Compact and Thin Multi-lens System for Machine Vision Applications

Edmund Y. Lam

Imaging Systems Laboratory, Department of Electrical and Electronic Engineering,
University of Hong Kong, Pokfulam Road, Hong Kong

ABSTRACT

Compact imaging devices are desirable in many different machine vision applications. For instance, in inspection for semiconductor manufacturing systems, the reduction in feature size demands lenses with a small working distance but a wide field of view. An integrated computational imaging system has proved to be advantageous in this respect, as it integrates the optics, optoelectronics, and signal processing together in the system architecture. This allows for unconventional optical systems that require further image processing to reconstruct the images, but can be made to satisfy more stringent design constraints such as size, power, and cost. In this paper, we focus on a multi-lens optical architecture. We explain the possible designs and discuss the reconstruction of images, as we need to combine the multiple low-resolution images formed from the different optical paths into a high-resolution image. We will also explore its applicability in various machine vision applications.

Keywords: Imaging systems, multi-lens architecture, computational imaging, super-resolution, image reconstruction, machine vision

1. INTRODUCTION

In a typical imaging system design, the process is traditionally divided into two distinct aspects: the optical front-end, which is responsible for producing a crisp image, and the post-detection processing that aims at enhancing the quality or extracting information from the image. This is especially true for machine vision applications, where we often perform tasks such as edge detection or feature extraction on the images. In many designs, the optics often imposes significant physical constraints such as size and weight, and incurs substantial cost among the entire system.

However, recent advances in imaging technology have shown that an integrated approach combining the design of the optics, optoelectronics, and image processing can lead to more effective imaging systems.[1] This has its roots in the earlier days of optical signal processing,[2] but the key here is to tap the power of image post-processing to permit a more flexible optical architecture. The image formed may not be visually appealing, but it contains useful information that becomes apparent after the processing. Wavefront coding, which we will explain further below, is a successful realization of this design philosophy, where an imaging system with a suitable phase mask can achieve an extended depth of field compared to the traditional design.[3,4]

This integrated approach has also been applied to multi-lens architectures. Over the last few years, various groups worldwide have developed "compound-eye" systems inspired by the biological vision of arthropods.[5–9] They can be further classified into apposition and superposition systems. In the former, each lenslet produces images on separate regions of the sensor, while in the latter, light from multiple lenslets combines on the sensor. Various researchers have demonstrated that the apposition system can enjoy a wide field of view, but often at the expense of low spatial resolution and sometimes low light sensitivity.[7] The superposition system is more light sensitive, but spherical aberrations often affect the combination of light and lead to a resolution far from the diffraction limit.[10] In applying such multi-lens architecture for machine vision, a wide field of view is very desirable[11] and most "compound-eye" systems follow the apposition design. Image reconstruction techniques are improving to produce higher quality images to ameliorate the loss in spatial resolution.[12]

Further author information: (Send correspondence to Edmund Lam)
Edmund Lam: E-mail: elam@eee.hku.hk, Telephone: 852 2859 7093

Image Processing: Machine Vision Applications, edited by Kurt S. Niel, David Fofi,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6813, 681305, © 2008 SPIE-IS&T · 0277-786X/08/$18

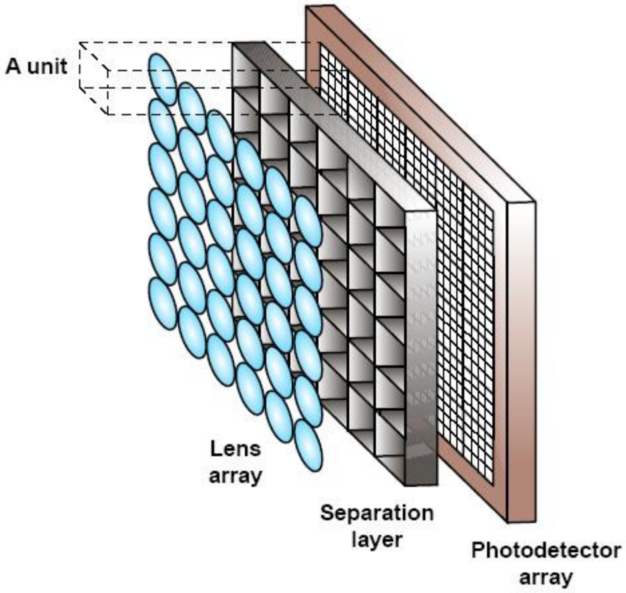


Figure 1. The TOMBO system.[13]

