.*

artifacts [LH96, MP04], most IBR approaches make additional use of epipolar constraints [MB95, SD96, MBR*00, VBK05], scene depth [CW93, GGSC96, IMG00, BBM*01, ZKU*04], or full 3D geometry information [DBY98, WAA*00, CTMS03, SSS06]. The quality of IBR techniques is strongly dependent on the accurate camera calibration, scene geometry, and/or time synchronized acquisition. These limitations make data acquisition for IBR a time-consuming and delicate endeavour which typically requires a controlled environment and special hardware.

3. Image Deformation Model for Time and View Interpolation

The relation between two projections of a 3D plane can be directly described via a homography between the homogeneous 2D coordinates of the projections (e.g. [HZ]). Such a homography can for example describe the relation between a 3D plane seen from two different cameras, the 3D rigid motion of the plane between two points in time seen from a single camera or a combination of both. Thus, the interpolation between images depicting a dynamic 3D plane can be achieved by a per pixel deformation according to the homography directly in image space without the need to reconstruct the underlying 3D plane, motion and camera parameters (cf. Figure 1). The relation between the corresponding pixels of images from a typical dynamic real-world scene on the other hand is of course far more complex. However, Computer Graphics has been very successful in creating photo-realistic images from approximations of natural scenes and objects with meshes consisting of simple planar triangles. For each such triangles the relation of the corresponding pixels is again exactly described via local homographies.

Our proposed image deformation model is motivated by these observations. We assume that natural images can be decomposed into regions, for which the deformation of each element is sufficiently well described by a homography. Specifically, we introduce *translets* which are homographies that are spatially restricted. That is, a translet is described by a 3×3 matrix H and an image segment. To obtain a dense deformation we enforce that the set of all translets is a complete partitioning of the image and thus each pixel is part of exactly one translet. Note, that since the deformation model is defined piecewise, it can well describe motion discontinuities as for example resulting from occlusions.

3.1. Connection to Human Motion Perception

Human vision is a very powerful system, adept at extracting meaningful patterns so that we can understand, navigate through, and interact with our surroundings rapidly and efficiently. Interestingly, the brain seems thereby not to be restricted by the exact limitations of physics and focuses only on certain aspects to achieve this goal. For example, animation movies and special effects often bend the physically possible to achieve realistic but impossible effects. Studying the literature about human motion perception reveals what the aspects of motion are that the visual system focuses upon. Especially the motion of edges, homogeneous regions and an the overall coherence are such important aspects. In [SLW*08] we elaborate on the connection of human vision to our interpolation approach and verify this in a user study.

4. Estimating the Image Deformation

Here we consider the estimation of dense correspondences between two images using our image deformation model. Therefore, a partitioning of the images into regions that can be approximated by 3D planes and the corresponding homographies from correspondences are computed. In this paper we estimate the homographies from matches between edge pixels of the images. Note, that in general our model is not restricted to this approach and other correspondences or methods for the estimation of the homographies can be applied. While the optimal partitioning of the images into translets is not known a priori this has great influence on the solution. A small number of translets will result in a very robust but restrictive solution, while a larger number increases the flexibility at the cost of decreased robustness against outliers in the match. To obtain an optimal result we follow a bottom up approach. We start from a large number of translets and merge neighboring translets in a greedy manner until the optimal ratio is achieved. In the following sections, we discuss the steps in estimating the image deformation between two images in detail.

4.1. Matching of Edge Pixels

The first step in establishing the image deformation model is to find a sparse set of correspondences between the images that can be used to estimate the homographies. Edges

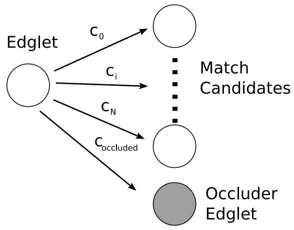

Figure 2: *Subgraph of the weighted bipartite graph matching problem for a single edglet. Each edglet has an edge to its possible match candidates and an additional edge to its occluder edglet.*

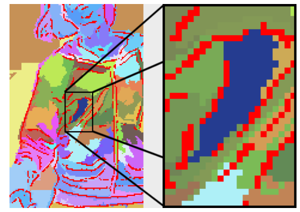

Figure 3: *The translets of an image are found by partitioning the image according to a superpixel segmentation and computing local homographies from point correspondences to the target image.*

and corners are especially suited for this as they are both relatively stable over time and viewpoints and are the image parts where motion is most apparent. For our experiments we used the Compass operator [RT99] as it has the advantage to directly make use of color information and in general outperforms the Canny operator [Can86]. After non-maximal suppression, we obtain a set of edge pixels or *edglets*.

