

Projected Pixel Localization and Artifact Removal in Captured Images

Himanshu Arora, Anoop Namboodiri

Center for Visual Information Technology, IIT, Hyderabad - 500032, India
{himanshu@research.,anoop@}iit.ac.in

Abstract—Projector-Camera systems are extensively used for various applications in computer vision, immersive environments, visual servoing, etc. Due to gaps between neighboring pixels on the projector’s image plane and variations in scene depth, the image projected onto a scene shows pixelation and blurring artifacts. In certain vision and graphics applications, it is desirable that a high quality composition of the scene and the projected image, excluding the artifacts, is captured, while retaining the scene characteristics. Localization of projected pixels can also help in dense estimation of scene shape. In this paper, we address the problem of localizing each of the projected pixels from a captured scene and restoring the captured image so that the pixelation and blurring artifacts of the projector are removed. We improve the quality of the captured image further by virtualizing a high resolution projector. i.e., we modify the captured image as if the scene were illuminated with a high-resolution projector. We propose robust solutions to these problems and demonstrate their effectiveness on scenes of different complexities.

I. INTRODUCTION

Projector-camera systems are extensively used in computer vision [1], HCI [2], immersive environments [3], and improving projection quality and versatility [4], [5], [6]. Image projected on a scene suffers from pixelation artifacts due to gaps between the neighboring pixels on the projector’s image plane, defocus artifacts due to varying scene depth and color transformation artifacts. If the camera is sufficiently close to the scene the pixelation artifacts are clearly visible. However, pixelation artifacts are also useful in certain applications.

The *first problem*, we address is to accurately localize each of the projected pixels. Detection of the projected pixels in the captured image can facilitate applications such as recomposition of a projected image and a scene, which is useful during post production stage. Procams are also useful for capturing surface properties. Considering the pixelation and blurring artifacts improves the accuracy of such estimations. Feature computations, such as SIFT, in a captured image can be inaccurate due to these artifacts. The relative spatial configuration of the localized pixels help in computing a dense shape for dynamic scenes. Usually, either a time shifted stripe pattern [1] or stereo image pairs [7] are used for static scenes.

The *second problem*, we address is the restoration of the captured image having pixelation and defocus artifacts. Public capturing of images of various projector scene composition such as presentation slides, immersive environments [3] etc. requires restoration. Projector-scene composition is useful in

movies for special effects. Images are rendered on real objects and the video is captured [8].

The localization and restoration of captured images is difficult due to a variety of factors such as spatially varying blur, background texture, noise, shapes of scene objects, and color transformations of projector and camera. The clue for the identification of projected pixels is that the sinusoids describing them share the same frequency with the neighboring pixels. The frequency, describing the sinusoids, can change over the whole image depending on the scene shape, and need to be estimated locally. Gabor filter is used to find the frequency of the repeating sinusoid in a window. Gabor filter is widely used because of its frequency selective properties. We extend the usage of local phase, computed using the Gabor filter, to isolate each of the projected pixels distinctively. Local phase is robust to noise and intensity variations and recently Arora *et al.* [9] have theoretically shown its invariance to a class of blur kernels. For restoration of captured images, we reproject the projected pixels such that these artifacts are absent. To improve the quality further we propose a mechanism to virtualize a high-resolution projector.

Related Work: To our best knowledge, this is the first work focusing on localizing projected pixels accurately and on enhancing captured images. Limited amount of work exist on display systems that reduce projector artifacts. However, they are applicable to static scenes and requires careful calibration and elaborate hardware setups. Zhang and Nayar [10] proposed a mechanism to project defocused images using a co-axial camera-projector. By slightly defocusing the projector and using the defocus compensation algorithm the pixelation artifacts are removed. Venkata and Chang [6] proposed simulating high-resolution projector using super-imposition of multiple low resolution projectors. Bimber and Emmerling [5] used multiple projectors, each having different focal planes, to project focused images at multiple depths. In all these cases, the main objective is to display a high-quality image on a surface. Note that in some situations it is not necessary or practical to display a high-quality image to improve the captured image. A seemingly related problem is restoring the halftone images. However, the techniques used are not applicable as the halftone images are binary images where different configurations of dots are perceived as gray-scale rendition. In our case, the captured images have varying background texture and blur.

