CONFOCAL DISPARITY ESTIMATION AND RECOVERY OF PINHOLE IMAGE FOR REAL-APERTURE STEREO CAMERA SYSTEMS

Jangheon Kim and Thomas Sikora

Department of Communication System, Technical University of Berlin, Germany
E-mail : {j.kim, sikora}@nue.tu-berlin.de

ABSTRACT

A single dense depth estimation using stereo or defocus cannot produce a reliable result due to the ambiguity problem. In this paper, we propose a novel anisotropic disparity estimation embedding a stereo confocal constraint for real-aperture stereo camera systems. If the focal length of a real-aperture stereo camera is just changed, the depth range is localized in a focused object which can be discriminated from defocused blurring. The focal depth plane is estimated by the displacement of tensors which are derived from generalized 2-D Gaussian, since the point spread functions (PSF) in defocused blurring can be approximated by a shift-invariant Gaussian function. We localize the isotropic propagation in blurring over invariance by a sparse Laplacian kernel in Poisson solution. The matching of real-aperture stereo images is performed by observing the focal consistency. However, the isotropic propagation cannot exactly hold a non-parallel surface to the lens plane i.e. unequifocal surface. An anisotropic regularization term is employed to suppress the isotropic propagation near the non-parallel surface boundary. Our method achieves an accurate dense disparity map by sampling the disparities in focal points from multiple defocus stereo images. The pels in focal points are utilized to recover the pinhole image (i.e. an ideally focused image for all different depths).

Index Terms— stereo vision, image matching, optical transfer functions, diffusion processes.

1. INTRODUCTION

Recent years have seen a lot of advances in the problem to reconstruct a complex 3-D scene from a series of multiple images. Most algorithms choose a human visual feature perceiving depth e.g. disparity and DoF (depth of field). By establishing the correspondences in a series, objects are distinguished by depth information related to the respective position. However, the ill-posed correspondence problem cannot be perfectly solved by only one series.

For example, disparity estimation is difficult to solve the ambiguity problem in local image structures due to image noise, unbalanced brightness, similar texture and occlusion, etc. To achieve more reliable estimation performance, local appearance matching [1] with boundary constraints between features, edges and disparity discontinuity etc. is employed. The constraints are utilized as a landmark of the coherency of objects. However, the performance is not satisfactory due to the localization problem. The latest researches incorporate a regularization term which attempts to filter off the delocalized error. Isotropic regularization uses a convolution based on variance. However, the scale of linear transformation in the convolution leads to undesired smoothing of important discontinuity. Anisotropic diffusion methods prevent the important structure from smoothing by modifying the transforms at discontinuity edges by weighting intensity gradients. Although this method results a dense disparity map with local structure preservation, semantic information is needed to combine divided regions of an object.

Practical camera systems use a real aperture camera model which yields focus-related blurring (i.e. DoF). Depth from focus (or defocus) methods [2, 3] exploits the variation of the blurring in a number of images captured at different focus settings. If the camera is focused on an object at a depth, the other objects in a different depth are blurred. The relative blurring between the defocused images can be utilized as the stereo cue. In this paper, we propose a novel anisotropic disparity estimation embedding the defocus cue (i.e. confocal constraint) in a real-aperture 3-D camera system shown in Figure 1.

2. RELATED WORKS

Blurring of defocus by optic can be approximated by linear convolution between the focused image and the blurring function i.e. known as “point spread function” (PSF). Blurring between near- and far-focus images can be estimated by the second central moments of the blur circle because the PSF is a circularly symmetrical function. Subbarao and Surya [2] proposed the S-transform of Laplacian as a focus operator. A defocus function acts as a low-pass filtering e.g. 2-D Gaussian and a focus operator performs the inverse. The difference of the standard deviations i.e. the spread parameter between near- and far-focused images can be mapped to the respective depths. However, the operator is suitable only for equifocal surfaces since the operator is isotropic. In [3], Favaro employs an anisotropic diffusion to solve the problem. If a scene is highly textured, the method is sufficient to estimate a reliable depth. However, regions with weak textures are still ambiguous to distinguish from blurred regions.