2. THE TOMBO ARCHITECTURE

Short for thin observation module by bound optics, TOMBO is a multi-lens system design that has attracted a lot of attention.[9] A basic system is represented in Figure 1. It consists of a lens array with, say, $N \times N$ microlenses, and a photodetector array with $D \times D$ photosensitive cells. D is generally much larger than N. We can let $D = \beta N$, and each microlens would correspond to $\beta \times \beta$ cells. A microlens and its array of photosensitive cells together form an imaging unit. Each imaging unit of the system generates a low resolution (LR) sub-image of the object. To avoid cross-talk between the individual sub-images, a separation layer, which consists of partition blocks, is often placed between the lens array and the photodetector array.

The working principle of the TOMBO system is summarized in Figure 2.[14] The multi-lens camera first takes a snapshot of the target object. The result is an intermediate image consisting of an array of LR sub-images. This then undergoes a reconstruction algorithm to give a final HR image of the object.

There are several challenges involved in the design and analysis of this multi-lens system:

1. *Optical system modeling*
   The LR images are warped versions of one another because of differences in the lenslet positions. An accurate parameter estimation of the displacement and modeling of the aberrations in the optics are needed to fuse them. Information from the captured images is often needed for such a purpose.

2. *Optical system enhancement*
   As digital processing is inevitable for our multi-lens architecture, it is possible to accommodate other imaging methods that also require post-processing. For example, phase modification through wavefront coding has been shown to increase image depth of field and reduce color artifacts, and it is possible to take advantage of such development to construct a more powerful optical system as discussed below.

3. *Image restoration*
   The aberrations in the optics and the finite size of the sensor pixels both cause image blurring. Moreover, separators among the lenslets to avoid signal cross-talk can cause noticeable blocks in the HR image after the reconstruction. Sharpening through image restoration is therefore necessary.

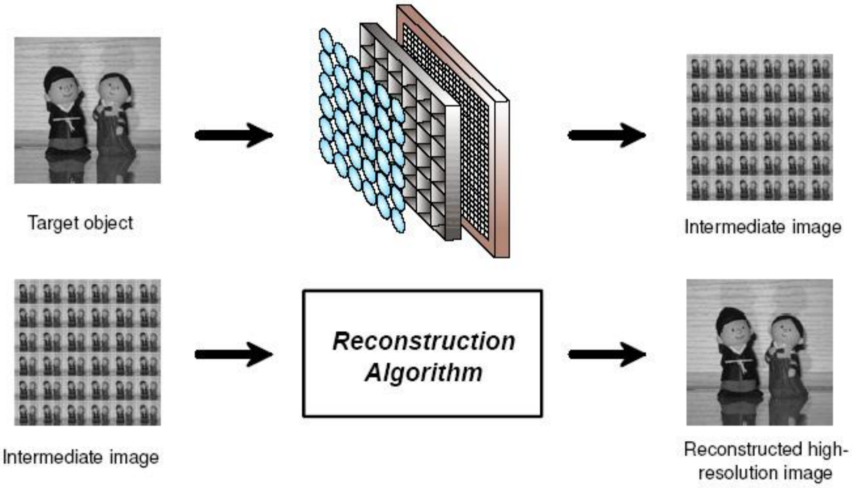


Figure 2. The working principle of the TOMBO system.[13]

4. *Speed of reconstruction*
   Iteration is common for the reconstruction and the restoration processes, so it is necessary to ensure numerical stability and fast convergence. Furthermore, in machine vision inspection there is often a strict time requirement, adding extra strain on the computational efficiency.