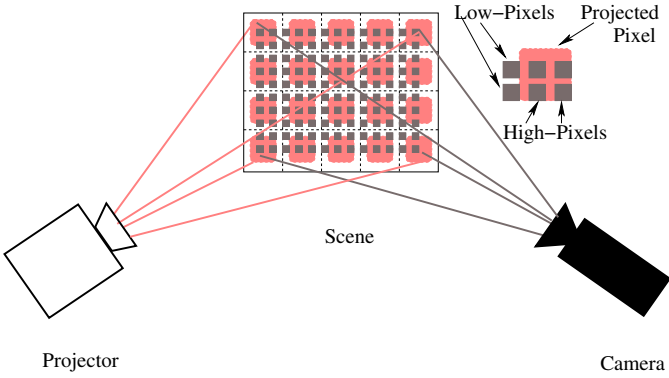


Fig. 1. A projector-camera system: Red squares correspond to the pixels of the projector, and black pixels correspond to the pixels seen by the camera. High-pixels and low-pixels in the captured image are also marked.

Our Contributions: We first analyze the structure of the projected pixels on the textured scene and propose a systematic approach to localize the projected pixels and remove the projector artifacts in the captured image. As our algorithm requires only one image, our system can work for dynamic scenes as well. No camera-projector calibration or co-axial camera-projector system is required. Specifically we propose: 1) an image re-formation model that describes the relationship between the display image, the projected scene, and the captured image with pixelation and defocus artifacts; 2) a robust algorithm for identification of the projected pixels seen in the captured image; 3) a method to remove the pixelation and blurring artifacts of the projector in the captured image; and 4) a mechanism to improve the quality of the captured image further by virtualizing a high-resolution projector, so that the captured image sees a larger number of projected pixels. Experiments have been performed on scenes of different complexities, under different projector settings, to show the robustness of the proposed approach.

II. PROBLEM FORMULATION

We refer the color image that is projected on the scene as the *display image*, and the image captured by the camera after the display image is projected on the scene as the *captured image*. We assume that the camera is sufficiently close to the scene so that each pixel projected on the scene is seen by more than one pixel in the camera's CCD. If this is not the case, pixelation would not be visible or can be treated as minor noise. The pixels of the captured image are classified into two categories: *high-pixels*, that see projected pixels of the display image, and *low-pixels*, which see the portions of the scene between neighboring projected pixels. High pixels have higher intensity values than the neighboring low pixels and hence the naming. Fig. 1 shows a typical projector camera system. The pixels shown are the locations seen by the pixels in the projector's LCD panel and the camera's CCD. We use the term *center-high-pixel* as the centroid of the group of high-pixels corresponding to a single display pixel. We frequently refer *projected pixel* as the group of high-pixels which corresponds to a single pixel of the display image.

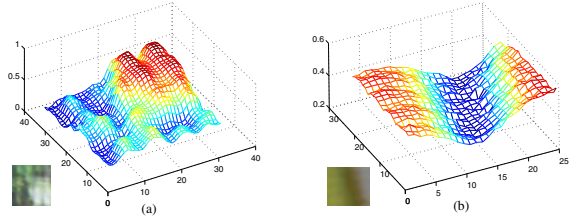


Fig. 2. Intensity plots of patches from captured images; with (a) scene texture and no blur and (b) without scene texture and blur.

We mathematically formulate the image re-formation model, which describes the relationship between the display image and the captured image. This is defined in terms of color transformations at image plane, blurring artifacts of both the camera and the projector, pixelation artifacts of the projector, scene deformation and radiance due to ambient light. Given a display image \mathbf{x} , represented as a column vector, the image is transformed by a matrix \mathbf{C}_p , which models radiometric response of the display device, projector brightness, and spectral response of the projector channel (for more details see [4]). This discrete color transformed input is converted into continuous domain with pixelation, due to the gaps between the neighboring pixels on the projector plane, by a discrete to continuous transformation, $p_p(\cdot)$. The output of this function is convolved with a blur kernel b_p of projector's lens, which is a function of the scene depth, and the image is mapped on to the screen by the transformation function $f_p(\cdot)$, which is with respect to camera's co-ordinate frame. α_s models various scene surface properties and k_s is the radiance due to ambient light. The scene is mapped to the CCD of the camera by the transformation $f_c(\cdot)$ and it is blurred by the camera's lens with the blur kernel b_c . The image is converted into digital form by $d_c[\cdot]$ and the color space is transformed by the matrix \mathbf{C}_c , which models various camera's CCD parameters. \mathbf{y} is the final image captured. The process is mathematically represented as:

