Hindawi Publishing Corporation
EURASIP Journal on Image and Video Processing
Volume 2008, Article ID 803231, 15 pages
doi:10.1155/2008/803231

Research Article

Occlusion-Aware View Interpolation

Serdar Ince[1,2] and Janusz Konrad (EURASIP Member)[1]

[1] *Department of Electrical and Computer Engineering, Boston University, 8 Saint Mary's Street, Boston, MA 02215, USA*
[2] *IntelliVid Corporation, Cambridge, MA 02138, USA*

Correspondence should be addressed to Janusz Konrad, jkonrad@bu.edu

Received 3 March 2008; Accepted 1 October 2008

Recommended by Peter Eisert

View interpolation is an essential step in content preparation for multiview 3D displays, free-viewpoint video, and multiview image/video compression. It is performed by establishing a correspondence among views, followed by interpolation using the corresponding intensities. However, occlusions pose a significant challenge, especially if few input images are available. In this paper, we identify challenges related to disparity estimation and view interpolation in presence of occlusions. We then propose an occlusion-aware intermediate view interpolation algorithm that uses four input images to handle the disappearing areas. The algorithm consists of three steps. First, all pixels in view to be computed are classified in terms of their visibility in the input images. Then, disparity for each pixel is estimated from different image pairs depending on the computed visibility map. Finally, luminance/color of each pixel is adaptively interpolated from an image pair selected by its visibility label. Extensive experimental results show striking improvements in interpolated image quality over occlusion-unaware interpolation from two images and very significant gains over occlusion-aware spline-based reconstruction from four images, both on synthetic and real images. Although improvements are obvious only in the vicinity of object boundaries, this should be useful in high-quality 3D applications, such as digital 3D cinema and ultra-high resolution multiview autostereoscopic displays, where distortions at depth discontinuities are highly objectionable, especially if they vary with viewpoint change.

1. INTRODUCTION

The generation of a novel (virtual) view of a scene captured by real cameras is often referred to as *image-based rendering*. The problem is illustrated for the case of two cameras in Figure 1. The goal is to reconstruct image J, that would have been captured by camera C_J had it been used, based on images I_L and I_R captured by cameras C_L and C_R, respectively. Generation of such views is an essential step in content preparation for multiview 3D displays [1–3], free-viewpoint video [4, 5], and multiview compression [6–8]. A very similar problem exists in frame-rate conversion of video, except that novel images are created from different-time snapshots rather than different views.

In order to render novel view, first a correspondence among known views needs to be established followed by an estimation of new intensity from the known intensities in correspondence. Depending on how the correspondence mapping is defined, two approaches are possible. One approach is based on *backward projection* of intensities (the term "backward projection" is borrowed from the field of video coding where it refers to predicting luminance/color from previous (in time) frames), where the mapping is defined in the coordinate system of the unknown view (J in Figure 1). Thus, this approach simplifies the final estimation to an interpolation problem. The other approach is based on *forward projection* of intensities, where the mapping is defined in the coordinate system of one of the known views (I_L or I_R in Figure 1), thus making the final estimation more difficult since projected intensities do not, in general, belong to the sampling grid of J. In fact, the problem cannot be solved in this case by interpolation, and the novel-view intensities must be approximated instead. Typically accomplished by means of additional constraints, this process is often referred to as view reconstruction. We will review these two approaches in more detail in Section 2.

One of the significant challenges in image-based rendering is dealing with occlusion areas (see Figures 1(b)–1(c)). By the term *occlusion area,* we mean an area in one input image disappearing from the other input image due

to scene structure, for example, area A in I_L is occluded in I_R (see Figure 1). Note that a disappearing area becomes an appearing area (also known as *uncovered* or *newly-exposed area*), and vice versa, if the order of views is reversed, that is, "right-to-left" instead of "left-to-right."

Many approaches to novel view generation have been proposed to date. Although some methods account for occlusions, few handle occlusions accurately. Consequently, occlusion areas are recovered inaccurately. Our motivation in this paper is to improve the novel image quality in occlusion areas. This is somewhat easier in approaches based on forward projection since the correspondence mapping is defined in the coordinate system of known images, and thus known luminance/color can be used to reason about the presence/absence as well as nature of occlusions. We have recently developed successful methods in this category [9, 10]. On the other hand, in backward-projection methods the mapping is defined in the coordinate system of a novel image and thus no luminance/color is available to reason about occlusions. We address this difficulty here. We propose a new occlusion-aware backward-projection view interpolation. The method first identifies pixel visibility in the intermediate image, that is, whether a particular pixel is visible in all input images or only in those to the left or to the right of the image to be reconstructed. These labels are incorporated into a variational formulation to adaptively choose different pairs of input images and reliably estimate disparity under anisotropic regularization constraint. The final view generation is accomplished by occlusion-adaptive linear intensity interpolation.

The paper is organized as follows. In Section 2, we review prior work on intermediate view interpolation and reconstruction as well as occlusion detection. In Section 3, we present the new occlusion-aware view interpolation, and in Section 4, we show experimental results. In Section 5, we discuss benefits and deficiencies of forward- and backward-projection approaches, and in Section 6, we summarize the paper and draw conclusions.

2. PRIOR WORK

Image-based rendering is concerned with creating an image at a specific 3D location and specific time. Adelson and Bergen [11] formulated a description for all possible images by means of the so-called *plenoptic function* that records light rays at every possible 3D location, in every possible direction, at every time instant, and for all wavelengths. In order to generate a new image, one simply needs to sample this 7-dimensional function. However, capturing a full plenoptic function is difficult, if not impossible, and thus various assumptions aiming at the reduction of this high dimensionality have been proposed. For example, if only static scenes are considered in grayscale, the number of dimensions reduces to five.

Although prior work can be classified based on the number of dimensions of a plenoptic function used [12], in the context of work proposed, here we prefer to classify prior methods based on their need for structure information and the number of input images.

(1) *Methods that rely on oversampling.* Among the most prominent methods that rely on scene oversampling are lightfield rendering [13] and lumigraph [14]. Both methods create a 4D representation of the scene using many input images. The novel views are created by slicing (sampling) this 4D representation. Since the scene is oversampled, the rendering process simply blends the input images, ignoring scene structure. The presence of occlusions is not a problem because, thanks to oversampling, occlusions between nearest cameras are negligible, and all texture in the scene is visible from several cameras.

(2) *Methods that use undersampled data sets with known structure.* Given the scene structure, it is possible to reduce the required number of images [15–18]. If the depth map or 3D model of a scene is available, it is possible to project pixels of the known images to a new viewpoint and reconstruct a new image. Obviously, it is not guaranteed that all pixels in the new image will be visible in the input images. However, since the scene structure is known, locations of occlusions are known, which is not the case considered in this paper.

(3) *Methods that use severely undersampled data sets with unknown structure.* These methods have no access to scene structure and use few input images, typically 2–4. The scene structure is computed implicitly (from disparity) either using correspondence matching or projective geometry. These methods can be categorized based on what approach they use to estimate the disparity: methods based on projective geometry or rectification [19–21], methods based on optical flow [5, 9, 22, 23], methods based on block correspondence: variable-size blocks [24], fixed-size blocks [25], sliding blocks [26], methods based on feature correspondence [27], and methods using dynamic programming [28]. Because of limited input data and unknown scene structure, the reconstruction problem is ill-posed and requires some from of the regularization, usually by means of additional constraints. The work presented in this paper is closest to this class of methods.

2.1. Forward- and backward-projection methods

When computing an intermediate view, the central role is played by a transformation between the coordinate systems of known images and the novel image. This transformation depends on camera geometry and scene structure, and is usually unknown. It can be estimated by solving the correspondence problem with two possible definitions of the transformation: from known to novel image coordinates, also called *forward projection*, or from novel to known image coordinates, called *backward projection*.

Let I_L and I_R be images of the same scene captured on 2D lattice Λ by two cameras. We assume the distance between the cameras is normalized to 1. Suppose we need to reconstruct intermediate view J, also defined on Λ, but at distance $0 < \alpha < 1$ from I_L. Clearly, for $\alpha = 0$, $J = I_L$, whereas for $\alpha = 1$, $J = I_R$ (see Figure 2). Due to this simple stereo setup, the transformation mentioned above simplifies to a disparity field between I_L and I_R.

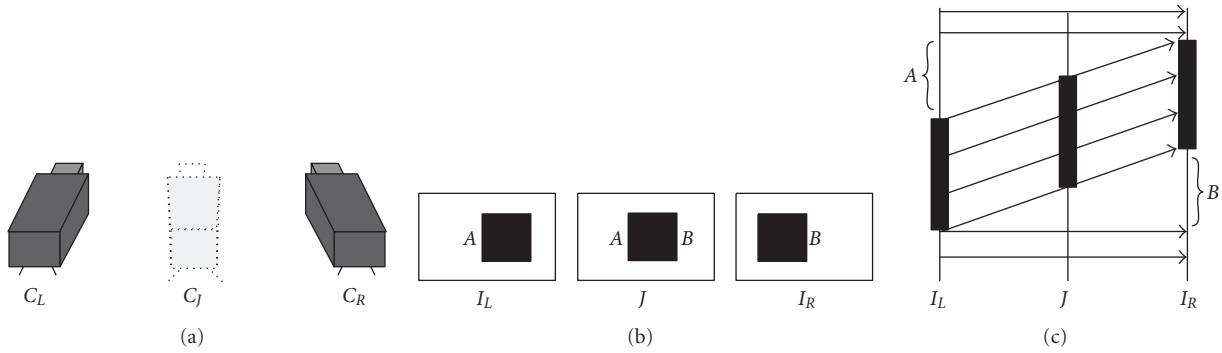


FIGURE 1: Illustration of intermediate view reconstruction from two cameras: (a) camera setup (C_J is a virtual camera, while C_L and C_R are real cameras), (b) occlusion effect in captured images, and (c) occlusion effect in one row of pixels from the images. Area A from I_L is being occluded in I_R by the object, while area B is being uncovered (area B would undergo occlusion had the direction of arrows been reversed).

