

# STOCHASTIC FUSION OF MULTI-VIEW GRADIENT FIELDS

*Aswin C. Sankaranarayanan and Rama Chellappa*

Center for Automation Research and Department of Electrical and Computer Engineering  
University of Maryland, College Park, MD 20742  
{aswch,rama}@umiacs.umd.edu

## ABSTRACT

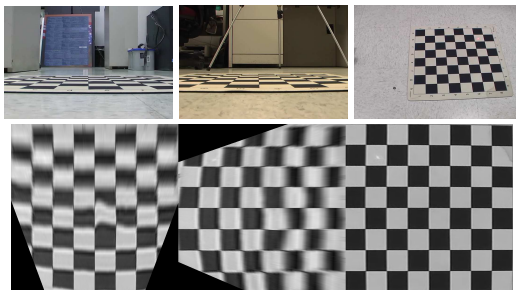
Image gradients form powerful cues in a host of vision and graphics applications. In this paper, we consider multiple views of a textured planar scene and consider the problem of estimating the scene texture map using these multi-view inputs. Modeling each camera view as a projective transformation of the scene, we show that the problem is equivalent to that of studying the effect of noise (and the projective imaging) on the gradient fields induced by this texture map. We show that these noisy gradient fields can be modeled as complete observers of the scene radiance. Further, the corrupting noise can be shown to be additive and linear, although spatially varying. However, the specific form of the noise term can be exploited to design linear estimators that fuse the gradient fields obtained from each of the individual views. The fused gradient field forms a robust estimate of the scene gradients and is useful in many applications.

**Index Terms**— Multi-view estimation, Image fusion, Gradient fields, Image restoration

## 1. INTRODUCTION

Imaging of a scene with a camera is well approximated by a projective transformation and an understanding of the geometry introduced by this imaging process is useful in many estimation problems. In this context, the constraints induced by the projective geometry greatly influences the choice and design of statistical estimators [1].

In this paper, we study the problem of robust estimation of gradient fields using inputs from multiple (projective) views. Gradient fields play an important role in many vision and graphics applications. Estimation of optical flow and shape recovery using shading information, both classical vision problems, involve estimation using gradient fields. Illumination invariant image analysis is tied heavily to the properties of image gradients [2]. The properties of image gradients have been used heavily in many image editing [3] and fusion [4, 5] algorithms. A weakly related concept is that of *scene-flow* estimation [6, 7]. Scene flow refers to the 3D



**Fig. 1.** (top row) Camera views (bottom) Registered top view of the plane as seen from each camera. We pose the following problem: Is it possible to construct a high resolution image of the plane like the one seen in the right-camera, from views such as the left and center camera.

motion flow of the scene and can be computed using optical flow inputs from multiple views. The use of gradients for image fusion has been explored in the context of multi-modal image registration. A distance metric based on Normalized image gradients is used in [8] to fuse multi-modal images for medical applications. In [9], multi-spectral images of varying resolutions are fused by minimizing the differences between their gradients, and using a reconstruction to integrate the fused gradient field.

Consider the three images of a planar scene in Figure 1. We pose the problem of estimating the scene view from a camera looking vertically down (or, equivalently a metric rectification of the images). However, depending on the specific orientation/placement of the camera to the plane, such rectified views can look quite different (see Figure 1). In particular, we draw attention to the way the gradients in the three cases are distorted. In this paper, we study the effect of projective transformations of noisy images and the induced image gradient fields. We show that these image gradients are corrupted with anisotropic noise whose statistics depend heavily on the specific projective transformation (and hence, the specific camera view). Finally, given multiple such image gradient fields, each arising from a different camera, a global estimate can be computed using linear filtering techniques.

This work was partially supported by NSF-ITR Grant 0325119 and DARPA Flexiview Grant HR001107C0059.

