

Coded Aperture and Coded Exposure Photography

Martin Wilson
University of Cape Town
Cape Town, South Africa
Email: Martin.Wilson@uct.ac.za

Fred Nicolls
University of Cape Town
Cape Town, South Africa
Email: Fred.Nicolls@uct.ac.za

Abstract—This article presents an introduction to the field of coded photography, with specific attention given to coded aperture and coded exposure theory, methods, and applications. A coded aperture is optimized for the task of defocus deblurring, and constructed using a simple cardboard occluding mask. Furthermore, a series of coded aperture photographs are used to capture a 4D light field with a standard SLR camera, and the captured light field is then used for depth-estimation and refocusing applications. Finally, a coded exposure pattern is optimized for motion deblurring, and coded exposure photographs are captured by controlling the incident illumination of the scene. The coded aperture and exposure methods are shown to be superior to traditional photographic methods in terms of deblurring accuracy, and the captured light field is used successfully to produce a stereo depth estimate from a stationary camera, and to produce post-exposure refocused photographs.

I. INTRODUCTION

Digital photography currently has applications in almost every area of industrial and scientific research. However, the limitations of traditional photographic techniques and equipment (e.g. defocus blur, motion blur, image noise, and finite resolution) continue to restrict its usefulness and flexibility. Furthermore, traditional photography is not able to capture all of the available information within a scene’s visual appearance (e.g. scene depth, surface reflectance properties, and ray-level structure), and often this extra information would be very valuable to the application in question. Computational photography is a recently established area of research concerned with overcoming these disadvantages by utilizing computational techniques during image capture and post-processing.

Coded photography is a branch of computational photography that attempts to capture additional visual information by reversibly encoding the incoming optical signal before it is captured by the camera sensor. The encoding process can occur at multiple points in the photographic model, including at the generation of the incident scene illumination and within the camera itself. Two popular in-camera methods include using specially engineered aperture shapes and high-frequency exposure patterns.

When analyzing computational photography methods, it is often useful to represent the incoming optical signal as a set of light rays rather than merely as a 2D array of intensity pixels. The set of incoming light rays can be defined as a 4D function known as a *light field*, which specifies the intensity of the ray passing through each location on a 2D surface, at each possible 2D angular direction. Recently it has been shown that

a subset of the full light field can be practically captured by making relatively minor modifications to a standard camera, and that the captured light field can have a sufficient resolution for a variety of useful applications such as synthesizing virtual photographs with arbitrary camera parameters.

In section II a selection of related work is presented, specifically in the fields of coded aperture, coded exposure, and light field photography. In section III the implementation details regarding our own experiments with coded aperture and coded exposure photography are described. The experiments performed include reducing defocus and motion blur, as well as capturing and applying a light field. The results of the experiments are presented in section IV, and finally conclusions are drawn in section V.

II. RELATED WORK

Our work is inspired by related work in the fields of coded aperture photography, coded exposure photography, and light field acquisition. Recently a number of comprehensive surveys have been published regarding these fields of research [1], [2], and in this section we briefly summarize a selection of related work in order to provide a context for our own experiments.

A. Coded Photography

A popular reason for employing coded apertures within an optical system is the fact that it allows the point-spread-function (PSF) to be engineered for specific purposes. By replacing the traditionally round aperture found in most current cameras with a carefully designed coded aperture, the PSF can be engineered to enhance the performance of depth estimation techniques such as depth-from-defocus [3], [4], or it can be engineered to preserve frequency information in out-of-focus images, thereby increasing the performance of deblurring techniques [3], [5]. While most coded apertures are implemented using only occluding optics, some have been developed using reflective elements that allow multiple apertures to be used in a single exposure [6]. The success of coded aperture methods relies on selecting the optimal coded aperture for a particular application, and therefore this non-trivial task has received a significant amount of attention from the research community [5].