$$\mathbf{y} = \mathbf{C}_c d_c [b_c * (f_c(\alpha_s f_p(b_p * p_p(\mathbf{C}_p \mathbf{x})) + k_s))]. \quad (1)$$

Aperture value of the camera is set to the lowest so that we can get wide depth of the scene in focus and the blurring due to camera lens is negligible. The goal is to rectify a single captured image such that the deblurring and pixelation artifacts of the projector are not present. Mathematically, the problem can be described as: given an image captured using the model in equation 1, restore it such that it were captured using the following model:

$$\mathbf{y} = \mathbf{C}_c d_c [f_c(\alpha_s f_p(n_p(\mathbf{C}_p \mathbf{x})) + k_s)], \quad (2)$$

where $n_p(\cdot)$ is a function that converts the discrete image pixels into the continuous space without any gap between the neighboring pixels on the projector plane. To restore the captured image, the main algorithm involved is the identification of center-high-pixel. We also propose a solution for virtualizing a high-resolution projector such that given an image which has been captured with the image re-formation equation 1, restore the captured image such that \mathbf{x} is of high resolution and it has been captured using the image formation equation 2.

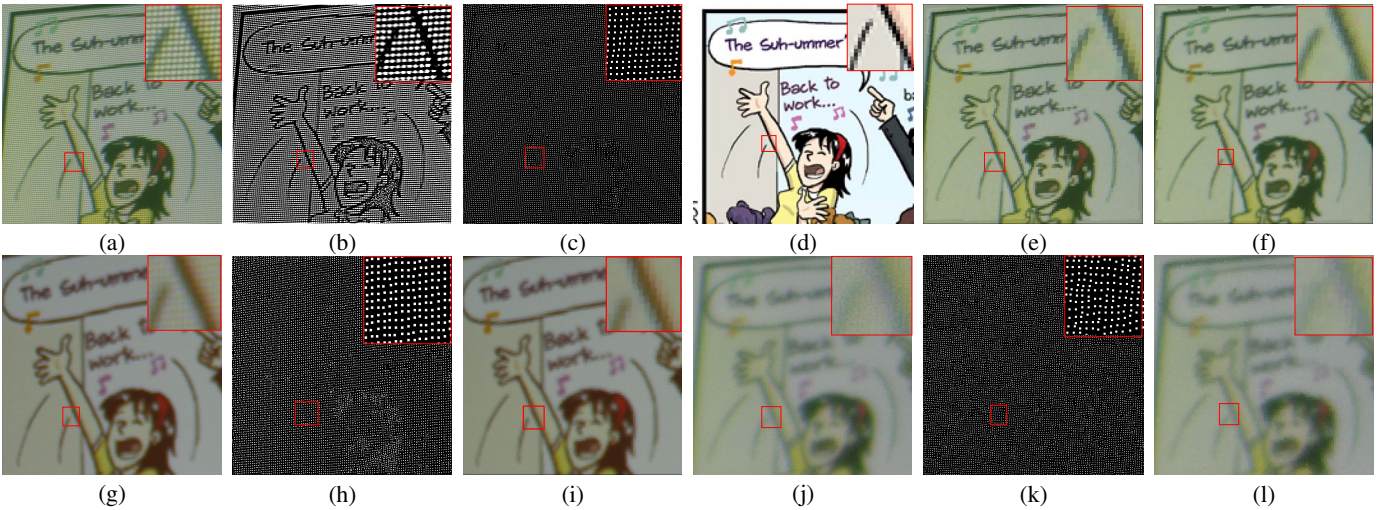


Fig. 3. (all images to be zoomed in) (a) captured image (patch) with pixelation artifacts; (b) local contrast normalized image; (c) center-high-pixel location map; (d) display image patch; (e) pixelation artifacts restored image; (f) high-resolution projector virtualized image; (g) captured image, projected using a different projector with slight defocus; (h) center-high-pixel location map of image in (g); (i) restored image (j) captured image with high blurring artifacts; (k) center-high-pixel location map of the image in (j); (l) defocus artifact restored image.

