

View Synthesis from Wide-Baseline Views Using Occlusion Aware Estimation of Large Disparities

Ahmed S. Elliethy^a, Hussein A. Aly^a and Gaurav Sharma^b

^aComputer Dept., Military Technical College, Cairo, Egypt;

^bECE Dept., University of Rochester, Rochester, NY, USA

ABSTRACT

Accurate disparity estimation is a key ingredient required when generating a high fidelity novel view from a set of input views. In this paper, a high quality disparity estimation method is proposed for view synthesis from multiple input images with large disparities and occlusions. The method optimally selects one out of three image pairs to estimate the disparity map for different regions of the novel view. The novel view is then formed using this disparity map. We introduce two novel elements: a) an enhanced visibility map that is able to segment the scene accurately near object boundaries and b) a backward unilateral and bilateral disparity estimation procedure using the Gabor transform on an expandable search window to tackle large disparities. The quality of the interpolated virtual views produced by the proposed method is assessed and compared against two of the prominent previously-reported methods. The proposed method offers a significant improvement both in terms of visual quality of the interpolated views as well as the peak signal-to-noise ratio (PSNR) and structured similarity image index (SSIM) metrics.

Keywords: Large disparity estimation, backward projections Gabor transform, variational regularization, view interpolation

1. INTRODUCTION

In multi-view and 3D imaging, the generation of a virtual view of the scene from any viewpoint that is located at an intermediate position between the actual camera viewpoints is important in several applications. This view-interpolation step, for instance, allows viewer navigation within the scene by enabling continuous generation of a virtual view of the scene based on the position and orientation of the viewer, functionality that is required for multi-view 3D displays^{1,2} and free viewpoint video.^{3,4} View-interpolation is also a key component in a number of multi-view compression schemes.⁵⁻⁸

View interpolation can be performed for multi-depth or single-depth scenes. Single-depth scenes are encountered either in the distant imaging scenario, or when the scene imaged is inherently flat. In these situations, the relation among different views of the scene can be represented by a global geometric transformation between the input images, and there are no occluded or new exposed areas in the synthesized view (since all pixels are visible in all input images).⁹⁻¹³ On the other hand, view interpolation for multi-depth scenes is much more arduous, because no single global transformation can relate the pixels of the different input views and there are some areas that are occluded in one or more views. For instance, in translationally-displaced views of a multi-depth scene, as is common in stereoscopic and 3DTV applications, between the different views, nearer objects (to the camera) exhibit larger displacements than further objects. In general, the quality of the synthesized view generated by view interpolation techniques depends heavily upon: a) the accuracy of estimated disparity map that relates the input views' pixels, and b) reasonable assignment of pixel values to the occluded parts in the synthesized view by the view interpolation algorithm.

In this paper, we consider the disparity estimation problem for general multi-depth scenes under the setting typical for stereoscopic and 3DTV applications where the captured views and the desired views to be synthesized all correspond to viewpoints located along a line as shown in Fig. 1. The dense motion field (*disparity field*)

Further author information: (Send correspondence to Ahmed S. Elliethy)
A.S.E.: E-mail: ahmed.s.elliethy@rochester.edu, Telephone: 1 585-275-8122

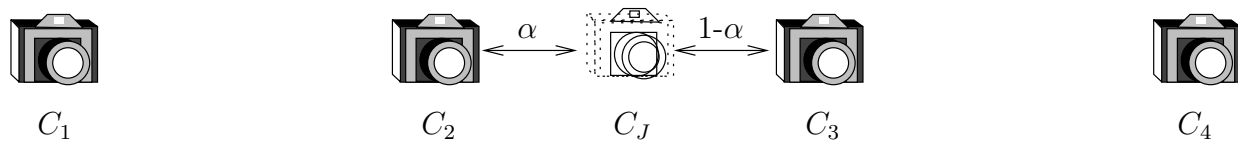


Figure 1. Problem definition : Given 4 captured views of 3D scene by 4 cameras C_1, C_2, C_3 and C_4 and required to get a novel view as if it is captured by a virtual camera C_J .

for this scenario is denoted by $\mathbf{d}(\mathbf{x}) = [u(\mathbf{x}) \ v(\mathbf{x})]^T$, where u and v are the displacements in the horizontal and vertical directions respectively. Unlike applications such as video frame rate up-conversion, where small displacements are encountered between frames, here we deal with a set of sparse views. The disparity estimation process therefore has to tackle much larger displacements¹⁴ and faces new challenges such as large area occlusions and significant changes between the input views.