## 2. PROBLEM FORMULATION

Consider a function  $f : \mathcal{X} = \mathbb{P}^2 \rightarrow \mathbb{R}$ , characterizing the texture map over the plane in its Euclidean frame of reference  $\mathcal{X}$ . We denote  $\mathbf{x} = (x, y)^T$ ,  $\mathbf{u} = (u, v)^T$  and  $\tilde{\mathbf{x}} \sim (x, y, 1)^T$ ,  $\tilde{\mathbf{u}} \sim (u, v, 1)^T$  represent  $\mathbf{x}$  and  $\mathbf{u}$  in their homogeneous coordinates respectively.

We now consider a set of  $C$  cameras observing this planar scene. Given that the scene is planar, we can map points on the image plane of the camera uniquely to the world plane. This transformation is projective is equivalent to the metric rectification of the plane [10]. Each camera is identified by a projection matrix  $H_i, i = 1, \dots, C$ . Defining,  $\mathcal{U}_i = \mathbb{P}^2$  as the coordinate reference on the image plane of camera  $i$ , we define  $H_i$  as the projective transformation mapping  $\mathcal{U}_i$  onto  $\mathcal{X}$ .

$$\begin{aligned} H_i : \mathcal{X} &\rightarrow \mathcal{U}_i \\ \mathbf{x} &\mapsto \tilde{\mathbf{u}} \sim H_i \tilde{\mathbf{x}} \end{aligned} \quad (1)$$

When the imaging is perfect (continuous, infinite resolution and noiseless), each camera observes a projective transformed image of the scene, i.e, each camera observes a functional  $f_i : \mathcal{U}_i = \mathbb{P}^2 \rightarrow \mathbb{R}$ , such that,

$$f_i(\mathbf{u}) = (f \circ H_i^{-1})(\mathbf{u}) \quad (2)$$

However, imaging parameters such as finite resolution and sensor noise corrupt the observation of the function  $f$ . Further, a host of illumination related issues (camera gain, reflectance of the ground plane) will modify this observation. We assume that the illumination model is known and compensated for. Allowing for noise, the imaging of  $f$  changes as follows.

$$f_i(\mathbf{u}) = (f \circ H_i^{-1})(\mathbf{u}) + n_i(\mathbf{u}) \quad (3)$$

where  $n_i$  is a noise process that accounts for the discrepancies including pixelation.

An immediate problem of interest is that of estimating  $f$  from its noisy projective counterparts, namely  $f_i, i = 1, \dots, C$ . This would correspond to traditional image based mosaicing/restoration problem, which finds use in several applications. An alternate problem that we would like to consider is estimating the gradient field induced by  $f$ , as  $f$  can then be recovered by integrating the gradient field appropriately [11].

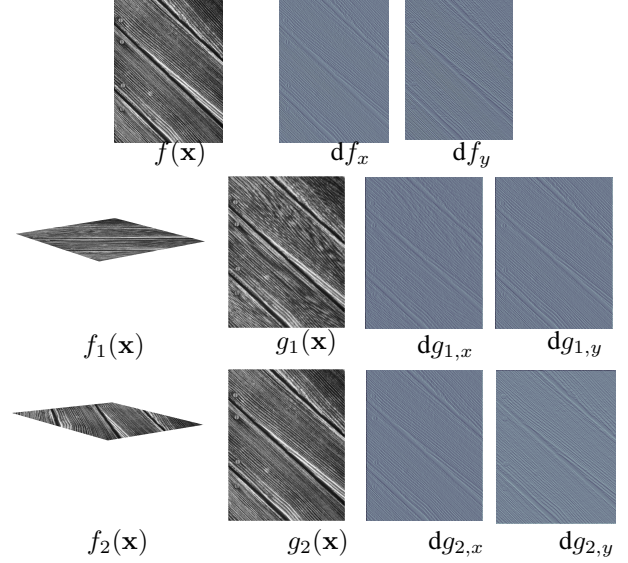
To begin with, we can project  $f_i$  back to the world plane to obtain  $g_i$ .

$$g_i(\mathbf{x}) = (f_i \circ H_i)(\mathbf{x}) = f(\mathbf{x}) + (n_i \circ H_i)(\mathbf{x}) \quad (4)$$

From (4), we can obtain the corresponding equation linking the gradient fields of  $g_i$  and  $f$ .

$$dg_i(\mathbf{x}) = df(\mathbf{x}) + \nabla H_i dn_i \big|_{\mathbf{u}=H_i(\mathbf{x})} \quad (5)$$

Further, practical imaging considerations allow us to restrict the both  $\mathcal{X}$  and  $\mathcal{U}_i$  from  $\mathbb{P}^2$  to  $\mathbb{R}^2$ .



