Generalized Assorted Camera Arrays: Robust Cross-channel Registration and Applications

Jason Holloway*, *Student Member, IEEE,* Kaushik Mitra, *Member, IEEE,* Sanjeev Koppal *Member, IEEE,* and Ashok Veeraraghavan, *Member, IEEE*

***Abstract*—One popular technique for multi-modal imaging is Generalized Assorted Pixels (GAP), where an assorted pixel array on the image sensor allows for multi-modal capture. Unfortunately GAP is limited in its applicability because of the need for multi-modal filters that are amenable with semiconductor fabrication processes and results in a fixed multi-modal imaging configuration. In this paper, we advocate for Generalized Assorted Camera (GAC) arrays for multi-modal imaging—i.e., a camera array with filters of different characteristics placed in front of each camera aperture. GAC provides us with three distinct advantages over GAP: ease of implementation, flexible application dependent imaging since filters are external and can be changed and depth information that can be used for enabling novel applications (e.g. post-capture refocusing). The primary challenge in GAC arrays is that since the different modalities are obtained from different viewpoints, there is a need for accurate and efficient cross-channel registration. Traditional approaches such as SSD, SAD, and mutual information all result in multi-modal registration errors. Here, we propose a robust cross-channel matching cost function, based on aligning normalized gradients, that allows us to compute cross-channel sub-pixel correspondences for scenes exhibiting non-trivial geometry. We highlight the promise of GAC arrays with our cross-channel normalized gradient cost for several applications such as low light imaging, post-capture refocusing, skin perfusion imaging using RGB+NIR and hyperspectral imaging.**

I. Introduction

Traditional methods for multi-modal acquisition–such as using hyperspectral imagers–trade-off temporal resolution in order to improve spectral resolution by the use of either a spectrally-tunable filter [1] or by using line-scan cameras and a method for spatial scan of the scene [2]. While these methods are popular techniques for hyperspectral acquisition, they are often limited to static scenes and result in objectionable motion-related artifacts in dynamic scenes. Generalized Assorted Pixels (GAP) is slowly gaining popularity as a method for acquiring spectral [3], [4], [5], polarization [3], [4], and angular information [6] on a single image sensor. Unfortunately, several challenges to its widespread adoption remain: Fabrication–GAP requires nano-scale manufacturing techniques that can produce filters in a manner that is compatible with semiconductor fabrication processes, Cost–currently available GAP sensors such as those offered by PixelTeq and IMEC are expensive, and Resolution–the use of GAP results in loss of spatial resolution and often produces low-resolution hyperspectral images (e.g., 256×256 resolution in IMEC).

The GAP strategy is to place filters on a single sensor. Generalized camera arrays, where multi-modal filters are attached to the aperture of multiple cameras, are an alternative solution. We advocate Generalized Assorted Camera (GAC) arrays, which mitigate the fabrication, cost and resolution challenges faced by GAP and enable high resolution, flexible (application-dependent), multi-modal imaging without making restrictive assumptions about scene geometry.

In this paper we make two important advances, we improve cross channel registration and highlight several applications of GAC arrays.

Cross-channel Registration: Traditional multi-view stereo methods assume brightness constancy across the images from each camera in the array, which is violated when the views observe different modalities. One solution is to repeat cameras for each modality which increases the number of cameras in the array. For example, Pelican Imaging [7], uses an array with 16 cameras to obtain 3 channels. Instead of this inefficient use of cameras, we propose a cross-channel multi-modal registration cost function that is based on aligning normalized gradients. This allows us to make efficient use of the cameras in our array enabling us to capture similar fidelity registration across channels with far fewer cameras.

Applications: We highlight the promise of GAC arrays on several applications such as (a) low light imaging, (b) post-capture refocusing, (c) skin perfusion imaging using RGB+NIR, and (d) hyperspectral imaging.

A. Motivating Applications

Camera Arrays for Smartphones: Manufacturers of smartphone cameras have begun using camera arrays to reduce the thickness of the camera module [7], [13]. By distributing channel measurements to separate sensors, the GAC array eliminates demosaicing artifacts, and the additional viewpoints enable post-capture digital refocusing (similar to the capabilities of the Lytro camera [14], [15]). Unlike the work presented in [7], where more than one array element measures the same channel, we demonstrate RGB fusion where each camera records a unique channel, significantly reducing the number of cameras in the array.

Flexible Application-dependent Imaging: The flexibility gained by using GAC arrays allows for targeted deployment tailored to the requirements of specific systems. Applications such as measuring blood oxygenation and heart rate [16],

Manuscript received April 15, 2014.

J. Holloway, K. Mitra, and A. Veeraraghavan are with the Department of Electrical and Computer Engineering, Rice University, Houston, TX, 77005, USA e-mail: jh25@rice.edu.

S. Koppal is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA

TABLE I
COMPARISON OF SPECTRAL ACQUISITION METHODS WITHIN THE VISIBLE SPECTRUM

				Camera Arrays			
	Time seq. Prog filter [8]	Time Seq. scanline [2]	Snapshot (GAP) [3], [4], [5]	Co-location [9], [10]	Repeated cameras* [7]	Planar assumption [11], [12]	**Ours (GAC)**
Cost	$$	$$	$$$	$$	$	$	**$**
Handles Motion	No	No	Yes	Yes	Yes	Yes	**Yes**
Spatial Resolution	Low	High	Low	Low [9] High [10]	Low	High	**High**
Complex Geometry	Yes	Yes	Yes	Yes	Yes	No	**Yes**
# Cams / # Spectra	$1/33$	$1/175$	$1/7$ [3], $1/8$ [4], $1/45$ [5]	$1/9$ [9], $8/8$ [10]	$16/3$	$5/5$ [11], $12/12$ [12]	$\mathbf{21/28}$
Light Throughput	Low	Low	Low	Low	High*	Low	**High**

*Repeated cameras is implemented only for RGB fusion

material classification [17], [18], viewing veins [19], monitoring the ripeness of fruit [20], shadow detection [21], and natural image retouching [22] only require information from a few specific channels to achieve the desired outcome. We demonstrate an implementation of a GAC array to naturally retouch portraits and to enhance the visibility of veins in arms.

Hyperspectral Imaging: Hyperspectral (HS) cameras provide spectral signatures that can be used for object classification [23], retinal imaging [24], [25], environmental imaging [26], and surveillance [27]. Current HS cameras have limited spatio-temporal resolution and must record numerous measurements of the scene in a serial manner. While this is acceptable for static scenes, such HS systems cannot be used to capture dynamic scenes. GAC arrays have been proposed to circumvent these limitations and record video with high spatial, temporal, and spectral resolution [10], [11], [12], [28]. Previous solutions involving GAC arrays suffer from poor signal-to-noise ratios and require either simple scene geometries [11], [12], complicated optics to co-locate images [10], or inflexible design parameters [28]. Table I compares the qualities of existing hyperspectral imaging systems. We propose a solution using a GAC array with inexpensive commodity broadband filters placed before each camera. The final hyperspectral image is recovered by demultiplexing the captured data.

B. Contributions

Specifically, we present the following contributions:

- Design and test a novel normalized gradient cost metric to compute point correspondences across color channels.
- Reduce the camera-to-channel ratio of GAC arrays by removing repetitive channel measurements without sacrificing spatial resolution or light throughput.
- Use our cross-channel registration technique to show that GAC arrays can be used for RGB fusion, post-capture refocusing and low light imaging.
- Incorporate side-band near-infrared information to perform application-specific tasks such as image retouching and enhancing the visibility of veins.
- Demonstrate that GAC arrays can be used for capturing hyperspectral video. Each camera captures multiplexed broadband spectral measurements, which increases the light throughput.

II. RELATED WORK

RGB Fusion: Conventional color cameras use a color filter array (CFA) followed by demosaicing to sample the visible spectrum of light. While the Bayer pattern [29] is the most common CFA, other patterns have been proposed to improve the rendered image quality [30], [31]. These approaches sacrifice spatial resolution in favor of higher spectral sampling.

To avoid reduction in spatial resolution, multiple sensors may be used to capture color channels independently. Foveon [32] created a camera which optically separates color channels through dispersing prisms onto three separate sensors. PiCam [7] uses a camera array and duplicates each color channel to allow intra-modality depth reconstruction. Our approach does not require complicated and thick optics or repetitious sampling of the same modality.

Hyperspectral Imaging: A common approach for hyperspectral imaging of static scenes is to capture a sequential series of images using narrowband filters [1]. Hyperspectral video can be realized by adding elements along the optical path such as a dynamic mirror [33], Lyot filter [34], aperture filters [35], novel mirror based hardware [9], diffraction grating [36], or a dispersing prism [37], [38], [39], to create smaller, low-resolution images on a single sensor. The spectral imaging system by Wagadarikar et al. [40] uses a coded aperture to make sparse measurements of a scene which is then recovered using compressive sensing techniques; cumbersome optics limits the flexibility of the system. Efforts to extend CFAs to enable hyperspectral imaging have led to reduced spatial resolution and demosaicing artifacts [3], [4], [5].

Hyperspectral GAC arrays, such as our design, can produce full resolution images but require image registration. Previous work has performed this registration assuming distant or planar scenes [12] or by using hand marked key points [11]. An alternate approach is to directly apply regular stereo metrics on filters with adjacent center wavelengths [28] and apply depth fusion for the final scene structure, but this imposes limits on array configuration and composition.

Multiplexing hyperspectral illumination is yet another approach which has proven successful [41], [42]. To achieve high quality reconstruction these methods temporally multiplex many illumination sources, each with a different spectral profile. Therefore, this approach suffers from motion artifacts

when imaging dynamic scenes. Our work achieves passive, high quality reconstructions for dynamic scenes and does not require *a priori* knowledge of the illumination spectrum.