Coded apertures can also be used to capture images in highly unconventional ways. For example, by using an aperture consisting of multiple programmable attenuating layers it is possible to capture an image without using a lens, and the

parameters of this lensless camera (e.g. focus, pan and tilt) can be adjusted without any physical movement [7]. Alternatively, by taking a sequence of photographs, each with a different coded aperture, a 4D light field can be separated into 2D slices and captured using an unmodified image sensor [8].

Coded exposure photography is conceptually very similar to coded aperture photography, in that the traditional box-shaped exposure window is replaced with a coded pattern in order to engineer the PSF of moving objects. Coded exposures can be captured using a standard camera with an additional high-speed electronic shutter, and this has been shown to be useful for improving the performance of motion deblurring [9].

B. Light Field Capture and Applications

The 4D light field (also known as a *lumigraph*) is defined as the radiance of every light ray passing through a 2D surface, at every possible angle. An ideal light field cannot be physically captured in its entirety, but was first proposed as a useful data representation for image based rendering [10], [11]. However, due to the increasing availability of inexpensive, high quality digital cameras, a variety of methods for partially sampling a light field have been developed, and these partial light fields have since found applications in many image processing fields.

The first practical methods for capturing light fields used either an array of cameras [12], or a single camera on a gantry [11], to record multiple exposures of a scene from slightly different locations. Despite their conceptual simplicity, camera arrays are difficult to use in practice due to their large size and mechanical complexity. For this reason, building small, portable, single-camera devices for capturing light fields is currently a popular area of research. Capturing a light field in a single exposure requires placing additional optical elements into the camera, in order to prepare the 4D light field for measurement on a 2D sensor. One method is to place a micro-lens array between the image sensor and the lens, thereby modulating the image formed on the sensor according to each ray's angular direction [13]. Another alternative is to use a high-frequency attenuating mask, which creates spectral 'tiles' of the 4D light field on the 2D image sensor, in a process similar to *heterodyning* in radio electronics [14].

Once a light field has been successfully captured, it can be used for a number of practical applications including glare reduction, depth estimation, and refocusing. Glare effects in a photograph are caused by a small subset of light rays, and therefore if the light field can be captured, the offending rays can be easily identified and ignored [15]. This is not possible for a conventional photograph due to the integrative process of traditional image capture. Light fields can also be used to generate stereo views of a scene without requiring that the camera be physically moved to a new position. This allows stereo depth methods to be used in what is essentially a monocular system [8]. Lastly, virtually refocused photographs can be synthesized from a light field by placing a virtual image plane into the model, and calculating the image formed using ray-tracing techniques [8], [13], [14].

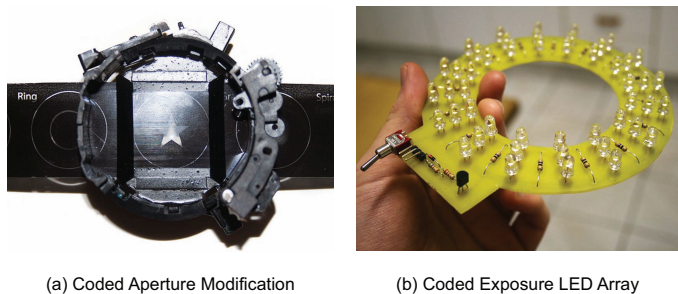


Fig. 1. Photographs showing the two crucial elements of the prototype coded photography camera.

III. IMPLEMENTATION

A. Coded Photography Prototype Camera

A prototype coded photography camera was constructed from a Canon 500D SLR camera and a Canon EF 50mm f/1.8 II prime lens. The standard aperture and autofocus modules were removed from the lens, and replaced with a plastic brace that allows aperture masks to be inserted from slits cut into the lens's external housing. Coded exposures were captured using coded illumination generated by a programmable LED array that was mounted around the lens. This method is far simpler than replacing the camera's shutter with a high-speed electronic shutter, and produces good results provided that the ambient lighting in the scene can be minimized. Figure 1 shows the modified aperture housing and programmable LED array.

