Overall Well-Focused Catadioptric Image Acquisition With Multifocal Images: A Model-Based Method

Weiming Li and Youfu Li, *Senior Member, IEEE*

***Abstract*— When a catadioptric imaging system suffers from limited depth of field, a single image cannot capture all the objects with clear focus. To solve this problem, a set of multifocal images can be used to extend the depth of the field by fusing the best-focused image regions to an overall well-focused composite image. In this paper, we propose a novel model-based method that avoids the computational cost problem of previous extended-depth-of-field algorithms. On the basis of the special optical geometry properties of catadioptric systems, the proposed model describes the shapes of the best-focused image regions in multifocal images by a series of neighboring concentric annuluses. Then we propose a method to estimate the model parameters. On the basis of this model, an overall well-focused image is obtained by combining the best-focused regions with a fast and reliable online operation. Experiments on catadioptric images of a variety of different scenes and camera settings verify the validity of the model and the robust performance of the proposed method.**

***Index Terms*— Catadioptric system, multifocal images, well-focused image.**

I. Introduction

BEING able to capture a wide field of view, catadioptric imaging systems consisting of a curved mirror and a dioptric camera have become popular in various applications [1]. Most existing work on catadioptric systems have focused on mirror design [2], calibration [3], or applications [4]. Relatively little attention has been paid to improving the image quality in terms of focus (or reducing defocus blur), which is the interest of this paper.

As in other image-forming optical devices, acquiring well-focused images is important in catadioptric systems. The issues on focusing with a traditional dioptric camera have been long studied in optics and it is known that only the objects located within the depth of field (DOF) appear well focused [5]. The difference with a catadioptric system is that, instead of capturing the real images of the world, the dioptric camera captures the virtual images of the world formed by the reflection of the curved mirror. The range of space where the virtual images reside has been studied [2], and has recently been explicitly derived [6] and named as caustic volume. In theory, if the DOF of the dioptric camera used in the catadioptric system contains the entire caustic volume, all objects in the world would appear clearly focused in a single catadioptric image. In practice, however, there are many situations where the camera's DOF is not sufficiently wide. One class of examples is the compact catadioptric systems [7]. In these systems, the cameras are typically mounted at a close distance to the mirror. In optics [5], a close object distance leads to a small DOF. Other cases include cameras that have to work with large apertures to allow efficient exposure, where a large aperture also leads to a small DOF [5]. In these situations, object points whose virtual features are beyond the DOF would appear out-of-focus. Such a phenomenon is illustrated by an example catadioptric image in Fig. 1(c) taken

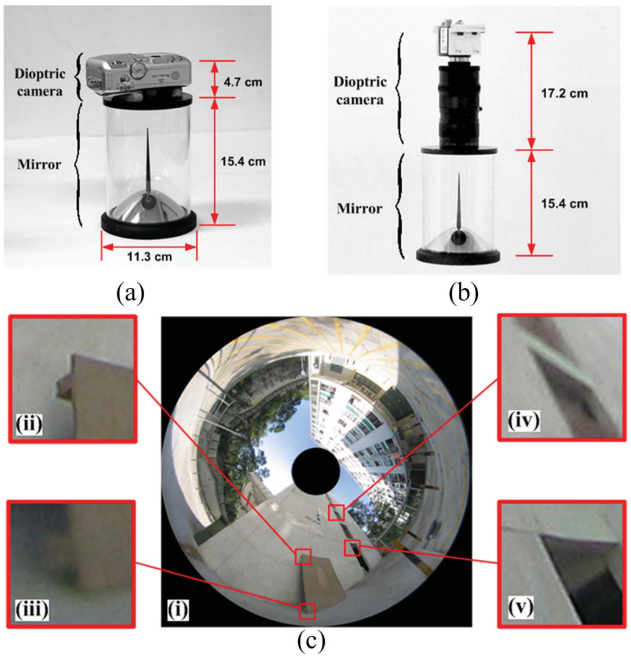

Fig. 1. (a) and (b) Two catadioptric systems used in this paper. (c) Example catadioptric image taken by the system in (a). As the DOF of the dioptric camera is not wide enough, a close inspection reveals that while some image regions are well focused, such as in (c.ii) and (c.v), some other image regions containing similar objects are defocused, such as in (c.iii) and (c. iv).

Manuscript received March 5, 2010; revised October 10, 2011; accepted March 3, 2012. Date of publication April 17, 2012; date of current version July 18, 2012. This work was supported in part by a grant from the Research Grants Council of Hong Kong under Project CityU118311. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Lina J. Karam.

W. Li was with the City University of Hong Kong, Kowloon, Hong Kong. He is now with the Samsung Advanced Institute of Technology, China Laboratory, Beijing 100125, China (e-mail: weiming.li@samsung.com).

Y. Li is with the City University of Hong Kong, Kowloon, Hong Kong (e-mail: meyfli@cityu.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2012.2195010

by a system shown in Fig. 1(a). It can be seen that while some image regions are well focused, some other image regions are not focused as well. This leads to difficulties in both subsequent computer processing procedures and human visual inspections. In particular, this hinders applications where well-focus property is desired over the entire view such as wide area inspections and large-scale 3-D measurements.

Since a single image acquisition with small DOF is not able to capture an overall well-focused image, a possible alternative is to take a stack of images with different focal settings and combine the well-focused pixels in each image through image fusion. These methods are often referred to as extended-depth-of-field (EDF) algorithms and have been investigated for dioptric cameras in bright-field microscopy and close distance photography [8]–[19]. Given a stack of multifocal images $I = \{I_i(x, y)\}$, the task of an EDF algorithm is to determine the best-focused image index i for every pixel (x, y) in the image domain. For this, a focus measurement is needed to evaluate the degree of focus at each pixel. Then, the best-focused pixels are selected and combined into a composite image. Depending on how the focus measurement is formed, the existing EDF methods can be roughly categorized into pixel-based methods, region-based methods, and multi-resolution-based methods. Pixel-based methods represent early EDF approaches [8], [9]. Despite some success in certain areas, their performances are shown to be limited [10], since the defocus effect is essentially defined on image region. Regions-based methods take an image region around each pixel into account and calculate the focus measurements that reflect high-frequency signal component in the region, such as local variance (or contrast) [11]–[13], high-order statistics [14], and edge filter outputs [15]. As they are easy to implement, these methods are still used in some software nowadays. To date, multi-resolution-based methods are among the most successful EDF methods. These methods first transform all the images using multi-resolution analysis methods such as steerable pyramids [16] or wavelet [10], [17], and then perform image fusion in the transformed space. Such methods automatically process image features at multiscales and avoid the artifacts induced by using fixed-size image windows. For most existing EDF methods, the constraint of local spatial continuity is applied to optimize the result in various forms such as low-pass filtering [11], consistency check [10], [17], or region merging [14]. Some other recent works also employ image segmentation techniques [12], [13], [18] or employ parametric image formation models [19] for further improvements.