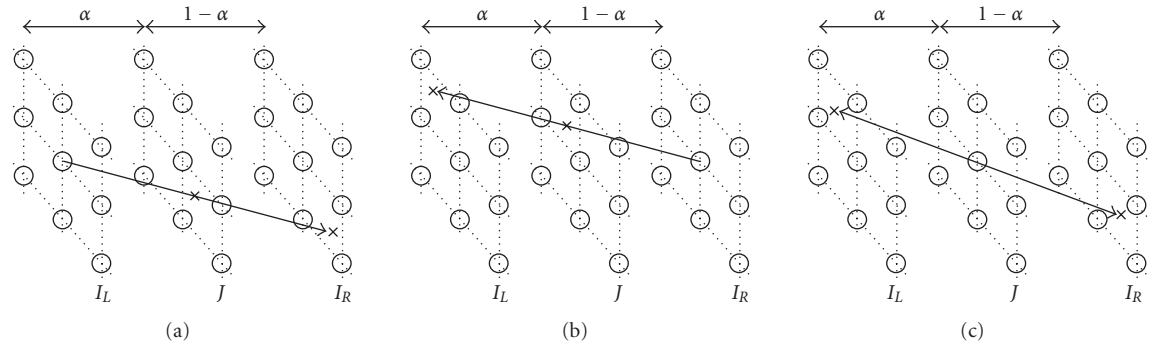


FIGURE 2: Disparity vectors defined (pivoted) in: (a) left (known), (b) right (known), and (c) intermediate (unknown) images.

2.1.1. Forward projection

Disparity vectors (transformation) are defined in the coordinate system of known images. Let $\mathbf{d}_L$ be a disparity field defined on lattice Λ of I_L (see Figure 2(a)), and let $\mathbf{d}_R$ be defined on lattice Λ of I_R (see Figure 2(b)). Under the constant-brightness assumption [29], the following holds

$$I_L(\mathbf{x}) = I_R(\mathbf{x} + \mathbf{d}_L(\mathbf{x})), \quad I_R(\mathbf{x}) = I_L(\mathbf{x} + \mathbf{d}_R(\mathbf{x})), \quad \forall \mathbf{x} \in \Lambda. \tag{1}$$

Assuming that brightness constancy holds along the whole disparity vector, also the following is true:

$$\begin{aligned} J(\mathbf{x} + \alpha\mathbf{d}_L(\mathbf{x})) &= I_L(\mathbf{x}), \\ J(\mathbf{x} + (1-\alpha)\mathbf{d}_R(\mathbf{x})) &= I_R(\mathbf{x}), \end{aligned} \quad \forall \mathbf{x} \in \Lambda. \tag{2}$$

Clearly, the reconstruction of intermediate-view intensities $J(\mathbf{x} + \alpha\mathbf{d}_L(\mathbf{x}))$ and $J(\mathbf{x} + (1-\alpha)\mathbf{d}_R(\mathbf{x}))$ can be as simple as substitution with $I_L(\mathbf{x})$ and $I_R(\mathbf{x})$, respectively. However, in general, $\mathbf{x} + \alpha\mathbf{d}_L(\mathbf{x}) \notin \Lambda$ and $\mathbf{x} + (1-\alpha)\mathbf{d}_R(\mathbf{x}) \notin \Lambda$, that is, the projected points are off lattice Λ. In fact, due to the space-variant nature of disparities, the above locations are usually irregularly spaced, whereas the goal is to reconstruct $J(\mathbf{x})$ regularly spaced ($\mathbf{x} \in \Lambda$). One option is to force the locations $\mathbf{x} + \alpha\mathbf{d}_L(\mathbf{x})$ and $\mathbf{x} + (1-\alpha)\mathbf{d}_R(\mathbf{x})$ to belong to Λ. For orthonormal lattices typically used, this means forcing $\alpha\mathbf{d}_L(\mathbf{x})$ and $(1-\alpha)\mathbf{d}_R(\mathbf{x})$ to be full-pixel vectors, that is, rounding coordinates to the nearest integer [21, 25]. Advanced approaches, such as those using splines to perform irregular-to-regular conversion, have also been proposed [9]. While simple rounding suffers from objectionable reconstruction errors, advanced spline-based methods produce high-quality reconstructions but require significant computational effort.

2.1.2. Backward projection

Disparity vectors are defined in the coordinate system of the intermediate image J, and bidirectionally point toward the known images [24, 30, 31]. As shown in Figure 2(c), $\mathbf{d}_J$ is defined on Λ in J thus forcing disparity vectors to pass through pixel positions of the intermediate view (i.e., vectors are *pivoted* in the intermediate view). The constant-brightness assumption now becomes

$$I_L(\mathbf{x} - \alpha\mathbf{d}_J(\mathbf{x})) = I_R(\mathbf{x} + (1-\alpha)\mathbf{d}_J(\mathbf{x})), \quad \forall \mathbf{x} \in \Lambda. \tag{3}$$

Compared to (1), each pixel in J is guaranteed to be assigned a disparity vector and, therefore, two intensities (from I_L and I_R) associated with it. Although usually $\mathbf{x} - \alpha\mathbf{d}_J(\mathbf{x}) \notin \Lambda$ and $\mathbf{x} + (1-\alpha)\mathbf{d}_J(\mathbf{x}) \notin \Lambda$, intensities at these points can be easily calculated from I_L and I_R using spatial interpolation.

In order to compute J at distance α, a disparity field pivoted at α is needed. Although this necessitates disparity

estimation for each α, it also simplifies the final computation of J. The reason is that view rendering becomes a byproduct of disparity estimation; once $\mathbf{d}_J$ that satisfies (3) is found, either left or right luminance/color can be used for the intermediate-view texture. An even better reconstruction is accomplished when weighted averaging (linear interpolation) of both intensities is applied [24, 32]

$$J(\mathbf{x}) = (1-\alpha)I_L(\mathbf{x} - \alpha\mathbf{d}_J(\mathbf{x})) + \alpha I_R(\mathbf{x} + (1-\alpha)\mathbf{d}_J(\mathbf{x})), \quad \forall \mathbf{x} \in \Lambda. \tag{4}$$

Clearly, all intermediate-view pixels are assigned an intensity, and postprocessing is not needed.

2.2. Occlusion-aware image-based rendering

In the case of oversampled data sets, if occlusions can be reliably identified, then selection of visible features is not difficult (many views are available). In fact, explicit detection of occlusions is not even necessary; robust photo-consistent measures embedded into the rendering algorithm are sufficient [33].

In the case of undersampled data sets, the situation is different, especially when scene structure (depth) is unknown. In fact, occlusions have dual impact in this case. First, correspondence (disparity) is not defined in occlusion areas, and thus some a priori assumptions must be made about correspondences (e.g., smoothness). Secondly, during the estimation of disparities unreliable estimates in occlusion areas impact the outcome at neighboring positions, thus spreading the occlusion-related errors. Knowing where occlusions take place can help correcting both problems.

In forward-projection methods, pixels from I_L (see Figure 2(a)) or I_R (see Figure 2(b)) that are occluded in the other image can be assigned a disparity based on depth constancy assumption [34], that does not work well at object boundaries, or by means of edge-preserving disparity inpainting [9], that has been shown to be more accurate. The latter approach is possible since disparities are defined in the coordinate system of known images (I_L or I_R), and thus their underlying gradients can be used to guide anisotropic disparity diffusion that improves the quality of estimated disparities (discontinuities) [35, 36].

In backward-projection methods, disparity is defined in the coordinate system of the unknown image J, and no underlying gradients are available to permit anisotropic diffusion. Therefore, the estimated disparities are usually excessively smooth. Although robust error metrics can be used in regularization [37], this is often insufficient. Moreover, it is unclear how to identify occlusions using a single disparity field. These are the main issues we address in this paper.

As for occlusion detection, it usually exploits one of several constraints. An ordering constraint preserves pixel order on corresponding rows of left and right images [38] but cannot handle thin foreground objects or narrow holes. A uniqueness constraint assures one-to-one mapping of pixels on corresponding rows [39]. In one implementation, it relies on the geometry of disparity fields; a significant difference between forward (e.g., left-to-right) and backward (e.g., right-to-left) disparity vectors is indicative of occlusions [40]. This constraint can also be thought of as a *geometric constraint* as it relies on the analysis of disparity field geometry. Some other geometric constraints assume that disparity varies smoothly everywhere except object boundaries (continuity constraint) [39], or that occlusion areas exhibit excessive disparity gradient [41]. Yet another geometric constraint seeks uncovered pixels in I_R by inspecting an irregular grid of forward disparity-compensated pixels of image I_L. This constraint has been shown to be very effective and noise resilient in occlusion detection [42]. A related, although weaker, visibility constraint [43] also assures consistency of uncovered pixels in one image with disparity of the other image, but it permits many-to-one matches in visible areas. Finally, a *photometric constraint* (or constant-brightness constraint [29]) ensures intensity match in visible areas. It is the simplest indicator of occlusions but prone to errors in presence of image noise and illumination changes. Methods based on multiple views compare intensity consistency along a path formed by displacement vectors in 3 or more frames [44–46]. Graph cuts have also been used in multiview occlusion detection [47].

3. OCCLUSION-AWARE BACKWARD-PROJECTION VIEW INTERPOLATION

In backward-projection methods, disparities estimated around occlusion areas are erroneous since no underlying image gradients are available. Lack of image gradient prevents the use of edge-preserving (anisotropic) diffusion. Below, we argue that by using a coarse estimate of the intermediate image the fidelity of disparity field can be significantly improved around occlusion areas. With this capacity to compute more accurate disparities, we then propose a new approach to occlusion-aware backward-projection view interpolation.

3.1. Edge-preserving disparity regularization using a coarse intermediate image

Edge-preserving (anisotropic) regularization preserves disparity edges better than isotropic diffusion [35, 48, 49] but requires an image gradient to guide the diffusion process. Since in backward-projection methods the disparity is defined on the sampling grid of the unknown image J, no such gradient is available. However, it turns out that simple backward-projection view interpolation described in Section 2.1.2 produces intermediate views with reliable edge information despite distorted texture in occlusion areas. Although this may seem counterintuitive, the reason is that visible edges are easily matched and thus are prominent in the interpolated view. Figure 3 shows an experimental result proving this point. Images shown in Figures 3(a) and 3(c) are the input left and right images, and the one in Figure 3(b)

is the true intermediate image. Disparity, estimated using simple isotropic regularization:

$$\arg\min_{\mathbf{d}(\mathbf{x})} \iint_{\mathbf{x}\in\Omega_J} \left(I_L(\mathbf{x}-\alpha\mathbf{d}(\mathbf{x})) - I_R(\mathbf{x}+(1-\alpha)\mathbf{d}(\mathbf{x}))\right)^2 + \lambda\left(\|\nabla u\|^2 + \|\nabla v\|^2\right) d\mathbf{x}, \tag{5}$$

where Ω_J is the domain of J, $\mathbf{d}(\mathbf{x}) = [\ u(\mathbf{x})\ \ v(\mathbf{x})\ ]^T$, and ∇ is the gradient operator, is shown in Figure 3(d). Clearly, it is excessively smooth. Figure 3(e) shows an intermediate image computed by using this disparity in (4). Although there are significant texture errors (as clear from Figure 3(f)), edge maps, obtained using the Canny edge detector, are very similar for the true and reconstructed intermediate images (see Figures 3(g) and 3(h)).

