Segmentation of Epipolar-Plane Image Volumes with Occlusion and Disocclusion Competition

Jesse Berent and Pier Luigi Dragotti
Communications and Signal Processing Group
Electrical and Electronic Engineering Department
Imperial College, Exhibition Road, London SW7 2AZ, United Kingdom
{jesse.berent, p.dragotti}@imperial.ac.uk

***Abstract*— Consider a dense array of cameras uniformly distributed along a line. A solid block of 3D data can be constructed by arranging the images into a stack. This volume, also known as the Epipolar-Plane Image volume, contains highly structured data that can be segmented for object removal, insertion and compression. In this paper, we propose a segmentation scheme that takes fully advantage of the known geometry in order to model occlusions explicitly as a result of disparity. Moreover, we include this knowledge into an energy minimization scheme based on region competition with active contours. Instead of extracting layers sequentially from front to back, each layer is made to compete with the regions it is going to occlude and the ones it is going to disocclude. This enables a virtually unsupervised segmentation.**

I. Introduction

The data acquired by multiple cameras from multiple viewpoints can be parameterized in a single function called the plenoptic function. It was first introduced by Adelson and Bergen in [1] with the goal of describing what one sees from an arbitrary viewpoint in space. It can therefore be characterized with seven dimensions namely the viewing location and direction, wavelength and time. A particular case is obtained by fixing time and wavelength and reducing the viewing position to a line. Under these constraints, the plenoptic function reduces to a 3-dimensional function also known as the Epipolar-Plane Image (EPI) volume [2]. Such a volume is constructed by collating multiple images taken from equidistant locations along a line as shown in Figure 1. Under a projective camera model, a point in space is projected onto a line in the EPI with a slope inversely proportional to its depth. This setup provides a coherent function for analyzing all the images simultaneously in 3D.

Epipolar-Plane Images were first analyzed in the seminal work of Bolles et al. in [2] were it was shown that 3D information of a scene can be recovered by finding lines in the EPI. In [3], it is suggested that the segmentation of the EPI into coherent regions can be beneficial for numerous applications such as scene interpretation, object manipulation, occlusion removal and compression [4]. In order to segment the data, the authors introduce the notion of EPI tube. Similarly to the volume carved out by an object in a video [5], an EPI tube is obtained by gathering a collection of lines in the EPI volume that have similar slope or belong to the same layer. The segmentation is done by incorporating the knowledge of the geometry of the camera setup in order to take into account occlusions explicitly as a result of disparity. Indeed, occlusions and disparity are closely related since a point with a larger disparity (i.e. closer to the camera plane) will occlude a point that is further away. Following this observation, most EPI analysis algorithms [2], [3], [6] extract layers separately in each horizontal slice and in a sequential manner by detecting lines. The frontmost areas are isolated and removed from further consideration. Subsequent occlusions are thus explained.

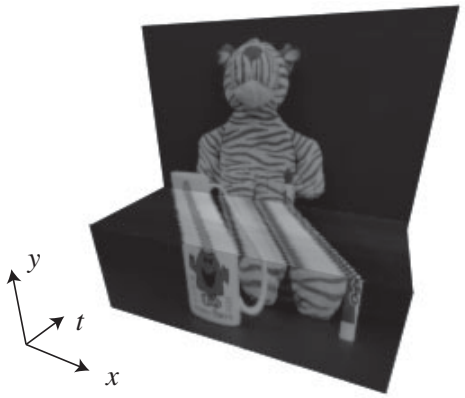

Fig. 1. Epipolar-Plane Image (EPI) volume. The 3D data is generated by arranging images taken from a uniformly distributed linear camera array into a stack. We cut the volume in order to show one Epipolar-Plane Image. According to the geometry, a point in space is mapped onto a line in the EPI. Points that are closer to the camera plane have larger slopes than the points that are further away.