3. IMAGE RECONSTRUCTION IN A MULTI-LENS SYSTEM

Substantial computation is needed to reconstruct the HR image from the various LR images. Various methods have been proposed in the literature, including the image-sampling method, backprojection method, pixel-rearrange method, and super-resolution-based reconstruction.

Both the image-sampling method and the backprojection method are proposed along with the early TOMBO system.[9] The former simply finds the signals from the various optical path. The pixel selection is based on the geometric relationship between the object and the individual photodetector arrays. The latter formulates the imaging process as a linear system and attempts to invert it via truncated singular value decomposition (TSVD). It was, however, later pointed out that neither delivered satisfactory performance in practice,[12] and the pixel-rearrange method is considered an improvement. The idea is to map the pixels in each imaging unit into a virtual image plane consisting of fine pixels, and neglect the physical size of the original pixels. Interpolation is then performed to fill in the remaining pixels, and image filtering is used to compensate for the blurring and diffraction of the lenses.[12]

This approach lends itself to the more general image super-resolution framework.[15] Consider the mathematical modeling in Figure 3. Let $g(x, y)$ represent the object sampled at a fine grid. It is the target of our reconstruction algorithm. Assume that the N^2 lenslets are indexed by the subscript k. Lenslet k has a relative displacement with respect to each other, represented by $m_k(x, y)$. Each lenslet also produces a blurring $h_k(x, y)$, which can encapsulate both the optical transfer function (OTF) from the optics and the finite pixel size of the sensors. An accurate computation of the OTF is critical for a good modeling of the imaging system, especially in a multi-lens system where aberration would cause much difficulty in image registration. For diffraction-limited imaging under incoherent light, the OTF is expressed as a function of the pupil, according to the formula[2]

$$\mathcal{H}(f_x, f_y) = \frac{\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} P\left(x + \frac{\lambda z_i f_x}{2}, y + \frac{\lambda z_i f_y}{2}\right) P\left(x - \frac{\lambda z_i f_x}{2}, y - \frac{\lambda z_i f_y}{2}\right) dx\, dy}{\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} P(x, y)\, dx\, dy}, \tag{1}$$

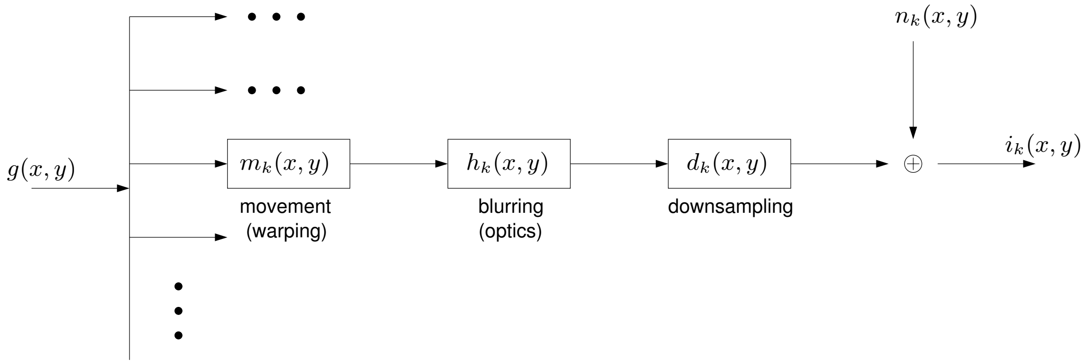


Figure 3. Image super-resolution framework of the multi-lens image acquisition.

where $P(x, y)$ represents the pupil function, λ is the wavelength of the illumination, and z_i is the image distance. Space-invariant aberrations can be incorporated through changing $P(x, y)$. As the image from each lenslet is at a lower resolution than the reconstructed image, we can represent this reduction in resolution through the downsampling $d_k(x, y)$. Noise $n_k(x, y)$ is also added before we capture each image as $i_k(x, y)$. Image reconstruction in a multi-lens architecture is equivalent to estimating $g(x, y)$ as accurately as possible based on the set of $i_k(x, y)$.