For the solution, fusions with stereo [4-6] were proposed. In [4], a probabilistic model of focus and stereo merges the depths by weighted averaging the local variances that are estimated by Cramer-Rao inequality in the unbiased estimator. However, this method cannot guarantee the accuracy of the estimated variances. A Markov random field method [5] integrates the depths by smoothness priors in an energy functional which is minimized by simulated annealing. Another method [6] employs graph-cut for a focus measure. However, these methods do not consider the local features to avoid the convergence into local minima. In our

previous works [7], the multiple-resolutions of the scale-space provide the best trade-off between the detection and the localization performance of features. The range of disparity matching is propagated by isotropic gradient weights following the streamline of the scale-space and localized in the spatially invariant sparse Laplacian kernel by Poisson solver. The dense disparity estimation is locally performed by a scheme of local appearance matching. Anisotropic diffusion globally regularizes the disparity map by suppressing the length of the isotropic propagations to the orthogonal direction of edges. The method gave us good dense disparity maps preserving the important local structures. However, it cannot combine divided regions in the main object.

This paper tries to combine the regional structure in a focal object in defocused images. A focus image offers semantic information such as the sharp boundary of a focusing object which is easily distinguished by the defocus blur. The defocus cue serves a constraint to localize the disparity estimation in a confocal boundary. However, the variances in PSF are isotropic and the propagation near the edges of unequifocal surfaces may not be accurate. We regularize the propagation in the dense disparity map by anisotropic filtering.

2. CONFOCAL DISPARITY ESTIMATION

2.1. Confocal localization for disparity estimation

If a focused point belongs to an object surface, the diameter σ of blurring in a series of defocus images is given by a lens law [5].

$$\sigma_n = \kappa r_n V_n \left(1/F_n - 1/Z_n - 1/V_n\right) \tag{1}$$

n is the number of focus images, r is the lens aperture, F is the focal length, V is the sensor plane-to-lens distance and Z is the distance between the object surface and the lens i.e. the object distance. We use a parameter κ from pre-calibration to fit the focal depth plane into the depth from disparity.

Since conventional stereo systems use a pin-hole camera model, we assume that an ideal focus (i.e. F_n=f) enables the infinite depth of field. The ideal focuses F_{far} and F_{near} on the surfaces which are located at respective depths Z_a and Z_b accompany defocus blurring as Figure 1a illustrates. The confocal disparity can be defined by an epipolar constraint.

$$d = bf / z \tag{2}$$

z is associated depth to the disparity d and b is the baseline. In the pin-hole model, the f in (2) should be the same with the V_n in (1). The relationship between the disparity and the focus of a stereo image pair is achieved by the confocal condition which should estimate a same depth (i.e. $z \approx Z$).

$$d_n = b\left(V_n / F_n - \sigma_n / \kappa r_n - 1\right) \tag{3}$$

Since all parameters except σ_n are given by camera setting, if the focal length of the stereo camera is just changed, the stereo matching can be localized by the confocal constraint within the range of d_n which are estimated by defocus blurring.

2.2. Confocal constraint from defocus

For an object surface S with a function s:$\mathbb{R}^2 \rightarrow [0,\infty]$ which assigns a depth value to each pixel coordinate, the irradiance I:$\Omega \subset \mathbb{R}^2 \rightarrow [0,\infty]$ with another function R:$\mathbb{R}^2 \rightarrow [0,\infty]$ on the surface is observed as the radius σ of the defocus blur. The defocus image J(x) which is dependent upon the camera optics is defined as

$$J(\mathrm{x}) = \int h\left(I(\mathrm{x}), \sigma(s, F_n)\right) R(\mathrm{x}) d\mathrm{x} \tag{4}$$