**Fig. 2.** Illustration of symbols using a synthetic example. The gradient images show signed magnitudes of the individual components. The reader is instructed to use the zoom tool to view the pictures better.

With this, we can define  $H_i$  and  $\nabla H_i$  as,

$$\begin{aligned} H_i : \mathbb{R}^2 &\rightarrow \mathbb{R}^2 \\ \mathbf{x} &\mapsto \frac{1}{H_{i,3}^T \tilde{\mathbf{x}}} \begin{bmatrix} H_{i,1}^T \tilde{\mathbf{x}} \\ H_{i,2}^T \tilde{\mathbf{x}} \end{bmatrix} \end{aligned} \quad (6)$$

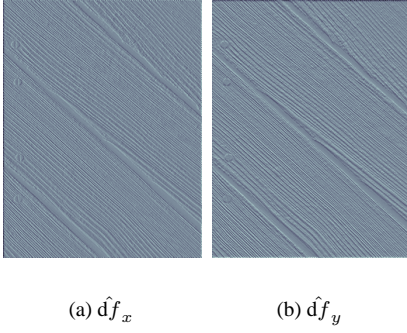
where  $H_{i,j}^T$  is the  $j$ -th row of the matrix  $H_i$ . The expression for  $\nabla H$  can now be derived using basic algebra.

$$\nabla H_i(\mathbf{x}) = \frac{1}{H_{i,3}^T \tilde{\mathbf{x}}} \left( \begin{bmatrix} H_{i,11} - uH_{i,31} & H_{i,12} - uH_{i,32} \\ H_{i,21} - vH_{i,32} & H_{i,22} - vH_{i,32} \end{bmatrix} \right) \quad (7)$$

A closer look at (5) suggests that the gradient field  $dg_i$  is a complete observer of the gradient field  $df$ . This is due to the invertible nature of projective imaging considered in the context of planar scenes. However, the key point to note is the way the noise term appears in (5). Note that the noise corrupting  $dg_i$  is spatially varying given the dependence of the matrix  $\nabla H$  on  $x$ . However, the nature of this *mixing* term is not only point-wise but also linear. We can exploit this property for recovering  $df$  robustly.

## 3. GRADIENT ESTIMATION USING MULTI-VIEW INPUTS

The overall properties of the effect of noise depends critically on the nature of the *mixing* matrix  $\nabla H_i$ . For example, if  $\nabla H_i$  has eigenvalues that are of unequal magnitude then noise gets amplified in anisotropy. Hence, given different views of the same scene, each top-view  $g_i$  is corrupted with noise that is



**Fig. 3.** Minimum variance estimate of the gradient field obtained by fusing the gradient fields induced by  $f_1$  and  $f_2$  (see Figure 2).

not only varying spatially, but is also anisotropic. However, if we look at the way the gradient field is corrupted, the properties of the corrupting noise are modified with a spatially varying linear matrix  $\nabla H_i$ .

Coupling the equation arising from each camera, we get the following equation

$$\begin{bmatrix} dg_1(\mathbf{x}) \\ \vdots \\ dg_C(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \mathbb{I}_2 \\ \vdots \\ \mathbb{I}_2 \end{bmatrix} df(\mathbf{x}) + \begin{bmatrix} \nabla H_1(\mathbf{x})n_1(\mathbf{u}_1) \\ \vdots \\ \nabla H_C(\mathbf{x})n_C(\mathbf{u}_C) \end{bmatrix} \quad (8)$$