Once the edglets of two images are found we seek a matching between them. For the given problem a good match is as complete as possible and considers the spatial context of each edglet. The shape context descriptor [BMP] performs well at capturing the spatial context of the nearest k neighbor edglets and is robust against the expected deformations. The matching problem based on the euclidean distance and shape contexts can be solved via an maximum weighted bipartite graph matching problem. The advantage is that these kinds of problems can be solved globally optimal in the matter of seconds for the problem sizes we are facing [Ber92]. One prerequisite for the reformulation is that for each edglet in the first set a match in the second set can be found. While this is true for most edglets, some will not have a correspondence in the other set due to occlusion. Thus we add a virtual occluder edglet for each edglet in the first edglet set to ensure this prerequisite. The graph for the matching problem is build as depicted in Figure 2. Each edglet of the first image has an edge to its possibly corresponding edglets of the second image and additionally to its occluder edglet. The cost function for a pair of edglets is defined as

$$C(e_i, e'_j) = C_{dist} + C_{shape} \qquad (1)$$

where the cost for the shape is the χ^2-test between the two shape contexts and the cost for the distance is defined as

$$C_{dist}(e_i, e'_j) = \frac{a}{(1 + e^{-b\,||e_i - e'_j||})} \qquad (2)$$

with $a, b > 0$ such that the maximal cost for the euclidean distance is limited by a. The cost $C_{occluded}$ is user defined and controls how aggressively the algorithm tries to find a match with an edglet of the second image. The lower $C_{occluded}$ the more conservative the resulting matching will be as more edges will be matched to their occluder edglets.

4.2. Estimating the Local Homographies

Once a match between the edglets is computed, we can use these to compute the local homographies according to our deformation model. To find the edglets that together can be used to estimate such transformations we first seek a partitioning of the images. For each such region we would assume that the motion of each region can be approximated via the relation of projections of a 3D plane as discussed in Section 3. From Gestalt theory [Wer] it is known that for natural scenes these regions share not only a common motion but in general also share other properties such as similar color and texture.

Felzenszwalb and Huttenlocher [FH04] proposed to partition images into so called superpixels based on color similarities. Thus, we can use these superpixels to find the initial translets for our image deformation model. For each translet a homography is estimated from at least four matches of the edglets that are part of the image segment (cf. Figure 3). As a least-squares estimation based on all matched edglets of a translet is sensitive to outliers we use a RANSAC approach to obtain a robust solution [HZ].

4.3. Optimization of the Deformation Model

From the point correspondences we have established dense correspondences between the images using our deformation model. However, in our experiments we observed that between 20% to 40% of the computed matches are outliers and thus some translets will have wrongly estimated transformations. We address this problem by optimizing the number of translets of our image deformation model to increase the robustness against these outliers. The initial solution of our model is generally very conservative, such that the spatial support of the translets can be too small for a reliable estimation. Using a greedy approach, we iteratively merge the most similar transformed neighboring translets into one, as depicted in Figure 4, until the ratio of outliers to inliers is lower than a user defined threshold. When two translets are merged, the resulting translet then contains both edglet sets and has the combined spatial support. The homographies are

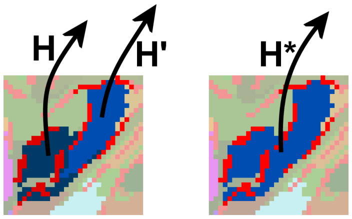

Figure 4: *During optimization similar transformed neighboring translets are merged into a single translet. After merging, the resulting translet consists of the combined spatial support of both initial translets (light blue and dark blue) and their edglets (light red and dark red).*

re-estimated based on the new edglet set and the influence of the outliers is reduced by the RANSAC filtering.