where h:$\Omega \times \mathbb{R}^2 \rightarrow [0,\infty]$ is the PSF which is defined as the impulse response by Green's function [3]. The radius σ is related on the surface s and the focus parameter F. Let a focus image be the irradiance map I_0 of the focal surface s(F_0) of a initial focus F_0. The depth Z_n=$s(F_0)$~$s'(F_n)$ is estimated by observing the variances of a defocus image I_n which is taken with a different focus F_n at a time t. The energy functional is

$$E(\sigma) = \int_{\Omega} \left(h\left(I_0(\mathrm{x}+\sigma(F_0), 0\right) - h\left(I_n(\mathrm{x}+\sigma(F_n), t\right)\right)^2 d\mathrm{x} \tag{5}$$

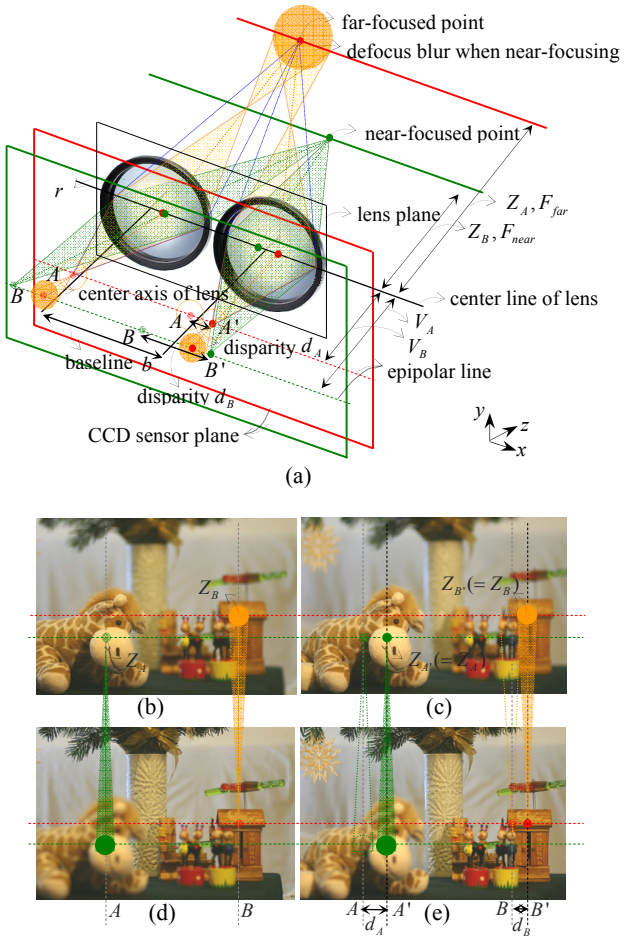

Figure 1. Real-aperture 3-D camera system, (a) definition of the parameters in the system, (b) left viewpoint in a near-focusing setup (f=F_{near}), (c) right viewpoint at F_{near}, (d) left viewpoint in a far-focusing setup (f=F_{far}), (e) right viewpoint at F_{far}.

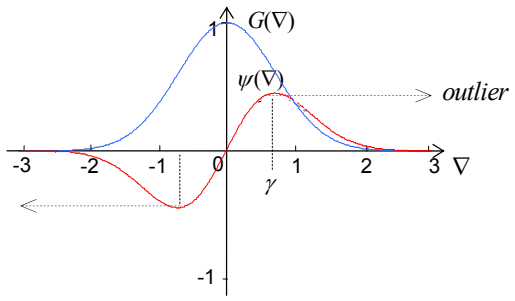

Figure 2. Anisotropic diffusion function G and flux function ψ.