In this paper, we include the disparity-occlusion correlation into an energy minimization with active contours. Instead of removing layers from front to back sequentially, we include the geometric properties in the segmentation of all the layers. Each object is competing with the objects it is going to occlude or disocclude. In this manner, the final result is not dependent on the segmentation of prior layers. Furthermore, we do not treat each EPI slice separately, but rather we perform the segmentation on the whole volume, thus preserving continuity along the vertical axis as well.

The paper is organized as follows: In Section II we give a brief introduction to image segmentation using active contours and the level set methodology. We then describe the EPI tube

extraction algorithm followed by the disparity estimation and initialization. Section III shows results for synthetic as well as real data. Finally, we conclude in Section IV.

II. Proposed method

In this section, we recall the speed function used in order to evolve an active contour towards the local minimum of an energy functional. We then show how the methodology can be applied to the segmentation of EPI volumes with region competition.

A. Preliminary: Image segmentation with level sets

Since the original work by Kass et al. [7], active contours have been used successfully in numerous image and video segmentation algorithms. A few examples can be found in [5], [8], [9], [10]. The idea is to evolve a curve with a speed function that is defined in such a way that it minimizes a certain energy functional. Usually the functional contains two terms, one attracting the contour towards the boundary of the object and one that regulates the smoothness of the curve. Consider an image domain separated in two regions Ω and $\overline{\Omega}$ by the curve Γ. The energy functional to minimize can be written in the form

$$E(\Gamma) = \iint_{\Omega(\Gamma)} f(x,y)dxdy + \iint_{\overline{\Omega}(\Gamma)} g(x,y)dxdy + \int_{\Gamma} \lambda ds,$$

where $f(x,y)$ and $g(x,y)$ are the functions to minimize in Ω and $\overline{\Omega}$ respectively and s is the arc length of the curve. The constant weight λ defines the influence of the curvature term. In order to derive an evolution equation that attracts the curve towards the local minimum of the functional, the boundary Γ is made dependent of an evolution variable τ. It can be shown either through the Green-Riemann theorem and Euler Lagrange equations [9], [11] or through Eulerian derivatives [10] that the steepest descent of the energy yields the partial differential equation (PDE)

$$\frac{\partial \vec{\Gamma}(\tau)}{\partial \tau} = [f(x,y) - g(x,y) + \lambda\kappa]\vec{N} = F\vec{N}, \qquad (1)$$

where $\vec{N}$ is the inward unit normal to the curve and κ is its curvature. The initial condition $\Gamma(0)$ is defined by the user. The PDE in (1) can be solved efficiently using the level set methodology [12]. The curve Γ is embedded as the zero level set of a higher dimension surface ϕ such that $\Gamma(\tau) = \{(x,y)|\phi(x,y,\tau) = 0\}$ and the evolution equation becomes $\frac{\partial \phi}{\partial \tau} = F|\nabla\phi|$. Usually, ϕ is chosen to be the signed distance to the curve therefore $|\nabla\phi| = 1$. The main advantages of using level sets are the capacity in handling topological changes and numerical stability. We refer to [12] for a detailed description.

B. Region competition for EPI tube segmentation

In classical EPI analysis, the viewpoints are uniformly distributed along a line. We denote a point in the EPI volume as $I(x,y,t)$ where x and y are the image coordinates and t denotes the position of the camera along the line. Assuming the scene is made of approximately fronto-parallel regions, each EPI tube is made of a collection of lines with the same slope p. Indeed, the disparity of a point between two consecutive cameras remains constant throughout the stack. Furthermore, the intensity along the lines remains constant assuming the scene is made of opaque Lambertian surfaces. Therefore, a good measure for a point's consistency with a particular layer is the variance along a line of the corresponding slope.