Image Alignment: Fusing views from a camera array is difficult because each image is taken from a different vantage point, which introduces depth-dependent disparity. Traditional methods for aligning two images using stereo matching on rectified image pairs are quite mature (see [43] for a review of relevant algorithms). Robust multi-view stereo algorithms (see [44] for a survey of the literature) can be used to reconstruct a 3-D model of the scene for arrays of many cameras provided that each camera is operating in the same modality.

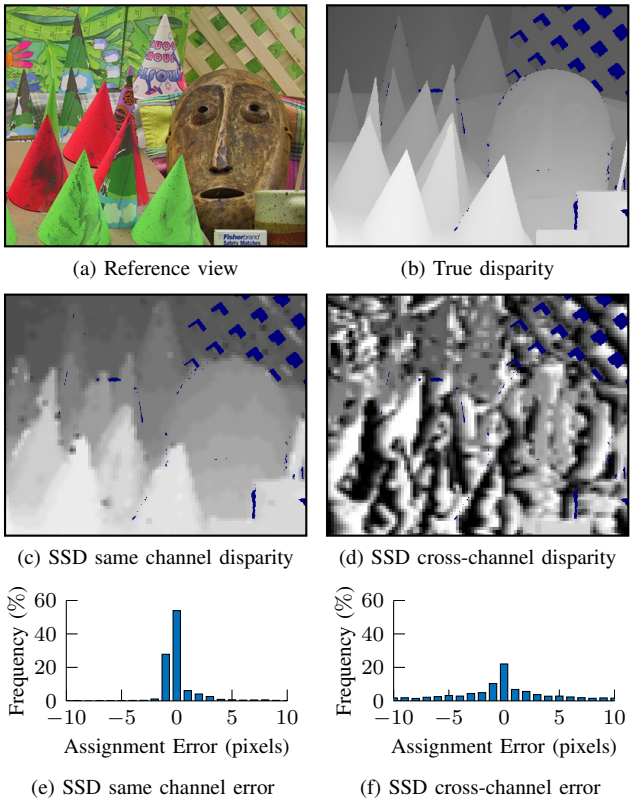
Image registration across modalities (e.g. cameras recording different spectral bands) is still difficult, due to the lack of shared information between views. Irani and Anandan [45] reduce registration complexity by assuming planar scenes and fuse images by matching edges within a global framework. Peng et al. [46] align linearly correlated images by minimizing the rank of the aligned image stack. Bando et al. [47] use a generalized normalized cross-correlation metric to register images captured in the red, green, and blue channels typical of color imaging. The use of mutual information as a registration metric was proposed by Woods et al. [48] and refined by Hill et al. [49] to align MRI and PET images, but the results suffer in the presence of image noise.

Gradient information has been successfully incorporated in traditional stereo-matching algorithms [50], [51], [52], [53], particularly to increase robustness to radiometric variation [54], [55]. A cross-channel stereo matching algorithm was proposed by Pinggera et al. [56] which matches dense histogram of gradient (HOG) descriptors between images taken in different spectral bands. In simulated results, matching using HOG descriptors performed well, but tests on real images lead to coarse and inaccurate depth maps. Using normalized gradients to compute global registration for medical images was proposed by Haber and Modersitzki [57] and extended Rühaak et al. [58] and by Hodneland et al. [59] among others. These works are intended to align medical images (e.g. MRI, PET, and CT scans) in a global framework with local deformation model. Such an approach is unable to align images with depth-based disparities where depth discontinuities are common. We propose an image alignment algorithm which extends normalized gradients to align cross-channel perspective images in the gradient domain, which we now explain in detail.

III. Image Alignment using Normalized Gradients

Computing the sum-of-squared differences (SSD) over a local window is one of the most popular methods for stereo matching. However, methods based on pixel intensities fail when used to compute correspondences across color channels.

To illustrate this, consider the rectified stereo dataset from [60] shown in the top row of Fig. 1. We consider three images from the dataset with a disparity search range of $1-27$ pixels. The distribution of erroneous disparity assignments shows that 87.8% of pixels in textured regions are within 1 pixel from their true value. However, if the input is not the intensity of the RGB image at each viewpoint, but is instead a unique color channel (red, green, and blue from the first, second, and third views respectively) the computed disparity map (Fig. 1(d)) is of significantly lower quality. Not only is the disparity map wrong, the distribution of error for textured regions is much wider than when matching within a channel. Only 39.1% of pixels are within 1 pixel of the correct disparity while over 25% are more than 6 pixels away from the correct disparity.



(a) Reference view (b) True disparity

(c) SSD same channel disparity (d) SSD cross-channel disparity

(e) SSD same channel error (f) SSD cross-channel error

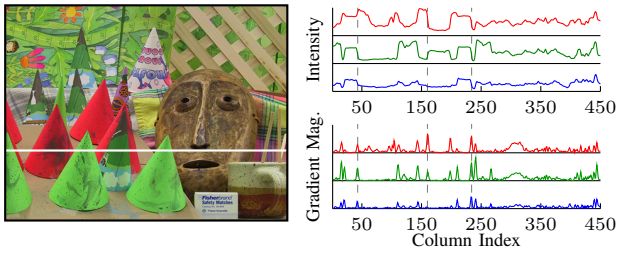
Fig. 1. **Failure of standard stereo algorithms to match across channels:** The sum-of-squared differences cost is used to find correspondence for (a) a scene from [60] with (b) known disparity. Matching is performed using (c) the luminance and (d) different color channels at each viewpoint. Accuracy is computed using the top 20% of pixels sorted by gradient magnitude. As seen in (e) SSD matching performs well within the same channel with 87.8% of correspondences accurate to within 1 pixel of the correct disparity; (f) performance across color channels suffers with an accuracy of only 39.1%.

A. Color Channel Gradients

While pixel intensities may vary between color channels, the location of edges are aligned as shown in Fig. 2. Thus, matching edges across color channels should yield a higher fidelity disparity map than matching pixel intensities.

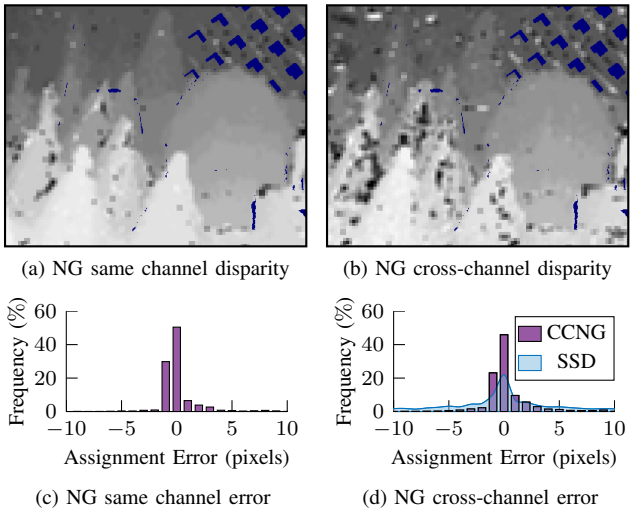
Given a hypothesized disparity, d, we generate an aligned stack of M images by shifting all of the views d pixels toward a reference image. Consider a single $8 \times 8 \times M$ image patch $I_{\{p,d\}}(u, v, \Lambda)$ centered at pixel p. Spatial dimensions are indexed by $u, v = 1, \ldots, 8$ while $\Lambda = 1, \ldots, M$ is the channel index. We compute gradient magnitudes in each channel by convolving the patch with a first-order centered difference kernel (after applying a Gaussian blur to reduce noise).

While the magnitude of image gradients serves as a good proxy for edge location, it is not expected that the gradient magnitudes are uniform across color channels. To account



(a) Reference view (b) Scanline profile

Fig. 2. **Edge alignment across color channels:** Consider one row of the reference image marked by the scanline in (a). Pixel intensity values and gradient magnitudes for the RGB channels are plotted in column (b) for the red, green, and blue channels. The intensity plot shows that rapid changes in intensity correlate across color channels, but the direction and magnitude is not consistent. Dashed gray lines on the plots show that gradient magnitudes tend to align across multiple color channels.



(a) NG same channel disparity (b) NG cross-channel disparity

(c) NG same channel error (d) NG cross-channel error

Fig. 3. **Improved cross-channel correspondence using normalized gradients:** The stereo-matching experiment from Fig. 1 is performed with the cross-channel normalized gradient cost (CCNG). When matching within the same channel (a) CCNG is comparable to SSD with (c) 86.7% of pixels within 1 pixel of the ground truth. (b) Unlike SSD, the CCNG cost is also able reliably to match across channels with (d) 78.6% accuracy. As seen in (d), cross-channel matching using CCNG has a much narrower distribution of disparity assignments than the SSD result.

for these differences we normalize magnitudes independently for each channel. Let $G_{u,\{p,d\}}$ be the gradient magnitudes of $I_{\{p,d\}}$ in the u-direction. The normalized gradients, $\hat{G}_{u,\{p,d\}}$ in the u-direction are then given by

$$\hat{G}_{u,\{p,d\}}(u,v,\Lambda) = \frac{G_{u,\{p,d\}}(u,v,\Lambda)}{\|G_{u,\{p,d\}}(\cdot,\cdot,\Lambda)\|}, \tag{1}$$

where the denominator is the L_2 norm of the patch (computed independently for each spectral channel). Following an analogous process in the v-direction yields $\hat{G}_{v,\{p,d\}}(u,v,\Lambda)$. Concatenating both measures produces $\hat{G}_{\{p,d\}}(u,v,\Lambda)$, which represents the normalized gradients of $I_{\{p,d\}}$. In Eq. (1), a problem arises for textureless image patches where gradient magnitudes are zero or nearly zero. To avoid numerical instabilities caused by dividing by values close to zero, a small epsilon is added to each channel in $G_{u,\{p,d\}}$ and $G_{v,\{p,d\}}$.