A method for estimating large disparities from two views is presented in Ref. 14 for stereoscopic rendering. The method uses the Gabor transform along with regularization and can handle large disparities but cannot effectively deal with occlusions. Methods capable of handling occlusions using 4-views have been presented in Ref. 15,16. These methods are based on minimization of a regularized objective function and guided by the so-called visibility map or labels. However, these methods are not suitable for estimating large disparities even when implemented using multi-resolution pyramid. This is due to the fact that at the very coarse scale the objects becomes blurry and the regularized estimation tends to favor the coherence of the disparity field over data fidelity. Moreover, these techniques do not accurately estimate the visibility map and some areas in the new virtual view are interpolated from a non-optimal view as we clarify in Section 3.1.

In this paper, we present a new method for accurately estimating a dense disparity field in the presence of large displacements between the input views that also effectively handles occlusions. We use this disparity field to synthesize virtual views and demonstrate the advantages of the proposed method over prior techniques in this application. The method has two ingredients: an enhanced visibility map and a backward disparity field for large displacements.

The rest of the paper is organized as follows. In Section 2 we describe the problem formulation and introduce our notation. Our proposed method is given in Section 3, where we present our main contributions in two subsections: an enhanced visibility map algorithm and a method for backward large disparity estimation. We present experimental results and comparisons in Section 4. Finally, we give concluding remarks in Section 5.

2. PROBLEM DEFINITION

The problem setting is shown in Fig. 1. Four acquired images are shown two to the left and two to the right of the desired view, with the inter-view spacing normalized to unity for the two captured views adjacent to the

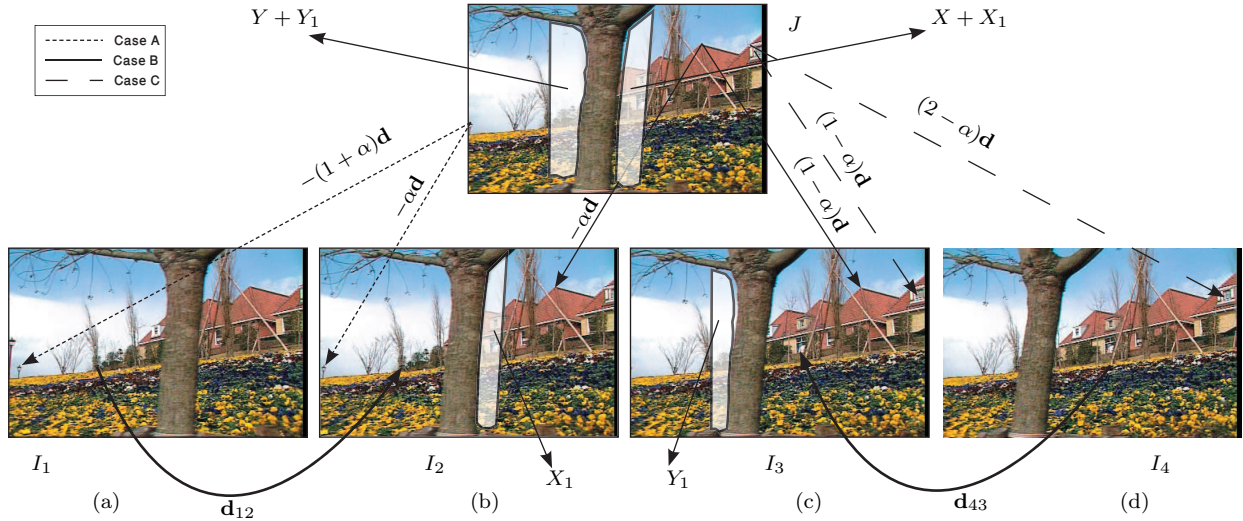


Figure 2. Problem formulation depicted using frames from the flower-garden test sequence using frame number: (a) 1 for I_1 , (b) 13 for I_2 , (c) 25 for I_3 , and (d) 37 for I_4 . Using these four views, the view J intermediate between I_2 and I_3 , and located at a normalized distance of α from I_2 , is to be synthesized. \mathbf{d}_{12} and \mathbf{d}_{43} denote the forward disparities from I_1 to I_2 and from I_4 to I_3 , respectively and \mathbf{d} denotes the backward disparity from the coordinate system of J , at each pixel, pointing to frames from which the pixel is interpolated, categorized in turn by three cases A, B, and C. Specifically, cases A, B, and C correspond to situations where interpolated values for pixels are obtained from (I_1, I_2) , (I_2, I_3) , and (I_4, I_3) , respectively. For cases A and C the backward disparities are extrapolated, whereas for case B these are interpolated. The proposed approach reduces the regions of extrapolation from $X + X_1$ to X and from $Y + Y_1$ to Y , improving the resulting view interpolation.