III. CHARACTERIZING HIGH-PIXELS

We first analyze the nature of the captured image to gain an insight on kind of algorithms suitable for robust characterization of high pixels. Problem is also analyzed from the perspective of equation 1 and 2. Assumption is made that the blurring is not extreme. Algorithm is described for computing center-high-pixel locations.

Fig. 2 shows the intensity plots of patches from the captured image. Local intensity sinusoidal peaks are visible in Fig. 2(a) and sinusoids are visible along left and right edges in Fig. 2(b). Another observation is that the intensity values is slightly higher between two consecutive projected pixels along horizontal or vertical direction than along diagonal direction. It allows us to decompose the problem into the detection of sinusoids in two orthogonal direction separately for robustness rather than modeling a single pixel.

The clue for the identification of projected pixels is that the sinusoids describing their shape share the same frequency with the neighboring pixels. The repeating dominant frequency in a small window is estimated and the local phase corresponding to these frequencies are used to isolate each of the projected pixels. In equation 1, for any general scene, the parameters C_c, C_p and k_s remains almost constant in a small window. The change in frequency content due to this is minor. The shape parameters, f_c and f_p , is also assumed to be smoothly varying and the computation of frequencies is windowed. As the Gabor filter is non-ideal band pass filter, it handles minute change in frequencies and orientations easily. Aperture of the camera is set to the lowest to bring the scene in focus. Local phase is invariant to a class of blur kernels which are even functions(Arora *et al.* [9]) and is also known to be invariant to illumination and robust to noise [11]. This property helps for robust isolation of projected pixels. Background texture can change the frequency content causing errors at low intensity projection but in general scenarios the change is small. All this

properties help to easily estimate $p_p(\cdot)$ robustly in presence of all the artifacts.

A. The Algorithm

The projected pixels of the display image, as seen in the captured image, can be thought as the intersection of two sets of equally spaced parallel lines which are approximately orthogonal. By calculating the orientation and frequency of these lines, and then taking the intersection of them, we identify the location of the projected pixels in the captured image. Before that the captured color image is converted into a gray level and the local contrast is normalized.

Local Contrast Normalization: The captured image is locally normalized to a zero mean and a unit variance. This is done so as to separately highlight the high-pixels and the low-pixels uniformly in the image. Each pixel value of the captured image, $\mathcal{I}(i, j)$, is reinitialized as

$$\mathcal{P}(i, j) = (\mathcal{I}(i, j) - \mu_w(i, j)) / \sigma_w(i, j), \quad (3)$$

where $\mu_w(i, j)$ and $\sigma_w(i, j)$ are the mean and standard deviation in a local window of size $w \times w$ at (i, j) .

Orientation and frequency estimation: We use Gabor filters [12] to calculate the orientation and frequency of these lines. Gabor filter is a band-pass filter which has frequency selective and orientation selective properties. The captured image is convolved with a bank of even symmetric Gabor filters at equally spaced angular directions and at multiple frequencies. The reason behind convolving with even-symmetric Gabor filter rather than complex Gabor filter is that they respond high to ridge like structure for the same sinusoid. The image is divided into blocks of considerable size. Along each direction, in each of the blocks, we select that frequency for which the sum of Gabor filter response is the maximum. At the next level we select two directions that has responded the maximum for any frequency based on the constraint that these two directions are at least some angle apart. To speed up the whole process,

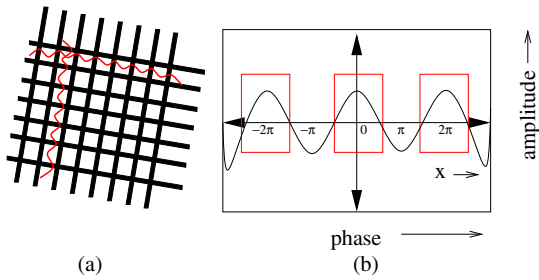


Fig. 4. (a) captured image can be seen as super-imposition of two sets of approximate orthogonal directional sinusoids; (b) high-pixels are robustly extracted by thresholding on phase information instead of amplitude because of robustness against noise, intensity, blur.

we first identify the line direction using a limited number of filters, and then refine the frequency estimate in the two orientations.