B. Cross-channel Normalized Gradient (CCNG) Cost Metric

If an image patch is to be consistent according to our gradient metric, then edges must be aligned across different channels. Therefore, our metric should favor edges that exist in multiple channels. We compute a cost $C'(p,d)$ for each pixel and disparity by collapsing along the channel dimension. By taking an element-wise product, edges that are consistently present in each channel will be magnified. Summing these responses and taking the M^{th} root produces our metric,

$$C'(p,d) = -\sqrt[M]{\sum_{u,v}\prod_{\Lambda=1}^{M}\hat{G}_{\{p,d\}}(u,v,\Lambda)}. \tag{2}$$

We further weight the cost for each disparity, $d = 1,\ldots,D$, relative to the average cost at p across all disparities. The final cost, $C(p,d)$, is given as

$$C(p,d) = \frac{C'(p,d)}{\frac{1}{D}\sum_{d=1}^{D}C'(p,d)}. \tag{3}$$

A disparity map E can be created by finding the disparity d which minimizes Eq. (3) for each pixel location p,

$$E(p) = \arg\min_{d} C(p,d). \tag{4}$$

The cross-channel normalized gradient (CCNG) cost from Eq. (3) is used to compute a disparity map for the stereo sequence in Fig. 1, and the CCNG results are shown in Fig. 3. When matching across different color channels, as shown in Fig. 3(d), our CCNG cost matches 78.6% of the pixels (recall that SSD matches 39.1% of the pixels).

C. Textureless Regions

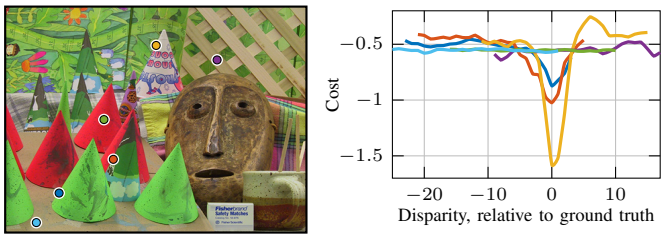
The proposed CCNG cost is well suited to match edges across channels. However, the accuracy reduces in textureless regions as shown in Fig. 4.

We address this issue with a two-step approach. First, we adaptively vary patch size to use large patches in flat regions and small patches near edges. Thus far the disparity maps were computed with 8×8 image patches, which is good for textured regions but poor for flat patches. Patch size is determined heuristically using the gradient magnitudes from a reference view. We choose the the smallest patch size n ($n = 8,\ 16$) which has 6 or more pixels with strong magnitudes in both the u and v directions. We define strong magnitudes to be above the 25th percentile of all magnitudes in the reference image. If the test is failed for both patch sizes, the patch is deemed a textureless region and n is set to 32.

Results of using the varying patch sizes to find correspondence across color channels can be seen in Fig. 5(a). Many of the holes present in Fig. 3(b) have been filled and the distribution of error is much tighter around 0, with 80.4% of all pixels within 1 pixel of the true disparity.

In addition to varying patch size, we also impose a smoothness penalty $S(p,d)$ when finding a depth map using Eq. (4),

$$E(p) = \arg\min_{d} C(p,d) + \mu S(p,d), \tag{5}$$



(a) Selected image patches (b) Normalized gradient (NG) cost

Fig. 4. **Stability of normalized gradients cost for textured regions:** The CCNG cost metric is designed to robustly align image edges across channels. Three patches are taken from textured and and three from smooth regions as shown in (a). The textured patches strongly favor the true disparity (0 relative to the ground truth) while the smooth patches have nearly uniform costs. The CCNG cost alone is insufficient to find correspondence in flat regions.

where μ is a weighting term. A simple but powerful smoothing term is to penalize differences in disparity assignments between neighboring patches. Let pixel q be in the neighborhood, $\mathcal{N}_p$, of pixel p in the aligned image stack. By assuming that all connections between neighboring pixels are equal, the smoothing term could taken as the L_1 distance,

$$S(p,d) = \sum_{q \in \mathcal{N}_p} |d - d_q|, \tag{6}$$

where d_q is the disparity assigned to the patch centered at q.

Using a global smoothing term such as the one in Eq. (6) improves accuracy in textureless regions but also leads to severe blurring at depth discontinuities. To combat this, we implement a bilateral smoothing term similar to one proposed by Yoon and Kweon [61]. We assume that neighboring patches with similar median intensity values belong to the same object and thus should have the same disparity. A non-uniform weighting is applied to the smoothing term based on this difference so that depth edges are preserved.

Let $\widetilde{\{\cdot\}}$ denote the median operator. We enforce spatial weighting using the difference between the median value of a patch in the reference view, $\widetilde{I}_{\{p,d\}}(u,v,r)$, and the median value of a neighboring patch, $\widetilde{I}_{\{q,d\}}(u,v,r)$,

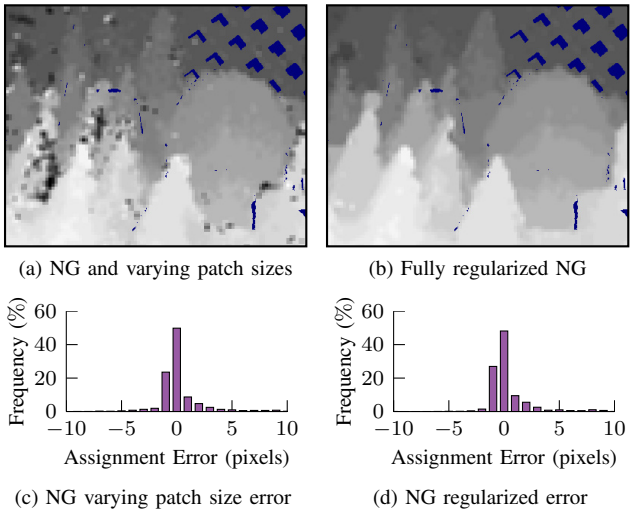
$$\delta(q) = |\widetilde{I}_{\{p,d\}}(u,v,r) - \widetilde{I}_{\{q,d\}}(u,v,r)|. \tag{7}$$

In Eq. (7), the difference is taken within a single reference channel r, thus the comparison of intensity values is valid.

In order to restrict smoothing to regions of low texture, we use a sigmoid function to weight the smoothing term. We set the center of the sigmoid's transition band, a, to 0.15 and the width, b, to 0.01 in our experiments. Combining the L_1 penalty in Eq. (6) and bilateral weighting from Eq. (7) gives the complete smoothing term,

$$S(p,d) = \sum_{q \in \mathcal{N}_p} |d - d_q| \frac{1}{1 + e^{\frac{\delta(q) - a}{b}}}. \tag{8}$$

We solve Eq. (5) with the smoothness given by Eq. (8) using α-expansion graph cuts with supplied code from [62], [63], [64]. The result of finding depth maps across color channels with Eq. (5) is shown in Fig. 5(b).



(a) NG and varying patch sizes (b) Fully regularized NG

(c) NG varying patch size error (d) NG regularized error

Fig. 5. **Accurate cross-channel correspondence in flat regions using a smoothness penalty:** Local correspondence metrics are unable compute disparity in flat regions. We add a varying patch size (a) to increase the likelihood of having gradients to match within a patch. A bilateral smoothness penalty is added to regularize the disparity assignment in Eq. (4). Flat patches are required to have a disparity similar to neighboring patches. Simply varying patch sizes yields 82.1% accuracy in textured regions and 80.4% overall. The smoothness term further improves accuracy with textured accuracy of 86.2% and overall accuracy of 87.6%.

D. Finding Depth Maps for Unrectified Data

The previous discussion described how the proposed CCNG algorithm is implemented when provided rectified data, where the correspondence search space is restricted to a single dimension. For unrectified multi-view stereo, we employ a plane-sweep stereo algorithm using our proposed CCNG cost to compute a depth map. Cameras in the array are assumed to be fixed relative to a reference camera view which may be arbitrarily chosen, though for convenience will be considered to coincide with a centrally located array element. It is further assumed that the internal and external camera calibration parameters are known.

The depth map is computed by first hypothesizing a set of depths, $\Omega \in (0\text{mm}, \infty)$, measured along the optical axis of the reference view for the plane sweep step increments. An aligned image stack is generated by assuming that all scene elements lie on a plane at distance $d \in \Omega$. Should the projected pixels fall in between pixel coordinates, the value is bicubically interpolated. A depth map is computed using Eq. (5) as before.

E. Cost Metric Validation Through Simulations

We simulate finding pixel correspondences between channels of the rectified stereo image sequence shown in Fig. 1(a). Images from the multi-view stereo sequence are taken from different viewpoints and objects at different scene depths have different disparities. The red channel (taken from the first viewpoint) is used as a reference.

In addition, we simulate using GAC for taking direct hyperspectral measurements using the "pompoms" hyperspectral data provided by Yasuma et al. [3]. The 550nm channel is chosen to be the reference view while the remaining channels are virtually positioned 5 pixels away from the reference view along equally distributed orientations.

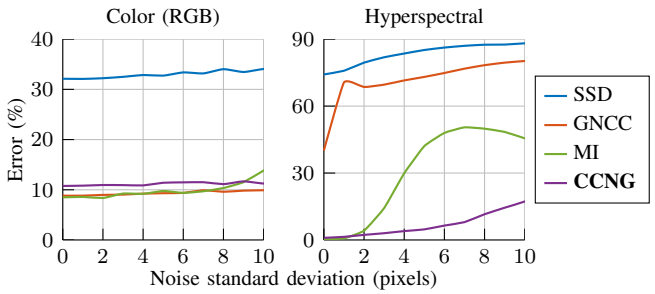


Fig. 6. **Correspondence accuracy for noisy images:** Disparity estimation performance as a function of image noise for color and hyperspectral images. The color stereo pair from Fig. 1 and the "pompoms" hyperspectral dataset from Yasuma et al. [3] are used for the simulations. The proposed NG cost degrades gracefully as image noise increases in both imaging regimes whereas the accuracy of GNCC and MI drop significantly for noisy hyperspectral images. Accuracy measurements are the average of 10 trials.