desired view. We need to estimate the disparity fields at the virtual view \hat{J} that will be used to interpolate a virtual view \hat{J} from the four ordered input images I_1, I_2, I_3 , and I_4 all having pixels with spacing determined by a lattice Γ .¹⁷ In this paper, all images have spatially limited support as opposed to the mathematically infinite extent of a lattice. For brevity, however, we will implicitly assume this throughout our description without mentioning this point repeatedly. As a result of the displacement of the camera between the captured views, there are regions of the of the 3-D scene that are visible in some views and occluded by objects closer to the camera in other views. Therefore some pixels in \hat{J} are “visible” in I_2 but “invisible” in I_3 and vice versa. Sample occlusions are shown in Fig. 2 and labeled as cases A and C, which are explained later. To handle occlusions, we use information from the remaining views I_1 and I_4 as in Ref. 15. In general, we have three options as shown in Fig. 2: case A: left-sided extrapolation from (I_1, I_2) , case B: bi-sided interpolation from (I_2, I_3) , and case C: right-sided extrapolation from (I_4, I_3) . Cases A and C are only used in occlusion situations. In this paper, we use multiple disparity fields and denote these by:

1. \mathbf{d}_{nm}^M for the forward disparity field defined on image n and pointing toward image m and estimated using method $M = \{G, A\}$, where G stands for Gabor, and A for the anisotropic regularized estimation method as in Ref. 15. In forward projection, the disparity field is defined on the sampling lattice of the known view and points towards the interpolated virtual view,¹⁸ where the end points of the disparity vectors need not necessarily correspond to the sampling lattice of the interpolated view \hat{J} .
2. \mathbf{d}^M for the backward disparity field defined onto the interpolated image \hat{J} . The backward disparity field is defined on the sampling lattice of the interpolated image \hat{J} and points either left to the pair (I_2, I_1) , bi-directionally to (I_2, I_3) , or right to (I_3, I_4) . Each pixel location in the interpolated virtual view \hat{J} has a disparity vector. We estimate this backward disparity field using the method $M = \{G, I, A\}$, where G stands for the backward Gabor transform, I for isotropic regularization, and A for anisotropic

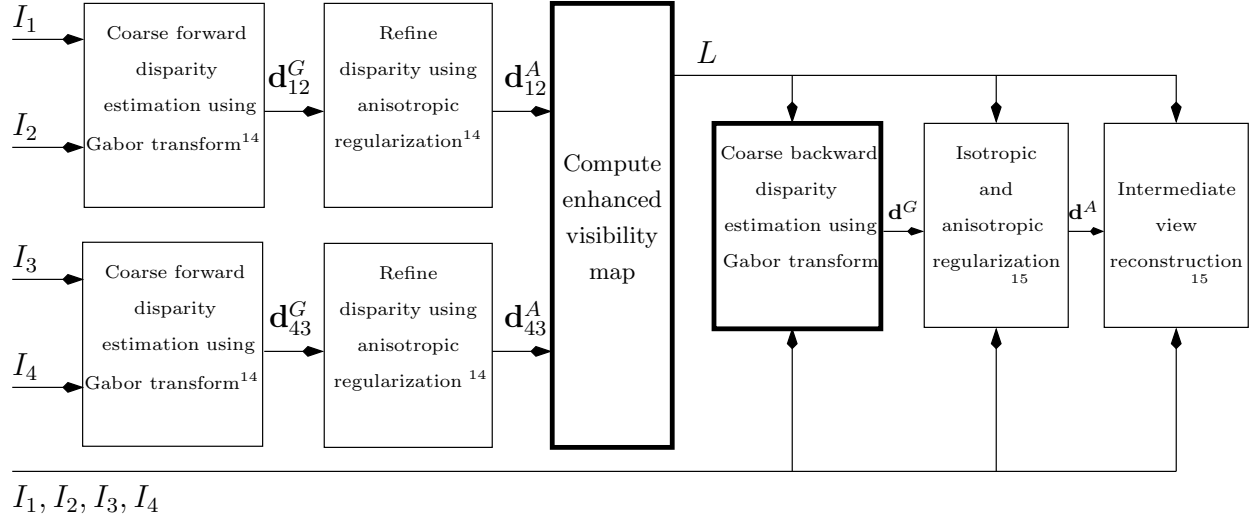


Figure 3. Block diagram of interpolation aware view interpolation using the proposed method for estimation of large disparities in a wide-baseline setting.

regularization. The backward disparity field is associated with a visibility map to determine, for each pixel, a corresponding image pair out of the 3 aforementioned cases.

3. PROPOSED METHOD

Our proposed method uses Gabor transform based estimation of large disparities followed by both isotropic and anisotropic regularization to obtain a coherent disparity field that in turn allows accurate estimation of the visibility map for identification of occlusions. The method has two main stages as shown in Fig. 3.