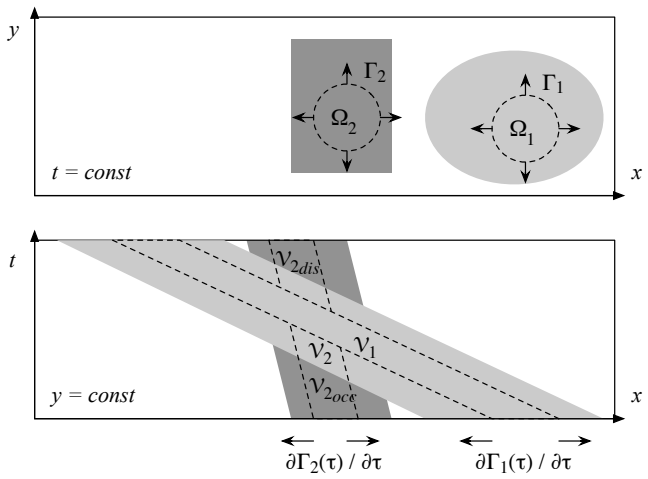

Fig. 2. Occlusion and disocclusion region competition. The front tube $\mathcal{V}_1$ is in competition with the rear tube $\mathcal{V}_2$. One side of $\mathcal{V}_1$ is competing with the 'to be occluded' volume $\mathcal{V}_{2_{occ}}$ and the other side is competing with the 'disoccluded' volume $\mathcal{V}_{2_{dis}}$.

Consider a scene made of N layers, each region Ω_n delimitated by its contour Γ_n and assume, for the moment, that the slopes p_n are known. Each region corresponds to an EPI tube denoted $\mathcal{V}_n$. According to the occlusion ordering, the volumes are arranged from font ($n = 1$) to back ($n = N$) and orthogonalized by removing occluded regions such that

$$\mathcal{V}_n^{\perp} = \mathcal{V}_n \cap \sum_{i=1}^{n-1} \overline{\mathcal{V}_i^{\perp}}. \qquad (2)$$

The global energy we seek to minimize $E_{tot}(\Gamma_1, \ldots, \Gamma_N) = \sum_{n=1}^{N} E_n(\Gamma_n)$ is the sum of the energies for each region. Notice that a point in one image can have multiple depth values and therefore individual curves are propagated for each region. We define the individual energies as

$$E_n(\Gamma_n) = \iint_{\Omega_n} \sigma_n^2 dxdy + \iint_{\overline{\Omega}_n} \overline{\sigma}_n^2 dxdy + \int_{\Gamma_n} \lambda ds,$$

with

$$\sigma_n^2(x,y) = \frac{1}{\mathcal{L}_n} \int_{t \in \mathcal{V}_n^{\perp}} [I(x + p_n t, y, t) - \mu_n]^2 dt, \qquad (3)$$

and

$$\mu_n(x,y) = \frac{1}{\mathcal{L}_n} \int_{t \in \mathcal{V}_n^{\perp}} I(x + p_n t, y, t) dt,$$

with $\mathcal{L}_n = \int_{t \in \mathcal{V}_n^{\perp}} \sqrt{p_n^2 + 1} dt$. A first approach for the choice of $\overline{\sigma}_n^2(x,y)$ is to consider it a constant threshold T.

Indeed, the speed function for the evolving contours reduces to $F_n = \sigma_n^2 - T + \lambda\kappa_n$. Ignoring the curvature term, a line with a variance smaller than T will result in a negative force, thus it will be included in Ω_n. Inversely, a line with a variance larger than T will induce a positive force thus causing Ω_n to reject it. In this case, each epipolar tube can be extracted sequentially starting with the frontmost one. The layer is consequently removed from further consideration and the following tubes can be extracted thereafter. This method suffers from two drawbacks. First, the segmentation accuracy relies on the choice of the threshold. Second, as the layers are extracted sequentially, errors propagate to the segmentation of further layers. In order to resolve these issues, we propose to perform the segmentation on all the layers simultaneously using a region competition. Figure 2 illustrates the reasoning behind the competition formulation. For clarity, we describe the method for two layers. The extension to N layers follows the same reasoning.