Theoretically, a HR image can be obtained from such a collection of LR images because each LR contributes additional information. This is the essence of super-resolution, a topic receiving much attention in recent years.[16] Using lexicographic ordering of the images, we can write the following matrix equation for the image acquisition process:

$$I_k = D_k H_k M_k G + N_k \qquad k = 1, \ldots, N^2. \tag{2}$$

We want to recover G based on the collection of I_k's. If we let $A_k = D_k H_k M_k$, we can see that equation (2) is the same as a conventional model used for image restoration. We can therefore conclude that, just with image restoration, the image reconstruction from multi-lens is also ill-conditioned and requires regularization. Furthermore, a solution to equation (2) can be written as

$$\hat{G} = \arg\min_G \sum_{k=1}^{N^2} \left\{ \|I_k - D_k H_k M_k G\| + \lambda_k \|C_k G\| \right\}, \tag{3}$$

where $\|\cdot\|$ can be L_1 or L_2 or other norm operations such as total variation (TV), λ_k is a regularization parameter and C_k defines the regularization operation.

Many techniques have been developed to actually solve the above equation. Most of them are iterative in nature, such as using alternating projections. A lot of attention has also be paid to the numerical stability and computational efficiency in the iterations. Readers are referred to articles such as Ref. 16 for details. As an example, Figure 4 shows a simulation with our implementation of the super-resolution algorithm described in Ref. 13, compared with the pixel-rearrange method.

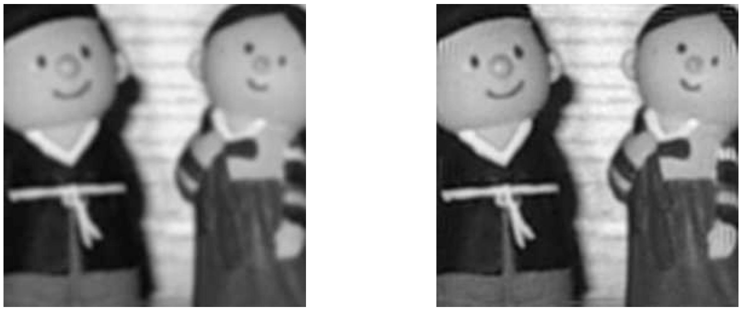
4. APPLICATIONS

Several applications have been explored for the multi-lens systems. The following is by no means exhaustive:

- *Target tracking*
  The goal is to track a well-defined small object that transverse a large field of view. This object may move from far field to near field. A multi-lens system has been built and evaluated for this purpose.[8]

- *Fingerprint identification*
  A multi-lens system has the advantage of being very thin with a large field of view, and this is ideally

(a) Original object (b) Captured image (c) One sub-image



(d) Pixel-rearrange reconstruction (e) Super-resolution reconstruction

Figure 4. Comparison of super-resolution reconstruction and pixel-rearrange method.

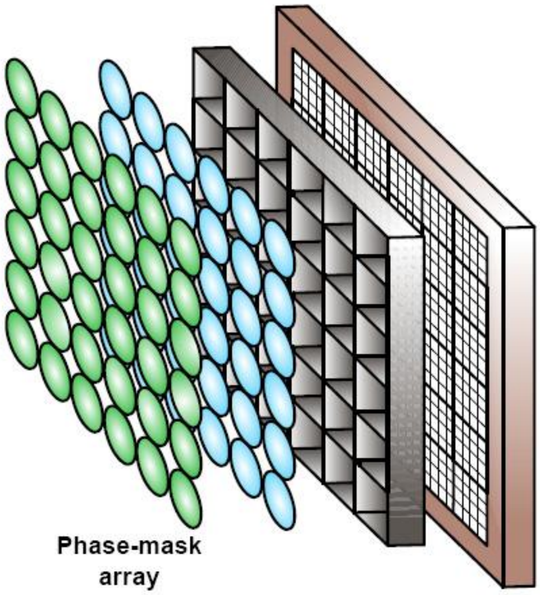
suitable for fingerprint identification systems. However, the object is also located very near the lenses, and therefore each imaging unit cannot capture the entire object. Instead, the system has to operate much like the apposition compound-eye structure,[17] i.e., each unit observes a different part of the object. The separation layer depicted in Figure 1 continues to function as avoiding cross-talk, and the images from each unit are pieced together, after removing the peripheral of the images that result from shading of the separation layer, to form the entire fingerprint image.[18] It is expected that some distortion corrections may be necessary.[19]