- Stage 1: occlusion estimation at the interpolated view \hat{J} ; this is done by estimating disparities \mathbf{d}_{12}^G and \mathbf{d}_{43}^G , then refining these using anisotropic regularization¹⁵ to obtain \mathbf{d}_{12}^A and \mathbf{d}_{43}^A . Both \mathbf{d}_{12}^A and \mathbf{d}_{43}^A are used to estimate a visibility map L which is used to guide stages 2 and 3. L is a dictionary that defines for each pixel in \hat{J} a corresponding image pair out of the three cases, as explained in Section 3.1.
- Stage 2: estimate the disparity field \mathbf{d}^A at the virtual view \hat{J} ; first we estimate \mathbf{d}^G as explained in Section 3.2, then refine it to get \mathbf{d}^I and use it to obtain a coarse estimate of the interpolated view J_C . This coarse estimate is used to guide the anisotropic regularization to obtain the final disparity field \mathbf{d}^A .

The minimization of the isotropic and anisotropic energy functionals for disparity estimation¹⁵ is obtained by solving their associated Euler-Lagrange equations. We use the gradient descent method to solve them as in Ref. 15. The forward disparity estimation using Gabor transform is obtained by an exhaustive search in a limited window size in the underlying images as in Ref. 14. In addition to integrating Gabor transform¹⁴ and occlusion handling method,¹⁹ we introduce an enhanced visibility map L for large disparities in Section 3.1, and a backward projection disparity field by Gabor transform in Section 3.2 as our main contributions. Both are necessary for dealing with large disparities.

3.1 Enhanced visibility map for large disparities

In Ref. 15, pixel locations in \hat{J} which have the number of accumulated visits, of the forward compensated disparity vectors originating from a far view ($\mathbf{d}_{12}(\mathbf{x})$ or $\mathbf{d}_{43}(\mathbf{x})$), less than a predefined threshold are identified as invisible from that view. Thus newly exposed areas, such as those labeled by $X + X_1$ in Fig. 2, are identified as invisible from the side of I_1 by the forward compensated disparity vectors $\mathbf{x} + (1 + \alpha)\mathbf{d}_{12}^A(\mathbf{x})$, where $\mathbf{x} \in \Gamma$. Similarly, areas such as regions labeled $Y + Y_1$ in Fig. 2, are identified as invisible from the side of I_4 by the compensated

disparity vectors $\mathbf{x} + (2 - \alpha)\mathbf{d}_{43}^A(\mathbf{x})$, where $\mathbf{x} \in \Gamma$. This is directly adopted from Ref. 15 as the initial steps 2 and 3 in our proposed Algorithm 1 for enhanced visibility map estimation. We can interpolate these pixels from the opposite view, where they are visible. For instance, interpolate the pixels of \hat{J} in areas such as ones labeled $X + X_1$ in the figure from the image pair (I_3, I_4) and similarly the pixels in areas such as the ones labeled $Y + Y_1$ from (I_1, I_2) .

Since the estimation of L in Ref. 15 is based on small disparities among the input images, its direct reuse for large disparities (sparse views) can lead to non-optimal decisions. For instance, the area slices labeled X_1 and Y_1 in the vicinity of the tree in Fig. 2, exhibiting large displacements among the views, can still be viewed in images I_2 and I_3 , respectively. Hence we have the option to synthesize area X_1 from the pair (I_4, I_3) or (I_2, I_3) . The method in Ref. 15 uses (I_4, I_3) , accumulating errors due to extrapolation from far views. On the other hand interpolating from the pair (I_2, I_3) is bi-sidedly bounded yielding better estimation and results (Step 4 in Algorithm 1) for areas such as X_1 . Similarly areas like the region labeled Y_1 , are identified in Step 5 of Algorithm 1. Areas which are only visible in the image pair (I_1, I_2) (such as regions labeled X in Fig. 2) can

Algorithm 1: Enhanced visibility map

Input: I_1, I_2, I_3, I_4

Output: Enhanced Visibility map L for view \hat{J} to be determined via interpolation