4.4. Per-Pixel Correspondences

So far smoothness and discontinuitiy is handled on the translet level. However, when only a part of a translet boundary is at a true motion discontinuity, noticeably incorrect discontinuities still produce artifacts along the rest of the boundary. An example is for example a motion of an arm in front of the body. Here the motion is discontinuities along the silhouette of the arm, while the motion at the shoulder changes continuously. This can only be solved on a per pixel basis. Since the translets partition the image, each pixel in the image is uniquely associated with a translet t. The deformation vector for a pixel x is thus computed as

$$d(x) = H_t \cdot x - x. \tag{3}$$

We can then resolve the per pixel smoothing by an anisotropic diffusion [PM90] on this vector field using the diffusion equation

$$\delta I/dt = div(\ g(min(|\nabla d|, |\nabla I|)\ \nabla I) \tag{4}$$

which is dependent on the image gradient ∇I and the gradient of the deformation vector field ∇d whichever is smaller in magnitude at the observed pixel. The function g is a simple mapping function as defined in [PM90]. Thus, the deformation vector field is smoothed in regions that have similar color or similar deformation, while discontinuities that are both present in the color image and the vector field are preserved. This improves the smoothness of the deformations on a per-pixel level while preserving important motion discontinuities. During the anisotropic diffusion, edglets that have an inlier match are considered as boundary conditions of the PDE to ensure exact edge transformations. The total timings for the computation of the deformation field for different resolutions and scenes are listed in Table 1

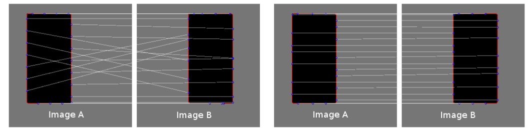

Figure 5: *Local matching minima (left) can be avoided by multiple iterations. In a coarse to fine manner in each iterations the number of translets increases avoiding local matching minima by using the previous result as prior (right).*

4.5. Multiple Iterations and User Interaction

Since our matching energy function (Eq. 1) is based on spatial proximity and local geometric similarity, we can introduce a motion prior by pre-warping the edglets with a given deformation field. The estimated dense correspondences described in the last sections can be used as such a prior. We can then implement a coarse to fine iterative approach to overcome local matching minima, as for example depicted in Figure 5, as follows: In the first iteration, we optimize the number of translets until we obtain the coarsest possible deformation model with only one translet and thus approximate the underlying motion by a single perspective transformation. During consecutive iterations the threshold is decreased to allow for more accurate deformations as the number of final translets increases while we overcome local matching minima since the previous solution is used as a motion prior

While multiple iterations with coarse to fine motion models avoid local matching minima, some scenes still can not be matched automatically sufficiently well. For example when similar structures appear multiple times in the images the matching is ambiguous and can only be resolved by high level reasoning. However, these regions can be selected in both images by the user and the automatic matching is computed again only for the so selected subset of edglets. Due to this restriction of the matching the correct match is found and used to correct the solution. Typically such user interactions take a total of two to ten minutes to achieve the desired results. Note also, that when interpolating a set of images, e.g. from a video sequence, the results of the previous motion can be extrapolated and used as a motion prior for the next frame which generally resolves the ambiguity for the following frames.

5. Interpolation Rendering

Rendering in-between images is achieved by applying the correspondence field estimated with our image deformation model to the images and blending these warped images. This can be implemented on graphics hardware using per-vertex mesh deformation and alpha blending with real time rendering performance. To get the deformations for the in-between images we linearly interpolate the deformation vector field.

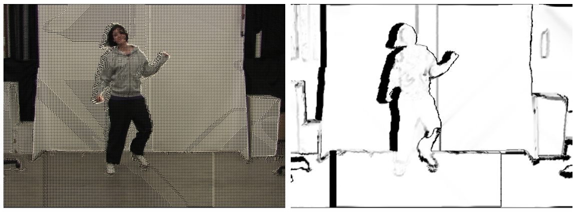

Figure 6: *Left: Per-vertex mesh deformation is used to compute the forward warping of the image, where each pixel corresponds to a vertex in the mesh. The depicted mesh is at a coarser resolution for visualization purposes. Right: The connectedness of each pixel that is used during blending to avoid a possibly incorrect influence of missing regions.*

5.1. Warping with Occlusions

We implemented the forward warping by a per-vertex deformation of a regular planar triangle mesh of the image plane, where each pixel in the image is represented by a vertex with appropriate texture coordinates. Two problems arise with forward warping at motion discontinuities: Fold-overs and missing regions.