Therefore, we propose to use a coarse intermediate image J_c, computed using isotropically-diffused disparities (5), to guide edge-preserving regularization as follows:

$$\arg\min_{\mathbf{d}(\mathbf{x})} \iint_{\mathbf{x}\in\Omega_J} \left(I_L(\mathbf{x}-\alpha\mathbf{d}(\mathbf{x})) - I_R(\mathbf{x}+(1-\alpha)\mathbf{d}(\mathbf{x}))\right)^2 + \lambda\left(F_{\mathbf{x}}(u, J_c) + F_{\mathbf{x}}(v, J_c)\right) d\mathbf{x}. \tag{6}$$

Above, $F_{\mathbf{x}}(\cdot)$ assures anisotropic regularization [50] and is defined as follows:

$$F_{\mathbf{x}}(u, J_c) = \nabla^T u(\mathbf{x}) \begin{bmatrix} g(|J_c^x(\mathbf{x})|) & 0 \\ 0 & g(|J_c^y(\mathbf{x})|) \end{bmatrix} \nabla u(\mathbf{x}), \tag{7}$$

where $g(\cdot)$ is a monotonically decreasing function, and J_c^x, J_c^y are horizontal and vertical derivatives of J_c at $\mathbf{x}$. If $|J_c^x(\mathbf{x})| = |J_c^y(\mathbf{x})|$, then isotropic smoothing takes place ((6) simplifies to (5), except for different λ). However if, for example, $|J_x^x(\mathbf{x})| \gg |J_c^y(\mathbf{x})|$ then stronger smoothing takes place vertically, and the vertical edge is preserved.

The disparity field shown in Figure 3(i) was computed using formulation (6). It is clear that the object shape is very well preserved. The intermediate view obtained using this disparity field in backward projection (4) and its interpolation error are shown in Figures 3(j) and 3(k), respectively. As is clear from error images, distortions along the horizontal boundaries of the square are suppressed compared to Figure 3(f) because the excessive smoothness of disparity field is eliminated. Although these are nonoccluding boundaries, they were assigned incorrect disparities due to isotropic regularization (5). Edge-preserving regularization (6) corrected the problem, and these areas are now assigned accurate disparities. Consequently, the intermediate image is properly reconstructed there. Significant errors, however, persist in occlusion areas (vertical boundaries, see Figure 3(k)). This is due to occlusion unawareness of the algorithm; a point is visible only in one of the images, but reconstruction based on backward projection (4) averages intensities from both images. Therefore, next we propose to use additional images to solve for occlusion areas.

3.2. Backward-projection view interpolation using multiple images

In order to improve reconstruction in occlusion areas, we first need to estimate their locations, and then figure out what intensities belong there. Without loss of generality, let us consider four input images as shown in Figure 4. Although this is a simple scenario, it does convey the main idea we intend to pursue. While the top row shows images containing a black square against background containing areas A and B, the bottom row shows their horizontal cross-sections (rows of pixels). The goal is to reconstruct the intermediate image J using input images I_1, I_2, I_3, and I_4. Note that areas A and B are being occluded/exposed between the four images.

In occlusion-unaware interpolation (4), I_2 and I_3 would be the input images, and a disparity field defined on J would be estimated. For most points in J, it is possible to estimate accurate disparities because the corresponding points are visible in both I_2 and I_3. However, areas A and B are occluded in either I_2 or I_3, and it is not possible to estimate disparities there. Note that areas A and B are visible in additional images to the left of I_2 and to the right of I_3. Thus, it should be possible to estimate disparities in area A using I_1 and I_2, and disparities in area B using I_3 and I_4. Therefore, a formulation is needed to estimate disparities of J by choosing between three image pairs: (I_1, I_2), (I_3, I_4), or (I_2, I_3).

In order to implement switching between image pairs, one first needs to identify areas A and B. We propose to use a method that we had developed earlier [42]. Given a disparity field between two images, this method identifies areas that will be exposed between images, and such areas are equivalent to occluded areas when target and reference images are interchanged. The method is based on the fact that pixels in the target image, that did not exist in the reference image (i.e., newly-exposed pixels), have no relationship to the reference image and, as such, cannot be pointed to by disparity vectors. Thus, when pixels of reference image are forward disparity compensated onto target image, these areas are empty and can be easily detected.

Since we need to identify areas that disappear to the left and to the right of J, we must estimate two disparity fields: $\mathbf{d}_{12}$ defined in I_1 and pointing to I_2, and $\mathbf{d}_{43}$ defined in I_4 and pointing to I_3. We use formulation (6) with I_1 (I_4) used for edge-preserving regularization when computing $\mathbf{d}_{12}$ ($\mathbf{d}_{43}$).

Our occlusion detection method [42] yields the area B by using $(1+\alpha)\mathbf{d}_{12}$. The coefficient $(1+\alpha)$ is needed to normalize the disparity field so that it is correctly mapped onto J (see Figure 4). The estimated area B is exposed between I_1 and J, and, therefore, visible in I_3 and I_4. Similarly, using $(2-\alpha)\mathbf{d}_{43}$ yields area A, which is visible in I_1 and I_2. Let $L(\mathbf{x})$ be a visibility label at location $\mathbf{x}$ in J that we wish to estimate. Clearly, by using $\mathbf{d}_{12}$ and $\mathbf{d}_{34}$, we can label all points in J as visible in I_1 and I_2 only ($L(\mathbf{x}) = -1$), visible in I_3 and I_4 only ($L(\mathbf{x}) = 1$), or visible in I_2 and I_3 ($L(\mathbf{x}) = 0$). (The actual label values have no importance; other values, such as 1, 2, and 3, could have been chosen.)

With the visibility of points in J identified, we can now reliably compute each point's disparity from a suitable pair of images and also prevent oversmoothing via edge-preserving

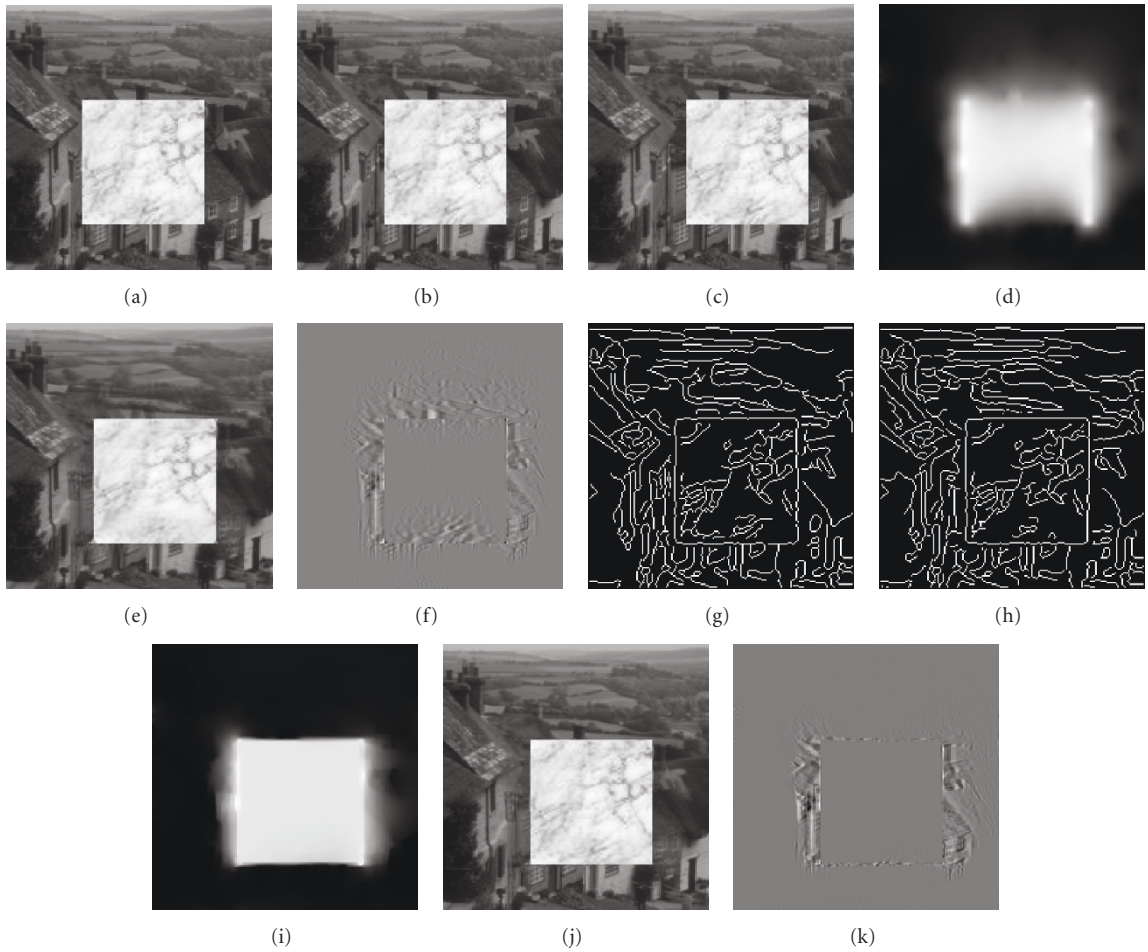


Figure 3: Results of backward-projection view interpolation for synthetic sequence no.1 with horizontal disparity: (a) I_L, (b) ground-truth J, (c) I_R, (d) disparity estimated using isotropic diffusion (5), (e) J interpolated using (4) with disparity from (d), (f) interpolation error for J from (e), (g) edge map of ground-truth image J, (h) edge map of interpolated image J, (i) disparity estimated using anisotropic diffusion (6) with J_c from (e), (j) J interpolated using (4) with disparity from (i), (k) interpolation error for J from (j). See Table 1 for PSNR values of the interpolation error.

regularization. We first define matching errors for image pairs (I_1, I_2), (I_2, I_3), and (I_3, I_4) as follows:

$$\begin{aligned}\theta_{12}(\mathbf{x}) &= I_1(\mathbf{x} - (1+\alpha)\mathbf{d}(\mathbf{x})) - I_2(\mathbf{x} - \alpha\mathbf{d}(\mathbf{x})),\\ \theta_{23}(\mathbf{x}) &= I_2(\mathbf{x} - \alpha\mathbf{d}(\mathbf{x})) - I_3(\mathbf{x} + (1-\alpha)\mathbf{d}(\mathbf{x})),\\ \theta_{34}(\mathbf{x}) &= I_3(\mathbf{x} + (1-\alpha)\mathbf{d}(\mathbf{x})) - I_4(\mathbf{x} + (2-\alpha)\mathbf{d}(\mathbf{x})).\end{aligned} \tag{8}$$

The coefficients $(1-\alpha)$, $(1+\alpha)$, $(2-\alpha)$ adjust disparity vectors depending on the distance to J. For locations $\mathbf{x} \in \Omega_J$ outside of A and B, all three errors yield small magnitudes. However, in occlusion areas only one of them will have a small magnitude. For example, for $\mathbf{x}$ in area A, $\theta_{12}(\mathbf{x})$ will have a small magnitude, whereas for $\mathbf{x}$ in area B the magnitude of $\theta_{34}(\mathbf{x})$ will be small.