To demonstrate the efficacy of the proposed algorithm in noisy environments we simulate cross-channel matching for RGB and HS imaging in the presence of additive Gaussian white noise. We compare the computed correspondence of the proposed CCNG algorithm with the correspondences returned by using the SSD, generalized normalized cross correlation (GNCC), and mutual information (MI) metrics. The average accuracy over 10 trials is shown in Fig. 6. In both RGB (Fig. 6(a)) and HS matching (Fig. 6(b)) our proposed cost decays gracefully as more noise is added to the system. The GNCC metric is unable to extend to hyperspectral images and the performance of mutual information deteriorates with increasing levels of noise.

IV. Camera Calibration

To facilitate finding accurate correspondences, geometric constraints imposed by the array configuration are used to limit the search space along epipolar lines. This requires knowledge of the internal parameters of each camera and the relative position between cameras. Once a depth map has been computed, these parameters are used to warp outlying views to the viewpoint of the reference camera, forming an aligned stack of images.

Internal camera parameters are computed using the Caltech calibration toolbox [65]. The calibration is performed by imaging a planar checkerboard pattern in 15 different orientations with each camera in the array.

The external relationship between cameras is modeled as a rigid body transformation. Initial estimates of the rotation and translation are computed independently for each outlying camera-reference camera pair using the toolbox provided in [65]. The estimates for the internal and external parameters for each camera are then aggregated and optimized en masse using the sparse bundle adjustment toolbox [66]. The average reprojection error of a 3-D point onto any camera in the arrays presented in this paper is between $0.1 - 0.4$ pixels.

V. Application I: Consumer Imaging

Three important challenges in consumer imaging today are (a) color cross-talk and demosaicing, (b) low light imaging performance, and (c) full resolution light-field acquisition and post-capture refocusing. We show that a simple GAC array consisting of 4 cameras, each sensing a unique color channel (red, green, blue, and panchromatic as shown in 7(a)), can tackle all three of these challenges and provides superior performance compared to traditional cameras.

Four monochrome Point Grey Flea3 machine vision cameras (model FL3-U3-13E4M-C) are used to build a prototype GAC array. A Flea3 color camera having the same sensor with a CFA is placed adjacent to the reference camera and provides comparison images. Each Flea3 camera is 29mm wide and 29mm tall and, accounting for the width of the lenses, the baseline between neighboring cameras is 29.5mm. As the sensors in the cameras are responsive to wavelengths up to 1000nm, IR cut filters are attached to all four monochrome cameras. Three color filters (R, G, B) from Edmund Optics are mounted to the lenses in front of three of the cameras. It is important to note that the spectral profile of these external filters are different than the CFA of the color camera which may result in some differences in appearance.

For each scene all of the cameras are set to the largest aperture which permits the entire scene to be in focus. Once the cameras have been set, we perform camera calibration as outlined in Sec. IV. The images presented in this paper have been gamma corrected for display purposes.

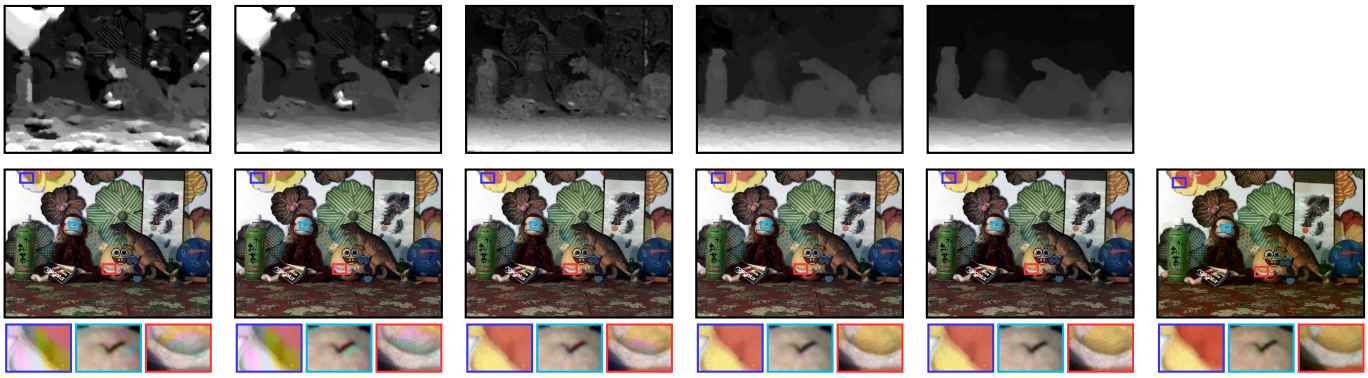
Color Imaging Without Color Cross-talk: Traditional color cameras use a CFA to interleave filters with different pass bands on neighboring pixels to sample the visible spectrum. As a consequence, cross-talk between adjacent pixels leads to color desaturation and the need for color correction during post-processing [67]. Furthermore, interpolation artifacts are introduced when demosaicing the CFA to compute the tristimulus values for each pixel. In our GAC array each camera is sensitive to a single channel, eliminating color cross-talk and the need for demosaicing; however, the four channels must be fused to create a color image.

We first compute the depth map using our normalized gradient cost metric as described in Sec. III. The camera with the panchromatic (Y) channel is used as the reference camera and the red, green, and blue channels are aligned according to the depth map. After alignment, we avoid a direct fusion of the RGBY channels because errors in the estimated depth map could lead to a degradation in image quality. Instead, we reconstruct the color image in the YCbCr color space using the luminance values from the unfiltered Y channel and computing chrominance from the aligned R, G, and B channels. We smooth the chrominance channels with a 5×5 Gaussian convolution kernel to reduce color artifacts which arise from warping errors. The final color image is found by converting to the RGB color space.

In Fig. 7, we show a comparison between (e) the color image recovered using our array and (f) the image captured from the color Flea3 camera. The comparison color image is created by demosaicing the raw image offline using the method suggested in [68]. Outsets show that unlike our image, demosaicing artifacts are present in the image captured using a color filter array. During image acquisition, exposure times are allowed to vary between cameras, but the analog gain

(a) GAC array (b) Input images (c) Direct fusion (d) Depth map (e) RGB image from GAC array (f) Bayer RGB image

Fig. 7. **RGB image capture and fusion:** (a) Three monochrome cameras in a 2×2 array record red, green, and blue spectral channels while the fourth captures the luminance (Y). The captured images are shown in (b). (c) Directly layering the R, G, and B channels results in significant artifacts. (d) A depth map, which is used to align the four viewpoints, is found using our CCNG metric. (e) Color images are fused in the YCbCr color space using luminance values from the Y channel and chrominance from the aligned R, G, and B channels. (f) A fifth camera having the same sensor with a Bayer mosaic is used to capture a comparison color image. Our GAC array is able to produce a color image with comparable quality to a Bayer RGB image. The color filters used in the array and the Bayer filters have different spectral profiles, object color may differ between the images. Please zoom in on the PDF for image details.



(a) Fusion using SAD (b) Fusion using SSD (c) Fusion using MI (d) Fusion using GNCC (e) Fusion using CCNG (f) Bayer RGB image

Fig. 8. **Depth estimation for RGB fusion:** Traditional stereo matching methods such as (a) the sum-of-absolute differences (SAD) and (b) SSD produce inaccurate depth maps (top row) which lead to color artifacts in the fused color image (bottom row). Cross-channel methods: (c) mutual information, (d) generalized NCC, and (e) our normalized gradient algorithm yield better depth estimates, though mutual information is inferior to GNCC and CCNG. A color image captured with a Bayer mosaic is presented in (f) for comparison. The color image outsets exhibit noticeable artifacts in (a), (b), and (c).

is fixed at 3dB. The camera array, Fig. 7(a), captures four channels as shown in Fig. 7(b). Due to the wide baseline of the cameras, significant parallax is introduced and a color image cannot be recovered by simply fusing the red, green, and blue channels into a single image (Fig. 7(c)). Instead, a depth map is computed using our proposed algorithm, Fig. 7(d), and the views are aligned using this depth map.

Comparison With State of the Art: Accurate depth maps are a crucial component for recovering a color image using our GAC; we compare the performance of our CCNG algorithm with stereo matching algorithms in Fig. 8. Traditional intensity-based stereo matching such as (a) sum-of-absolute differences (SAD) and (b) SSD produce inaccurate depth maps and the corresponding color images are unappealing. Algorithms designed for finding cross-channel correspondence perform much better than intensity-based metrics. Mutual information (c) produces a depth map of decent quality, though errors persist into the final color image. Both (d) generalized NCC and (e) our CCNG metric yield high-quality depth maps. An image captured with a conventional Bayer pattern is included in Fig. 8(f) for comparison. Overall, our method produces RGB images that are comparable in quality to conventional color cameras which employ a Bayer pattern as shown in detail in Fig. 7(e) and (f).

Low Light Imaging: Demosaicing artifacts are not the only disadvantage of using a traditional color cameras. Traditional cameras have a global exposure, even though light throughput in individual color channels is not uniform. In low light environments, this necessitates larger analog gains in channels with low light throughput which amplifies image noise. Using the proposed GAC array, the greater light throughput of the Y effectively reduces image noise.