Fold-overs occur when two or more pixels in the image end up in the same position during warping. This is the case when the foreground occludes parts of the background. Consistent with motion parallax we assume that the faster moving pixel is closer to the viewpoint to resolve this conflict. When on the other hand regions get dis-occluded during warping the information of these regions is missing in the image and must be filled in from the other image. This leaves two options in this case: cutting the mesh at the motion discontinuities *before* warping or detecting triangles that span over these discontinuities *after* rendering. Mark et al. [MMB97] pointed out that the second approach performs better and proposed a connectedness criterion evaluated on a per-pixel basis after warping. We adapt this measure and compute it directly from the divergence of the deformation vector field such that

$$c_A = 1 - div(d_{AB})^2. \tag{5}$$

Table 1: *Timing results of our method on a AMD Athlon(tm) 64 X2 Dual Core Processor 4800+, 4GB RAM, NVIDIA GeForce 7800 GTX to compute the dense correspondences. If users interact to improve the solution only parts are recomputed which reduces the response times.*

Scene	Edglets	Res.	Matching	Optim.
Dancer	2570	960x540	1.94 s	5.67 s
Dimet.	8604	584x388	11.27 s	16.54 s
Rub.Whale	13474	584x388	19.04 s	29.87 s
Hair	17560	960x540	27.77 s	35.57 s

Figure 7: *Jaggy artifacts due to aliasing artifacts can get visible at motion discontinuities. These are however easily discriminated by a threshold on the motion field. In a second rendering pass we correct the previously detected artifacts.*

with c_a the connectedness and d_{AB} vector field between the images A and B (cf. Figure 6). The connectedness is computed on the GPU during blending to adaptively reduce the alpha values of pixels with low connectedness. Thus in missing regions only the image which has the local information has an influence on the rendering result.

5.2. Feathering

At fold-overs, the warped images can have jaggy artifacts due to aliasing problems of the rendering. Opposed to recordings with cameras, rendered pixels at the boundaries are not a mixture of background and foreground color but are either foreground or background color. However, these artifacts occur only at large motion discontinuities, which can be robustly discriminated by the local change in the motion vectors by simple thresholding (cf. Figure 7). In a second rendering pass, we model the color mixing of foreground and background at boundaries using a small selective low-pass filter applied only to the detected motion boundary pixels. This effectively removes the artifacts with a minimal impact on rendering speed and without affecting rendering quality in the non-discontinuous regions.

6. Results

First, we compared our results to interpolation results based on state-of-the-art optical flow methods using the Middlebury examples [BSL*07] (cf. Table 2, Figure 8). Since these methods do not allow for user interaction we compare the results of our unimproved automatic results. As can be seen our approach is best when looking at the interpolation errors and best or up to par in the sense of the normalized interpolation error. We also like to point out that from a perception point of view the normalized error is less expressive than the unnormalized error since discrepancies at edges in the image (e.g. large gradients) are dampened. Interestingly, relatively large angular errors are observed with our method emphasizing that the requirements of optical flow estimation and image interpolation are different.

In addition, we recorded dynamic scenes with conventional, unsynchronized, and uncalibrated video cameras. The results can be seen in the accompanying video. We used off the shelf Canon HDV camcorders that have a horizontal field of view of $\approx 50^\circ$ and were spaced apart by $\approx 15^\circ$. We additionally interactively corrected local errors to further improve the results in cases where the automatic estimation failed as discussed in Section 4.5. The shown scenes also contain surfaces that are hard to reconstruct in 3D and are thus problematic for typical image based rendering methods, such as the flame of the fire-breather and the flying hair of the woman.

7. Conclusions and Future Work

In this paper, we have presented a novel interpolation method for view and time interpolation directly in image space. We introduced our image deformation model that is used to enforce relaxed physical constraints on the estimation of dense correspondences based on homographies between planes in 3D space. Favorable properties of this model are that edges are transformed exact and that motion discontinuities are preserved, and thus occlusions are handled appropriately. The benefit of our approach is that we can robustly estimate correspondences between images recorded with unsynchronized and uncalibrated cameras. We compared the results with other general state-of-the art interpolation methods and showed that our method performs best in terms of interpolation error on a set of standard examples.

Since the image deformation model is based on the sparse point correspondences, additional point correspondences such as e.g. SIFT features [Low04] can also be considered. When interpolating longer sequences, establishing correspondences over more than two images will also increase the robustness for complex and cluttered scenes.

Acknowledgements We like to thank Christian Wallraven and Douglas Cunningham for the fruitful discussions.