B. Optimizing Apertures and Exposures for Deblurring

Both defocus blur and motion blur can be modelled as the convolution between an ideal sharp image and a non-ideal PSF, or in the frequency domain, as the multiplication of their frequency spectra [5]. Minima and zeros in the PSF's spectrum cause information to be irreversibly lost in the observed blurred images, thereby making the deblurring process ill-defined. Therefore, intuitively it can be proposed that aperture shapes and exposure patterns that have PSFs with large minimum values in their frequency spectra will perform well in deblurring applications. Two specific performance metrics were used in our experiments, R_{raskar} and R_{zhou} , which were proposed by Raskar et al. [9] and Zhou et al. [5] in their respective papers. R_{raskar} is a formalized version of the intuitive proposal described above, and is defined as

$$R_{raskar}(f_k) = \alpha \cdot \min(|F_k|) + \beta / \text{variance}(|F_k|), \quad (1)$$

where f_k is the aperture or exposure pattern under consideration, F_k is the Fourier transform of f_k , and α and β are tunable scalars. The variance term is included to account for the inaccuracies that would result from errors in the estimation of f_k . R_{zhou} is a more complicated performance metric that also takes into account the level of noise present in the observed image and statistics of natural images. It is defined as

$$R_{zhou}(f_k) = \sum_{\omega} \frac{\sigma^2}{|F_k(\omega)|^2 + \frac{\sigma^2}{A(\omega)}}, \quad (2)$$

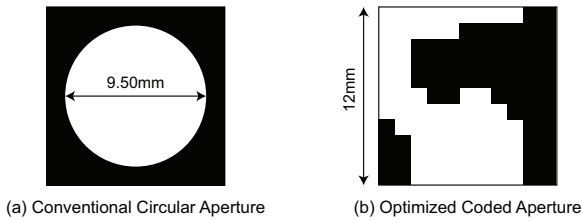


Fig. 2. The two aperture shapes used in the defocus deblurring experiments.

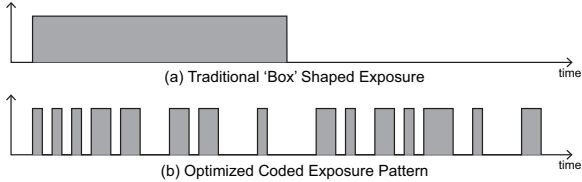


Fig. 3. The two exposure patterns used in the motion deblurring experiments.

where σ is the standard deviation of the assumed white Gaussian noise and $A(\omega)$ is an approximation of the power spectrum of natural images. For each frequency, ω , the metric indicates the degree to which noise at that frequency will be amplified. To simplify the optimization problem, aperture shapes were restricted to an 11×11 binary grid, and exposure patterns were restricted to a 52-bit binary string. The solution spaces are too vast for an exhaustive search for an optimum to be feasible, and therefore a genetic algorithm was used to find well-performing local optima. Figure 2 shows the traditional circular aperture and the optimized coded aperture that were used in the defocus deblurring experiments, while figure 3 shows the traditional ‘box’ exposure and the optimized coded exposure that were used in the motion deblurring experiments.

C. Deblurring by Deconvolution

Since blur can be modelled as convolution, deblurring a photograph requires deconvolving the observed image with the appropriate PSF. This requires that the PSF be estimated from the camera and scene parameters. For defocus deblurring, the shape of the PSF is equal to the shape of the aperture, and the scale is a function of depth, while for motion deblurring the shape of the PSF is determined by the velocity of the scene’s relative motion and the total exposure time. In our experiments the deconvolution implementation developed by Levin et al. [3] was used, which assumes a heavy-tailed, sparse derivative distribution for natural images.