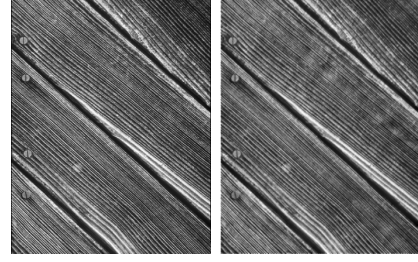
where  $\mathbb{I}_2$  is the rank 2 identity matrix. Equation (8) provides the basic filtering equation for estimating  $df$ . The exact nature of the solution depends on the specific properties of the noise fields  $n_i$ . For our purposes, we assume that the only corrupting noises are from pixelation and mis-registration errors. An accurate model of these noise processes is in general difficult. Toward this end, we assume that the noise is *zero mean*, stationary with a standard deviation of 1 pixel<sup>1</sup>.

$$\forall i, \mathbf{u} \in \mathcal{U}, \mathbf{E}(n_i(\mathbf{u})) = \mathbf{0}, \mathbf{E}(n_i(\mathbf{u})n_i(\mathbf{u})^T) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (9)$$

As a result, we can suitably model this into an estimator for the gradient field induced by  $f$  by efficiently incorporating this information in the fusion scheme. Under this setting, it is possible to achieve a minimum variance estimate using a linear estimator. The minimum variance estimate for  $df(\mathbf{x})$ , denoted as  $\hat{df}(x)$ , is computed as,

$$\hat{df}(\mathbf{x}) = \sum_{j=1}^C \Sigma_j^{-1}(x)\Sigma(x)dg_j(\mathbf{x}) \quad (10)$$

<sup>1</sup>An error of 1 pixel standard deviation is a reasonable approximation of the pixelation error. However, under extreme perspective imaging, artifacts due to misregistration can be quite severe. Such error can be approximated with a Gaussian model with higher variance, or by explicitly modeling the misregistration error.



**Fig. 4.** (left) Ground truth of the scalar field  $f(\mathbf{x})$  (right) Scalar field obtained by reconstructing the minimum variance estimate of Figure 3

where

$$\Sigma_j(\mathbf{x}) = \nabla H_j(\mathbf{x})\nabla H_j^T(\mathbf{x}) \quad (11)$$

and

$$\Sigma = \left( \sum_{j=1}^C \Sigma_j^{-1}(\mathbf{x}) \right)^{-1} \quad (12)$$

Figure 3 shows the minimum variance estimate for the problem depicted in Figure 2. Such an estimator is also optimal (over the class of all estimators) when the noise process are assumed to be Gaussian.

Further, the estimated gradient field  $\hat{df}$  can now be integrated to obtain an estimate of  $f(\mathbf{x})$  (see Figure 4). We use the Poisson reconstruction algorithms described in [11] for the reconstruction.

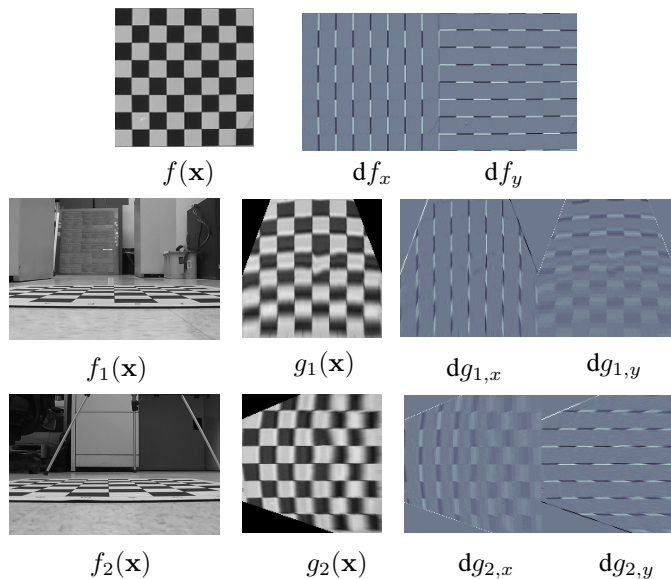
## 4. RESULTS

We applied the proposed theory of multi-view gradient estimation on the chessboard images shown in Figure 1. Figures 5 and 6 summarize the results for this dataset. With the exception of the region that is outside the fields of view of the two cameras, it is seen that  $\hat{df}$  is indeed a good approximation to the ground truth  $df$ . The reason for this comes from the specifics of camera placement. The two cameras are placed so that each view resolves the gradient fields accurately only along one orientation. The fused estimate appropriately weighs the individual estimates to obtain low variance estimates for both orientations.