In order to estimate disparities either bidirectionally (visible pixels) or unidirectionally (occlusion areas), we propose the following variational formulation that controls intensity matching using labels L under edge-preserving regularization:

$$\min_{\mathbf{d}} E, \quad E = \iint_{\mathbf{x}\in\Omega_J} e_P(\mathbf{x}) + \lambda e_S(\mathbf{x})\, d\mathbf{x} \tag{9}$$

with

$$\begin{aligned}e_P(\mathbf{x}) &= P_{12}(\mathbf{x}) + P_{23}(\mathbf{x}) + P_{34}(\mathbf{x}),\\ e_S(\mathbf{x}) &= F_{\mathbf{x}}(u, J_c) + F_{\mathbf{x}}(v, J_c),\end{aligned} \tag{10}$$

$$\begin{aligned}P_{12}(\mathbf{x}) &= \delta(L(\mathbf{x}) + 1)\,(\theta_{12}(\mathbf{x}))^2,\\ P_{23}(\mathbf{x}) &= \delta(L(\mathbf{x}))\,(\theta_{23}(\mathbf{x}))^2,\\ P_{34}(\mathbf{x}) &= \delta(L(\mathbf{x}) - 1)\,(\theta_{34}(\mathbf{x}))^2,\end{aligned} \tag{11}$$

where $F_{\mathbf{x}}$ is defined in (7), J_c is a coarse intermediate image reconstructed using disparity estimation with isotropic regularization, as proposed in Section 3.1, and $\delta(x)$ is the Kronecker delta function. Clearly, e_P adaptively selects different pairs of input images depending on L. For example,

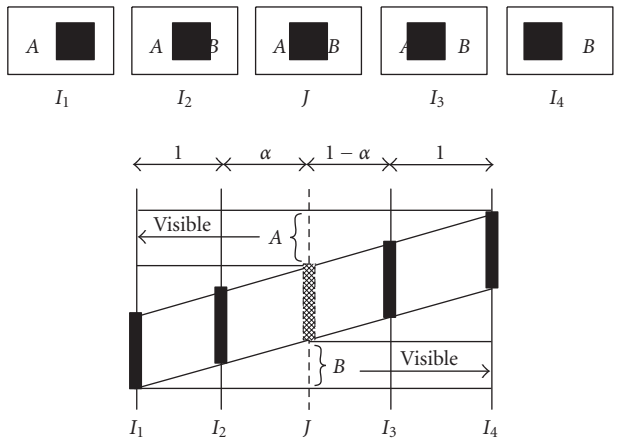


Figure 4: Illustration of how to use four images in backward-projection intermediate view interpolation. Areas A and B can be estimated in J using (I_1, I_2) and (I_3, I_4), respectively, while points outside of A or B can be estimated using (I_2, I_3).

if $L(\mathbf{x}) = -1$, then $P_{12}(\mathbf{x})$ is used. Since Kronecker delta $\delta(x)$ is not differentiable, we use an approximation, such as $\delta(x) = \lim_{k\to\infty} e^{-kx^2}$ ($k = 10^{10}$ gives good approximation). The derivation of Euler-Lagrange equations for the above variational formulation is included in the appendix.

Once the disparity field has been estimated, it is possible to reconstruct J by using any intensity value along the disparity vector, but averaging leads to better results (noise suppression). We propose to reconstruct the intermediate view J as follows:

$$J(\mathbf{x}) = \delta(L(\mathbf{x})+1)\xi_{12} + \delta(L(\mathbf{x}))\xi_{23} + \delta(L(\mathbf{x})-1)\xi_{34} \quad \forall \mathbf{x} \in \Omega_J, \tag{12}$$

where $\xi_{\cdot}$ are intensity averages along disparity vector $\mathbf{d}(\mathbf{x})$ defined as follows:

$$\begin{aligned}
\xi_{12} &= \frac{1}{2}\left[I_1(\mathbf{x} - (1+\alpha)\mathbf{d}(\mathbf{x})) + I_2(\mathbf{x} - \alpha\mathbf{d}(\mathbf{x}))\right],\\
\xi_{23} &= \frac{1}{2}\left[I_2(\mathbf{x} - \alpha\mathbf{d}(\mathbf{x})) - I_3(\mathbf{x} + (1-\alpha)\mathbf{d}(\mathbf{x}))\right],\\
\xi_{34} &= \frac{1}{2}\left[I_3(\mathbf{x} + (1-\alpha)\mathbf{d}(\mathbf{x})) - I_4(\mathbf{x} + (2-\alpha)\mathbf{d}(\mathbf{x}))\right].
\end{aligned} \tag{13}$$

Note that at every $\mathbf{x}$, only one of the values in (13) contributes to $J(\mathbf{x})$ (12) because of the $\delta(\cdot)$ terms.

4. EXPERIMENTAL RESULTS

We solve partial differential equations derived in the appendix using explicit discretization with a small time step $dt = 1.5 \times 10^{-5}$ and 11×10^3 iterations. We employ a 4-level hierarchical implementation in order to avoid local minima, and bicubic interpolation to estimate subpixel intensities. In all experimental results shown in the paper, we use $\lambda = 2000$. Compared to the disparity estimation step (9), the final view interpolation (12) is very simple and requires little computation.

In order to gauge gains due to the use of 4 images, we have compared the proposed algorithm with view interpolation based on 2-image backward projection with isotropic as well as edge-preserving regularization of disparities (see Section 3.1). We have also compared our algorithm with equivalent forward-projection reconstruction using the same 4 images [9]. The method uses occlusion-aware edge-preserving estimation of 3 disparity fields (from (I_1, I_2), (I_2, I_3), and (I_3, I_4)), followed by occlusion detection, and spline-based image reconstruction. A listing of tested algorithms along with corresponding objective metrics (PSNR of interpolation error, i.e., difference between the ground-truth and computed intermediate images) can be found in Table 1.

In the first test, we generated two additional images for the synthetic test sequence shown in Figure 3. The four input images are shown in Figures 5(a)–5(d), and the ground-truth disparity, intermediate image, and label map are shown in Figures 5(e)–5(g). A label field L estimated using the method proposed in [42] is shown in Figure 5(h). In all label fields in this paper, black is used to denote $L(\mathbf{x}) = -1$, that is, a point is visible in, and interpolation is performed on (I_1, I_2). Similarly, gray is used to denote $L(\mathbf{x}) = 0$ and thus interpolation from (I_2, I_3), while white is used to denote $L(\mathbf{x}) = 1$ and interpolation from (I_3, I_4). Although there are false positives at the top and bottom boundaries of the square, since these areas are visible in all images, they can be predicted from any pair and do not contribute to the interpolation error.

Results for the 4-image occlusion-aware forward and backward projection are shown in the first row and the second row of Figure 6, respectively. While the disparity field from Figure 6(a) (one of 3 disparity fields estimated in forward projection) was estimated using one of the original images to guide edge-preserving regularization and implicit occlusion detection to prevent intensity mismatches, the disparity shown in Figure 6(d) was estimated using a coarse image J_c and occlusion labels from Figure 5(h). In comparison with disparity from Figure 3(i), computed from two images using edge-preserving regularization, the improvement in occlusion areas is clear in both 4-image results. Although it is difficult to judge the estimated intermediate images J, the interpolation errors in Figures 6(c) and 6(f) are clearly smaller than those in Figures 3(f) and 3(k). This is confirmed by numerical results shown in Table 1, with the 4-image occlusion-aware backward projection outperforming 4-image forward projection by over 1 dB. Interestingly, the proposed edge-preserving regularization using a coarse intermediate images offers over 2 dB improvement over isotropic regularization, both using two images.

In order to verify this performance, we have prepared another synthetic sequence with more complex occlusions (see Figure 7); two objects displace by 4 and 20 pixels, respectively, between each two views, therefore occluding both the background and each other. The original input images I_1–I_4 are shown in Figure 7 along with the ground truth: disparity, intermediate image, and label map. Also, a visibility label map estimated using the method proposed in [42] is shown in Figure 7(h). Figure 8 shows the estimated

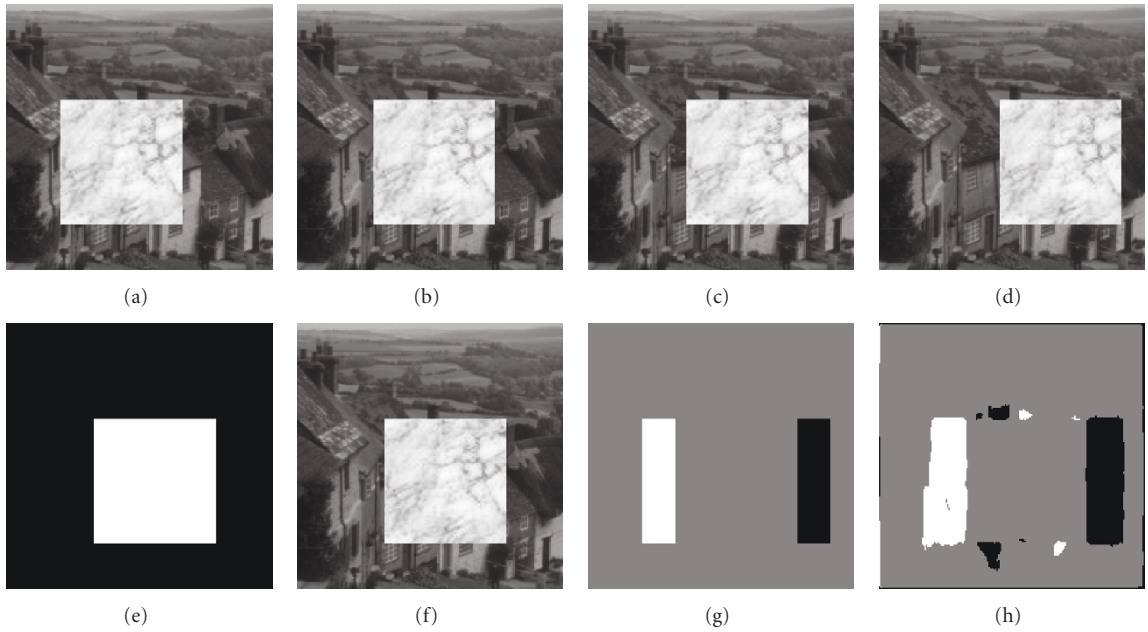


Figure 5: Extended synthetic sequence no.1: (a)–(d) I_1–I_4 ground-truth, (e) disparity, (f) intermediate image, and (g) label map, (h) estimated label map (black, gray, and white colors indicate (I_1, I_2), (I_2, I_3), and (I_3, I_4) image pairs to be used, resp.).