If we approximate the PSF by a generalized 2-D Gaussian as

$$G_{V_n}(\mathrm{x}) = \left(\sqrt{2\pi|V_n|^{1/2}}\right)^{-1} e^{-(x_n^T V_n^{-1} x_n/2)} \tag{6}$$

where V is the symmetric 2×2 variance matrix with the determinant |V|, the eigenvectors determine the principal axes η and ξ from the partial derivatives of a Gaussian and the eigenvalues (μ_+,μ_-) determine the scalar variance along these axes.

$$V(r) = \mu_+ \eta\eta^T + \mu_- \xi\xi^T \quad \text{and} \quad \eta \perp \xi \tag{7}$$

r is the displacement in a polar coordinate $r(\theta)=\sqrt{x^2+y^2}$. This derives an ellipsoid which has a combination of value-weighted orthogonal orientations $\mu_+\theta_+\theta_+^T+\mu_-\theta_-\theta_-^T$ with $\theta_+ \perp \theta_-$ centered at a point. This is more general to represent a partial homogeneous region in a defocus images by convolution $I_n(\mathrm{x})=I_0(\mathrm{x})\otimes G_v(\mathrm{x})$. In the n defocus images, the smoothly varying structure is defined by a positive definite tensor $\mathcal{D}$ that denotes a gradient flux.

$$\mathrm{x}(r\,|\,t,\mathcal{D}) = \left(\sqrt{2\pi^n|2t\mathcal{D}|}\right)^{-1} e^{-r^T\mathcal{D}^{-1}r/4t} \tag{8}$$

The variances in the PSF is observed by the displacement of tensors at each image point x. It represents a difference in propagations to the direction (θ_+,θ_-). If the space of the functions is continuous with the partial derivatives in $\mathbb{R}^2$, (5) can be minimized in the PDEs by harmonic solution $I(\mathrm{x})\rightarrow J:\mathbb{R}^2\times[0,\infty]\rightarrow\mathbb{R}$

$$\begin{cases} J(I(\mathrm{x}),0) = \sigma & {}^{\forall}\mathrm{x}\in\Omega \\ J(I(\mathrm{x}),t) = div\left(\mathcal{D}(\mathrm{x})\nabla(I_n(\mathrm{x}),t)\right) & t>0 \end{cases} \tag{9}$$

where $J(\mathrm{x})=I(\mathrm{x})+\varepsilon$ where ε denotes a correction gradient symbol related to the amount of blurring and the divergence term *div* is implemented by a sparse Laplacian kernel in the Poisson solution $\mathcal{D}(\mathrm{x})\nabla(I_n(\mathrm{x}),t)\cdot n=4\pi\rho$ with $\rho=0$. The unit vector n is orthogonal to $\partial\Omega$.

2.3. Anisotropic disparity estimation with confocal constraint and recovery of pinhole image

With the confocal constraint the d_n from (3) for a focusing stereo object, we obtain the maximum window size $\mathcal{W}$ for the disparity matching.

$$\max_D \mathcal{W}_{l,n}(\mathrm{x}) \le \sum_{i=0}^{d_n} I_{r,n}(\mathrm{x}+i) \tag{10}$$

where the subscripts l and r respectively denote the left and right image. The sampling of pels in the focal plane in defocus stereo images recovers a non-blurring image for all objects i.e. pinhole image. The confocal disparity is achieved by iteratively updating the estimated disparities in the window $\mathcal{W}$ as

$$E_\Omega(d) = \int_\Omega \left(I_l(\mathrm{x}) - \mathcal{P}_r(\mathrm{x}+d_{l\rightarrow r}(\mathrm{x}+\sigma)\in\mathcal{W})\right)^2 d\mathrm{x} + \lambda\int_\Omega e_\sigma(d)d\mathrm{x} \tag{11}$$

The subscripts denote the matching direction, e.g. $l\rightarrow r$ for left-to-right direction and e_σ is a global regularization term which minimizes the matching error with Lagrange multiplier λ. The range of matching is restricted within the confocal depth range. The window using isotropic PSFs is suitable to avoid the splitting problem in a homogenous equifocal surface. However, the edges of unequifocal surfaces may not be exactly localized. Hence, we globally regularize the dense disparity map by an anisotropic diffusion.