Finally, we reconstruct the scalar field corresponding to the gradient estimate obtained using fusion. We suppress the false gradients that arise because of field of view lines by assigning a high variance to those regions. This is easy done, as the field of view lines corresponds to pixels near the boundary of the captured image at each view. Figure 7 shows the reconstruction results.

The fusion algorithm presented in this paper works under the assumption of perfect registration, or equivalent when the errors in registration can be captures as additive noise. This,

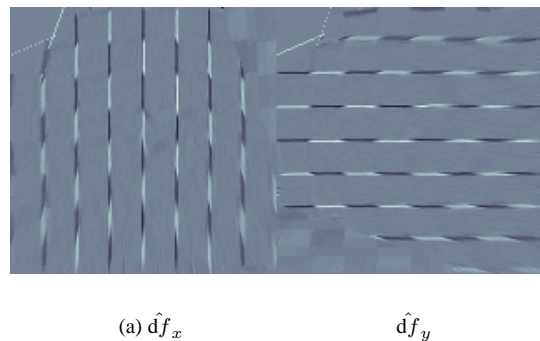
however, is not necessarily true, especially when the perspective distortions are severe. Further, when the scene structure deviates significantly from a plane, the errors due to parallax (off the plane) introduces additional errors in registration. Such errors need to be explicitly addressed, and forms avenues for future research.



**Fig. 5.** Visual representations of the various components for the chessboard example of Figure 1. Note the high distortion in  $dg_{1,y}$  and  $dg_{2,x}$ .

## 5. CONCLUSION

In this paper, we show that it is possible to estimate gradient fields robustly from noisy projective transforms. The ability to estimate gradient fields robustly is useful in many vision and graphics problems. While we consider the planar scene



**Fig. 6.** Minimum variance estimate of the gradient field obtained by fusing the gradient fields induced by  $f_1$  and  $f_2$  (see Figure 5).



**Fig. 7.** (left) Ground truth of the scalar field  $f(\mathbf{x})$  (right) Scalar field obtained by reconstructing the minimum variance estimate of Figure 6

setting (thereby allowing for invertible projective transforms) in this paper, we plan on extending the fusion methodology for motion field estimation without any stringent scene constraints (similar to the work on scene flow estimation).

## 6. REFERENCES

- [1] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier Science Inc. New York, NY, USA, 1996.
- [2] H.F. Chen, P.N. Belhumeur, and D.W. Jacobs, “In search of illumination invariants,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2000.
- [3] A. Agrawal, R. Raskar, and R. Chellappa, “Edge Suppression by Gradient Field Transformation Using Cross-Projection Tensors,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.
- [4] J. Domke and Y. Aloimonos, “Multiple View Image Reconstruction: A Harmonic Approach,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.
- [5] R. Raskar, A. Ilie, and J. Yu, “Image fusion for context enhancement and video surrealism,” in *International Conference on Computer Graphics and Interactive Techniques*, 2005.
- [6] S. Vedula, S. Baker, P. Rander, R. Collins, and T. Kanade, “Three-dimensional scene flow,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 475–480, 2005.
- [7] Y. Zhang and C. Kambhamettu, “On 3d scene flow and structure estimation,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.
- [8] E. Haber and J. Modersitzki, “Intensity gradient based registration and fusion of multi-modal images,” *Methods of Information in Medicine*, vol. 46, no. 3, pp. 292–299, 2007.
- [9] Jianting Wen, Yan Li, and Haifeng Gong, “Remote sensing image fusion on gradient field,” in *International Conference on Pattern Recognition*, 2006, pp. 643–646.
- [10] R. Hartley and A. Zisserman, *Multiple View Geometry in computer vision*, Cambridge Univ. Press, 2003.
- [11] A. Agrawal, R. Raskar, and R. Chellappa, “What is the range of surface reconstructions from a gradient field,” in *European Conference on Computer Vision*, 2006.