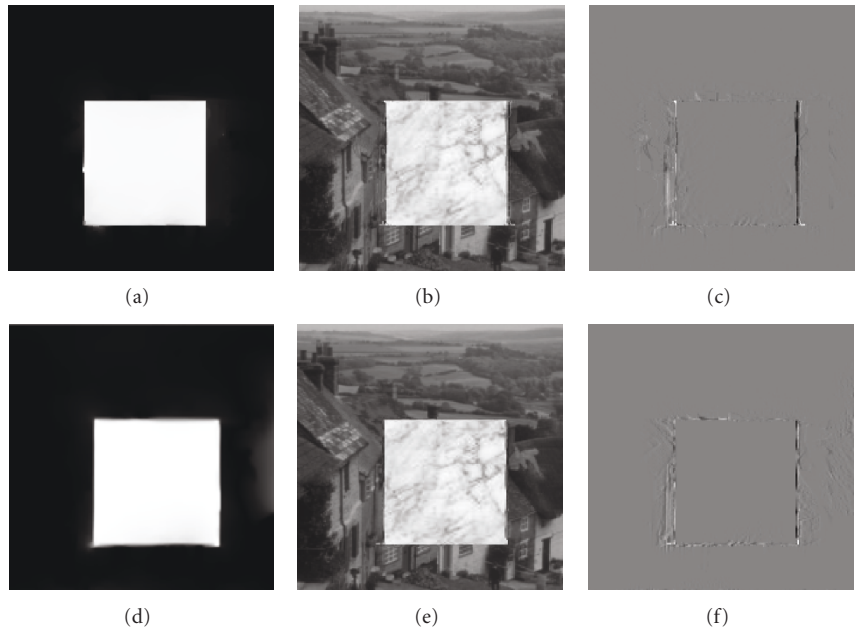


Figure 6: Comparison of view interpolation methods for synthetic sequence from Figure 5 (disparity, interpolated view, and interpolation error are shown). (a)–(c) 4-image occlusion-aware forward projection, (d)–(f) 4-image occlusion-aware backward projection. See Table 1 for algorithm description and PSNR values.

disparity, interpolated intermediate image, and interpolation error for the 4 methods described in Table 1. From error images and PSNR values, it is clear that the method proposed here outperforms 2-image backward-projection methods and also the 4-image forward-projection method.

Visually, the two 4-image methods stand out, the estimated disparity fields are most accurate, and the computed intermediate images carry little error. Numerically, however, the proposed method has a clear edge over the 4-image forward projection (1.6 dB gain). The somewhat inferior performance of the forward-projection method stems from the fact that if a single intensity is projected to an incorrect location due to erroneous disparity estimate, it will affect neighboring pixels during irregular-to-regular conversion

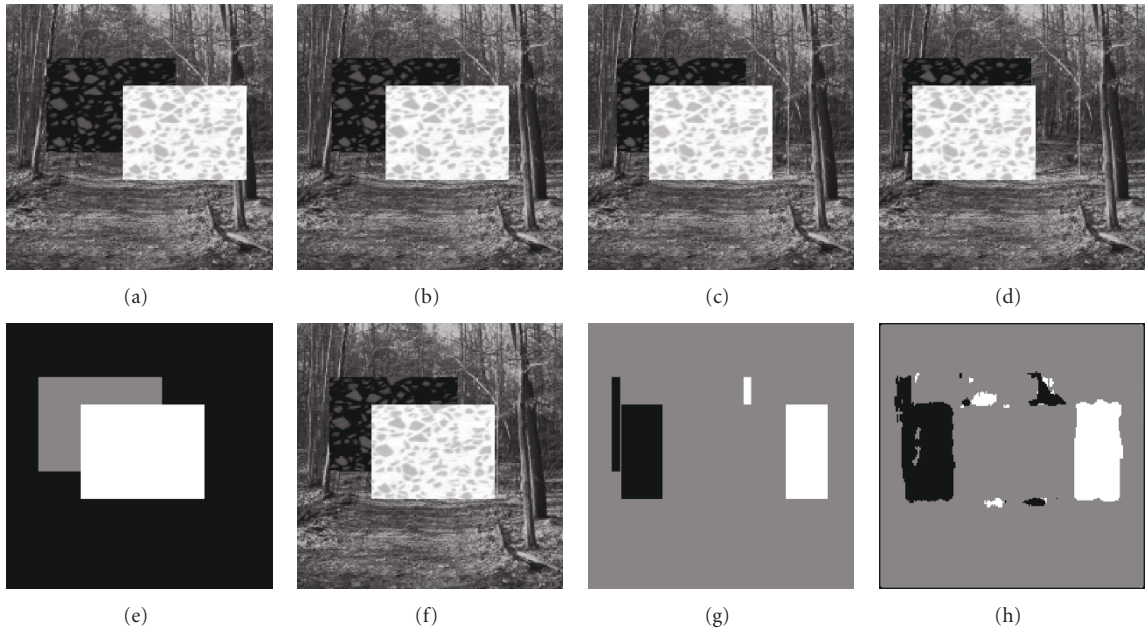


Figure 7: Synthetic sequence no. 2 with horizontal disparity. (a)–(d) I_1–I_4 ground-truth, (e) disparity, (f) intermediate image, (g) label map, and (h) estimated label map.

Table 1: Description of four-view interpolation methods tested and PSNR values [dB]of the corresponding interpolation error for synthetic test sequences from Figures 5 and 7, and natural sequence from Figure 10.

Method	Description	Figure 5	Figure 7	Figure 10
2-image isotropic BP	Backward projection (BP) using 2 images (I_2, I_3) Isotropic disparity regularization (5) Linear interpolation (4)	30.77	26.58	33.72
2-image edge-preserving BP	Backward projection using 2 images (I_2, I_3) Edge-preserving disparity regularization (6) Linear interpolation (4)	32.84	27.16	34.74
4-image occlusion-aware FP	Forward projection (FP) using 4 images (I_1, I_2, I_3, I_4) Occlusion-aware edge-preserving disparity regularization [10] Spline-based reconstruction [9]	33.05	27.41	35.89
4-image occlusion-aware BP	Backward projection using 4 images (I_1, I_2, I_3, I_4) Occlusion-aware edge-preserving disparity regularization (9) Occlusion-aware linear interpolation (12)	34.15	29.03	36.35

using splines. The reason is that spline-based reconstruction is performed globally; every pixel contributes to the reconstruction of all other pixels. This is not the case for backward projection, where neighboring interpolations are solved independently (except for disparity estimation). Although there are some artifacts around edges in the proposed approach, they are isolated as opposed to spline-based reconstruction.

This test sequence, however, has revealed one weakness of the proposed method. As it can be noticed in Figure 8(j), the disparity to the right of the objects is distorted. This is due to the weak gradient between the object and the background. Since edge preserving regularization fails in this case, the disparity of the object leaks into the background. This is a common problem in edge-preserving regularization. Nevertheless, the proposed method improves the results for this synthetic sequence by 2.5 dB in comparison with 2-image backward projection with isotropic disparities.

We also tested the proposed method on natural images. We used four frames (nos. 10, 16, 22, 28) of the *Flowergarden* sequence to reconstruct frame no. 19. The four original images are shown in Figures 9(a)–9(d). Note how the tree trunk occludes the house in the background. Disparity estimated, using backward projection with isotropic regularization based on 2 images (nos. 16 and 22), is shown in Figure 9(e). Note how smooth this disparity field is. The