- 1 Initialize $L(\mathbf{x}) = 0 \forall \mathbf{x} \in \Gamma$ to indicate that they are generally visible in the near image pair (I_2, I_3) ;
 - 2 Identify the set F_{12} of samples that are invisible in I_1 and partially invisible in I_2 (such as regions labeled $X + X_1$ in Fig. 2) by thresholding the pixel revisit count at \hat{J} of the compensated irregular samples $\{\mathbf{x} + (1 + \alpha)\mathbf{d}_{12}^A(\mathbf{x})\}$ as in Ref. 15;
 - 3 Identify the set F_{43} of samples that are invisible in (I_4) and partially invisible in (I_3) (such as regions labeled $Y + Y_1$ in Fig. 2) by thresholding the pixel revisit count at \hat{J} of the compensated irregular samples $\{\mathbf{x} + (2 - \alpha)\mathbf{d}_{43}^A(\mathbf{x})\}$ as in Ref. 15;
 - 4 Identify the set N_{12} of samples that are invisible in I_1 (such as regions labeled X_1 in Fig. 2) by thresholding the pixel revisit count at I_2 of the compensated irregular samples $\{\mathbf{x} + \mathbf{d}_{12}^A(\mathbf{x})\}$;
 - 5 Identify the set N_{43} of samples that are invisible in I_4 (such as regions labeled Y_1 in Fig. 2) by thresholding the pixel revisit count at I_3 of the compensated irregular samples $\{\mathbf{x} + \mathbf{d}_{43}^A(\mathbf{x})\}$;
 - 6 Identify areas which are only visible from the image pair (I_4, I_3) (such as regions labeled X in Fig. 2) by the set difference and labeled by $L(F_{12} - N_{12}) = 1$;
 - 7 Identify areas which are only visible from the image pair (I_1, I_2) (such as regions labeled Y in Fig. 2) by the set difference and labeled by $L(F_{43} - N_{43}) = -1$;
-

be obtained by the set difference $F_{12} - N_{12}$ as shown in step 6 of Algorithm 1, where F_{ij} is a set of samples that are invisible in I_i and partially visible in I_j and N_{ij} is a set of samples that are invisible in I_i and completely visible in I_j w.r.t the interpolated frame \hat{J} . Similarly in step 6 of Algorithm 1, areas which are only visible in the image pair (I_4, I_3) (such as regions labeled Y in Fig. 2) can be obtained as the set difference $F_{43} - N_{43}$. We use three labels to build L as in Ref. 15 : 0 for the pixels visible in (I_2, I_3) , 1 for pixels which are only visible in (I_4, I_3) , and -1 for pixels which are only visible in (I_1, I_2) . The set differences computed in Steps 3 \rightarrow 7 of Algorithm 1, provide our enhanced visibility map estimate that is better-suited for large disparities. For the case of small disparities, areas such as the regions labeled X_1 and Y_1 will not occur and the algorithm will operate in a manner analogous to the method in Ref. 15. It should be noted that the enhanced feature of Algorithm 1 is not applied to the leftmost and rightmost portion of the virtual view image where only steps 2 and 3 are applied. In practice, these two portions are selected to correspond to the maximum disparity value in pixels.

3.2 Backward disparity estimation by Gabor transform

Estimating \mathbf{d}^A using an energy minimization method²⁰ is not capable of correctly estimating the large disparities seen in a wide baseline setting, even when using coarse to fine resolutions framework. This is due to the fact that the coarse scale lacks fine details that are important to guide the matching procedure, and the scene may contain objects that are smaller than their disparity, and these objects will typically be over-smoothed at the

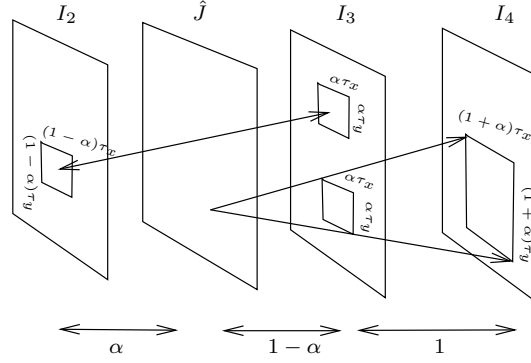


Figure 4. Search windows for bilateral and unilateral Gabor transform-based backward projection disparity estimation.

coarse scale.²⁰ Hence, we introduce a new Gabor transform-based disparity estimation \mathbf{d}^G as an initial step to obtaining a backward disparity \mathbf{d}^A defined on the sampling grid of \hat{J} that accommodates potentially large disparities. For each pixel in \hat{J} , we selectively search an image pair for a disparity vector that minimizes the sum of squared difference of the magnitudes of the Gabor filtered image pairs. This is done based on the visibility map L by using the terms $\delta(L+1)$, $\delta(L)$ and $\delta(L-1)$ in our proposed objective function :

$$\begin{aligned} \mathbf{d}^G(\mathbf{x}) = \arg \min_{\hat{\mathbf{d}}(\mathbf{x})} \sum_i & \\ & \delta(L+1) \left| G_1^i(\mathbf{x} - (1+\alpha)\hat{\mathbf{d}}(\mathbf{x})) - G_2^i(\mathbf{x} - \alpha\hat{\mathbf{d}}(\mathbf{x})) \right|^2 \\ & + \delta(L) \left| G_2^i(\mathbf{x} - \alpha\hat{\mathbf{d}}(\mathbf{x})) - G_3^i(\mathbf{x} + (1-\alpha)\hat{\mathbf{d}}(\mathbf{x})) \right|^2 + \\ & \delta(L-1) \left| G_3^i(\mathbf{x} + (1-\alpha)\hat{\mathbf{d}}(\mathbf{x})) - G_4^i(\mathbf{x} + (2-\alpha)\hat{\mathbf{d}}(\mathbf{x})) \right|^2, \end{aligned} \quad (1)$$