Consider a scene made of two fronto-parallel regions Ω_1 and Ω_2 that correspond to two EPI tubes $\mathcal{V}_1$ and $\mathcal{V}_2$. Since $\mathcal{V}_1$ is closer to the camera plane, it is not occluded and $\mathcal{V}_1^{\perp} = \mathcal{V}_1$. The $\mathcal{V}_2$ however will be occluded and $\mathcal{V}_2^{\perp} = \mathcal{V}_2 \cap \overline{\mathcal{V}_1^{\perp}}$. Notice that this is the fundamental difference with the sequential threshold method as σ_2^2 and therefore the evolution of $\mathcal{V}_2^{\perp}$ depends on the evolution of $\mathcal{V}_1^{\perp}$. Hence, we use an iterative approach by fixing Ω_2 while propagating Ω_1 and inversely fixing Ω_1 while propagating Ω_2. From Figure 2, it is clear that the evolution of the left side of Ω_1 modifies σ_2^2 in the so-called 'to be occluded' region [5]. Similarly, the evolution of the right side of Ω_1 modifies σ_2^2 in the 'disoccluded' region. Therefore, $\mathcal{V}_2^{\perp}$ is divided into two sub-tubes $\mathcal{V}_{2_{occ}}$ and $\mathcal{V}_{2_{dis}}$ as shown in Figure 2. In order to make Ω_1 and Ω_2 compete, we define $\overline{\sigma}^2_{1_{occ}}$ as

$$\overline{\sigma}^2_{1_{occ}} = \frac{1}{\Omega_2}\int_{\Omega_2} \sigma^2_{2_{occ}}(u,v)dudv,$$

where $\sigma^2_{2_{occ}}$ is as defined in (3) and $\mathcal{V}_2^{\perp}$ is given by (2). The $\overline{\sigma}^2_{1_{dis}}$ is obtained in the same way. Recall that the level set function ϕ is chosen to be the signed distance to the curve. Therefore the left hand side of the curve corresponds to $\frac{\partial\phi_1}{\partial x}(x,y) < 0$ and the right hand side corresponds to $\frac{\partial\phi_1}{\partial x}(x,y) > 0$. This observation leads to the PDE

$$\frac{\partial\phi_1}{\partial\tau} = (\sigma_1^2 - \overline{\sigma}_1^2 + \lambda\kappa_1)|\nabla\phi_1|,$$

with

$$\overline{\sigma}_1^2(x,y) = \begin{cases} \overline{\sigma}^2_{1_{occ}} & \frac{\partial\phi_1}{\partial x}(x,y) < 0 \\ \overline{\sigma}^2_{1_{dis}} & \frac{\partial\phi_1}{\partial x}(x,y) > 0. \end{cases}$$

Since $\mathcal{V}_2$ is the rearmost tube, it does not occlude or disocclude any other layers. The $\overline{\sigma}_2^2$ is therefore chosen to be unity when the speed is normalized. In the case of N layers, each region competes the average variances along the lines of all the layers it is occluding or disoccluding.

Notice that our EPI tube segmentation algorithm shares some concepts like 'to be occluded' and 'disoccluded' regions with the segmentation of object tunnels and occlusion volumes proposed in [5] for video segmentation. However, there are several important differences due to the fact we are considering multi-view images instead of moving ones. Indeed, while in [5] the minimization is performed in 3D with active surfaces, we use 2D active contours. In our case, two dimensions are sufficient to segment tubes thanks to epipolar geometry [13]. Furthermore, we include a disparity-occlusion rule that takes into account the disparities of layers.

C. Disparity estimation and initialization

In our current implementation, the disparities of each layer are computed using a standard least squares minimization. A set of blocks are tracked throughout the EPI volume. We then keep only the blocks and respective disparities when the trajectory corresponds to a line. In order to initialize a set of layers, blocks where $\frac{p_n}{p_i} < \beta$ are merged. Their contours are used as initial values for the curve evolution.