- *Depth of field extension*
  The idea is that the different imaging units, which give rise to the respective LR images, can be separately considered and modified. This gives rise to a lot of extra flexibility in the optical system. One instance of which is to incorporate the wavefront coding technique. It has been shown that by deliberately introducing phase masks, of which a cubic one has been shown to be quite effective, it is possible to alter both the in-focus and defocus OTF so that they resemble each other. Thus, the same restoration filter can be applied to bring them back in focus. This can work for a much wider range of defocus than is normally possible with conventional optical design, and thus is considered a method for extending the depth of field of the optical system.[3,4]

  We can apply the phase masks to the imaging units in a multi-lens architecture as shown in Figure 5(a). As a result, the modeling of each of the optical path is changed as shown in (b). The extra $\hat{H}_k$ represents the phase mask, and the optical path produces an intermediate image $\hat{I}_k$. Restoration filter represented by R_k is needed for post-processing. Figure 6 is a simulation of a scene where a cup is normally in-focus a book is out-of-focus. In (a), no phase mask has been used, and we can see that the book is severely blurred. In comparison, in (b) both objects appear sharp.

5. CONCLUSIONS

In this paper we have highlighted the recent efforts in integrated computational imaging systems pertaining to the design of multi-lens architecture. The primary advantage of such a system is flexibility and wide field of

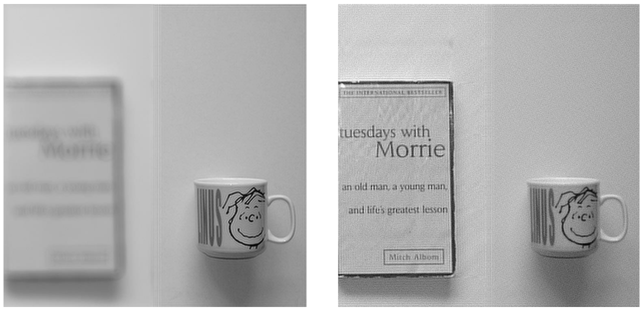


(a) Modification of the TOMBO system with the incorporation of phase masks.[20]

$$\begin{tikzcd}
 & & & & & N_k \arrow[d] & & & \\
G \arrow[r] & \boxed{M_k} \arrow[r] & \boxed{\hat{H}_k} \arrow[r] & \boxed{H_k} \arrow[r] & \boxed{D_k} \arrow[r] & \bigoplus \arrow[r] & \hat{I}_k \arrow[r] & \boxed{R_k} \arrow[r] & I_k
\end{tikzcd}$$

(b) Modeling of an optical path of the phase-mask-incorporated multi-lens system.

Figure 5. A multi-lens architecture with phase masks in the optical paths.



(a) No phase mask (b) With phase mask

Figure 6. Comparison of multi-lens system with and without phase masks using super-resolution reconstruction.[21]

view, but further improvements in image reconstruction will be needed to advance this technology. Some machine vision applications have been explored, and it is expected that more can be found with specially adapted designs.

ACKNOWLEDGMENTS

The work described in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Number 714305). The effort of W.-S. Chan in analyzing multi-lens systems and supplying the simulation results is also gratefully acknowledged.