where δ is the unit sample function and G_i^j is the filtered image I_i by the j^{th} Gabor filter. We use 12 Gabor filters * as in Ref. 14 . Based on the assumption that the maximum displacement between (I_1, I_2) , (I_2, I_3) , and (I_4, I_3) is constrained by a value of τ_x in horizontal and τ_y in vertical directions, we propose to use two different types of search windows to solve (1). Referring to Fig. 4 these two search windows correspond to:

- *bilateral search*: in which \hat{J} is enclosed between (I_2, I_3) . The disparity vector is assumed to be a line passing through a pixel in \hat{J} , which at one end searches in a window of size $((1-\alpha)\tau_x \times (1-\alpha)\tau_y)$ in I_2 and at the other end searches in a window of size $(\alpha\tau_x \times \alpha\tau_y)$ in I_3 .
- *unilateral search* in which \hat{J} is on one side of either (I_1, I_2) or (I_3, I_4) image pairs. The disparity vector originates at a pixel in \hat{J} and searches in a window of size $(\alpha\tau_x \times \alpha\tau_y)$ at the near image ($(I_2$ or $I_3)$), and searches in a window of larger size $((1+\alpha)\tau_x \times (1+\alpha)\tau_y)$ in the further image ($(I_1$ or $I_4)$).

The size of the search windows in both cases are dependent on the value of τ_x and τ_y , in addition to the value of α .

4. EXPERIMENTAL RESULTS

We tested our proposed method on images of several natural scenes captured with translational camera displacement. Here we present results for 3 datasets: the *Flower-Garden* and the *Lab*²¹ video sequences captured with

*The Gabor transform is implemented using a set of modulated limited-band Gaussian filters at different tuning frequencies $\{\pi/16, \pi/8, \pi/4\}$ and orientations $\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ as in Ref. 14 .



Figure 5. Sample view interpolation result obtained using the proposed method: (a), (e), and (g) shows a ground truth image J from each of the 3 sequences (Flower-Garden, Lab, and Ballroom), respectively. (b), (f), (h) is the result obtained with the proposed method, and (c)-(d) are close-ups from (a)-(b), respectively.

translational camera motion, and multi-view *Ballroom*²² video. To test the effectiveness of the proposed method, we simulated the sparse views by using one frame out of every twelve frames from the *Flower-Garden* sequence to produce I_1, I_2, I_3 and I_4 . An example of these four images is depicted in Fig. 2. The maximum horizontal disparity among the images is approximately 72 pixels and the maximum vertical disparity is approximately 16 pixels. Therefore, we set $\tau_x = 72$ and $\tau_y = 16$. Fig. 5 (a) shows the ground truth image J , (b) is the proposed method result \hat{J} , and (c)-(d) are closeups from (a)-(b) respectively. The figure shows that the proposed method reconstructs \hat{J} with high quality, and this high quality is maintained for the challenging regions around the tree that are occluded in one of the adjacent views. Table 1 lists the improvements in peak-signal-to-noise ratio Δ PSNR (in dB) and in the structured similarity index Δ SSIM²³ for the proposed method compared with the results obtained with: a) interpolation using the Gabor filter without occlusion awareness, which is similar to the method in Ref. 14, and b) the method in Ref. 15. The table lists the improvement across a range of disparities. For input images having disparities under 24 pixels, the gain over the method in Ref. 15 is insignificant but for larger disparities the gain is quite significant. This is due to the fact that the method in Ref. 15 behaves well in the presence of small disparities between the input images, but as the disparities get larger, the proposed method performs much better.

We also evaluated the proposed method on the *Lab* sequence using one out of 16 frames to obtain the four captured views in our large disparity setting. For the *Ballroom* sequence, which is used as a test sequence for multi-view video coding, the available data represents a simultaneous capture of the scene with 8 cameras; we select a view from each odd numbered camera and evaluate the performance on the even numbered camera locations that are spanned by a camera on either side. Figure 5 (a), (e), and (g) shows a ground truth image J from each of the 3 sequences, (b), (f), and (h) are the corresponding estimates (\hat{J}) obtained with the proposed method, and (c)-(d) are close-ups from (a)-(b), respectively. Table 2 summarize the PSNR and the structured similarity index SSIM²³ for the interpolated image \hat{J} obtained using the proposed method. The higher PSNR for the *Lab* sequence compared to *Flower-Garden* sequence is due to the high complexity of the latter sequence. It contains repeated patterns (flowers) and smooth areas (sky). These similar areas would introduce ambiguities in the disparity estimation and affect the coherence of the disparity field generated by the proposed method.