III. Experimental results and discussion

In this section, we show some preliminary results for the analysis of both synthetic and real EPI volumes. For both scenes, we use the same parameters $\lambda = 0.2$ and $\beta = 0.05$. The disparities are estimated using 8 by 8 pixel blocks. No interpolation was performed for these experiments as integer disparities were used. Due to the lack of a ground truth, we provide a qualitative assessment of the results.

The synthetic *checker* sequence consists of three perfectly fronto-parallel and Lambertian layers. Images 1 and 32 of the stack are depicted in Figures 3(a) and 3(b) respectively. Figure 3(c) illustrates one slice of the EPI volume and the extracted layers are shown in Figure 3(d). In this case, all the assumptions are satisfied and the segmentation can be very accurate. Using a constant threshold T for the region propagation can also provide a similar results if T is well chosen.

The 15 images of the *tiger* sequence were acquired by translating a camera along a linear axis. The first and last images are shown in Figures 3(e) and 3(f). We have not performed any calibration, however the black background was thresholded prior to analysis since it is completely textureless. In spite of the poor contrast and non fronto-parallel regions, the four layers depicted in Figure 3(h) were detected and extracted. An interesting point to notice is that the crude initialization process causes erroneous layers to be initialized. However, these layers vanish thanks to the capacity of the level set method in handling topological changes. The segmentation using a constant threshold is also shown in Figure 3(g). Here we show the best results obtained from a batch of experiments with different values for T. In this case, using the region competition drastically improves the result specially for the mug and the pen layers. Notice also the that paws and the nose of the tiger are at a different depth than the body and arms. This distinction is lost when using the threshold method.