All the EDF algorithms outlined above are applied with dioptric cameras. In dioptric cameras, the shape of a well-focused image region is dependent on the unknown 3-D scene structure and may appear at any unpredictable image location. Therefore, the focus measurement needs to be evaluated on every pixel in all the images and the optimization procedure needs to be performed over the entire image. This leads to a prohibitive computational load and memory consumption. In practical applications, catadioptric systems (especially the compact ones) are often required to perform fast with limited computational resources. In such situations, these EDF methods encounter limitations due to their computational cost.

In general, this paper also takes an image processing approach using multifocal images. However, unlike the previous EDF algorithms, we explore the special optical geometry properties of catadioptric systems and propose a novel model-based method that can avoid the computational cost problem. To the best of the authors' knowledge, this is the first attempt to employ such properties in combining multifocal catadioptric images, which makes the work different in the following ways.

1) *On the basis of the spatial distribution property of virtual features in catadioptric systems, we propose that the shapes of the best-focused regions in a set of M multifocal catadioptric images can be modeled by M neighboring concentric annuluses, all of which are centered at the image center:* This model is essentially different from that of the traditional dioptric cameras. Remember that for images acquired by dioptric cameras, the shapes of well-focused regions are dependent on the specific 3-D structure of the scene and cannot be predicted using such primitive geometric shapes. On the basis of this model, determination of the best-focused image regions in catadioptric systems is simplified to estimating the model parameters (the radii of the set of annuluses). In this paper, a set of partly well-focused (PWF) images are automatically identified. Then we propose a method to estimate the model parameters by evaluating the degree of focus for a set of image circles.
2) *The parameters of the proposed model are independent of the 3-D scene structure:* This feature allows the same model parameters to be used across different scenes. Based on this model, our method is implemented in two stages: an offline model construction stage and an online image operation stage. The model construction stage is first performed and the obtained model parameters are stored. Then in the image operation stage, the known model parameters are directly used to extract the best-focused image regions and combine them into an overall well-focused image. Since no additional computation is involved in the image operation stage, the online operation is very fast in practical applications. In contrast, all previous EDF methods need to perform all the computation steps online for each new scene.

Our method does not need the mirror shape or the system parameters to be known. The focal distance settings for the set of multifocal images are not involved in the computation either, which makes this algorithm highly automatic. Experiments on catadioptric images of a variety of scenes verify the validity of the proposed model. Experimental comparisons with previous EDF methods using a variety of DOF settings show consistent results. Yet our method is much faster in computation. These results suggest an efficient way to obtain the overall well-focused catadioptric images and they will inspire potential works in designing high-quality compact catadioptric image sensors.

The remainder of this paper is organized as follows. Section II discusses how to model the well-focused image regions in catadioptric systems and how to combine them to acquire an overall well-focused image. Section III describes a method for extracting the set of PWF images and a method for

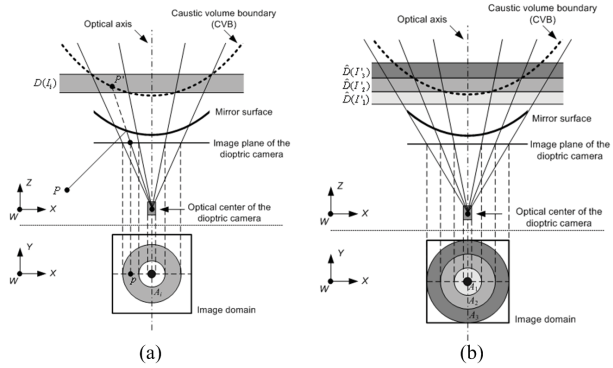

Fig. 2. Image formation process of well-focused image regions in a catadioptric system. (a) Shape of a well-focused image region in a single catadioptric image. (b) Shapes of the best-focused image regions in multifocal catadioptric images.

estimating the model parameters. Experiments on catadioptric images are presented in Section IV. Section V gives some concluding remarks.

II. Models of Well-Focused Image Regions in Catadioptric Systems

The basic principle of our method is to combine the well-focused image regions in the set of multifocal images into an overall well-focused composite image. For this purpose, a key issue is to determine the shapes, sizes, and locations of the well-focused image regions. By exploring the optical geometry properties in catadioptric image formation, we find that the shapes of well-focused image regions in catadioptric systems can be modeled by well-defined geometry shapes. In the following, we first propose a model that describes the shape of a well-focused image region in a single catadioptric image. Then we extend this model to describe the best-focused image regions in multifocal images. On the basis of the proposed model, we describe a method to combine these best-focused image regions into an overall well-focused image. Finally, a brief discussion on the non-coaxial installation issue is given.

A. Shape of a Well-Focused Image Region in a Single Catadioptric Image

The image formation process of a typical catadioptric system consisting of a curved mirror and a dioptric camera is illustrated in Fig. 2(a). Assume a Euclidean world coordinate $X-Y-Z$, where the Z-axis is parallel with the optical axis of the catadioptric system. As the system is rotationally symmetric with the optical axis, the analysis can be conducted in a 2-D profile as shown in Fig. 2. The upper part of Fig. 2(a) (separated by a horizontal dash line) shows the 2-D profile of the system on the $X-Z$ plane. The dioptric camera looks upward at the mirror surface. The lower part of Fig. 2(a) shows another view of the image plane of the dioptric camera on the $X-Y$ plane. An object point P in the world is first reflected by the mirror surface to form a virtual feature P'. Then the dioptric camera captures this virtual feature to form an image point p. Other terms and annotations in Fig. 2(a) are explained later in this section.