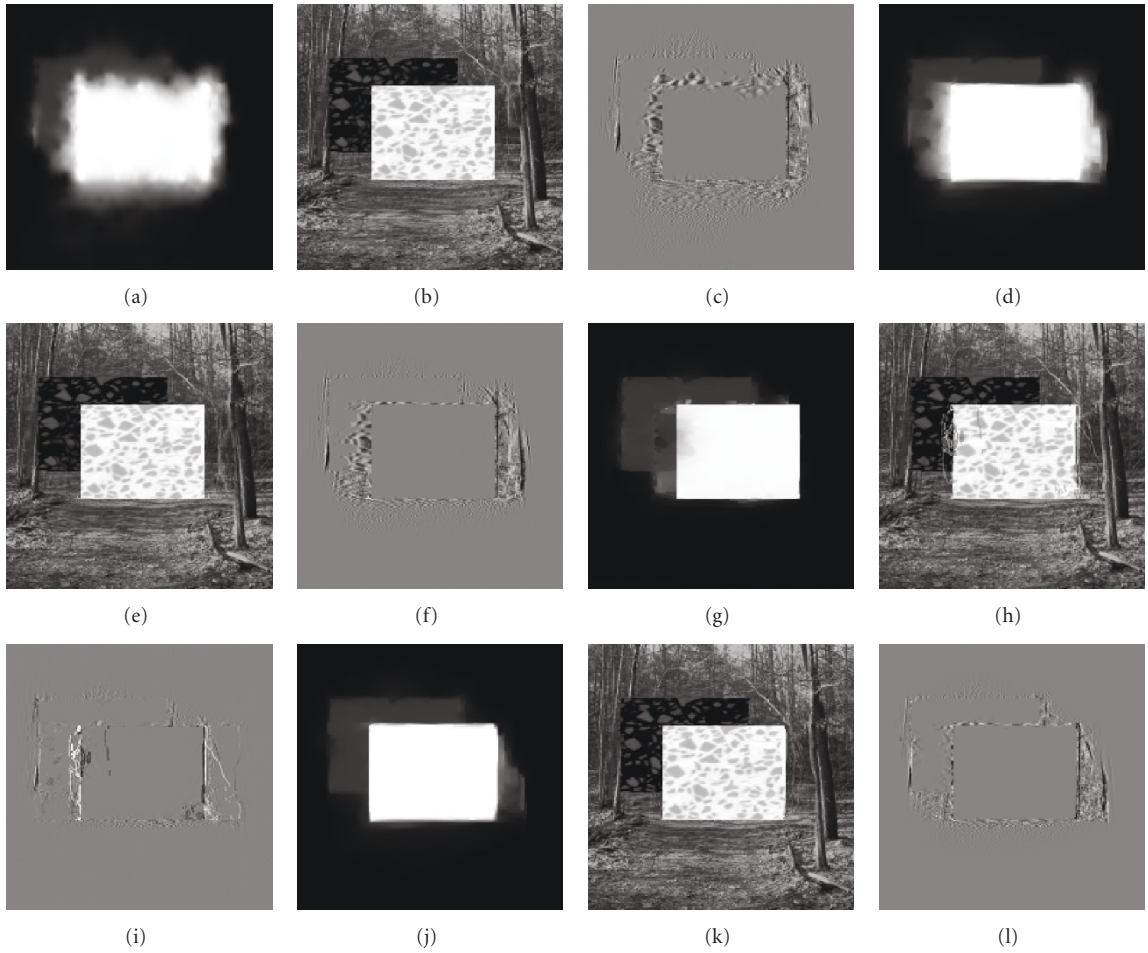


Figure 8: Comparison of view interpolation methods for synthetic sequence from Figure 7 (disparity, interpolated view, and interpolation error are shown). (a)–(c) 2-image isotropic backward projection, (d)–(f) 2-image edge-preserving backward projection, (g)–(i) 4-image occlusion-aware forward projection (only disparity estimated from I_2 to I_3 is shown), (j)–(l) 4-image occlusion-aware backward projection. See Table 1 for algorithm description and PSNR values.

interpolation of image no. 19 using this disparity field and images no. 16 and no. 22 is shown in Figure 9(f). Note that occlusion areas are poorly reconstructed; the texture around the tree trunk is highly distorted, especially on the flowerbed, house walls, and roof (see the closeup in Figure 9(k)).

A label field estimated using the method proposed in [42] is shown in Figure 9(g), while a disparity estimated using 4-image occlusion-aware edge-preserving regularization is shown in Figure 9(h). Compared to the 2-image isotropic result from Figure 9(e), the new disparity exhibits sharp tree trunk boundaries. The interpolated intermediate view is shown in Figure 9(i). Since the input sequence is actually a video sequence, we can compare the reconstructed view to the original frame no. 19. Closeups of the original frame no.19 and of both reconstructions are shown in Figures 9(j)–9(l). The texture of the flowerbed is not smeared in the new reconstruction and very similar to the original frame. Also, the windows of the house cannot be identified in Figure 9(k) as they are severely smeared. However, they are sharp and clear in Figure 9(l). Similarly, tree branches behind the house are distorted in Figure 9(k), but are more accurately reconstructed in Figure 9(l).

Finally, we tested our algorithm on an image from the Middlebury College's Vision Group (*Midd1* [51], Figure 10). Figure 11 compares the results of the proposed approach to the other three methods, while PSNR values are presented in Table 1. Compared to the isotropic case, the 2-image edge-preserving regularization sharpens the disparity field that, in turn, leads to 1 dB gain in PSNR. However, occlusions are still not handled well; the closeup of occlusion area shows severe artifacts. Since forward-projection with spline-based interpolation accounts for occlusions, we see an increase in PSNR value as well as proper reconstruction of texture in occlusion areas. The proposed method adds another 0.5 dB to the PSNR and produces intermediate image very close to the original closeup.

5. DISCUSSION

The focus of this work was on severely undersampled 3D data sets with unknown scene structure, and, more specifically, on

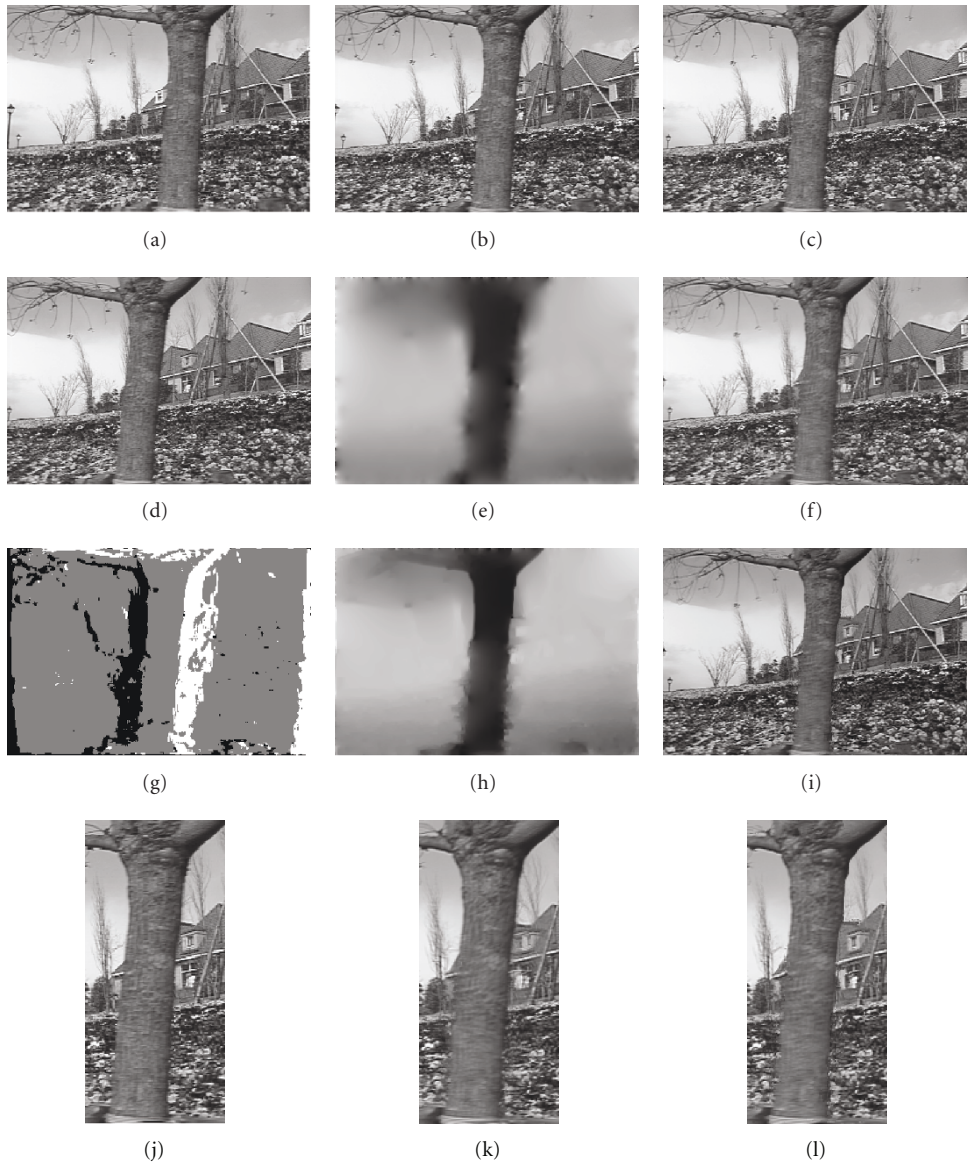
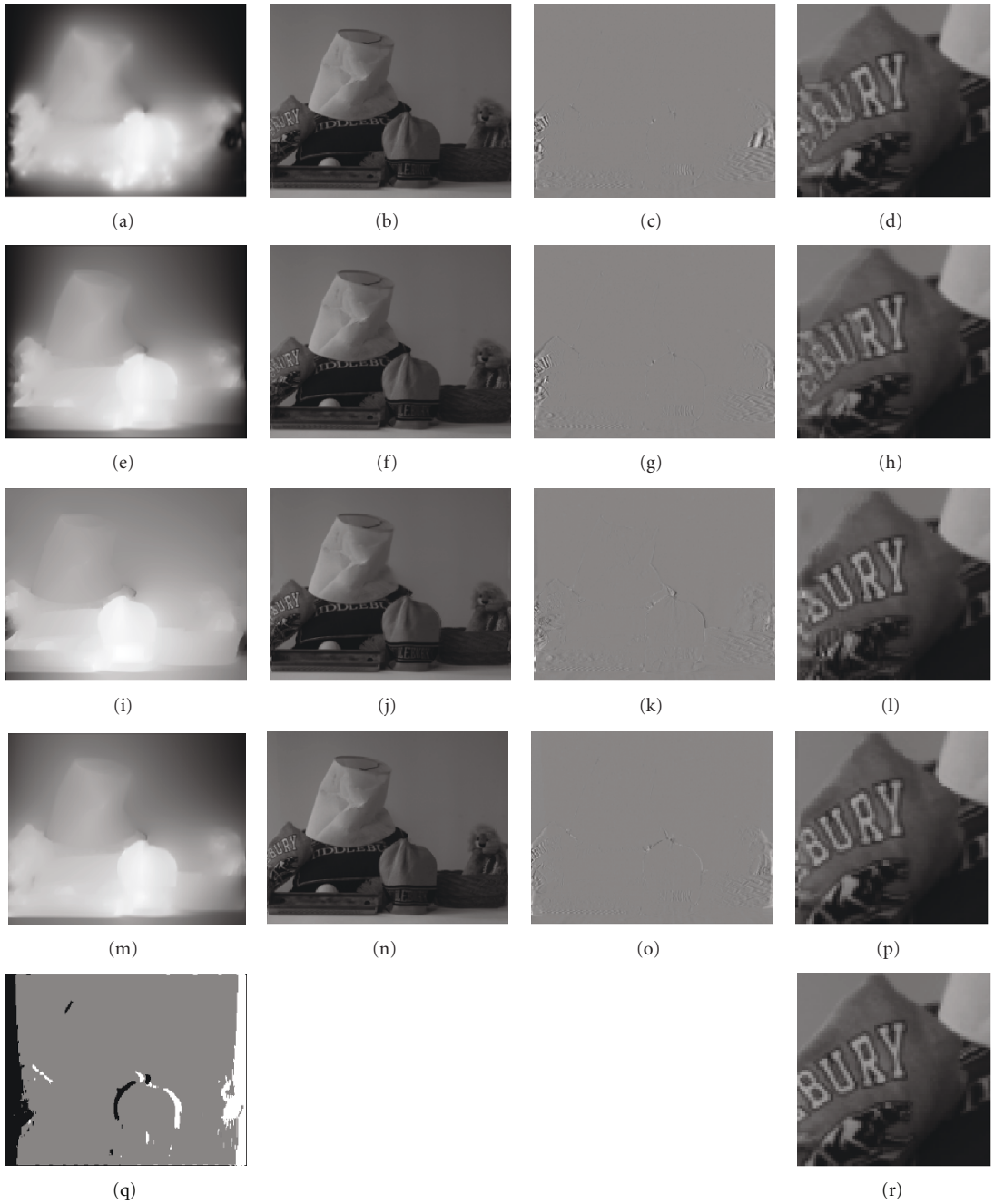