D. Light Field Capture

Figure 4 shows a popular light field representation that defines a ray by its intersection with two parallel planes, (u, v) and (s, t) [11]. If the (u, v) plane is defined at the physical aperture plane, and the (s, t) plane is defined at the physical image sensor plane, then the pixel values in a captured photograph represent the intensities of the rays intersecting with the (s, t) plane, integrated over all the (u, v) locations allowed by the particular aperture shape. Therefore, if rays are only allowed to pass through a single (u, v) location,

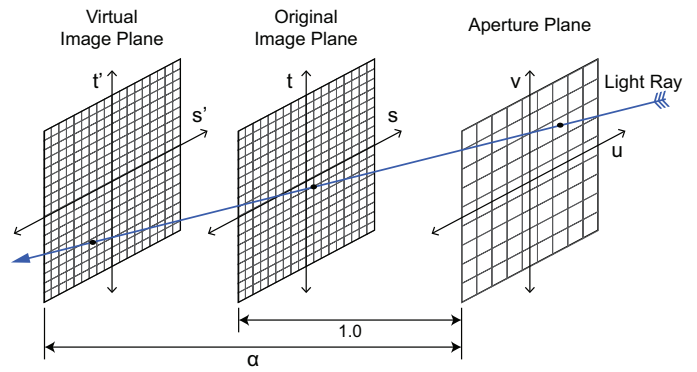


Fig. 4. Diagram showing the two-plane light field representation, as well as the virtual image plane used to synthesize virtually refocussed photographs.

then the captured image represents a single 2D slice of the 4D light field. If multiple slices are captured, each with a different constant (u, v) value, then the full 4D light field can be reconstructed in software [8]. In our experiments the light field was captured from 81 separate exposures, each one taken with a different block of a 9×9 binary coded aperture open at a time.

IV. RESULTS

A. Code Aperture Results

In this section our defocus deblurring and light field experiments are described, and their results are presented. Where possible, the performance of the coded apertures is compared to that of a traditional circular aperture.

1) *Defocus Deblurring Results:* Figure 5 shows a sample of the results obtained from defocus deblurring experiments, using both a planar resolution chart and a human face as target scenes. The observed images were taken with the camera 2.0m away from the scenes, and the lens was focused at 1.0m. Since the scene objects lie far outside the focal plane, the observed images predictably show a significant level of blurring. In the image of the resolution chart, the text is completely unrecognizable and only the coarsest lines are distinguishable, while in the image of the human face, all the hard edges have been lost and only large facial features are identifiable. It is also interesting to note that the blur caused by the conventional aperture is visually smoother than the blur caused by the coded aperture, and this supports the claim that the optimized coded aperture is able to preserve more high frequency information within the defocused areas.

The superiority of the coded aperture can clearly be seen when comparing the deblurred images in figure 5. The deblurred conventional aperture images contain a significant amount of ringing, and they have failed to recover even moderately fine details. Only the coarsest lines in the resolution chart are distinguishable, and while the contrast at hard edges in the face has been improved, the edges are distorted and over-simplified. In contrast, the deblurred coded aperture images contain far less ringing, and significant high frequency information has been recovered. The medium-to-fine lines in