In Fig. 9, we consider a typical low light imaging scenario; a diffuse incandescent light source is used to illuminate an indoor scene. All of the cameras have an exposure duration of 125ms, a (generous) limit of a photographer's ability to take images free of motion blur using a hand-held camera. Gain levels for each channel are set independently, the Y channel has the lowest gain of 8dB while the gain of red (11dB), green (13dB) and blue (15dB) channels is significantly higher. The color camera used for comparison requires a global value for the analog gain which is set to 13dB in this experiment. Our camera array records an image with far less noise (Fig. 9(a)) than the color camera (Fig. 9(b)). Notice that fine details around the eyes are retained by the GAC array while the noise in the Bayer image obliterates nearly all detail around the eye. Fine edge structure is also preserved using the GAC array as shown in the outset of the ball cap.

(a) RGB image from GAC array (b) Noisy Bayer RGB image

Fig. 9. **RGB fusion for improved low-light imaging:** By allowing one camera to remain unfiltered in the GAC array, higher light throughput is achieved which improves image quality in low-light environments. Exposure durations were fixed at 125ms for all cameras. (a) The image captured with the GAC. The light throughput of the Y channel greatly reduces noise in the image compared to (b) traditional color image capture using a Bayer mosaic. The image from the GAC array retains fine image detail around the dog's eyes which is lost in the Bayer image on the right.

(a) Bayer RGB image (b) Enhanced image from GAC array

Fig. 11. **NIR fusion for face smoothing:** Near-infrared light penetrates deeper into skin than visible light. The resulting subsurface scattering effectively smooths away surface blemishes. Combining NIR information with RGB images allows for "natural" image retouching for human skin. (a) A color image captured with a Bayer CFA and (b) the NIR enhanced image recorded using our GAC array. Notice the removal or softening of freckles and blemishes.

Post-capture Refocusing: A recent trend in consumer imaging is to coarsely sample the light-field of a scene which allows for digital refocusing after capturing an image.Current light-field cameras capture images with low spatial resolution and must super resolve a high resolution image. Our GAC array enables post-capture refocusing on color images with spatial resolution equal to that of the cameras in the array.

Recovered RGB images, the resulting depth map, and refocused images for two scenes are shown in Fig. 10. The scenes shown in Fig. 10(a) were captured with an apertures of $f/4$. Using the depth maps the in-focus images were digitally refocused using a synthetic aperture of $f/1$ to create the images in Fig. 10(b)-(d). The refocused images show the effect of changing the focal plane from near, to middle, to far focus in each of the scenes. Please view the image digitally and zoom in to see details.

VI. Application II: Skin Perfusion Imaging

The flexibility of GAC arrays allows additional channels to be incorporated into the array by simply adding additional cameras. In this way GAC arrays can be leveraged to augment RGB color images with near infrared (NIR) information. NIR

(a) Bayer RGB image

(b) Enhanced RGB image from GAC array

Fig. 12. **NIR fusion for vein viewing:** By adjusting the weight of the NIR channel, an effective vein viewing system can be devised. (a) Vasculature can be difficult to see with unaided traditional RGB cameras. (b) With enhanced RGB+NIR images from our GAC array, veins near the surface of the skin are readily visible.

light penetrates deeper into human skin than light in the visible spectrum [22], which in turn can be used to naturally retouch color images by smoothing blemishes or to help medical professionals easily see veins in the arm. By adding a fifth monochrome camera to the GAC array described in Sec. V, the R, G, B, Y, and NIR channels can be captured simultaneously. An IR pass filter is affixed to the lens of the fifth camera. Following image capture and alignment, a further processing step is performed before generating the final RGB composite image.

We follow the technique of Susstrunk et al. [22], to incorporate NIR information into the Y channel via bilateral filtering. The bilateral filter [69] separates an image into its base and detail components. The two components may loosely be considered to represent localized low- and high-frequency content, respectively. Using the fast bilateral filter approximation suggested by Paris and Durand [70], the Y and NIR channels are decomposed into their base and detail representations. Some, or all, of the detail of the Y channel is replaced with the detail in the NIR image using a weighting parameter α,

$$\mathrm{Y}_{\text{fused}} = \mathrm{Y}_{\text{base}} + ((1-\alpha)\mathrm{Y}_{\text{detail}} + \alpha\mathrm{NIR}_{\text{detail}})\,. \qquad (9)$$

Following the fusion of the Y and NIR channels, the chrominance from the R, G, and B channels are used to recover the composite color image as before.

Natural Retouching for Portraiture: Imperfections in the skin on the face, such as freckles, blemishes, and wrinkles, are less noticeable when viewed in NIR. Using the five element GAC array, NIR information is fused with the Y channel to smooth facial imperfections.

We show the effects of this "natural" image retouching in Fig. 11. Shown in Fig. 11 is (a) an image captured with a conventional Bayer pattern and (b) an enhanced image captured with the GAC array ($\alpha = 0.75$). Many of the undesirable surface skin features have been removed and strong blemishes have been softened significantly.

Vein Viewing: In addition to applications in portrait photography, the deeper skin penetration of NIR light can be used in medical applications. For example, veins near the surface

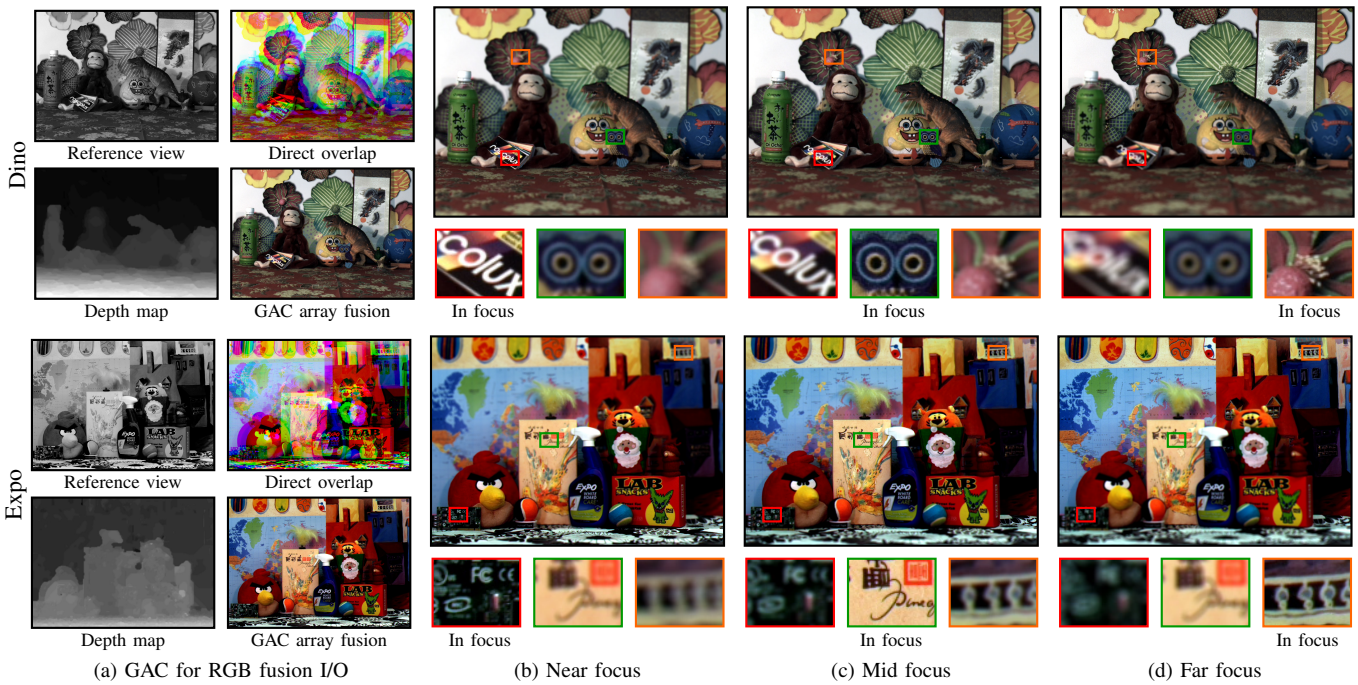


Fig. 10. **GAC array RGB fusion and post-capture refocusing:** (a) The depth map computed for RGB fusion may be used to digitally refocus the image post-capture. (b)-(d) The scenes are refocused with a synthetic aperture of $f/1$ and focal plane positioned in the near-, middle-, and background. Zoomed-in outsets of objects in the near- (red), middle- (green), and background (orange) show the effect of changing the focal plane position. Please view digitally.

of the arm may be difficult to see in the visible spectrum but are prominent in NIR. Using the GAC array it is possible to enhance RGB images to make veins more prominent while retaining color information to provide familiar visual cues to the user. An example of enhanced vein viewing is shown in Fig. 12. The top image is taken with a conventional RGB camera and the bottom image, captured with our GAC array, has additional NIR information ($\alpha = 1$). Veins are more pronounced and the intricate pathways are easily seen in the image acquired with the GAC array.

VII. Application III: Hyperspectral Imaging

Commercial hyperspectral cameras are able to achieve high spectral resolution by trading off spatial or temporal resolution. Furthermore, they suffer from poor light throughput which requires long exposure times and high analog gain which leads to low SNR in individual spectral bands. We demonstrate how a GAC array can be used to recover hyperspectral video with 10nm resolution, high light throughput, and no loss of spatial resolution. This represents a 10x improvement in spectral resolution over RGB cameras and at a fraction of the cost of currently available HS cameras.