It is known that only the object within the DOF of a dioptric camera can be well focused. In a catadioptric system, the well-focused object points are the ones whose virtual features are located within the DOF. Following this, the shapes of the well-focused regions are dependent on the locations of the virtual features and the range of DOF. Recently, Swaminathan [6] explicitly derives the locations of virtual features formed by the conic curved mirrors and finds that the virtual features of all object points are located within a finite space, named as caustic volume. As presented in the work [6], caustic volume is the space between the mirror surface and a virtual surface. The virtual surface is referred to as the caustic volume boundary (CVB), as shown in Fig. 2(a).

In this paper, we further find that the spatial density distribution of virtual features within the caustic volume is not uniform, and the great majority of the virtual features are located on the CVB. On the basis of the math formulation of image formation [6], we explicitly compute the locations of virtual features in a variety of quadric mirror–based central catadioptric systems, whose mirror eccentricities range from 0.8 to 1.2, heights from the mirror apex to the camera lens range from 5.0 to 115.0 mm, and the diameters of the mirror surface range from 5.0 to 135.0 mm. This set of parameters is typical for catadioptric sensor products, such as those produced by ACCOWLE Co. Ltd [20]. According to our computation, as an object point moves farther away from the system, its corresponding virtual feature would rapidly approach the CVB. For the tested set of system parameters, it is found that when an object point is more than 1 m away from the system, it will have its virtual feature located within a narrow neighborhood of the CVB, whose width is 5% of the distance from CVB to the mirror surface. In real applications with catadioptric systems, as the points of interests are located at a distance, the great majority of the virtual features can be considered to be located on the CVB.

Therefore, in a catadioptric system, the well-focused virtual features are located on the intersection of the CVB and the DOF of the dioptric camera. Denote the CVB as $V \subset \mathbb{R}^3$ and the DOF of an image I_i as $D(I_i) \subset \mathbb{R}^3$. Let the focal distance of image I_i be $f(I_i)$. In optics [5], we know $D(I_i)$ can be modeled as a space between the two parallel planes that are perpendicular to the optical axis, where $D(I_i)$ is represented as a gray region in Fig. 2(a). When the DOF is wide enough to contain the entire CVB, I_i would appear to be overall well focused as long as $f(I_i)$ is properly set, so that $V \subset D(I_i)$.

However, as mentioned in Section I, there are still many cases where $D(I_i)$ is not wide enough. In these cases, only the part of CVB that is within $D(I_i)$ is well focused. As the system is rotationally symmetric, the set of pixels corresponding to the mirror surface is a circular image region centered at the image center. We denote this circular image region as the image domain $s(I_i)$. Within the image domain, the projection of the space $D(I_i) \cap V$ onto $s(I_i)$ is an annulus as shown in the lower part of Fig. 2(a). Therefore, this paper uses this annulus to model the shape of a well-focused image region in I_i and denotes it as $A(I_i)$. The annulus $A(I_i)$ can be defined as the image region between the two concentric circles with radii $\hat{r}_{i-1}$ and $\hat{r}_i$, so that $A(I_i) = \{p | \hat{r}_{i-1} \leq r(p) \leq \hat{r}_i, p \in I_i\}$, where

$r(p)$ represents the distance from an image pixel $p \in I_i$ to the image center.

Let $F(p; I_i)$ be a function that indicates the degree of focus of a pixel $p \in I_i$. For dioptric cameras, it is known by optics [5] that the image feature of an object point becomes more defocused as the point moves away from the focal plane along the optical axis. Though an explicit form for this variation of focus measurement can be accurately deduced by optical geometry, in practice it is often modeled by a Gaussian function peaked at the focal plane [21], [22]. Our work also follows such an approximation. In catadioptric systems, as the images are formed through mirror reflection, the variation of focus along the optical axis is mapped onto the radius of the image plane through the mirror. Since the mirror surface is smooth and the DOF considered in this paper is small, this mapping can be locally approximated by a linear mapping. Following this, we still adopt a Gaussian function to model the variation of focus measurement along the image radius and give the form of $F(p; I_i)$ as follows:

$$F(p; I_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{(r(p) - \mu_i)^2}{2\sigma_i^2}\right). \quad (1)$$

As the variation is only along the radius, we rewrite (1) as

$$F(r; I_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{(r - \mu_i)^2}{2\sigma_i^2}\right) \quad (2)$$

where r is the distance from the pixel p to the image center. Note that since CVB and DOF are determined only by the optical configuration of the system, $A(I_i)$ is independent of the 3-D scene structures. This feature is essentially different from that in a dioptric system. In a dioptric system, the shapes of the well-focused image regions are closely related to the object depths, which is the theoretical basis for the class of shape/depth from focus/defocus methods [21], [22].

It can be seen that when the focal distance $f(I_i)$ increases, $D(I_i)$ moves from the front to the back of the CVB, during which $A(I_i)$ moves from the image central part to the peripheral part. However, as long as $D(I_i)$ cannot contain the entire CVB, the image regions beyond $A(I_i)$ would still appear defocused. In order to obtain an overall well-focused image with narrow DOF images, we next study the shapes of the best-focused image regions in multiple-focal images.

B. Shapes of the Best-Focused Image Regions in Multifocal Catadioptric Images

Let $I = \{I_i\}$ be a set of N multifocal catadioptric images, which are taken at the same view point for the same scene, yet with different focal distance settings $\{f(I_i)\}$, where $f(I_1) < f(I_2) < \cdots < f(I_N)$. Here, we assume that the dioptric camera is equipped with the image-space (rear) telecentric feature [23], which guarantees that the scene content in each image is exactly the same though the image focal distances are different. Note that this assumption is not a must to use our method. When the camera does not satisfy the image-space telecentric feature, the multifocal images can be aligned to the same coordinate by a standard image registration procedure as a preprocessing step.

In order to make all objects clearly recorded in at least one of the images, the following two conditions are required while taking the set of multifocal images I.