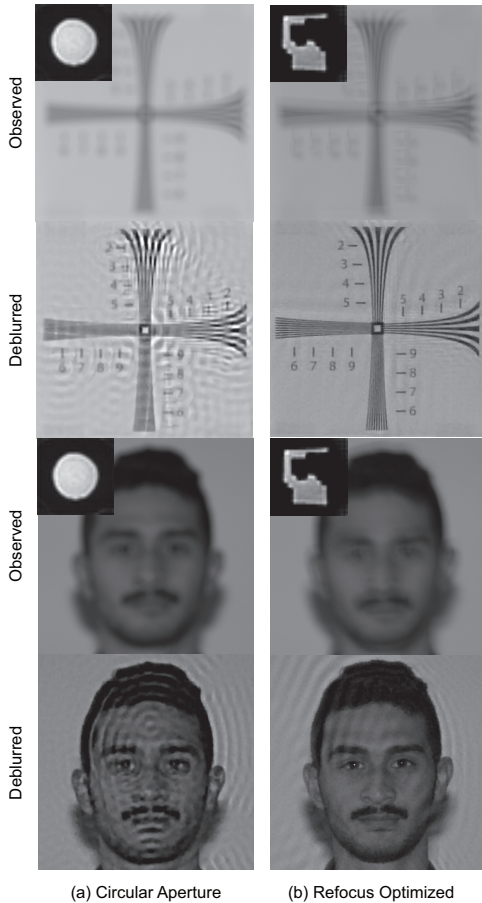


Fig. 5. A selection of results obtained from the defocus deblurring experiments. The camera was focused at 1.0m and placed 2.0m away from the scene.

the resolution chart are now clearly distinguishable and the text has been accurately reconstructed. In the deblurred face image, even the fine details such as the specular highlights in the eyes and the texture of the facial hair have been recovered.

While results are only shown for a camera distance of 2.0m, experiments were performed for distances ranging from 1.0m to 2.0m in 10cm increments. At all of these distances the results obtained with the coded aperture were superior to those obtained with the traditional aperture. However, the difference between the performances of the two apertures becomes less noticeable for distances near to the focal plane, and it is speculated that this is due to the fact that at these distances the scale of the PSF becomes too small to properly define its carefully engineered shape.

2) *Light Field Results*: Two visualizations of the light field captured in our experiments are shown in figure 6. Both represent a subset of the full light field as a 2D array of images, one with uv -major indexing and the other with xy -major indexing. In the case of uv -major, each (u, v) coordinate defines a single 2D image that represents a unique angular view of the scene. In the case of xy -major indexing, each (x, y) coordinate defines a 2D image that represents the angular intensity distribution of the light rays falling on a specific

sensor pixel.

Figure 7 shows the results of an experiment in which we attempt to calculate stereo disparity from a light field. The two input images were extracted from the captured light field by setting $(u, v) = (1, 5)$ and $(u, v) = (9, 5)$ respectively. The images represent two horizontally spaced views of the test scene, and so could be directly input into a stereo disparity algorithm without any further processing. Ignoring the noisy values in the background (which are due to the absence of texture in the input images), the output disparity map has accurately determined the relative depths of the objects in the test scene. While calculating stereo disparity from a pair of horizontally spaced input images is fairly commonplace, our result was obtained from single, stationary camera, and is therefore an extension of the standard method.

Results from a virtual image refocusing experiment are shown in figure 8. In (a) a photograph with the original unaltered focal plane is shown. At the time of exposure the camera was focussed at 1.0m and placed 1.0m away from the centre of the scene. Therefore the image of the metronome (which is located at the centre of the scene) is in sharp focus, while the objects in front and behind the focal plane have a significant amount of defocus blur. This is more clearly seen in sub-figure (b), which shows cropped and magnified images of the Rubik's cube, metronome, and mannequin. Refocused photographs were synthesized by placing a virtual image plane into the light field model at various depths from the aperture plane (shown in figure 4). The original image-to-aperture distance was normalized as $\alpha = 1.000$, and refocussed photographs were produced for α values ranging from 0.980 to 1.020 in 0.001 increments. Sub-figures (c) and (d) show cropped images taken from the refocussed photographs obtained using $\alpha = 0.995$ and $\alpha = 1.005$ respectively. In (c) the focal plane has been moved further away from the camera, thereby bringing the mannequin into focus, while in (d) the focal plane has been brought closer to the camera, thereby bringing the Rubik's cube into focus.

B. Coded Exposure Results

This section describes a series of motion deblurring experiments that were performed, and presents the results obtained. The performance of the coded exposure results are also compared to the results obtained with a traditional box-shaped exposure.