Identifying high-pixels: For each block, the exact orientations of lines and their frequencies are now available. To separate out the parallel lines (treated as a sinusoidal signal in one direction), we calculate the local phase at each point of the sinusoid and pixels with phase in $[-\pi/2, +\pi/2]$ is set to be belonging to a line. This way we separate different lines distinctively. Note that the local intensity peaks of the projected pixels correspond to the local phase value of 0. The same process is repeated for the orthogonal set of parallel lines. Intersection of these two line maps distinctively isolate the high pixels. Local phase is computed by convolving the image $i(x, y)$ with a Gabor wavelet, $g(x, y, f, \theta)$, having frequency f and orientation θ ,

$$\phi(x, y, f, \theta) = \arg[i(x, y) * g(x, y, f, \theta)], \quad (4)$$

where $\arg[\cdot]$ is a complex argument in $(-\pi, \pi]$. As the phase information is independent of the magnitude, the limits of threshold is fixed in advance.

Center-high-pixel locations: The high pixels calculated occur as a set of connected components, where each connected component correspond to one pixel of the display image (see Fig. 1). The center-high-pixel of each of these connected components is calculated by taking the mean of the co-ordinate locations.

IV. CAPTURED IMAGE ENHANCEMENTS

The pixelation and defocus artifacts are removed from the captured image. A mechanism is described to improve the quality of captured image by virtualizing a high-resolution projector. Before that we build the 8 neighborhood for each of the center-high-pixel. Instead of finding the eight closest points, we compute the 4 neighborhoods by utilizing the frequency and direction information computed in the previous section and then expand it to form the 8 neighborhood. For example, the north neighbor of the pixel in east would correspond to the north-east neighbor of the current pixel.

A. Depixelation and Deblurring

The intensity values of all the center-high-pixels is computed by taking the mean of the intensity values in a 3×3

window at that location. The pixels of the captured image are grouped around the center-high-pixels in the form of quadrilaterals. These quadrilaterals refers to the projected pixels which has been captured when the projector does not have pixelation and blur artifacts. They are computed by utilizing the center-high-pixel locations of neighbors for consistency. Each pixel in the quadrilateral is then assigned the value of the corresponding center-high-pixel. For textured scenes, the value of each of the pixel of the captured image is re-initialized to the weighted mean of its original value and the corresponding center-high-pixel location value.

B. Virtualizing a High Resolution Projector

The high resolution virtualization is defined in Section II. After calculating the 8 neighborhood for each of the center-high-pixel, we compute the location of new pixels to be embedded such that they lie uniformly with the neighborhood. The intensity values of the new projected pixels can be assigned either by using various interpolation techniques or by using the one pass learning based super-resolution techniques. The restoration process is same as described before.

V. EXPERIMENTS AND RESULTS

The proposed algorithm was tested on the images captured of scenes having different characteristics. The projector used was a HITACHI CP-S210, and the images were captured using a CANON EOS 350D camera. The aperture of the camera is set to the minimum. Gabor filters at uniform angles in eight directions were used at 3 different frequencies (0.17, 0.2, 0.23), for the initial estimation of line orientation and frequency. For refining the frequency estimates, Gabor filters with frequencies in the range $[0.16, 0.24]$ at an interval of 0.01 were used. We now describe the results with the display image projected on scenes of various properties.