A straight-forward GAC implementation would be to place narrow-band filters in front of each camera element. However, the use of narrow-band filters results in a low light throughput. We propose a multiplexing strategy, similar to [71] and [72] to improve SNR. We demonstrate a proof-of-concept system using the 5×5 ProFUSION color camera array from Point Grey as the imaging platform and Roscolux broadband filters. 21 cameras were selected to be in the GAC array (4 were unused due to quality issues). A Headwall Photonics Micro-Hyperspec VNIR imaging spectrometer was used to capture

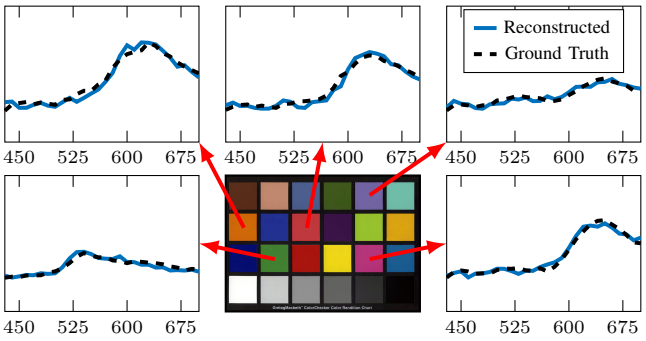
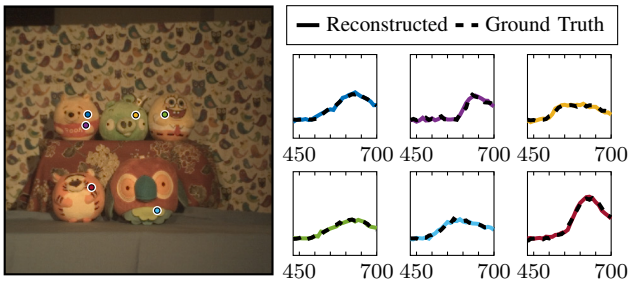


Fig. 13. **Recovering hyperspectral measurements using calibration charts:** Spectral profiles of a color calibration target are recovered using a dictionary of known hyperspectral measurements. Ground truth data (dashed, black plots) was captured with a HS camera and the recovered spectrum (solid, blue plots) using the proposed GAC. The average SNR of the reconstructed curves is 23.7 dB. Ground truth was captured using Headwall Photonics Micro-Hyperspec VNIR imaging spectrometer.

spectral measurements that are then used as ground truth for evaluating the performance of our GAC hyperspectral estimates in Fig. 13, Fig. 14, and Fig. 16.

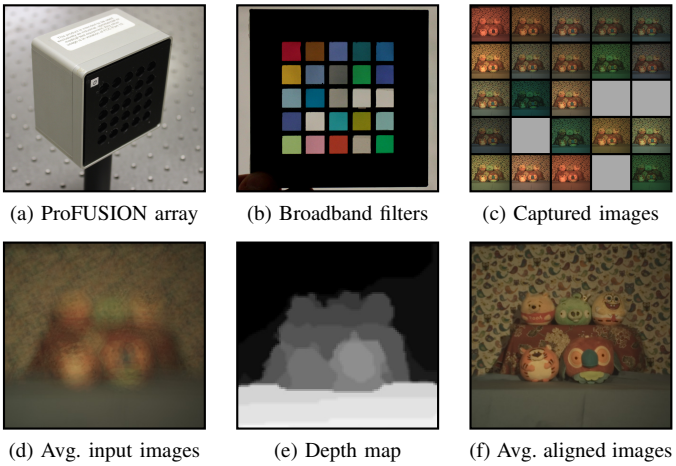
Recovering Hyperspectral Data: We use a set of broad-band filters, each with high transmittance, and whose transmission power for a particular discrete wavelength λ is $F_c(\lambda)$ where c is the camera index (since each camera has only one filter). Post capture, we demultiplex the intensities to produce the scene spectral reflectance $R(\lambda)$.

If the spectral response of the Bayer pattern is $B_f(\lambda)$, where f is the index of the red, green, and blue filters, then the intensity measurement I of this scene point at camera c and



(a) GAC array reference view (b) Recovered spectral profiles

Fig. 14. **Hyperspectral recovery for real scenes:** Hyperspectral measurements computed from the aligned image stack using 82 known spectral profiles. The average SNR of the reconstruction is 26.7 dB.



(a) ProFUSION array (b) Broadband filters (c) Captured images

(d) Avg. input images (e) Depth map (f) Avg. aligned images

Fig. 15. **Multiplexed hyperspectral imaging:** (a) A 5×5 camera array is outfit with (b) broadband optical filters. (c) 21 cameras record multiplexed images. (d) Direct fusion of the images leads to misaligned images, so (e) a depth map is found using the CCNG cost. (f) The average of the aligned images shows the quality of the depth map.

for color channel f is:

$$I_{cf} = \sum_{\lambda} B_f(\lambda) F_c(\lambda) J(\lambda) R(\lambda), \tag{10}$$

where J is the (unknown) spectral response of the illumination. By demosaicing the raw Bayer image data, we can obtain all three color channels at every pixel. Therefore, if there are $C = 21$ cameras and 3 color channels we get a total of $M = 3C = 63$ measurements at each scene point.

If we combine the known filter and Bayer spectral responses into an effective camera filter F_m ($m = 1, \ldots, M$) and include the illumination J into the reflectance R, summing over the S discrete spectral bins gives,

$$I_m = \sum_{i=1}^{S} Fm(\lambda_i) R(\lambda_i), \tag{11}$$

which in vector notation is $I = \mathbf{F}R$. We propose a calibration step that allows us to recover the unknown spectral reflectance of a scene point. Our method does not require recovery of the $M \times S$ mixing matrix $\mathbf{F}$.

We first take images of a color chart and capture an $M \times N$ matrix $\mathbf{I}$, where N is the number of color chart squares, such that $\mathbf{I} = \mathbf{FR}$. True spectral profiles are recorded using the Headwall HS camera and stored in a $S \times N$ matrix $\mathbf{T}$. Given a length-M vector of measurements X for some scene point obtained by the imaging process of Eq. (11), we learn a sparse set of coefficients ω that reconstruct X using $\mathbf{I}$ as a dictionary,

$$\arg\min_{\omega} \|X - \mathbf{I}\omega\| \quad \text{where} \quad \|\omega\|_0 \leq \Delta, \tag{12}$$

where Δ is a threshold on sparsity. Assuming linearity in the camera responses, the weighting coefficients ω can be used to reconstruct the unknown reflectance of X, $\hat{R}$, using T,

$$\hat{R} = \mathbf{T}\omega. \tag{13}$$

Therefore, given a calibration color-chart image, taken under the same unknown illumination J as the scene, and given the measurements X at a scene point, we find the sparse weights ω to reconstruct the measurements X from the multiplexed color chart image data $\mathbf{I}$. We use the *same* weights directly on the known spectral responses of the color chart to recover the spectral response $\hat{R}$ of the scene point. We choose the filters from a Roscolux booklet using a greedy algorithm to minimize the condition number of $\mathbf{F}$. To begin, we select the single filter which yields the smallest condition number and iteratively choose additional filters in this manner until we have the 21 filters that are used in the array.

Static Scene Recovery: We demonstrate our method first with a static calibration scene (Fig. 13) consisting of a Gretag Macbeth digital SG color chart with 140 squares and 28 spectra levels. Repetitive color samples and the 24 classic color checkers are not considered, leaving $N = 58$ unique color samples. Limited energy in the blue end of the visible spectrum restricted accurate spectral measurements to the range 430nm-700nm. We then took a second image of the classic color checker with 24 color swatches and recovered the squares' true spectra using Eq. (12) and Eq. (13). Fig. 13 shows the real and recovered spectra for a selection of the squares. The spectral response of each square was recovered with an average SNR of 23.7dB.

In the second experiment, we imaged the static scene shown in Fig. 14. The digital SG color checker is again used for calibration, and the recovered spectral response of selected scene points is shown in Fig. 14. Fig. 15 gives an overview of the imaging process and shows the need for image alignment.

Hyperspectral Video: An advantage of hyperspectral GACs is their ability to capture video data. Using the same procedure as Fig. 14, 40 frames of a dynamic scene were captured at 15 fps. In the video, two arms move throughout the FOV of the array. Camera elements within the array are synchronized and share the same exposure and gain settings. 5 frames of the video and the spectral reconstructions for two manually specified scene points are shown in Fig. 16.

Reconstruction Quality: In addition to providing color checker calibration data, the Micro-Hyperspec imager is used to provide ground truth spectral measurements. Ground truth spectral data acquisition took 10 seconds per scene, precluding the use of the Micro-Hyperspec for video capture. Using our GAC we were able to capture the scene in 30ms and recovered the spectral profiles with an average reconstruction SNR of 25.7dB for the scene points in Fig. 14 and 27.8dB for the

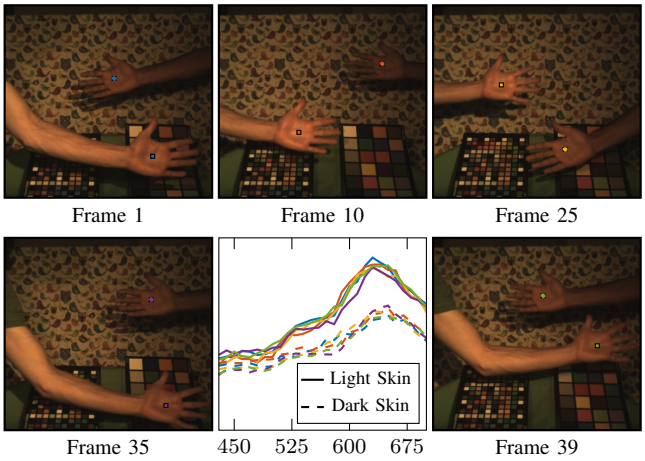


Fig. 16. **Hyperspectral Video using a GAC:** Average recon SNR is 27.8dB relative to ground truth collected with static scene elements. Aside from the difference in magnitude, the light skin patches have a peak at 630nm, while the dark skin patches peak at 650nm.