1) *Motion Deblurring Results*: Figure 9 shows the results of motion deblurring experiments involving moving scene objects. The observed photographs were captured using a 0.5s shutter time, and during exposure the scenes were manually moved vertically at an approximately constant speed (estimated to be $\approx 0.4\text{m/s}$). The relative motion between the scene and the camera has produced a significant amount of motion blur in the observed images, and almost all detail has been lost in the vertical direction. The horizontal lines and text in the resolution chart are completely unrecognizable, and only the basic elongated shape of the face is discernible. The motion

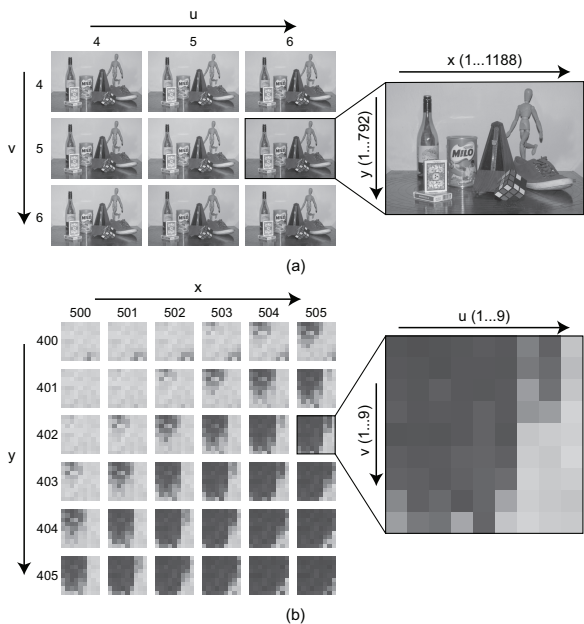


Fig. 6. Diagram showing a subset of the captured light field as a 2D array of 2D images: (a) using uv-major indexing, and (b) using yx-major indexing.

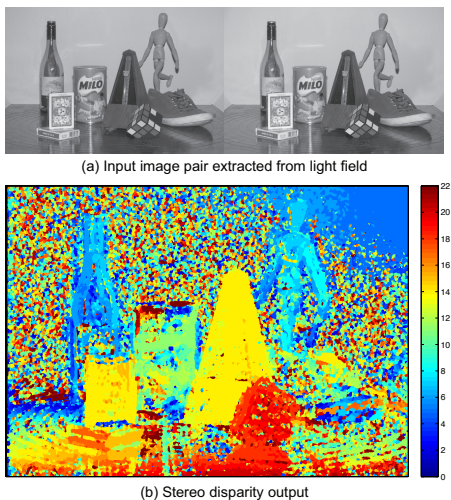


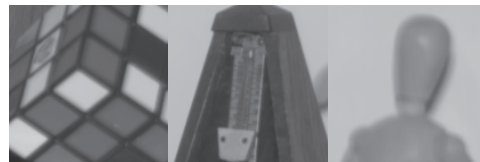
Fig. 7. Resulting stereo disparity map calculated from the stationary light field.

blur caused by the coded exposure pattern also seems to contain more vertical structure than the smooth blur caused by the traditional exposure, which suggests that it has preserved more high-frequency information than the traditional exposure.

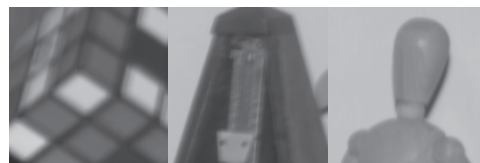
Deblurring the photographs captured using the traditional exposure has strengthened some of the vertical contrast, but most of the detail remains unrecovered and the background noise has been significantly amplified. The position of the text has been recovered, but the characters themselves remain unrecognizable, and despite a slight improvement in the large facial features (e.g. the forehead, chin, and eyebrows), the identity of the face remains unrecognizable. In contrast, a substantial amount of the original detail in the coded exposure



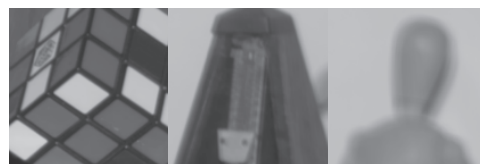
(a) Photograph with unaltered focal plane