Table 2: *Interpolation, Normalized Interpolation and Angular errors computed on the Middlebury Optical Flow examples by comparison to ground truth with results obtained by our method and by other methods taken from [BSL*07].*

Venus	Interp.	Norm. Interp.	Ang.
Our Method	**2.88**	**0.55**	**16.24**
Pyramid LK	3.67	0.64	14.61
Bruhn et al.	3.73	0.63	8.73
Black and Anandan	3.93	0.64	7.64
Mediaplayer	4.54	0.74	15.48
Zitnick et al.	5.33	0.76	11.42
Dimetrodon	Interp.	Norm. Interp.	Ang.
Our Method	**1.78**	**0.62**	**26.36**
Pyramid LK	2.49	0.62	10.27
Bruhn et al.	2.59	0.63	10.99
Black and Anandan	2.56	0.62	9.26
Mediaplayer	2.68	0.63	15.82
Zitnick et al.	3.06	0.67	30.10
Hydrangea	Interp.	Norm. Interp.	Ang.
Our Method	**2.57**	**0.48**	**12.39**
RubberWhale	Interp.	Norm. Interp.	Ang.
Our Method	**1.59**	**0.40**	**23.58**

References

[BBM*01] BUEHLER C., BOSSE M., MCMILLAN L., GORTLER S., COHEN M.: Unstructured lumigraph rendering. In *Proc. ACM Conf. Computer Graphics (SIGGRAPH)* (2001), pp. 425–432.

[Ber92] BERTSEKAS D.: Auction algorithms for network flow problems: A tutorial introduction. *Computational Optimization and Applications 1* (1992), 7–66.

[BMP] BELONGIE S., MALIK J., PUZICHA J.: Matching Shapes. In *Proc. IEEE International Conf. Computer Vision (ICCV)*.

[BN92] BEIER T., NEELY S.: Feature-based image metamorphosis. In *Proc. ACM Conf. Computer Graphics (SIGGRAPH'92), Chicago* (1992), ACM, pp. 35–42.

[BSL*07] BAKER S., SCHARSTEIN D., LEWIS J., ROTH S., BLACK M., SZELISKI R.: A Database and Evaluation Methodology for Optical Flow. In *Proc. IEEE International Conf. Computer Vision (ICCV)* (2007), pp. 1–8.

[Can86] CANNY J.: A Computational Approach To Edge Detection. *IEEE Trans. Pattern Analysis and Machine Intelligence 8* (1986), 679–714.

[CTMS03] CARRANZA J., THEOBALT C., MAGNOR M., SEIDEL H. P.: Free-viewpoint video of human actors. In *Proc. ACM Conf. Computer Graphics (SIGGRAPH)* (2003), pp. 569–577.

[CW93] CHEN S., WILLIAMS L.: View interpolation for image synthesis. In *Proc. ACM Conf. Computer Graphics (SIGGRAPH)* (1993), pp. 279–288.

[DBY98] DEBEVEC P., BORSHUKOV G., YU Y.: Efficient view-dependent image-based rendering with projective texture-mapping. In *Proc. Eurographics Rendering Workshop (EGRW)* (1998), pp. 105–116.

[FH04] FELZENSZWALB P., HUTTENLOCHER D.: Efficient Graph-Based Image Segmentation. *International Journal of Computer Vision 59* (2004), 167–181.

[GGSC96] GORTLER S., GRZESZCZUK R., SZELISKI R., COHEN M.: The Lumigraph. In *Proc. ACM Conf. Computer Graphics (SIGGRAPH)* (1996), pp. 43–54.

[HS81] HORN B., SCHUNCK B.: Determining Optical Flow. *Artificial Intelligence 17* (1981), 185–203.

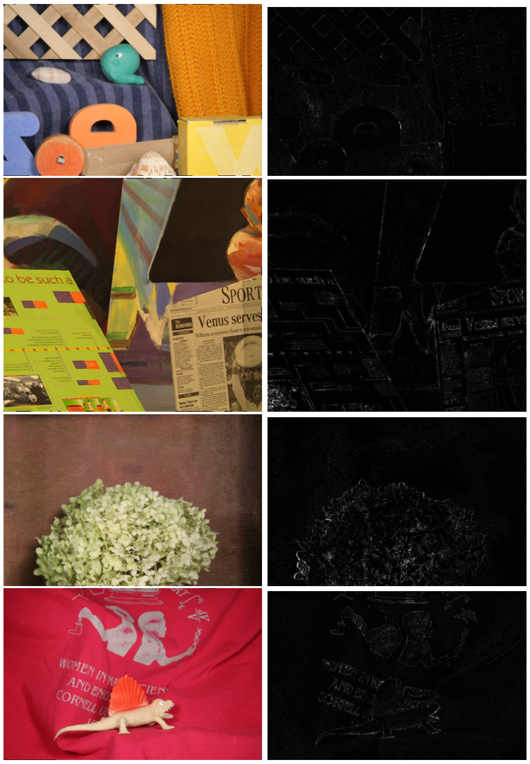

Figure 8: *Results on the Middlebury dataset [BSL*07]. (Left Column) In-between image automatically computed with our method. (Right Column) Contrast-stretched difference to ground truth.*

[HZ] HARTLEY R., ZISSERMAN H.: *Multiple View Geometry in Computer Vision*.