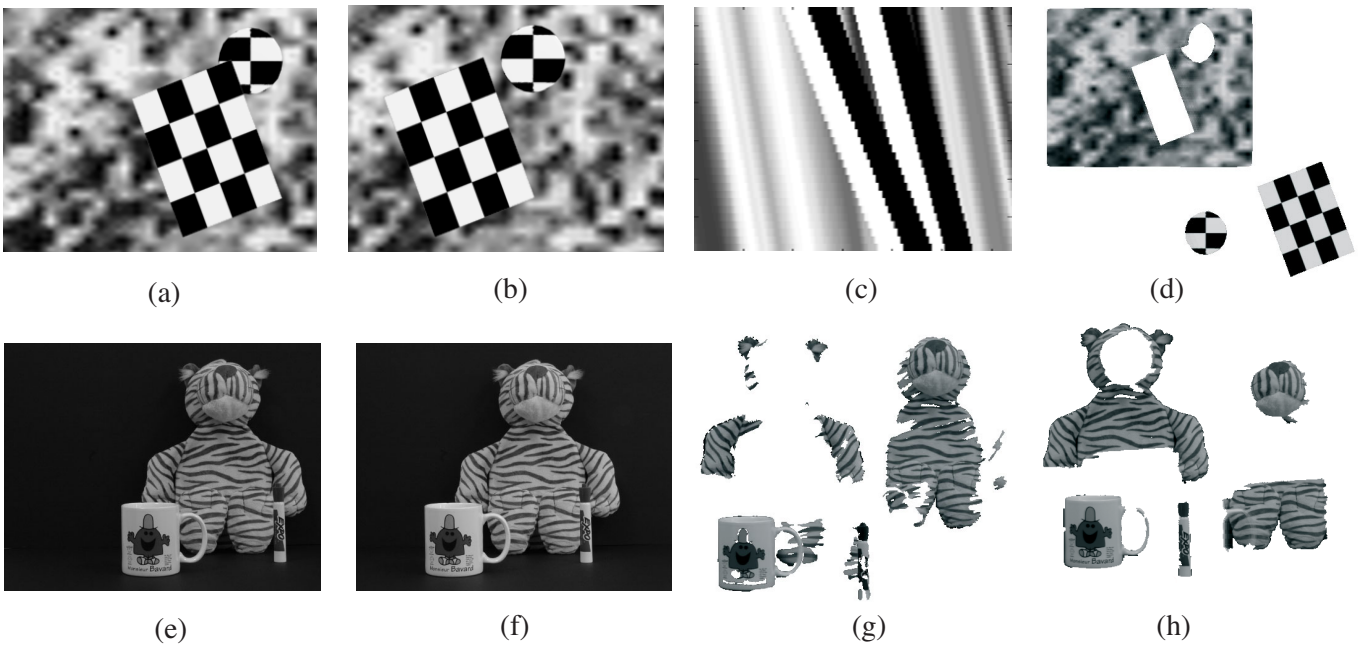

Fig. 3. Experimental results. The synthetic *checker* image sequence (320x240). (a) Image 1 of 32. (b) Image 32 of 32. (c) Epipolar-plane image at slice $y = 70$. (d) Three layers are detected and extracted. The holes in the background are occluded areas in all the images. The *tiger* image sequence (512x400) was acquired by translating a camera along a linear axis. (e) Image 1 of 15. (f) Image 15 of 15. (g) Extracted layers using a constant threshold for each layer. (h) Extracted layers with occlusion and disocclusion competition.

IV. Conclusion and future work

We have proposed a segmentation algorithm for the Epipolar-Plane Image volume that is based on 3D space continuity and explicitly takes into account occlusions given the disparity of each plane. Using epipolar geometry, we reduce the 3D problem of segmenting the EPI volume into a 2D curve evolution. The speed that governs the curve evolution however is computed using the whole stack of images. The main contribution of the scheme presented lies in the competition formulation that enables a global energy minimization instead of extracting layers individually. Furthermore, there are only two parameters in the segmentation process. The first one λ regulates the smoothness of the boarders and the second one β regulates the number of layers to be extracted.

We believe however that the segmentation scheme has potential to be more accurate. Indeed, the fronto-parallel region assumption is quite limiting. Therefore, we are currently investigating lifting the constraint by using affine or piecewise smooth disparity models such that they can be better fitted to the scenes. We are also looking into deriving a more precise evolution equation following in spirit the derivation presented in [10] for region dependent active contours. Finally, while we have not yet spent time on the optimization of the process, we note that there are efficient fast level set methods.

References

[1] 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, MA, 1991, pp. 3–20.

[2] R. Bolles, H. H. Baker, and D. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *Int. Journal of Computer Vision*, vol. 1, pp. 7–55, 1987.

[3] A. Criminisi, S. B. Kang, R. Swaminathan, R. Szeliski, and P. Anandan, "Extracting layers and analyzing their specular properties using epipolar-plane-image analysis," Microsoft Research, Microsoft Corporation, Redmond, WA 98052, Tech. Rep. MSR-TR-2002-19, March 2002.

[4] J. Y. A. Wang and E. H. Adelson, "Representing moving images with layers," *IEEE Trans. on Image Processing Special Issue: Image Sequence Compression*, vol. 3, no. 5, pp. 625–638, September 1994.

[5] M. Ristivojevic and J. Konrad, "Space-time image sequence analysis: object tunnels and occlusion volumes," *IEEE Trans. on Image Processing*, vol. 15, no. 2, pp. 364–376, February 2006.

[6] I. Feldmann, P. Eisert, and P. Kauff, "Extension of epipolar image analysis to circular camera movements," in *Int. Conf. on Image Processing*, September 2003, pp. 697–700.

[7] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, Jan. 1988.

[8] T. F. Chan and L. Vese, "Active contours without edges," *IEEE Trans. on Image Processing*, vol. 10, no. 2, pp. 266–277, February 2001.

[9] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. Journal of Computer Vision*, vol. 1, no. 22, pp. 61–79, 1997.

[10] S. Jehan-Besson, M. Barlaud, and G. Aubert, "Video object segmentation using Eulerian region-based active contours," in *Int. Conf. on Computer Vision*, 2001, pp. 353–361.

[11] S. C. Zhu and A. Yuille, "Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 884–900, September 1996.

[12] J. Sethian, *Level Set Methods*. Cambridge University Press, 1996.

[13] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.