$$e_\sigma = G\left(\|\mathcal{D}_l(\mathrm{x})\|\right)\nabla d_{l\rightarrow r}(\mathrm{x},t) \tag{12}$$

$G(\nabla)$ is an anisotropic diffusion weight which suppresses the propagation of only the edge direction θ_-.

$$G(\mu_+,\mu_-) = (e^{-(\mu_+^2/\gamma^2)},\mu_-) \tag{13}$$

where a positive constant γ controls the level of contrast of edges affecting the regularization process as Figure 2 shows. It enhance the discontinuities in θ_- by the flux function $\Psi(\nabla)=G(\mu_+,\mu_-)\nabla$. The detail numerical solution of energy functional in (12) is described in [7].

3. SIMULATION RESULTS

Figure 3a and 3b show a pair of stereo images which are respectively focusing on the most far and the nearest object surfaces in 10 defocus stereo images. The image resolution is 720×480. First, a focal consistency window from defocus images is estimated to localize the range of disparity matching. Figure 3c represents the window which is estimated from 10 defocus images. The disparity map is iteratively estimated by updating the fine (i.e. focusing) structure in the window. The sampling of the fine structure derives a pinhole image with minimum aperture. The dense disparity map and its recovered pinhole image respectively shown in Figure 3d and 3e show the excellent performance.

For the subjective evaluation of the results, an image-based modeling and an anaglyph rendering are given in Figure 3f and 3g. The image-based modeling is efficient to evaluate the localization performance by mapping the texture image to the depth. Anaglyph is a stereoscopic rendering method that physically recovers the depth by separately projecting red- and blue-coded stereo image into left and right eye. The physical depth can be recovered by red-blue glasses. The anaglyph image represents the accuracy of the recovered physical depth.

We compare the performance with the anisotropic disparity estimation method of our previous work [7]. Since the previous method needs a single well-focus image pair, disparity of

defocusing regions in the background cannot be estimated as Figure 3h represents. The novel method produces a dense disparity map preserving structure of focusing foreground and defocusing background.

In a far-focusing stereo camera setup, we exemplify the fact which the disparity estimation of near-defocusing object is difficult in Figure 4. Reliable Markov Random Fields method [8] and graph-cut optimization [9] cannot estimate the disparity in the blurred regions due to the ambiguity as Figure 4b and 4c shows. However, the proposed method overcomes the problem as Figure 4d represents since the ambiguity in the blurred regions can be avoided by the consistency in another focus images.

4. CONCLUION AND SUMMARY

Conventional stereo algorithms use a pair of ideal focus stereo image with a pinhole camera assumption (i.e. focusing for all objects at different depths). However, in practical camera setup with optical real-apertures, disparity cannot be recovered on unfocused regions due to the increased ambiguity. This paper solved the problem by a novel anisotropic disparity estimation and pinhole image recovery embedding a stereo confocal constraint.

5. ACKNOWLEDGMENT

This work was developed within 3DTV, a European Network of Excellence (*http://www.3dtv-research.net*) funded under the European Commission IST FP6 program.

6. REFERENCES

[1] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: theory and experiment," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 920-932, 1994.

[2] M. Subbarao and G. Surya, "Depth from defucos: a spatial domain approach," *Int. Jour. of Computer vision,* 13(3), pp.271-294.

[3] P. Favaro, S. Osher, S. Soatto, and L. Vese, "3d shape from anisotropic diffusion," in Proc. of *IEEE Int. Conf. on CVPR,* pp179-186, 2003.

[4] V. Bove, "Probabilistic method for integrating multiple sources of range data," *Jour. of the Optical Society of America*, vol. 7, no. 12, pp. 2193–2198, 1990.

[5] A. Rajagopalan, S. Chaudhuri, and U. Mudenagudi, "Depth estimation and image restoration using defocused stereo pairs," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1521–1525, 2004.