Condition 1: Any two images whose focal distance settings are adjacent should have overlapping DOFs so that

$$D(I_p) \bigcap D(I_q) \neq \varnothing \quad \text{for} \quad \forall I_p, I_q \in I, \quad q = p + 1.$$

Condition 2: The union DOF of the images should contain the entire CVB so that $V \subset \bigcup_{i=1}^{N} D(I_i)$.

Condition 1 can be satisfied by making the focal distance step sufficiently small. Condition 2 can be met by letting $D(I_1)$ locate before V and $D(I_N)$ locate behind V. This can be examined by visually checking whether both I_1 and I_N are overall out-of-focus images. Note that the above image acquisition procedure does not require the system parameters to be known. The focal distances of the images do not need to be accurately set either.

As the goal is to obtain an overall well-focused image, the images of interests are a subset of K PWF images $I' = \{I'_k\} \subset I$ that satisfy $D(I'_k) \cap V \neq \varnothing$, where $D(I'_k)$ denotes the DOF of I'_k. Also assume that $f(I'_1) < f(I'_2) < \cdots < f(I'_K)$. According to condition 1 and condition 2, it can be seen that $V \subset \bigcup_{k=1}^{K} D(I'_k)$. Remember that the image domain $s(I')$ is a circular image region centered at the image center. Therefore, any part in $s(I')$ would be clearly recorded in at least one of the images in I'. Notice that, as CVB is a limited space [6], K can be expected to be a limited number.

The number of images in the PWF image set is determined by the value of the focal distance step. In general, a larger focal distance step would lead to a smaller number of multifocal images, which is efficient to capture. However, when the focal distance step is too large, there would be gaps between the DOFs. To avoid this, the focal distance step should be set small enough so that neighboring multifocal images share the overlapping DOFs as required in condition 1. As long as condition 1 and condition 2 are satisfied, a decrease of the focal distance step (with a larger number of images in the PWF image set) would not further improve the final image quality.

It is known from optics [5] that the degree of focus within $D(I'_k)$ decreases monotonically as an object point moves away from $f(I'_k)$. Therefore, each $D(I'_k)$ contains a subspace $\hat{D}(I'_k) \subset D(I'_k)$, where a point is best focused in I'_k among all the K images in I'. The set of $\{\hat{D}(I'_k)\}$ satisfy: 1) $\hat{D}(I'_p) \bigcap \hat{D}(I'_q) = \varnothing$ for $\forall I'_p, I'_q \in I'$ and 2) $V \subset \bigcup_{k=1}^{K} \hat{D}(I'_k)$. Therefore, the projection of $\hat{D}(I'_k) \cap V$ forms the best-focused image region in image I'_k. As the system is rotationally symmetric, this region is an annulus, which we denote by A_k. Consider the K images in I', the best-focused image regions thus can be modeled by a set of K neighboring concentric annuluses $A = \{A_k\}$, where $\bigcup_{k=1}^{K} A_k = s(I')$ and $A_p \cap A_q = \varnothing$, for $\forall A_p \neq A_q$, $A_p, A_q \in A$. The annuluses $A = \{A_k\}$ are shown in Fig. 2(b).

Each annulus A_k can be described by the two concentric circles enclosing it, whose radii are r_{k-1} and r_k. Therefore, the model $A = \{A_k\}$ can be parameterized as a set of $(K + 1)$ radii $\{r_0, r_1, r_2, \ldots, r_K\}$, where $r_0 < r_1 < r_2 < \cdots < r_K$.

Here, r_0 is the radius of the circular area in the image central part where the scene is occluded by the camera itself, which is plotted as the black circle in Fig. 2(b). r_K is the radius of the circular view boundary of the mirror surface in the image. As r_0 and r_K are determined by the system setup, there are $(K-1)$ variable parameters, which are denoted as $R = \{r_1, r_2, \ldots, r_{K-1}\}$. As introduced in Section II-A, the degree of focus in image I'_k can be described by the Gaussian function $F(r; I'_k)$. Therefore, it can be seen that r_k is the solution to the equation $F(r; I'_k) = F(r; I'_{k+1})$.

In this paper, we propose a practical approach in Section III to compute the model parameters R, which only require observing a set of multifocal images and image processing computations.

C. Combining Multifocal Images Into a Single Overall Well-Focused Image

Given a set of multifocal images I' and the model parameter R, an overall well-focused image $\hat{I}$ can be simply obtained by combining the best-focused regions of I'_k, so that $\hat{I}(A_k) = I'_k(A_k)$. In practice, this can be simply implemented by a cropping and replacing operation, which does not require additional image processing computations and thus is very fast. Note that as R is only determined by the system setup and focal distance settings of I'_k, the same R is applicable for images of arbitrary scenes despite of their different 3-D structures, which is quite different from the situations in a conventional dioptric camera. We verify this property by experiments on catadioptric images in various environments in Section IV-B and with various camera DOF settings in Section IV-C.

D. Non-Coaxial Installation Issue

In the above discussions, we have assumed that the camera and the mirror are installed to be coaxial. Note that there are also situations when the installation is not coaxial. Furthermore, there are some systems where the mirror is intentionally installed with an angle to the camera such as the work [24]. When the camera's axis has a significant angle of deviation from the mirror's axis, the image formation process of the virtual features reflected in the mirror is not rotationally symmetric any more. The process of virtual feature formation must be reconsidered. Following this, the shape of a best-focused image region in a PWF image may not be an annulus. Then the proposed annulus combination method cannot be applied. To analyze the situations when the mirror axis and the camera axis are not aligned, the analysis needs to include the rotational motion parameters in modeling the virtual feature formation process. The shapes of the best-focused regions would be dependent on the mirror shape and the motion parameters.