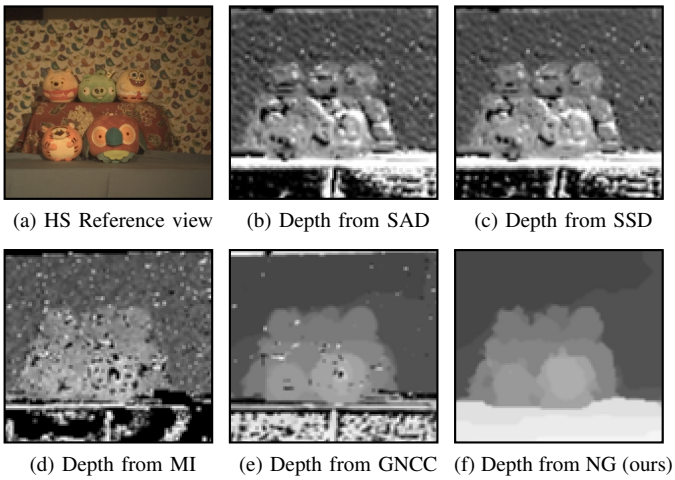


Fig. 17. **Comparison of Depth Estimation on Multispectral Data:** A depth map for the scene in (a) is computed using (b) SAD, (c) SSD, (d) MI, (e) GNCC, and (f) our NG cost. As with RGBY fusion (Fig. 8) the intensity-based metrics and mutual information (b)-(d) yield inaccurate depth maps. Generalized NCC (e) has spurious errors in the depth map, even within textured regions. The proposed method (f) offers superior depth estimates.

scene points in the video. Ground truth data for the video was collected with both hands resting stationary on the table.

Comparison With State of the Art Stereo Matching: Fig. 17 compares the performance of various stereo-matching algorithms. As with matching across color channels, the intensity based metrics and mutual information (b)-(d) yield poor depth estimates. Generalized NCC also returns a noisy depth map while our proposed CCNG algorithm is able to return a high quality depth map.

VIII. Conclusion

In this paper we presented a framework for using generalized assorted cameras and demonstrated examples of GAC arrays used to capture color images with four cameras, augment RGB images with NIR information, and capture multiplexed hyperspectral video. Robust cross-channel point correspondence using a normalized gradient cost allows all

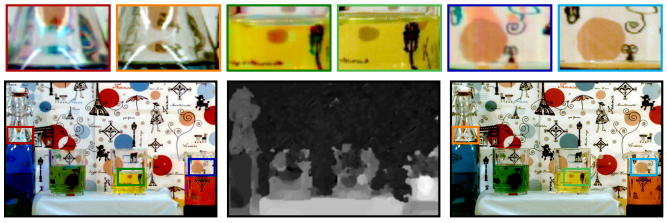


Fig. 18. **Failure for specular materials:** Like many stereo correspondence algorithms, our proposed method assumes scene elements are diffuse. When this assumption is violated, the depth map is resolved inaccurately, leading to garish image artifacts as shown in (a) the recovered image and (b) depth map. (c) An image captured with a Bayer CFA is provided for comparison.

of the camera views to be warped to an arbitrary viewpoint within the camera array. Moreover, the simultaneous capture of scene information allows for imaging of dynamic scenes and video capture.

Advantages: Generalized assorted camera arrays offer the following advantages compared to traditional imaging systems: (1) Adaptability: Cameras in the array can be controlled independently or as a whole to adapt to novel lighting and scene changes without compromising imaging quality. (2) Flexibility: By changing the filter arrangement in front of each camera it is easy to configure the camera array to fit application-specific imaging domains. (3) Increased resolution and SNR for hyperspectral imaging: Simultaneously capturing many multiplexed views with the camera array allows for greater temporal resolution in recorded images and video as well as higher light throughput.

Limitations: Camera arrays are not suitable for all imaging situations. GAC arrays will suffer from many of the same limitations as stereo camera systems. In addition to well-known problems with accurate depth estimation for thin structures at depth boundaries, GAC arrays face two main limitations. The first is proximity, and objects that are close to the sensors will not be present in the FOV of all cameras. Therefore, there exists a minimum operating distance which is a function of the camera array baseline. Adding more elements to the array (or using larger cameras) increases the minimum operating distance. The second limitation is common among stereo matching algorithms, namely that point correspondences for scenes with specularities and refractive media are not faithfully recovered as seen in Fig. 18. The figure shows significant errors in the recovered color image when compared with an image from a conventional color camera. Outsets highlight a few failure cases due to specularity on the glass bottle and refractions due to water and glass.

Acknowledgment

Jason Holloway, Kaushik Mitra and Ashok Veeraraghavan acknowledge support through NSF Grants NSF-IIS:1116718 and NSF-CCF:1117939. Sanjeev Koppal would like to thank Umit Batur at Texas Instruments for all of his helpful insights and fruitful discussions.

REFERENCES

[1] N. Gat, "Imaging spectroscopy using tunable filters: a review," in *Proceedings of the SPIE*, vol. 4056, 2000, pp. 50–64.

[2] H. Photonics, "Hyperspec® vs imaging spectrographs," *product brochure*, 2009.

[3] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum," *Image Processing, IEEE Transactions on*, vol. 19, no. 9, pp. 2241–2253, 2010.

[4] P. E. Buchsbaum and M. J. Morris, "Method for making monolithic patterned dichroic filter detector arrays for spectroscopic imaging," Oct. 28 2003, PixelTeq US Patent 6,638,668.

[5] B. Geelen, N. Tack, and A. Lambrechts, "A compact snapshot multispectral imager with a monolithically integrated per-pixel filter mosaic," in *SPIE MOEMS-MEMS*. International Society for Optics and Photonics, 2014, pp. 89 740L–89 740L.

[6] S. Sivaramakrishnan, A. Wang, P. Gill, and A. Molnar, "Enhanced angle sensitive pixels for light field imaging," in *Electron Devices Meeting (IEDM), 2011 IEEE International*, Dec 2011, pp. 8.6.1–8.6.4.

[7] K. Venkataraman, D. Lelescu, J. Duparré, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar, "Picam: an ultra-thin high performance monolithic camera array," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, p. 166, 2013.

[8] VariSpec™, "Liquid crystal tunable filter users manual," *Cambridge Research and Instrumentation*.

[9] A. Manakov, J. F. Restrepo, O. Klehm, R. Hegedüs, E. Eisemann, H.-P. Seidel, and I. Ihrke, "A reconfigurable camera add-on for high dynamic range, multispectral, polarization, and light-field imaging," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, p. 47, 2013.

[10] M. McGuire, W. Matusik, H. Pfister, B. Chen, J. F. Hughes, and S. K. Nayar, "Optical splitting trees for high-precision monocular imaging," *Computer Graphics and Applications, IEEE*, vol. 27, no. 2, pp. 32–42, 2007.

[11] D. Lau and R. Yang, "Real-time multispectral color video synthesis using an array of commodity cameras," *Real-Time Imaging*, vol. 11, no. 2, pp. 109–116, 2005.

[12] S. Heinhold (Tetracam Inc), "Tetracam Multi Camera Array (MCA) Installation and Operation," pp. 1–11, 2007. [Online]. Available: http://www.tetracam.com/PDFs/Mini-MCA%20Users%20Guide%20V2.3.pdf

[13] "Heptagon Advanced Micro Optics," http://www.hptg.com/products/imaging, 2014, [Online; accessed 31-March-2014].

[14] T. Georgiev, Z. Yu, A. Lumsdaine, and S. Goma, "Lytro camera technology: theory, algorithms, performance analysis," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2013, pp. 86 671J–86 671J.

[15] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Stanford University Computer Science Tech Report CSTR 2005-02*, 2005.

[16] L. F. C. Martinez, G. Paez, and M. Strojnik, "Optimal wavelength selection for noncontact reflection photoplethysmography," in *International Commission for Optics (ICO 22)*. International Society for Optics and Photonics, 2011, pp. 801 191–801 191.

[17] N. Salamati, C. Fredembach, and S. Susstrunk, "Material classification using color and NIR images," in *Color and Imaging Conference*. Society for Imaging Science and Technology, 2009, pp. 216–222.

[18] N. Salamati, D. Larlus, G. Csurka, and S. Süsstrunk, "Semantic image segmentation using visible and near-infrared channels," in *Computer Vision–ECCV 2012. Workshops and Demonstrations*. Springer, 2012, pp. 461–471.

[19] V. Paquit, J. R. Price, F. Mériaudeau, K. W. Tobin Jr, and T. L. Ferrell, "Combining near-infrared illuminants to optimize venous imaging," in *Medical Imaging*. International Society for Optics and Photonics, 2007, pp. 65 090H–65 090H.

[20] C. Camps and D. Christen, "Non-destructive assessment of apricot fruit quality by portable visible-near infrared spectroscopy," *LWT-Food Science and Technology*, vol. 42, no. 6, pp. 1125–1131, 2009.

[21] D. Rufenacht, C. Fredembach, and S. Susstrunk, "Automatic and accurate shadow detection using near-infrared information," 2012.

[22] S. Süsstrunk, C. Fredembach, and D. Tamburrino, "Automatic skin enhancement with visible and near-infrared image fusion," in *Proceedings of the international conference on Multimedia*. ACM, 2010, pp. 1693–1696.

[23] C.-I. Chang, *Hyperspectral data exploitation: theory and applications*. John Wiley & Sons, 2007.

[24] B. Khoobehi, J. M. Beach, and H. Kawano, "Hyperspectral imaging for measurement of oxygen saturation in the optic nerve head," *Investigative ophthalmology & visual science*, vol. 45, no. 5, pp. 1464–1472, 2004.

[25] W. R. Johnson, M. Humayun, G. Bearman, D. W. Wilson, and W. Fink, "Snapshot hyperspectral imaging in ophthalmology," *Journal of biomedical optics*, vol. 12, no. 1, pp. 014 036–014 036, 2007.

[26] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *International journal of Remote sensing*, vol. 28, no. 5, pp. 823–870, 2007.