Figure 6 presents the enhanced visibility map of the proposed method and that of Ince and Konrad¹⁵ (even initialized with Gabor-based disparity estimation of the proposed method). The black color represents $L = -1$ which means visible in the image pair (I_1, I_2) , the gray color represents $L = 0$ which means visible in the image pair (I_2, I_3) , and the white color represents $L = 1$ which means visible in the image pair (I_4, I_3) . These colors are overlaid in Fig. 6 on the ground truth image J . Interpolating areas labeled by $L = -1$ from (I_1, I_2) in Fig. 6

	Max. Disparity (pixels)	Without handling occlusions as in Ref. 14		Method in Ref. 15	
		Δ PSNR (dB)	Δ SSIM	Δ PSNR (dB)	Δ SSIM
Small disparity	24	0.2321	0.0027	0.1	0.001
Large disparity	48	0.3014	0.0246	3.3	0.21
	72	0.5075	0.0310	5.01	0.3

Table 1. Δ PSNR and Δ SSIM gain of the proposed method over the methods in Refs. 15 and 14 for the input scene views (*Flower-Garden*) from Fig. 2.

	<i>Flower-Garden</i>	<i>Lab</i>	<i>Ballroom</i>
PSNR (dB)	26.8	28.5	27.13
SSIM	0.8532	0.8451	0.8715

Table 2. PSNR and SSIM of the proposed method.

(a) has a major drawback which is that the pixels in this area are extrapolated from far views (I_1, I_2) which has much less accuracy than interpolating them from near views (I_2, I_3) when this is possible. Our proposed approach largely eliminates this drawback and minimizes the size of such areas as shown in Fig. 6 (b) using Algorithm 1.

5. CONCLUSION

We proposed a disparity estimation method from multiple views with large disparities. The method has 2 stages: estimation of an enhanced visibility map to handle occlusions, and estimation of a disparity field from an appropriate image pair using the Gabor transform for a coarse estimate that is refined by variational regularization. The enhanced visibility map guides the interpolation process to correctly select the appropriate image pair for use in interpolation and our proposed unilateral and bilateral search procedure for Gabor transform based backward disparity estimation tackles both large disparities and occlusions. The results show enhancements, both visually and quantitatively (PSNR and SSIM), over the method in Ref. 15 that is occlusion aware but does not accommodate large disparities and over non-occlusion aware Gabor transformation based estimation, which is similar to the method in Ref. 14.

Future research would involve the case of navigating close or out of the scene. This can be done by applying our view interpolation technique to get the required view then apply one of the regularized super-resolution techniques as in Refs. 24 and 25 using the image sequence with the guide of the enhanced visibility map L to interpolate from a subset of irregular samples from each image rather than interpolating blindly from all corresponding samples from all images. View interpolation from less structured wide-baseline datasets, where the cameras are not necessarily displaced along a line, also remains a problem that has received only limited attention.²⁶

REFERENCES

- [1] Dodgson, N. A., "Autostereoscopic 3D displays," *Computer* **38**(8), 31–36 (2005).
- [2] Matusik, W. and Pfister, H., "3D TV: A scalable system for real-time acquisition, transmission and autostereoscopic display of dynamic scenes," *ACM Trans. Graph.* **23**(3), 814–824 (2004).
- [3] Kanade, T., Rander, P., and Narayanan, P. J., "Virtualized reality: Constructing virtual worlds from real scenes," *IEEE Multimedia, Immersive Telepresence* **4**, 34–47 (January 1997).
- [4] Zitnick, C. L., Kang, S. B., Uyttendaele, M., Winder, S., and Szeliski, R., "High-quality video view interpolation using a layered representation," *ACM Trans. Graph.* **23**(3), 600–608 (2004).

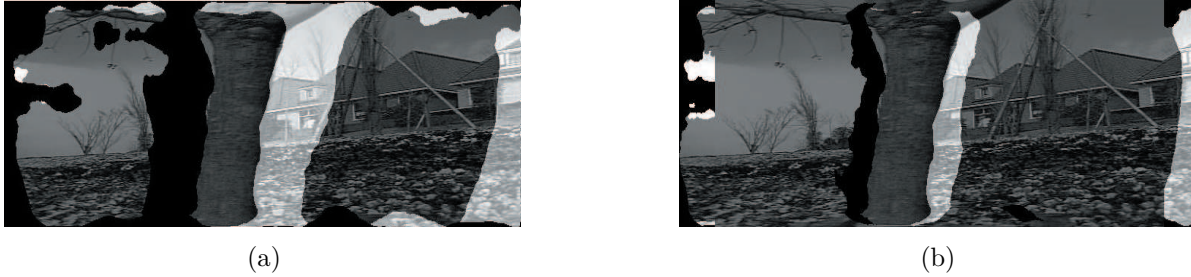


Figure 6. Comparison of the visibility map L produced by (a) method in Ref. 15 and (b) Proposed method (Algorithm 1).