III. Proposed Method

The proposed method acquires an overall well-focused image by combining the best-focused regions into multifocal catadioptric images using the parametric model introduced in Section II. Since the model parameters are independent of the environment, the model construction work can be performed offline. Therefore, the proposed method is implemented with an offline model construction stage and an online image operation stage, which is summarized by the diagram in Fig. 3. In the model construction stage, the camera first takes a set of multifocal images $I = \{I_i\}$, following condition 1 and condition 2 in Section II-B. Then a set of PWF images $I' = \{I'_k\}$ (defined in Section II-B) are identified using a method to be introduced in Section III-A. The camera parameters for $I' = \{I'_k\}$ are recorded, which include the number of PWF images K and their focal settings $\{f(I'_k)\}$. Based on $I' = \{I'_k\}$, the model parameters R are estimated using a method to be introduced in Section III-B. Later, the image operation stage is ready and it is performed whenever the user wants the system to capture an image with clear focus. In the image operation stage, the camera first takes the set of partly focused images according to K and $\{f(I'_k)\}$. Then the best-focused regions in them are cropped and combined into an overall well-focused image using the model parameters R as introduced in Section II-C. It can be seen that, once the model construction stage is done, the obtained parameters can be readily applied in the image operation stage that involves no additional computation. This makes the acquisition of the overall well-focused images extremely efficient.

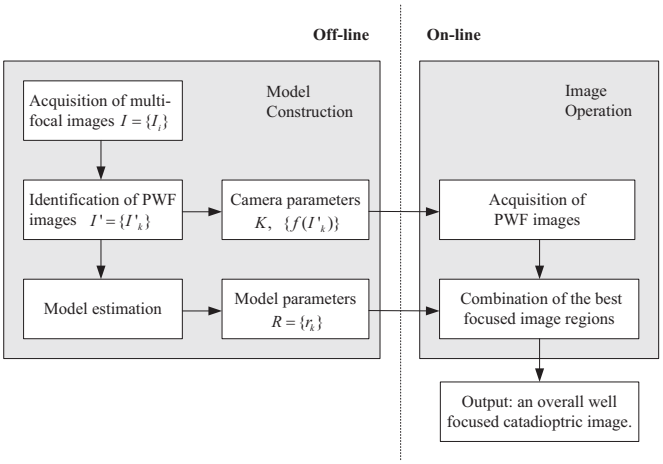

Fig. 3. Flow chart of the proposed method.

A. Identifying the PWF Images

Denote the set of multifocal images as $I = \{I_i\}$, where $i = 1, 2, \ldots, N$. Each image is an $M \times M$ square image, within which the region of interest (the image of the mirror) is a circular region $s(I)$ centered at $[x_c, y_c]$, where $x_c = M/2, y_c = M/2$. Let $s(I)$ be decomposed into a set of concentric circles $C = \{c^{(j)}\}$, $j = 1, 2, \ldots, M/2$, where $c^{(j)}$ is a circle centered at $[x_c, y_c]$ with a radius j. As defined by the model in Section II-B, this paper identifies a PWF image I_k by examining whether it contains a subset of circles that are the best focused in I_k among all the images in I.

To identify the set of PWF images automatically, the degree of focus of a circle $c^{(j)}$ in image I_i needs to be quantitatively evaluated with a focus measurement function, which is

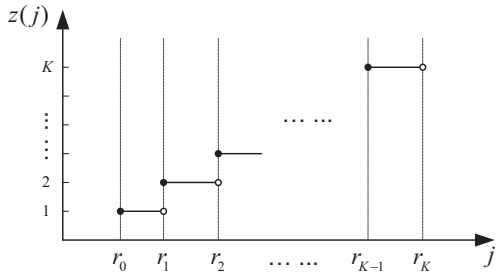

Fig. 4. Index function of the image where an image circle with radius j is the best focused. The discontinuous points correspond to the radii that define the annuluses in the proposed model.

denoted by $\hat{F}(c^{(j)}; I_i)$. In this paper, we use the sum of local gradient magnitude to measure the degree of focus, which is given as follows:

$$\hat{F}\left(c^{(j)}; I_i\right) = \sum_{x=1}^{M}\sum_{y=1}^{M} R\left(x, y; c^{(j)}\right) \cdot G_i\left(x, y\right) \tag{3}$$

where $R(x, y; c^{(j)})$ is a mask image that is defined as follows:

$$R(x, y; c^{(j)}) = \begin{cases} 1, & \text{if } (j-\Delta r) \le \sqrt{(x-x_c)^2+(y-y_c)^2} \le (j+\Delta r) \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Here Δr is the width of a neighborhood around the circle. $G_i(x, y)$ is the gradient magnitude image of image I_i.

On the basis of the above definitions, a circle $c^{(j)}$ is considered to be the best focused in I_k, if $\hat{F}(c^{(j)}; I_k) = \max_{i=1}^{N} \hat{F}(c^{(j)}; I_i)$. Hence, a set of circles that are the best focused in I_k can be identified as $C_k = \{c^{(w)}\} \subset C$. Denote the number of the best-focused circles (BFCs) in I_k as $N(C_k)$, then I_k is selected as a partly focused image if $N(C_k) > 0$.

B. Model Parameter Estimation

Denote the set of PWF images as $I' = \{I'_k\}$, where $k = 1, 2, \ldots, K$. Following the model in Section II-B, the best-focused image regions are K concentric annulus regions $A = \{A_k\}$, where $k = 1, 2, \ldots, K$. As introduced in Section II-B, the set of annuluses are defined by the $(K-1)$ radii $R = \{r_1, r_2, \ldots, r_{K-1}\}$.

Using a similar method as in Section III-A, let the image domain $s(I)$ be decomposed into a set of concentric circles $C = \{c^{(j)}\}$. Following (3), the image in which a circle $c^{(j)}$ is the best focused can be determined. Denote the index of the image, where a circle $c^{(j)}$ is the best focused as $z(j)$. Then $z(j)$ takes values from K integers $\{1, 2, \ldots, \mathrm{K}\}$. Following the model in Section II-B, the radii of the circles that are the best focused in the same image must fall in the same continuous interval. Then $z(j)$ can be considered as a function that represents the index of the image, where a circle with radius j is the best focused among the set of multifocal images. This index function $z(j)$ is illustrated in Fig. 4, where the circle radius range $[r_0\ r_K]$ is partitioned into K intervals. It can be seen that the partition points of these intervals are the parameters of the K annuluses $R = \{r_1, r_2, \ldots, r_{K-1}\}$. Following this, given a set of PWF images, we first obtain the function $z(j)$. Then the model parameters are detected as the discontinuous points of the index function $z(j)$.