(b) Cropped details: unaltered photograph ($\alpha = 1.000$)



(c) Cropped details: refocused on mannequin ($\alpha = 0.995$)



(d) Cropped details: refocused on cube ($\alpha = 1.005$)

Fig. 8. Diagram showing the results of synthesizing virtually refocused photographs from the light field.

photographs has been salvaged by deblurring. In the resolution chart, the text is now readable, and the medium-to-coarse horizontal lines can even be distinguished. Also, almost all the major facial features have been recovered, and the identity of the face has become clearly visible.

V. CONCLUSIONS AND FUTURE RESEARCH

A. Summary of Results

The results of the experiments clearly show that there are significant advantages to using coded apertures and exposures for applications such as defocus deblurring, motion deblurring, and light field capture. The coded photography techniques that have been covered require very simple and inexpensive hardware, and can be implemented easily by making small modifications to existing optical systems.

For defocus deblurring, coded apertures can be engineered to preserve high frequency information in the blurred regions, thereby improving the results of the deconvolution operation. The deblurred photographs obtained using the optimized coded aperture contained far less ringing than when traditional circular apertures are used, and the hard edges in the photographs were more accurately recovered.

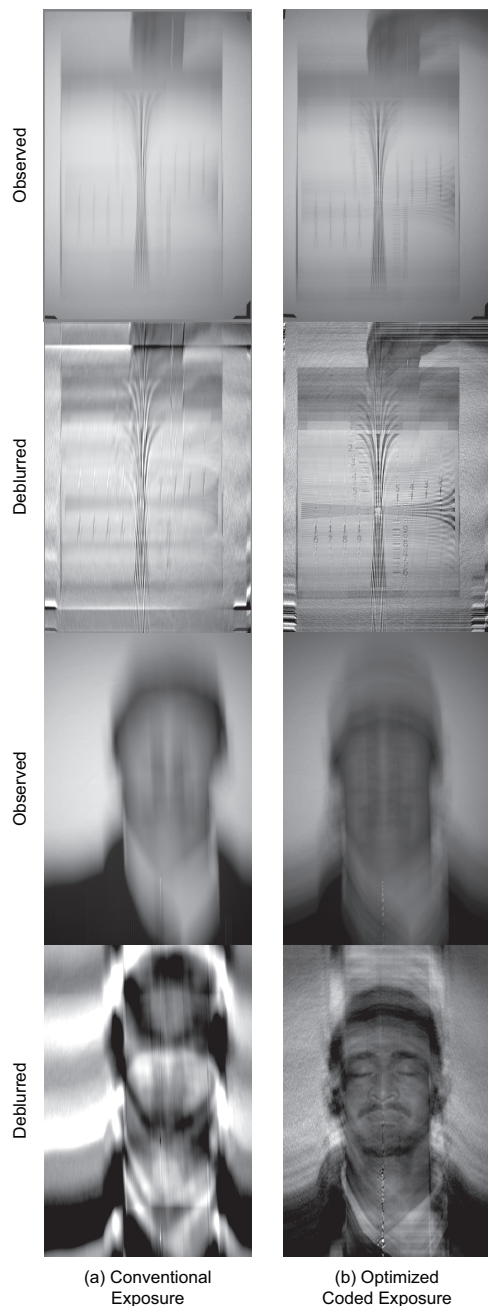


Fig. 9. A selection of results obtained from the motion deblurring experiments. The scenes were moved vertically at a constant velocity during the 0.5s exposure time.

Coded apertures were also used to capture a partial 4D light field of a 3D scene. While multiple exposures are required, this particular method can capture light fields with very fine spatial resolution and flexible angular resolution. The light field captured in our experiments was shown to be of practical use for calculating depth, and synthesizing virtual photographs with adjusted focus settings.