Planar Textureless Scene : Images were captured under three different settings. In the first setting, the pixelation artifacts of the projector is clearly visible (Fig. 3(a)) and the projector is in focus. The images were restored at very high quality. The restored image is comparable with the display image at pixel level but some artifacts due to color transformations and brightness values of the projector can be seen. The image restored with high resolution projector virtualization is smoother than the original restored image. Fig. 3(g) and Fig. 3(j) shows images that are captured with increasing amounts of projector blur. The pixel location were calculated with high accuracy for the lower blur case and is quite satisfactory for the case with severe blur. When the blur is severe, we can observe color noise, because of the mixing of illumination of neighboring projected pixels with the low-pixels. Color noise artifacts are reduced considerably in the restored image (Fig. 3(l)).

Planar Textured scene: A planar object with strong surface texture (Fig. 5(1a)) is used. The center-high-pixel locations are calculated at very high accuracy even in presence of the strong scene texture. By using the mechanism mentioned in section IV, the background information in the image is

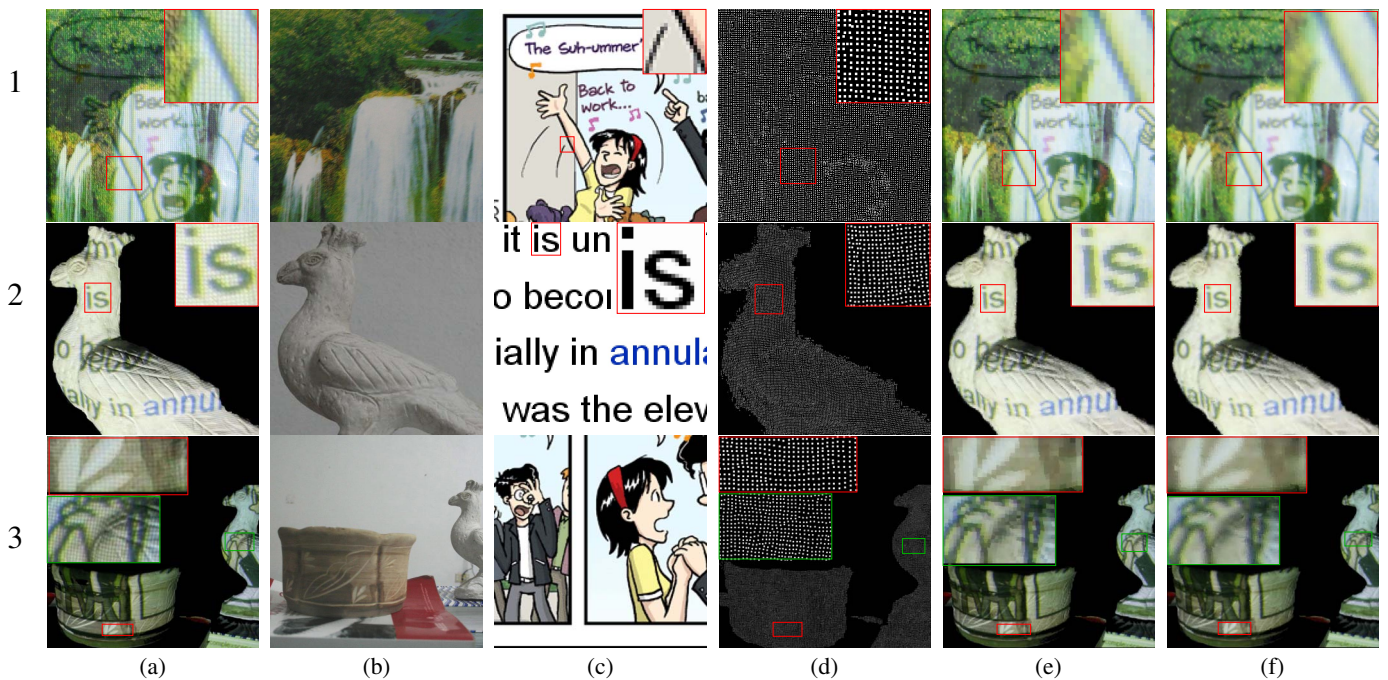


Fig. 5. (all images to be zoomed in) (a) composite captured image(patch); (b) background object; (c) display image patch; (d) center-high-pixel map; (e) restored image; (f) high-resolution projector virtualized image.

retained. With high resolution projector virtualization the quality of the restored image is further improved (Fig. 5(1f)).