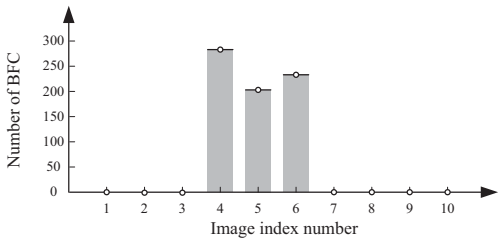

Fig. 5. Number of the BFCs in each input image.

IV. Experiments

A. Acquisition of an Overall Well-Focused Image

In this experiment, our method is applied to a catadioptric system as shown in Fig. 1(a). A Canon PowerShot S50 digital camera is mounted toward a hyperbolic mirror with its optical axis aligned with the mirror axis. The raw image has a resolution of 2592×1922 pixels and is cropped to 1701×1701 pixels to contain only the circular mirror surface, whose radius is 850 pixels. When the F number is 3.2 and the shutter speed is 1/400 s, it is difficult to capture an overall well-focused image with any focal distance setting, which indicates that the DOF is not wide enough to contain the CVB.

In the model construction stage, we first examine the two conditions in Section II-B and take a set of 10 multifocal images $I = \{I_i\}$, whose focal distances are manually set from 10 to 28 cm with an interval of about 2 cm. Note that other number of images and focal distance settings can also be used so long as the two conditions in Section II-B are satisfied. Guaranteed by the image-space telecentric feature [24], the scene contents in the images are identical despite their different focal distances.

The circular image region at the central part which is occluded by the lens has a radius of 130 pixels. Then a set of image circles was used to identify the set of PWF images using the method in Section III-A. The radii of the image circles range from 130 to 850 pixels. The number of the BFCs in each image is shown in Fig. 5. It can be seen that I_4, I_5, and I_6 contain a number of BFCs, and they are selected as the set of PWF images $I' = \{I'_1, I'_2, I'_3\} = \{I_4, I_5, I_6\}$.

Three local regions in the PWF image set I' (Local A, Local B, and Local C) are shown in Fig. 6. Their locations are indicated by square frames overlaid with image I'_1 in the first column. The image contents in I'_1, I'_2, and I'_3 are displayed respectively in the second, third, and fourth columns. It can be observed that local region A is the best focused in I'_1, local region B is the best focused in I'_2, and local region C is the best focused in I'_3. Although each PWF image contains some well-focused image regions, none of these images are overall well focused.

As the three PWF images are involved, our method uses three annuluses to model the best-focused image regions, in

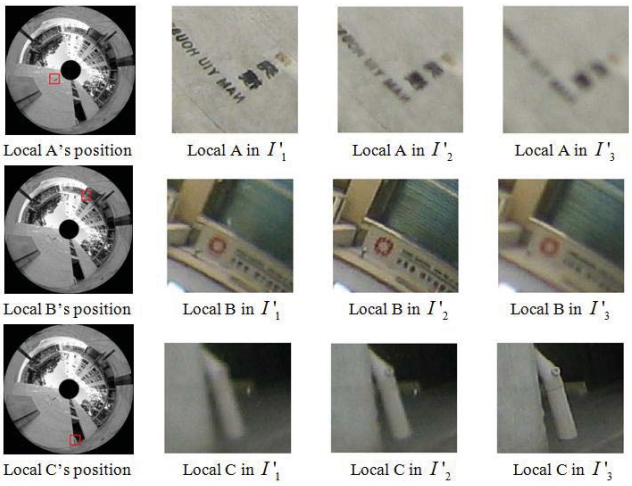

Fig. 6. Local image regions in the PWF image set.

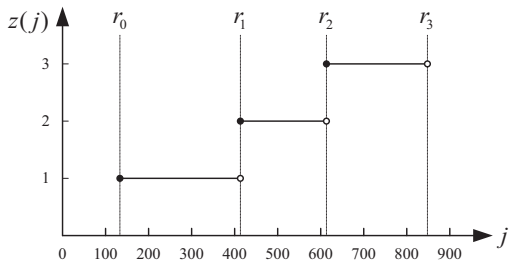

Fig. 7. Model parameters are determined as the discontinuous points of the function $z(j)$, which is the index function of the image, where an image circle with radius j is the best focused.

which two model parameters $\{r_1, r_2\}$ need to be determined. For an image circle, the index of the image, where this circle is the best focused, can be estimated with the method in Section III-A. The result for all the considered image circles is shown in Fig. 7, where $z(j)$ is an index function that represents the index number of the image, where a circle with radius j is the best focused. Then, by detecting the discontinuous points of $z(j)$, the model parameters are $r_1 = 410$ pixels and $r_2 = 609$ pixels.

After the model construction stage, the image operation stage directly takes the set of PWF images and combines the best-focused regions into an overall well-focused image. Using the same parameters, Harris corner extraction algorithm extracts at least 24.1% more corner points from this image than those from any single PWF image.

B. Model Validation in Different Environments

In order to validate the proposed model, the model construction stage is performed in eight different scenes with the same system setting and image capture procedure.

The same catadioptric system is used as in Section IV-A. The camera is set to the focal-bracketing mode. In this mode once the shutter button is pressed, three images are automatically taken with the focal settings changing at fixed intervals. The bracketing center is manually set to 18 cm according to the PWF image parameters obtained in Section IV-A.