- [5] Zwicker, M., Yea, S., Vetro, A., Forlines, C., Matusik, W., and Pfister, H., "Multi-view video compression for 3D displays," *Forty-First Asilomar Conf. on Sig., Sys. and Computers (ACSSC)*, 1506–1510 (2007).
- [6] Yea, S. and Vetro, A., "View synthesis prediction for rate-overhead reduction in ftv," *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008*, 145–148 (2008).
- [7] Yeo, C. and Ramchandran, K., "Robust distributed multiview video compression for wireless camera networks," *IEEE Trans. Image Process.* **19**(4), 995–1008 (2010).
- [8] Chung, T.-Y., Jung, I.-L., Song, K., and Kim, C.-S., "Multi-view video coding with view interpolation prediction for 2d camera arrays," *J. Vis. Commun. Image Represent.* **21**, 474–486 (July 2010).
- [9] Zokai, S. and Wolberg, G., "Image registration using log-polar mappings for recovery of large-scale similarity and projective transformations," *IEEE Trans. Image Process.* **14**, 1422–1434 (October 2005).
- [10] Vasileisky, A., Zhukov, B., and Berger, M., "Automated image coregistration based on linear feature recognition," *Proceedings of the Second Conference Fusion of Earth Data, Sophia Antipolis, France*, 59–66 (1998).
- [11] Foroosh, H., Zerubia, J. B., and Berthod, M., "Extension of phase correlation to subpixel registration," *IEEE Trans. Image Process.* **11**, 188–200 (Mar. 2002).
- [12] Wolberg, G. and Zokai, S., "Image registration for perspective deformation recovery," *SPIE Conference on Automatic Target Recognition, Orlando, Florida, USA*, 12 (April 2000).
- [13] Yang, Z. and Cohen, F., "Image registration and object recognition using affine invariants and convex hulls," *IEEE Trans. Image Process.* **8**, 934–946 (April 1999).
- [14] Huang, X. and Dubois, E., "Two-dimensional regularized disparity estimation based on the Gabor transform," *Visual Communications and Image Processing (San Jose, CA), SPIE* **6077**, 342–351 (Jan. 2006).
- [15] Ince, S. and Konrad, J., "Occlusion-aware view interpolation," *EURASIP Journal on Image and Video Processing* **2008** (Oct. 2008).
- [16] Jung, I.-L., Chung, T., Song, K., and Kim, C.-S., "Virtual view synthesis using multi-view video sequences," *Proc. IEEE Int. Conf. Image Processing*, 2341–2344 (2009).
- [17] Dubois, E., "The sampling and reconstruction of time-varying imagery with application in video systems," *Proc. IEEE* **73**, 502–522 (Apr. 1985).
- [18] Sun, J., Li, Y., Kang, S. B., and Shum, H.-Y., "Symmetric stereo matching for occlusion handling," *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 399–406 (2005).
- [19] Ince, S. and Konrad, J., "Geometry-based estimation of occlusions from video frame pairs," *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing* **2**, 933–936 (Mar. 2005).
- [20] Brox, T., Bregler, C., and Malik, J., "Large displacement optical flow," *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 41–48 (Jun. 2009).
- [21] "Computer vision test images." <http://www.cs.cmu.edu/~cil/v-images.html>.
- [22] "MERL." <ftp://ftp.merl.com/pub/avetro/mvc-testseq/orig-yuv>.
- [23] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.* **13**, 600–612 (Apr. 2004).
- [24] Ham, B., Min, D., Choi, J., and Sohn, K., "Virtual view rendering using super-resolution with multiview images," *Proc. IEEE Int. Conf. Image Processing*, 2345–2349 (2009).

- [25] Vázquez, C., Aly, H. A., Dubois, E., and Mitiche, A., “Motion compensated super-resolution of video by level sets evolution,” *Proc. IEEE Int. Conf. Image Processing* **3**, 1767 – 1770 (2004).
- [26] Caner, G., Tekalp, A. M., Sharma, G., and Heinzelman, W., “Multi-view image registration for wide-baseline visual sensor networks,” *Proc. IEEE Int. Conf. Image Processing* , 369–372 (Oct. 2006).