FIGURE 9: Comparison of backward-projection view interpolation for *Flowergarden*: original frames. (a) no. 10 (I_1), (b) no. 16 (I_2), (c) no. 22 (I_3), (d) no. 28 (I_4), (e)-(f) disparity and intermediate view for 2-image isotropic backward projection, (g) estimated label map, (h)-(i) disparity and intermediate view for 4-image occlusion-aware backward projection, (j) true frame no. 19, (k) closeup from (f), (l) closeup from (i).

handling of occlusion areas in novel view interpolation. As expected, view interpolation based on 4 images outperforms view interpolation using only 2 images, since occluded areas can be found in the additional images. Interestingly, experiments have shown that view interpolation from 4 images using occlusion-aware backward projection outperforms one using occlusion-aware forward projection. This result is somewhat surprising since the forward-projection approach uses known images for edge-preserving disparity diffusion, whereas the backward-projection approach uses a coarse intermediate image (estimated) for the same purpose. As seen in Figures 8(g) and 8(j), the disparity field estimated within forward projection has sharper discontinuities at object boundaries than the one estimated within backward projection. One possible explanation of this inconsistency is that disparity-compensated projections in the forward-projection case ($(1+\alpha)\mathbf{d}_{12}$ and $(2-\alpha)\mathbf{d}_{43}$) extend beyond their temporal support (magnifications $(1+\alpha)$ and $(2-\alpha)$), and thus any disparity errors get amplified in the coordinate system of intermediate image J. This is not the case in backward projection since disparities area anchored in J, and suitable image pairs are used for correspondence. Another likely factor is the spline-based reconstruction used in forward projection to recover on-lattice samples of J; it incorporates intensity smoothing in order to deal with irregularly-sampled projections. This is not the case in backward projection, where a simple averaging of two corresponding intensities is used.

(a) (b) (c) (d) (e)

Figure 10: *Midd1* test sequence [51]. (a)–(d) I_1–I_4, and (e) true intermediate image.



(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

(m) (n) (o) (p)

(q) (r)

Figure 11: Comparison of view interpolation methods for *Middl* sequence from Figure 10 (disparity, interpolated view, interpolation error, and closeup of occlusion area are shown). (a)–(d) 2-image isotropic backward projection, (e)–(h) 2-image edge-preserving backward projection, (i)–(l) 4-image occlusion-aware forward projection (only disparity estimated from I_2 to I_3 is shown), (m)-(p) 4-image occlusion-aware backward projection, (q) estimated label map, (r) closeup of the occlusion area from the true intermediate image (Figure 10(e)). See Table 1 for algorithm description and PSNR values.

The very simple occlusion-adaptive interpolation used in the backward-projection approach has the additional benefit of resilience to disparity errors. A single erroneous disparity vector affects luminance/color of a single pixel in the novel view J. This is unlike the case of forward projection where the spline-based reconstruction spreads this error due to the smoothness constraint used.

In terms of the computational complexity, it is clear from (4) that in backward projection the actual view interpolation is a byproduct of disparity estimation, and is a simple low-complexity operation. The main computational complexity of backward projection rests with the estimation of initial disparity fields $\mathbf{d}_{12}$ and $\mathbf{d}_{43}$ (to recover occlusions) as well as disparity field $\mathbf{d}$ (9); occlusion detection itself [42] has low computational complexity. Note, however, that a separate disparity field $\mathbf{d}$ needs to be computed for each novel image J with different α. On the other hand, in occlusion-aware forward projection in addition to the estimation of 3 disparity fields ($\mathbf{d}_{12}$, $\mathbf{d}_{23}$, $\mathbf{d}_{43}$), the main computational burden is in spline-based image reconstruction that is iterative and computationally complex. Herein lies a compromise, if only one novel view is needed between I_2 and I_3, backward projection should be more efficient computationally, however, if many novel views are needed between I_2 and I_3 (often the case in multiview autostereoscopic displays), forward projection may be more efficient.

6. SUMMARY AND CONCLUSIONS

In this paper, we overviewed different approaches to intermediate view reconstruction especially in the context of their occlusion awareness. We pointed out the fundamental difference between forward-projection and backward-projection approaches to view interpolation. We highlighted the limitations of backward-projection approaches, and specifically the absence of an underlying image for edge-preserving disparity regularization, and difficulties with occlusion handling. Then, we argued that although backward-projection reconstruction using two images creates distorted texture in the intermediate view, it reconstructs edges accurately. Exploiting this fact, we proposed to use a coarse intermediate image in disparity estimation for edge-preserving regularization purposes. We also proposed a new variational backward-projection view interpolation that works selectively on image pairs to handle occlusions. The basic idea is that when multiple images are available, a point in the intermediate image is visible in at least two images that can be used for accurate interpolation. Novel views computed using the proposed method show dramatic improvements over backward-projection interpolation based on two images, and a significant gain over 4-image forward-projection approach. Admittedly, the improvements are localized and affect image quality only in the immediate vicinity of object boundaries. However, in high-quality 3D applications, such as digital 3D cinema and ultra-high resolution multiview autostereoscopic displays, any distortions at depth discontinuities are highly objectionable, especially if they vary with viewpoint change.

APPENDIX

We need to carry out minimization (9) with respect to $\mathbf{d}$. Denote $e(\mathbf{x}) = e_P(\mathbf{x}) + \lambda e_S(\mathbf{x})$, where e_P and e_S are defined in (10). Using the calculus of variations, Euler-Lagrange equations for u and v can be found as follows:

$$\begin{aligned} e'(u) &= \frac{\partial e}{\partial u} - \frac{\partial}{\partial x}\frac{\partial e}{\partial u^x} - \frac{\partial}{\partial y}\frac{\partial e}{\partial u^y} = 0, \\ e'(v) &= \frac{\partial e}{\partial v} - \frac{\partial}{\partial x}\frac{\partial e}{\partial v^x} - \frac{\partial}{\partial y}\frac{\partial e}{\partial v^y} = 0, \end{aligned} \tag{A.1}$$

where u^x, v^x and u^y, v^y are the horizontal and vertical derivatives of horizontal and vertical components of disparity. Expanding these equations, we obtain

$$\begin{aligned} \frac{e_P}{\partial u} - \frac{\partial}{\partial x}\frac{\partial e_S}{\partial u^x} - \frac{\partial}{\partial y}\frac{\partial e_S}{\partial u^y} = 0, \\ \frac{e_P}{\partial v} - \frac{\partial}{\partial x}\frac{\partial e_S}{\partial v^x} - \frac{\partial}{\partial y}\frac{\partial e_S}{\partial v^y} = 0. \end{aligned} \tag{A.2}$$

Partial derivatives are defined as follows:

$$\begin{aligned} \frac{e_P}{\partial u} &= \frac{\partial P_{12}}{\partial u} + \frac{\partial P_{23}}{\partial u} + \frac{\partial P_{34}}{\partial u}, \\ \frac{e_P}{\partial v} &= \frac{\partial P_{12}}{\partial v} + \frac{\partial P_{23}}{\partial v} + \frac{\partial P_{34}}{\partial v}, \\ \frac{\partial}{\partial x}\frac{\partial e_S}{\partial u^x} + \frac{\partial}{\partial y}\frac{\partial e_S}{\partial u^y} &= \frac{\partial(2g(|J_c^x|)u^x)}{\partial x} + \frac{\partial(2g(|J_c^y|)u^y)}{\partial y}, \\ \frac{\partial}{\partial x}\frac{\partial e_S}{\partial v^x} + \frac{\partial}{\partial y}\frac{\partial e_S}{\partial v^y} &= \frac{\partial(2g(|J_c^x|)v^x)}{\partial x} + \frac{\partial(2g(|J_c^y|)v^y)}{\partial y}, \end{aligned} \tag{A.3}$$

with

$$\begin{aligned} \frac{\partial P_{12}}{\partial u} &= 2\delta(L(\mathbf{x})+1)\theta_{12}(\mathbf{x})\frac{\partial \theta_{12}(\mathbf{x})}{\partial u}, \\ \frac{\partial P_{12}}{\partial v} &= 2\delta(L(\mathbf{x})+1)\theta_{12}(\mathbf{x})\frac{\partial \theta_{12}(\mathbf{x})}{\partial v}, \\ \frac{\partial P_{23}}{\partial u} &= 2\delta(L(\mathbf{x}))\theta_{23}(\mathbf{x})\frac{\partial \theta_{23}(\mathbf{x})}{\partial u}, \\ \frac{\partial P_{23}}{\partial v} &= 2\delta(L(\mathbf{x}))\theta_{23}(\mathbf{x})\frac{\partial \theta_{23}(\mathbf{x})}{\partial v}, \\ \frac{\partial P_{34}}{\partial u} &= 2\delta(L(\mathbf{x})-1)\theta_{34}(\mathbf{x})\frac{\partial \theta_{34}(\mathbf{x})}{\partial u}, \\ \frac{\partial P_{34}}{\partial v} &= 2\delta(L(\mathbf{x})-1)\theta_{34}(\mathbf{x})\frac{\partial \theta_{34}(\mathbf{x})}{\partial v}, \end{aligned} \tag{A.4}$$

where

$$\begin{aligned}
\frac{\partial\theta_{12}(\mathbf{x})}{\partial u} &= -(1+\alpha)\widetilde{I}_1^x + \alpha\widetilde{I}_2^x,\\
\frac{\partial\theta_{23}(\mathbf{x})}{\partial u} &= -\alpha\widetilde{I}_2^x - (1-\alpha)\widetilde{I}_3^x,\\
\frac{\partial\theta_{34}(\mathbf{x})}{\partial u} &= (1-\alpha)\widetilde{I}_3^x - (2-\alpha)\widetilde{I}_4^x,\\
\frac{\partial\theta_{12}(\mathbf{x})}{\partial v} &= -(1+\alpha)\widetilde{I}_1^y + \alpha\widetilde{I}_2^y,\\
\frac{\partial\theta_{23}(\mathbf{x})}{\partial v} &= -\alpha\widetilde{I}_2^y - (1-\alpha)\widetilde{I}_3^y,\\
\frac{\partial\theta_{34}(\mathbf{x})}{\partial v} &= (1-\alpha)\widetilde{I}_3^y - (2-\alpha)\widetilde{I}_4^y,
\end{aligned} \tag{A.5}$$

and $I_\cdot^x$, $I_\cdot^y$ are horizontal and vertical derivatives of $I_\cdot$, while $\widetilde{I}_\cdot^x$ and $\widetilde{I}_\cdot^y$ are derivatives evaluated at a point off $\mathbf{x}$, for example, $\widetilde{I}_2^x = I_2^x(\mathbf{x} - \alpha\mathbf{d}(\mathbf{x}))$. Using an auxiliary time variable t, equations in (A.1) can be solved by discretizing the gradient descent equations:

$$\frac{\partial u}{\partial t} = -e'(u), \qquad \frac{\partial v}{\partial t} = -e'(v). \tag{A.6}$$