REFERENCES

1. J. N. Mait, R. Athale, and J. van der Gracht, "Evolutionary paths in imaging and recent trends," *Optics Express* **11**, pp. 2093–2101, September 2003.
2. J. W. Goodman, *Introduction to Fourier Optics*, Roberts and Company, Colorado, third ed., 2004.
3. E. R. Dowski and W. T. Cathey, "Extended depth of field through wave-front coding," *Applied Optics* **34**, pp. 1859–1866, April 1995.
4. W. T. Cathey and E. R. Dowski, "New paradigm for imaging systems," *Applied Optics* **41**, pp. 6080–6092, October 2002.
5. J. S. Sanders and C. E. Halford, "Design and analysis of apposition compound eye optical sensors," *Optical Engineering* **34**, pp. 222–235, January 1995.
6. K. Hamanaka and H. Koshi, "An artificial compound eye using a microlens array and its application to scale-invariant processing," *Optical Review* **3**(4), pp. 264–268, 1996.
7. J. Duparré, P. Schreiber, P. Dannberg, T. Scharf, P. Pelli, R. Völkel, H.-P. Herzig, and A. Bräuer, "Artificial compound eyes — different concepts and their application to ultra flat image acquisition sensors," in *MOEMS and Miniaturized Systems IV*, *Proceedings of the SPIE* **5346**, pp. 89–100, January 2004.
8. R. Hornsey, P. Thomas, W. Wong, S. Pepic, K. Yip, and R. Krishnasamy, "Electronic compound-eye image sensor: Construction and calibration," in *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications V*, *Proceedings of the SPIE* **5301**, pp. 13–24, January 2004.
9. J. Tanida, T. Kumagai, K. Yamada, S. Miyatake, K. Ishida, T. Morimoto, N. Kondou, D. Miyazaki, and Y. Ichioka, "Thin observation module by bound optics (TOMBO): Concept and experimental verification," *Applied Optics* **40**, pp. 1806–1813, April 2001.
10. M. Land, "The optics of animal eyes," *Contemporary Physics* **29**, pp. 435–455, September/October 1988.
11. N. A. Ahuja and N. K. Bose, "Design of large field-of-view high-resolution miniaturized imaging system," *EURASIP Journal on Advances in Signal Processing* **2007**, p. 59546, 2007.
12. Y. Kitamura, R. Shogenji, K. Yamada, S. Miyatake, M. Miyamoto, T. Morimoto, Y. Masaki, N. Kondou, D. Miyazaki, J. Tanida, and Y. Ichioka, "Reconstruction of a high-resolution image on a compound-eye image-capturing system," *Applied Optics* **43**, pp. 1719–1727, March 2004.
13. W.-S. Chan, E. Y. Lam, M. K. Ng, and G. Y. Mak, "Super-resolution reconstruction in a computational compound-eye imaging system," *Multidimensional Systems and Signal Processing* **18**, pp. 83–101, September 2007.
14. W.-S. Chan, E. Y. Lam, and M. K. Ng, "Investigation of computational compound-eye imaging system with super-resolution reconstruction," in *International Conference on Acoustics, Speech, and Signal Processing*, **4**, pp. 1177–1180, May 2006.
15. M. K. Ng, H. Shen, E. Y. Lam, and L. Zhang, "A total variation regularization based super-resolution reconstruction algorithm for digital video," *EURASIP Journal on Advances in Signal Processing* **2007**, p. 74585(16p), June 2007.
16. M. Ng, E. Y. Lam, and C.-S. Tong, "Superresolution imaging: Theory, algorithms and applications," *Multidimensional Systems and Signal Processing* **18**, pp. 57–58, September 2007.
17. J. Duparré, P. Dannberg, P. Schreiber, A. Bräuer, and A. Tünnermann, "Artificial apposition compound eye fabricated by micro-optics technology," *Applied Optics* **43**, pp. 4303–4310, August 2004.
18. R. Shogenji, Y. Kitamura, K. Yamada, S. Miyatake, and J. Tanida, "Bimodal fingerprint capturing system based on compound-eye imaging module," *Applied Optics* **43**, pp. 1355–1359, February 2004.

19. S. T. Suen, E. Y. Lam, and K. K. Wong, "Photographic stitching with optimized object and color matching based on image derivatives," *Optics Express* **15**, pp. 7689–7696, June 2007.
20. W.-S. Chan, "Design and analysis of integrated computational imaging systems," Master's thesis, The University of Hong Kong, 2007.
21. W.-S. Chan, E. Y. Lam, and M. K. Ng, "Extending the depth of field in a compound-eye imaging system with super-resolution reconstruction," in *International Conference on Pattern Recognition*, **3**, pp. 623–626, August 2006.