[6] C. Frese and I. Ghet, "Robust Depth Estimation by Fusion of Stereo and Focus Series Acquired with a Camera Array," in Proc. of *IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems*, 2006.

[7] J. Kim and T. Sikora, "Gaussian Scale-Space Dense Disparity Estimation with Anisotropic Disparity-Field Diffusion," in Proc. of *IEEE Int. Conf. on 3-D Digital Imaging and Modeling*, 2005.

[8] Y. Boykov, O. Veksler and R. Zabih, "Markov Random Fields with Efficient Approximations," In Proc. of *IEEE Conf. on Computer Vision and Pattern Recognition*, 1998.

[9] V. Kolmogorov and R. Zabih, "Computing Visual Corres pondence with Occlusions using Graph Cuts," In Proc. of *IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.

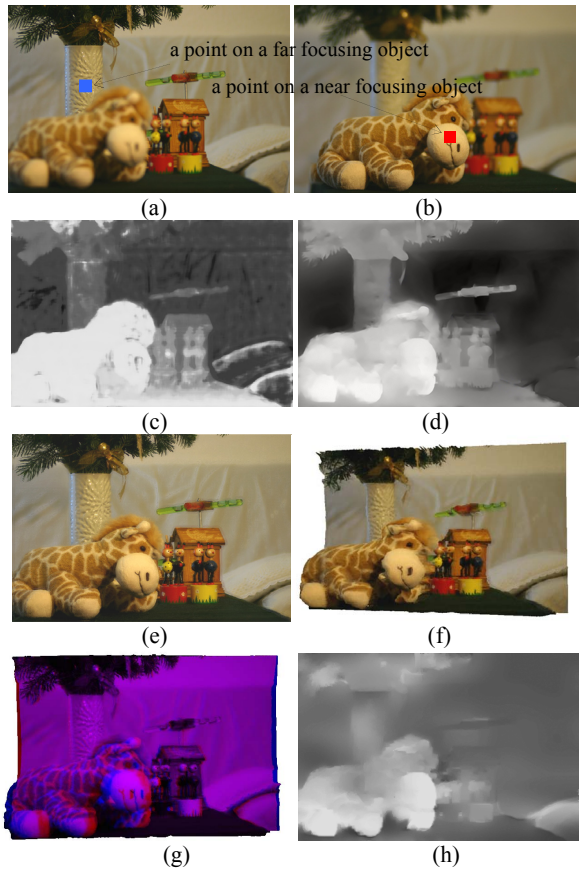

Figure 3. Recovery of disparity map and pinhole image - (a) left image of a far focusing object, (b) right image of a near focusing object, (c) focal consistency window from 10 defocus images for the left view, (d) dense disparity map by the proposed method, (e) recovered pinhole image, (f) 3D model using 3d and 3e, (g) anaglyph rendering using 3d and 3e, (h) a result from our previous work [8].

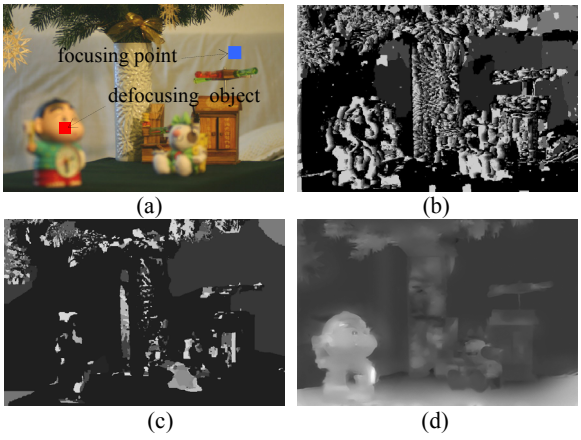

Figure 4. Disparity estimation of a defocusing near object in a far focusing camera - (a) left image, (b) disparity map by Markov Random Fields method [8], (c) disparity map by graph-cut method [9] (d) disparity map by the proposed method.