Finally, coded exposures can be optimized to preserve high frequency information in photographs with substantial motion blur. Our experiments with constant velocity motion showed

that coded exposure patterns produce far more accurate deblurring results than can be achieved with traditional exposures.

B. Recommendations for Future Research

Using an LCD-based aperture instead of the physical masks used in our experiments would allow for almost instantaneous aperture changes, which would reduce the time required to capture a light field, and offer the ability to capture video with a different aperture per frame. Also, using an LCD filter to control exposure rather than controlling the incident scene lighting would allow coded exposure photographs to be captured outside of the laboratory environment.

Another avenue for future investigation is the use of non-binary coded apertures and exposures. Using gradient apertures could allow for a greater number of possible aperture shapes without increasing the diffraction effects associated with hard edges. Also, since most digital cameras contain a Bayer-pattern colour mask, using apertures constructed out of RGB filters could allow each colour channel in a single exposure to be captured using a different aperture shape.

ACKNOWLEDGMENT

The authors would like to thank the National Research Foundation, and Armscor's PRISM program, managed by the CSIR, for their financial support.

REFERENCES

- [1] G. Wetzstein, I. Ihrke, D. Lanman, and W. Heidrich, 2011. Computational Plenoptic Imaging. *Eurographics State of the Art Report*, 1-24, 2011.
- [2] S.K. Nayar, 2011. Computational Camera: Approaches, Benefits and Limits. *DTIC Technical Report Document*, 2011.
- [3] A. Levin, R. Fergus, F. Durand, and W.T. Freeman, 2007. Image and depth from a conventional camera with a coded aperture. *Proceedings of ACM SIGGRAPH 2007* 26(3), July 2007.
- [4] C. Zhou, S. Lin, and S. Nayar, 2011. Coded Aperture Pairs for Depth from Defocus and Defocus Deblurring. *International Journal of Computer Vision* 93(1):53-72, 2011.
- [5] C. Zhou, and S. Nayar, 2009. What are Good Apertures for Defocus Deblurring? *ICCP 2009 (oral)*.
- [6] P. Green, W. Sun, W. Matusik, and F. Durand, 2007. Multi-aperture photography. *Proceedings of ACM SIGGRAPH 2007* 26(3), July 2007.
- [7] A. Zommet, and S. Nayar, 2007. Lensless Imaging with a Controllable Aperture. *IEEE Computer Society* 339-346, 2006.
- [8] C. Liang, T. Lin, B. Wong, C. Liu, and H. Chen, 2008. Programmable aperture photography: multiplexed light field acquisition. *Proceedings of ACM SIGGRAPH 2008* 27(3), August 2008.
- [9] R. Raskar, A. Agrawal, and J. Tumblin, 2006. Coded exposure photography: motion deblurring using fluttered shutter. *Proceedings of ACM SIGGRAPH 2006*, 25(3), July 2006.
- [10] S.J. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen, 1996. The Lumigraph. *Proceedings of ACM SIGGRAPH '96*, August 1996.
- [11] M. Levoy, and P. Hanrahan, 1996. Light Field Rendering. *Proceedings of ACM SIGGRAPH '96*, August 1996.
- [12] B. Wilburn et al, 2005. High Performance Imaging Using Large Camera Arrays. *Proceedings of ACM SIGGRAPH 2005*, 24(3), July 2005.
- [13] R. Ng, M. Levoy, M. Brdif, G. Duval, M. Horowitz, and P. Hanrahan, 2005. Light Field Photography with a Hand-Held Plenoptic Camera. *Stanford University Computer Science Tech Report CSTR*, 2005-02.
- [14] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, 2007. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *Proceedings of ACM SIGGRAPH 2007*, 26(3).
- [15] R. Raskar, A. Agrawal, C. Wilson, and A. Veeraraghavan, 2008. Glare aware photography: 4D ray sampling for reducing glare effects of camera lenses. *Proceedings of ACM SIGGRAPH 2008*, 27(3), August 2008.