[27] A. R. Harvey, J. E. Beale, A. H. Greenaway, T. J. Hanlon, and J. W. Williams, "Technology options for imaging spectrometry," in *International Symposium on Optical Science and Technology*. International Society for Optics and Photonics, 2000, pp. 13–24.

[28] I. Gheta, S. Höfer, M. Heizmann, and J. Beyerer, "A novel region-based approach for the fusion of combined stereo and spectral series," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2010, pp. 75 380G–75 380G.

[29] B. E. Bayer, "Color imaging array," Jul. 20 1976, US Patent 3,971,065.

[30] X. Li, B. Gunturk, and L. Zhang, "Image demosaicing: A systematic survey," in *Electronic Imaging 2008*. International Society for Optics and Photonics, 2008, pp. 68 221J–68 221J.

[31] K. Hirakawa and P. J. Wolfe, "Spatio-spectral color filter array design for optimal image recovery," *Image Processing, IEEE Transactions on*, vol. 17, no. 10, pp. 1876–1890, 2008.

[32] R. Lyon *et al.*, "Prism-based color separation for professional digital photography," in *IS and TS Pics Conference*. Society for Image Science & Technology, 2000, pp. 50–54.

[33] R. Bernstein, "Digital image processing of earth observation sensor data," *IBM Journal of Research and Development*, vol. 20, no. 1, pp. 40–57, 1976.

[34] A. R. Harvey, D. W. Fletcher-holmes, A. Gorman, K. Altenbach, J. Arlt, and N. D. Read, "Spectral Imaging in a Snapshot," *Proceedings of the SPIE*, vol. 5694, pp. 110–119, 2005.

[35] R. Horstmeyer, G. Euliss, R. Athale, and M. Levoy, "Flexible multimodal camera using a light field architecture," in *Computational Photography (ICCP), 2009 IEEE International Conference on*. IEEE, 2009, pp. 1–8.

[36] R. Habel, M. Kudenov, and M. Wimmer, "Practical spectral photography," in *Computer Graphics Forum*, vol. 31, no. 2pt2. Wiley Online Library, 2012, pp. 449–458.

[37] H. Du, X. Tong, X. Cao, and S. Lin, "A Prism-Mask System for Multispectral Video Acquisition." *IEEE transactions on pattern analysis and machine intelligence*, pp. 1–8, Apr. 2011.

[38] C. Ma, X. Cao, R. Wu, and Q. Dai, "Content-adaptive high-resolution hyperspectral video acquisition with a hybrid camera system," *Optics Letters*, vol. 39, no. 4, 2013.

[39] C. Ma, X. Cao, X. Tong, Q. Dai, and S. Lin, "Acquisition of high spatial and spectral resolution video with a hybrid camera system," *International Journal of Computer Vision*, pp. 1–15, 2013.

[40] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Applied optics*, vol. 47, no. 10, pp. B44–B51, 2008.

[41] C. Chi, H. Yoo, and M. Ben-Ezra, "Multi-spectral imaging by optimized wide band illumination," *International Journal of Computer Vision*, vol. 86, no. 2-3, pp. 140–151, 2010.

[42] A. Lam, A. Subpa-Asa, I. Sato, T. Okabe, and Y. Sato, "Spectral imaging using basis lights," 2013.

[43] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

[44] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1. IEEE, 2006, pp. 519–528.

[45] M. Irani and P. Anandan, "Robust multi-sensor image alignment," *Computer Vision, 1998. Sixth International Conference on*, pp. 959–966, 1998.

[46] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 11, pp. 2233–2246, 2012.

[47] Y. Bando, B.-Y. Chen, and T. Nishita, "Extracting depth and matte using a color-filtered aperture," in *ACM Transactions on Graphics (TOG)*, vol. 27. ACM, 2008, p. 134.

[48] R. P. Woods, S. R. Cherry, and J. C. Mazziotta, "Rapid automated algorithm for aligning and reslicing pet images," *Journal of computer assisted tomography*, vol. 16, no. 4, pp. 620–633, 1992.

[49] D. L. Hill, C. Studholme, and D. J. Hawkes, "Voxel similarity measures for automated image registration," in *Visualization in Biomedical Computing 1994*. International Society for Optics and Photonics, 1994, pp. 205–216.
[50] P. Vellanki and M. Khambete, "Enhanced stereo matching technique using image gradient for improved search time." *International Journal of Computer Science Issues (IJCSI)*, vol. 8, no. 3, 2011.
[51] D. Scharstein, "Matching images by comparing their gradient fields," in *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, vol. 1. IEEE, 1994, pp. 572–575.
[52] G.-Q. Wei, W. Brauer, and G. Hirzinger, "Intensity-and gradient-based stereo matching using hierarchical gaussian basis functions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 11, pp. 1143–1160, 1998.
[53] T. Twardowski, B. Cyganek, and J. Borgosz, "Gradient based dense stereo matching," in *Image Analysis and Recognition*. Springer, 2004, pp. 721–728.
[54] S. Hermann and T. Vaudrey, "The gradient-a powerful and robust cost function for stereo matching," in *Image and Vision Computing New Zealand (IVCNZ), 2010 25th International Conference of*. IEEE, 2010, pp. 1–8.
[55] X. Zhou and P. Boulanger, "Radiometric invariant stereo matching based on relative gradients," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*. IEEE, 2012, pp. 2989–2992.
[56] P. Pinggera, T. Breckon, and H. Bischof, "On cross-spectral stereo matching using dense gradient features," 2012.
[57] E. Haber and J. Modersitzki, "Intensity gradient based registration and fusion of multi-modal images," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*. Springer, 2006, pp. 726–733.
[58] J. Rühaak, L. Konig, M. Hallmann, N. Papenberg, S. Heldmann, H. Schumacher, and B. Fischer, "A fully parallel algorithm for multimodal image registration using normalized gradient fields," in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*. IEEE, 2013, pp. 572–575.
[59] E. Hodneland, A. Lundervold, J. Rørvik, and A. Z. Munthe-Kaas, "Normalized gradient fields for nonlinear motion correction of dce-mri time series," *Computerized Medical Imaging and Graphics*, vol. 38, no. 3, pp. 202–210, 2014.
[60] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 1. IEEE, 2003, pp. I–195.
[61] K.-J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 650–656, 2006.
[62] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 11, pp. 1222–1239, 2001.
[63] V. Kolmogorov and R. Zabin, "What energy functions can be minimized via graph cuts?" *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 2, pp. 147–159, 2004.
[64] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 9, pp. 1124–1137, 2004.
[65] J. Y. Bouguet, "Camera calibration toolbox for matlab." 2008.
[66] M. A. Lourakis and A. Argyros, "SBA: A software package for generic sparse bundle adjustment," *ACM Trans. Math. Software*, vol. 36, no. 1, pp. 1–30, 2009.
[67] K. Hirakawa, "Cross-talk explained," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008, pp. 677–680.
[68] H. S. Malvar, L.-w. He, and R. Cutler, "High-quality linear interpolation for demosaicing of bayer-patterned color images," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, vol. 3. IEEE, 2004, pp. iii–485.
[69] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Computer Vision, 1998. Sixth International Conference on*. IEEE, 1998, pp. 839–846.
[70] S. Paris and F. Durand, "A fast approximation of the bilateral filter using a signal processing approach," in *Computer Vision–ECCV 2006*. Springer, 2006, pp. 568–580.
[71] Y. Y. Schechner and S. K. Nayar, "Generalized mosaicing: High dynamic range in a wide field of view intensity range image mosaic," *International Journal on Computer Vision*, vol. 53, no. 3, pp. 245–267, 2003.
[72] J. Park, M. Lee, M. Grossberg, and S. Nayar, "Multispectral imaging using multiplexed illumination," in *International Conference on Computer Vision*, 2007.

Jason Holloway is a Ph.D. student working under the guidance of Ashok Veeraraghavan in the Electrical and Computer Engineering department of Rice University. His research interests are in computational imaging & photography and computer vision. He earned his B.S degrees in Electrical and Computer Engineering and Physics from Clarkson Univeristy in 2010 and his M.S. in Electrical and Computer Engineering from Rice University in 2013.

Kaushik Mitra is a postdoctoral research associate in the Electrical and Computer Engineering department of Rice University. His research interests are in computational imaging, computer vision and statistical signal processing. He earned his Ph.D. in Electrical and Computer Engineering from the University of Maryland, College Park, where his research focus was on the development of statistical models and optimization algorithms for computer vision problems.

Sanjeev J. Koppal is an assistant professor at the University of Florida's Electrical and Computer Engineering department. Prior to joining UF, he was a researcher at the Texas Instruments Imaging R&D lab. Sanjeev obtained his Masters and Ph.D. degrees from the Robotics Institute at Carnegie Mellon University. After CMU, he was a post-doctoral research associate in the School of Engineering and Applied Sciences at Harvard University. He received his B.S. degree from the University of Southern California in 2003. His interests span computer vision, computational photography and optics and include novel cameras and sensors, 3D reconstruction, physics-based vision and active illumination.

Ashok Veeraraghavan is currently an Assistant Professor of Electrical and Computer Engineering at Rice University, Tx, USA. At Rice University, Prof. Veeraraghavan directs the Computational Imaging and Vision Lab. His research interests are broadly in the areas of computational imaging, computer vision and robotics. Before joining Rice University, he spent three wonderful and fun-filled years as a Research Scientist at Mitsubishi Electric Research Labs in Cambridge, MA. He received his Bachelors in Electrical Engineering from the Indian Institute of Technology, Madras in 2002 and M.S and Ph.D. degrees from the Department of Electrical and Computer Engineering at the University of Maryland, College Park in 2004 and 2008 respectively. His thesis received the Doctoral Dissertation award from the Department of Electrical and Computer Engineering at the University of Maryland.