[IMG00] ISAKSEN A., MCMILLAN L., GORTLER S.: Dynamically reparameterized light fields. In *Proc. ACM Conf. Computer Graphics (SIGGRAPH)* (2000), pp. 297–306.

[LH96] LEVOY M., HANRAHAN P.: Light field rendering. In *Proc. ACM Conf. Computer Graphics (SIGGRAPH)* (1996), pp. 31–42.

[LK81] LUCAS B., KANADE T.: An iterative image registration technique with an application to stereo vision. In *Proc. Seventh International Joint Conf. Artificial Intelligence* (1981), pp. 674–679.

[Low04] LOWE D.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*, 2 (2004), 91–110.

[MB95] MCMILLAN L., BISHOP G.: Plenoptic modeling: An image-based rendering system. *Proc. ACM Conf. Computer Graphics (SIGGRAPH)* (1995), 39–46.

[MBR*00] MATUSIK W., BUEHLER C., RASKAR R., GORTLER S., MCMILLAN L.: Image-based visual hulls. In *Proc. ACM Conf. Computer Graphics (SIGGRAPH)* (2000), pp. 369–374.

[MMB97] MARK W., MCMILLAN L., BISHOP G.: Post-Rendering 3D Warping. In *Proceedings of the Symposium on Interactive 3D Graphics* (1997), pp. 7–16.

[MP04] MATUSIK W., PFISTER H.: 3D TV: A scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. In *Proc. ACM Conf. Computer Graphics (SIGGRAPH)* (2004), pp. 814–824.

[PM90] PERONA P., MALIK J.: Scale-space and edge detection using anisotropic diffusion. *Trans. Pattern Analysis and Machine Intelligence 12*, 7 (1990), 629–639.

[RT99] RUZON M., TOMASI C.: Color Edge Detection with the Compass Operator. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)* (1999), pp. 160–166.

[SD96] SEITZ S. M., DYER C. R.: View morphing. In *Proc. ACM Conf. Computer Graphics (SIGGRAPH)* (1996), pp. 21–30.

[SLW*08] STICH T., LINZ C., WALLRAVEN C., CUNNINGHAM D., MAGNOR M.: Perception-motivated image sequence interpolation. In *Proc. ACM Conf. Applied Perception in Graphics and Visualization (APGV)* (2008).

[SMW06] SCHAEFER S., MCPHAIL T., WARREN J.: Image deformation using moving least squares. In *Proc. ACM Conf. Computer Graphics (SIGGRAPH)* (2006), pp. 533–540.

[SSS06] SNAVELY N., SEITZ S., SZELISKI R.: Photo tourism: exploring photo collections in 3d. In *Proc. ACM Conf. Computer Graphics (SIGGRAPH)* (2006), pp. 835–846.

[VBK05] VEDULA S., BAKER S., KANADE T.: Image based spatio-temporal modeling and view interpolation of dynamic events. *ACM Trans. Graphics 24*, 2 (2005), 240–261.

[WAA*00] WOOD D., AZUMA D., ALDINGER K., CURLESS B., DUCHAMP T., SALESIN D., STUETZLE W.: Surface light fields for 3D photography. In *Proc. ACM Conf. Computer Graphics (SIGGRAPH)* (2000), pp. 287–296.

[Wer] WERTHEIMER M.: Laws of organization in perceptual forms. In *A Source Book of Gestalt Psychology*. pp. 71–88.

[Wol98] WOLBERG G.: Image morphing: A survey. *Visual Computer 14* (1998), 360–372.

[ZKU*04] ZITNICK C., KANG S., UYTTENDAELE M., WINDER S., SZELISKI R.: High-quality video view interpolation using a layered representation. In *Proc. ACM Conf. Computer Graphics (SIGGRAPH)* (2004), pp. 600–608.