ACKNOWLEDGMENT

This work was supported by the NSF under Grants ECS-0219224 and CNS-0721884, and by the NIH under Grant 1R21HD050655-01.

REFERENCES

[1] W. Matusik and H. Pfister, "3D TV: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 814–824, 2004.

[2] N. A. Dodgson, "Autostereoscopic 3D displays," *Computer*, vol. 38, no. 8, pp. 31–36, 2005.

[3] J. Konrad and M. Halle, "3-D displays and signal processing," *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 97–111, 2007.

[4] T. Kanade, P. Rander, and P. J. Narayanan, "Virtualized reality: constructing virtual worlds from real scenes," *IEEE Multimedia*, vol. 4, no. 1, pp. 34–47, 1997.

[5] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 600–608, 2004.

[6] Special Issue on 3-D Video Technology, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 2–4, March–June 2000.

[7] Special Issue on Multiview Video Coding and 3DTV, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, November 2007.

[8] Special issue on Multiview Imaging and 3DTV, *IEEE Signal Processing Magazine*, vol. 24, November 2007.

[9] S. Ince, J. Konrad, and C. Vázquez, "Spline-based intermediate view reconstruction," in *Stereoscopic Displays and Virtual Reality Systems XIV*, vol. 6490 of *Proceedings of SPIE*, San Jose, Calif, USA, March 2007.

[10] S. Ince and J. Konrad, "Occlusion-aware optical flow estimation," *IEEE Transactions on Image Processing*, vol. 17, no. 8, pp. 1443–1451, 2008.

[11] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*, M. Landy and J. A. Movshon, Eds., MIT Press, Cambridge, Mass, USA, 1991.

[12] C. Zhang and T. Chen, "A survey on image-based rendering—representation, sampling and compression," *Signal Processing: Image Communication*, vol. 19, no. 1, pp. 1–28, 2004.

[13] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '96)*, pp. 31–42, New Orleans, La, USA, August 1996.

[14] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '96)*, pp. 43–54, New Orleans, La, USA, August 1996.

[15] L. McMillan, *An image-based approach to three-dimensional computer graphics*, Ph.D. thesis, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 1997.

[16] P. E. Debevec, G. Borshukov, and Y. Yu, "Efficient view-dependent image-based rendering with projective texture-mapping," in *Proceedings of the 9th Eurographics Workshop on Rendering (EUROGRAPHICS '98)*, pp. 105–116, Vienna, Austria, June-July 1998.

[17] W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan, "Image-based visual hulls," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '00)*, pp. 369–374, New Orleans, La, USA, July 2000.

[18] C. Buehler, M. Bosse, L. McMillan, S. Gortler, and M. Cohen, "Unstructured lumigraph rendering," in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '01)*, pp. 425–432, Los Angeles, Calif, USA, August 2001.

[19] S. M. Seitz and C. R. Dyer, "View morphing," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '96)*, pp. 21–30, New Orleans, La, USA, August 1996.

[20] S. Avidan and A. Shashua, "Novel view synthesis in tensor space," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '97)*, pp. 1034–1040, San Juan, Puerto Rico, USA, June 1997.

[21] D. Scharstein, "Stereo vision for view synthesis," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '96)*, pp. 852–858, San Francisco, Calif, USA, June 1996.

[22] S. E. Chen and L. Williams, "View interpolation for image synthesis," in *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '93)*, pp. 279–288, Anaheim, Calif, USA, August 1993.

[23] E. Izquierdo and J.-R. Ohm, "Image-based rendering and 3D modeling: a complete framework," *Signal Processing: Image Communication*, vol. 15, no. 10, pp. 817–858, 2000.

[24] A. Mancini and J. Konrad, "Robust quadtree-based disparity estimation for the reconstruction of intermediate stereoscopic images," in *Stereoscopic Displays and Virtual Reality Systems V*, pp. 53–64, San Jose, Calif, USA, January 1998.

[25] J. S. McVeigh, M. W. Siegel, and A. G. Jordan, "Intermediate view synthesis considering occluded and ambiguously referenced image regions," *Signal Processing: Image Communication*, vol. 9, no. 1, pp. 21–28, 1996.

[26] J.-l. Park and S. Inoue, "Arbitrary view generation from multiple cameras," in *Proceedings of IEEE International Conference on Image Processing (ICIP '97)*, vol. 1, pp. 149–152, Santa Barbara, Calif, USA, October 1997.

[27] A. M. K. Siu and R. W. H. Lau, "Image registration for image-based rendering," *IEEE Transactions on Image Processing*, vol. 14, no. 2, pp. 241–252, 2005.

[28] A. Redert, E. Hendriks, and J. Biemond, "Synthesis of multi viewpoint images at non-intermediate positions," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '97)*, vol. 4, pp. 2749–2752, Munich, Germany, April 1997.

[29] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1–3, pp. 185–203, 1981.

[30] J. Konrad, "Enhancement of viewer comfort in stereoscopic viewing: parallax adjustment," in *Stereoscopic Displays and Virtual Reality Systems VI*, vol. 3639 of *Proceedings of SPIE*, pp. 179–190, San Jose, Calif, USA, January 1999.

[31] J. Zhai, K. Yu, J. Li, and S. Li, "A low complexity motion compensated frame interpolation method," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS '05)*, vol. 5, pp. 4927–4930, Kobe, Japan, May 2005.

[32] R. Franich, *Disparity estimation in stereoscopic digital images*, Ph.D. thesis, Delft University of Technology, Delft, The Netherlands, 1996.

[33] G. Vogiatzis, C. H. Esteban, P. H. S. Torr, and R. Cipolla, "Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2241–2246, 2007.

[34] H. Kim and K. Sohn, "3D reconstruction from stereo images for interactions between real and virtual objects," *Signal Processing: Image Communication*, vol. 20, no. 1, pp. 61–75, 2005.

[35] L. Alvarez, R. Deriche, J. Sánchez, and J. Weickert, "Dense disparity map estimation respecting image discontinuities: a PDE and scale-space based approach," *Journal of Visual Communication and Image Representation*, vol. 13, no. 1-2, pp. 3–21, 2002.

[36] C. Strecha and L. Van Gool, "PDE-based multi-view depth estimation," in *Proceedings of the 1st International Symposium on 3D Data Processing Visualization and Transmission (3DPVT '02)*, vol. 2, pp. 416–425, Padova, Italy, June 2002.

[37] M. J. Black and P. Anandan, "The robust estimation of multiple motions: parametric and piecewise-smooth flow fields," *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75–104, 1996.

[38] D. Geiger, B. Ladendorf, and A. Yuille, "Occlusions and binocular stereo," *International Journal of Computer Vision*, vol. 14, no. 3, pp. 211–226, 1995.

[39] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, vol. 194, no. 4262, pp. 283–287, 1976.

[40] M. Proesmans, L. J. Van Gool, E. J. Pauwels, and A. Oosterlinck, "Determination of optical flow and its discontinuities using non-linear diffusion," in *Proceedings of the 3rd European Conference on Computer Vision (ECCV '94)*, vol. 2, pp. 295–304, Stockholm, Sweden, May 1994.

[41] S. B. Pollard, J. E. Mayhew, and J. P. Frisby, "PMF: a stereo correspondence algorithm using a disparity gradient limit," *Perception*, vol. 14, no. 4, pp. 449–470, 1985.

[42] S. Ince and J. Konrad, "Geometry-based estimation of occlusions from video frame pairs," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '05)*, vol. 2, pp. 933–936, Philadelphia, Pa, USA, March 2005.

[43] J. Sun, Y. Li, S. B. Kang, and H.-Y. Shum, "Symmetric stereo matching for occlusion handling," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 2, pp. 399–406, San Diego, Calif, USA, June 2005.

[44] R. Depommier and E. Dubois, "Motion estimation with detection of occlusion areas," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '92)*, vol. 3, pp. 269–272, San Francisco, Calif, USA, March 1992.

[45] S.-L. Iu, "Robust estimation of motion vector fields with discontinuity and occlusion using local outliers rejection," *Journal of Visual Communication and Image Representation*, vol. 6, no. 2, pp. 132–141, 1995.

[46] K. P. Lim, A. Das, and M. N. Chong, "Estimation of occlusion and dense motion fields in a bidirectional Bayesian framework," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 712–718, 2002.

[47] V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts," in *Proceedings of the 7th European Conference on Computer Vision (ECCV '02)*, vol. 2352 of *Lecture Notes In Computer Science*, pp. 82–96, Copenhagen, Denmark, May 2002.

[48] H. H. Nagel and W. Enkelmann, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 5, pp. 565–593, 1986.

[49] A.-R. Mansouri, A. Mitiche, and J. Konrad, "Selective image diffusion: application to disparity estimation," in *Proceedings of IEEE International Conference on Image Processing (ICIP '98)*, vol. 3, pp. 284–288, Chicago, Ill, USA, October 1998.

[50] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[51] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, Minneapolis, Minn, USA, June 2007.