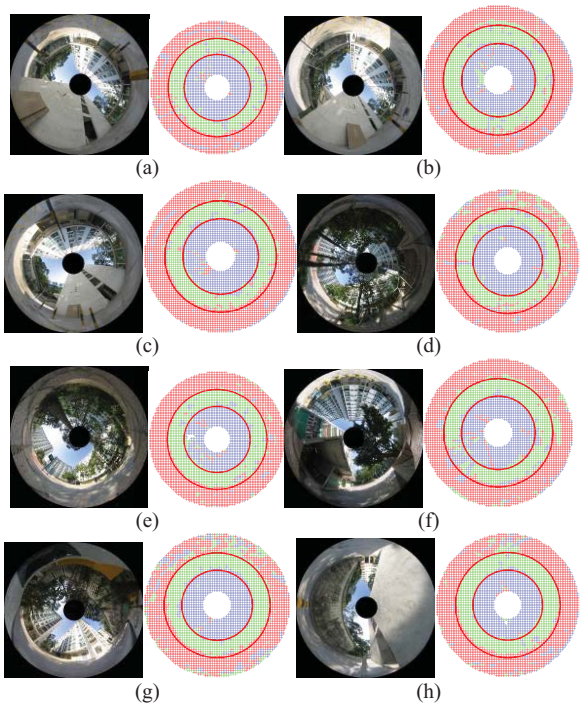

Fig. 8. Model estimation results. Subfigures (a) to (g) show results in eight different scenes respectively. In each subfigure, the left photo is one of the PWF images and the right plot shows the estimated model over a grid of sample points. The color of a sample point relates to a PWF image where the sample point is the best focused. (Best viewed in color.)

Using this mode, it can be guaranteed that the focal settings for each image set in different scenes are identical.

The results are displayed in Fig. 8. Each part figure shows the situation for one of the eight scenes, where the left image is one of the PWF images. The right-hand side image in each subfigure displays a grid of sample points in the image domain. For each of the sample points, the degree of focus of a square image window centered at this point is evaluated by calculating the variance of pixel intensities. Following this, the image where a sample point is the best focused is determined. Then the sample point is plotted to be blue if it is the best focused in I'_1, green if in I'_2, and red if in I'_3. These colors provide a visualization for a point-based estimation of the best-focused image features. Over the sample points, the annuluses of the proposed model determined by our method are displayed. It can be seen that our proposed model is consistent with the point-based estimations in all the eight scenes.

It can be seen that the eight scenes are quite different in the scene contents. Scenes 1–3 are taken with the system between the two neighboring buildings about 20 m apart. In Scene 4, a large part of the image is covered by trees. In Scene 5, the system is in the middle of an open area. Particularly, in Scene 8, the system is located very close to a high wall about 1.5 m apart. However, the shapes of the best-focused regions and the model parameters remain the same. As shown in Table I, the maximum model parameter difference from the average value is 1.33% for r_1 and 0.63% for r_2. Therefore, the model parameters can be deemed to be independent of the 3-D scene structures. This feature guarantees that, once estimated, the model parameters are still applicable for other

TABLE I
MODEL PARAMETER ESTIMATIONS OBTAINED UNDER DIFFERENT SCENES

	r_1	r_2
Scene 1	410	609
Scene 2	404	607
Scene 3	409	604
Scene 4	401	605
Scene 5	402	607
Scene 6	408	604
Scene 7	401	601
Scene 8	402	605
Average	404.6	605.3
Max. deviation	5.4 (1.33%)	3.8 (0.63%)

scenes as long as the imaging parameters do not change. Many modern cameras are able to store the imaging parameters on-chip, which facilitates the use of this method. Therefore, an overall well-focused image can be acquired by cropping and combining the annular regions without additional computation, which makes the procedure extremely efficient.

C. Model Validation With Different DOF Settings

To further validate the proposed model, we apply our method to different sets of multifocal images acquired with different DOF settings and compare the results with a classic EDF method. Here a different dioptric camera is used, which consists of a Prosilica 1394 color CCD camera and a FUJINON zoom lens that allow its optical parameters to be manually set in a more flexible manner than that in the Canon camera.

Four different DOFs are tested by setting the F value of the lens to 5.6, 4.0, 2.8, and 2.0. A set of multifocal images are taken for each DOF with the method in Section II-B. Our method is applied to the four image sets and the obtained overall well-focused images are shown in Fig. 9 as the left image of each part figure. Due to the limited angle of field of the employed dioptric camera, the image covers only a rectangular region inscribed to the circular mirror surface. With the lens apertures increasing, the DOF of the dioptric camera decreases, which aggravates the defocus blur phenomenon. The number of PWF images selected by our method is listed in Table II. It can be seen that, as the DOF becomes smaller, more PWF images are needed to obtain an overall well-focused image.

We compare our method with a classic EDF method based on the local variance (referred here as LV-EDF). For its low computational complexity and reasonable quality, LV-EDF is often used as a reference method for validation in EDF works [19]. Here the result of LV-EDF is shown by a topological map, in which each pixel is rendered by a color associated with the index of image, where the pixel is the best focused. As the local variance focus measurement best works in image regions with sufficient textures, pixels of interests are restricted to the ones whose local variance in the average image is above a threshold. This reduces the number of unreliable pixels in the topological map.

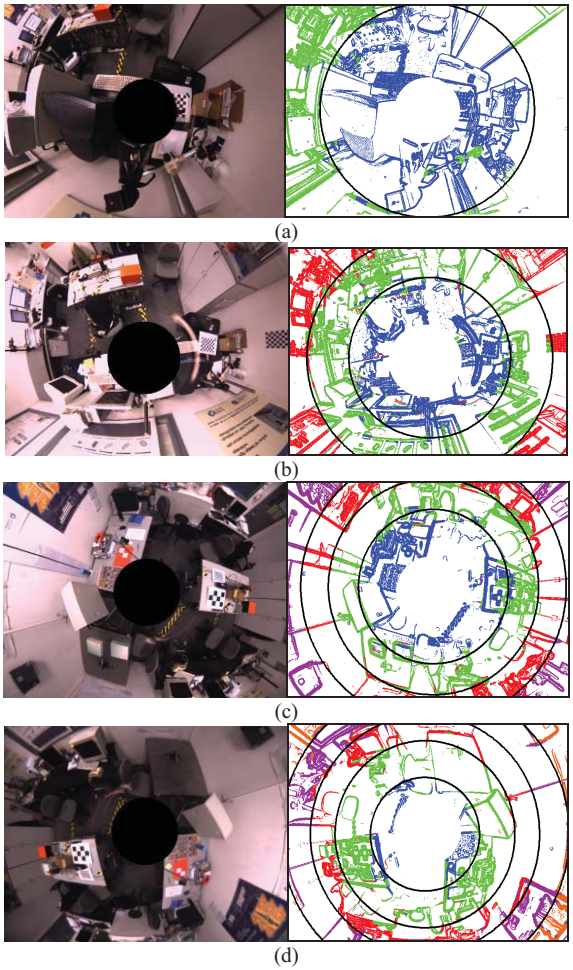

Fig. 9. Experiments on catadioptric images with different DOF settings. In each subfigure (a) to (d), the left photo is one of the PWF images and the right plot shows the estimated model over a set of sample points. The color of a sample point relates to a PWF image where the sample point is the best focused. (Best viewed in color.)