3D objects: A textureless 3D object, Fig. 5(2b) was chosen as the scene. The restored captured image looks blocky due to smoothing of fine details on the object. However, the restored image is much better than the captured one as the pixelation artifacts are removed. Fig. 5(3b) shows a scene with two 3D objects, placed 6 inches apart. The texture of the captured image also get pixelated. Again with high-resolution projector virtualization the captured image is restored at a higher quality.

VI. DISCUSSION AND CONCLUSIONS

The proposed algorithm for localizing projected pixels is robust against noise, intensity variations and blur. Algorithm fails to detect high-pixels accurately in low intensity regions because of high noise at low intensity values. But this does not affect the quality of restored images, even for high-resolution image generation. Virtualization of high-resolution projector is useful to obtain restored images at higher quality. As the intensity value of the new pixels are calculated using the neighboring pixels and the background region, most of the background information is retained. In restoration with textured scenes, the background sometimes becomes blocky but with high resolution projector virtualization most of these problems are removed. Complete restoration of highly blurred projected image is not possible although color noise is removed. Identification of center-high-pixel locations is not only useful for restoring the captured image but also for other applications such as calculating scene parameters, extraction of features in the captured scene etc. With increase in usage of projector camera systems, restoration of images with projected texture will find a wide variety of applications.

ACKNOWLEDGMENTS : Authors acknowledge Jorge Cham of phdcomics.com for comic strips.

REFERENCES

- [1] J. Davis, D. Nehab, R. Ramamoorthi, and S. Rusinkiewicz, "Spacetime stereo: A unifying framework for depth from triangulation," *IEEE PAMI*, vol. 27, no. 2, pp. 296–302, Feb. 2005.
- [2] Y. Kakehi, M. Iida, T. Naemura, Y. Shirai, M. Matsushita, and T. Ohguro, "Lumisight table: An interactive view-dependent tabletop display," *IEEE Comp. Graph. App.*, vol. 25, no. 1, pp. 48–53, 2005.
- [3] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs, "The office of the future: a unified approach to image-based modeling and spatially immersive displays," in *SIGGRAPH*, New York, NY, USA, 1998, pp. 179–188.
- [4] M. Grossberg, H. Peri, S. Nayar, and P. Belhumeur, "Making One Object Look Like Another: Controlling Appearance using a Projector-Camera System," in *IEEE CVPR*, vol. I, Jun 2004, pp. 452–459.
- [5] A. Emmerling and O. Bimber, "Multifocal projection: A multiprojector technique for increasing focal depth," *IEEE TVCG*, vol. 12, no. 4, pp. 658–667, 2006.
- [6] N. Damera-Venkata and N. L. Chang, "Realizing super-resolution with superimposed projection," in *PROCCAMS'2007*, Minnesota, USA, June 2007.
- [7] R. Raskar, M. Brown, R. Yang, W. Chen, G. Welch, H. Towles, B. Seales, and H. Fuchs, "Multi-projector displays using camera-based registration," in *Proceedings of the 10th IEEE Visualization Conference*, Washington, DC, USA, 1999.
- [8] R. Raskar, G. Welch, K. Low, and D. Bandyopadhyay, "Shader lamps: Animating real objects with image-based illumination," in *Proceedings of the 12th Eurographics Workshop on Rendering Techniques*. London, UK: Springer-Verlag, 2001, pp. 89–102.
- [9] H. Arora, A. M. Namboodiri, and C. V. Jawahar, "Robust image registration with illumination, blur and noise variations for super-resolution," in *ICASSP*, Las Vegas, USA, 2008, pp. 1301–1304.
- [10] L. Zhang and S. Nayar, "Projection defocus analysis for scene capture and image display," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 907–915, 2006.
- [11] D. J. Fleet and A. D. Jepson, "Stability of phase information," *PAMI*, vol. 15, no. 12, pp. 1253–1268, December 1993. [Online]. Available: citeseer.ist.psu.edu/fleet93stability.html
- [12] J.-K. Kämäräinen, "Feature extraction using Gabor filters," Ph.D. dissertation, Lappeenranta Univ. of Technology, 2003.