In Fig. 9, the topological map for each image set is shown as the right image of each part figure. Only the PWF images are used in the LV-EDF method, thus the number of colors in each topological map equals to the number of the PWF images. The best-focused regions given by our method are shown by the black circles superimposed on each topological map. It can be seen that the results of the two methods are consistent in all the experiments. The great majority of the pixels contained in an annulus have the same color. Notice that the scene consists of a number of cluttered 3-D structures and the camera also moves across the four image sets. This further verifies that the shapes of the best-focused regions can be well described by our model and they are independent of the 3-D scene structure and camera motion.

We evaluate the complexity of the two methods in terms of computational time as shown in Table II. As local focus measurements are computed for all the pixels of interests, the LV-EDF method requires a prohibitive computational load. More computational loads are involved in recent EDF methods that use more complex focus measurements and post-optimizations. In comparison, the computational load of stage 1 of our method is comparable with the LV-EDF method.

TABLE II
EXPERIMENTAL RESULTS OF LV-EDF AND OUR METHOD WITH DIFFERENT DOF SETTINGS

Aperture size (F)	Number of PWF images	Computation time of LV-EDF (s)	Computation time of our method	
			Stage 1	Stage 2
5.6	2	44.3	37.1 s	0.21 s
4.0	3	56.1	43.1 s	0.33 s
2.8	4	72.2	65.4 s	0.37 s
2.0	5	83.9	74.8 s	0.49 s

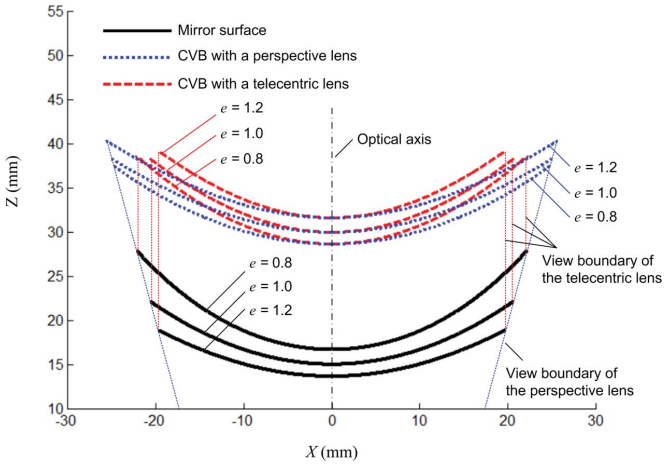

Fig. 10. CVB with different mirrors and cameras. The mirror shape is elliptic when $e = 0.8$, parabolic when $e = 1.0$, and hyperbolic when $e = 1.2$.

However, the computation time of stage 2 of our method is significantly small. Notice that the online computation of stage 2 is the only computation needed in our method after stage 1 is prepared offline. This makes our method perform fast in practical applications.

The experiments were performed on a PC with Intel Core Duo CPU (2.20 GHz) and 2-GB memory. The algorithms were implemented by MATLAB. The original image size is 1360 × 1024 pixels. As the dioptric camera here does not have image-space telecentric property, images are first aligned to the same coordinate by image registration [10], [22].

D. Note on Using Different Mirrors and Cameras

A simulation experiment is performed to examine the model when different kinds of mirrors and cameras are used in the studied catadioptric system.

As stated in Section II of this paper, the shape of the CVB is essential in the formation of the best-focused image regions. Here we simulate the CVB when different kinds of mirrors and cameras are used. The simulation follows the math formulation of work [6], which represents a quadric by the eccentricity e and the focal point of the conic section h. For a parabolic mirror, $e = 1$ and for elliptic and hyperbolic mirrors, $e < 1$ and $e > 1$, respectively. Therefore, by fixing h and changing e, different mirrors are obtained which all share the same focal point. Here three typical mirrors are studied, whose 2-D profiles are hyperbolic ($e = 1.2$), parabolic ($e = 1.0$), and elliptic ($e = 0.8$). To work with these mirrors, two types of cameras are studied: a perspective lens camera and a telecentric lens camera. In the simulation of CVB with each camera, we change the mirror shape by changing e and fix all the other imaging parameters.

The CVB with a perspective lens camera for different mirrors is shown in Fig. 10. In this 2-D profile, the x-axis and the z-axis are defined in the same way as in Fig. 2. The optical center of the perspective lens is set to $d = -50$ mm along the z-axis and the focal point of the mirror is set to $h = 35$ mm. As the same perspective lens camera is used, the CVBs share the same view boundary. Fig. 10 also shows the CVBs when a telecentric lens camera is used. For the telecentric lens cameras, the view boundaries are parallel with the optical axis and they are determined by the mirror's diameter.

The results show that for the same mirror, the CVB with a telecentric lens camera extends to a larger depth range compared to that with a perspective lens camera. Yet, for all the considered mirrors and cameras, the CVB shapes are similar. All the CVB shapes are consistent with the model analysis in Section II, which is the basis of the proposed method. We also test other system parameters. When the focal point p is changed from 2 to 100 mm and the lens center d is changed from 2 to 100 mm, the same result is observed. This indicates that for the catadioptric systems with these different mirrors and cameras, the proposed model-based method can be readily applied without any special modification.

V. CONCLUSION

As an improvement to image quality in terms of focus for catadioptric imaging systems, wherein a single image does not capture all objects with clear focus, this paper obtains an overall well-focused catadioptric image by combining multifocal images. In contrast to previous EDF methods, this paper does not evaluate a local focus measurement through the entire set of images as such approaches commonly lead to heavy computational load. Rather, the special optical geometry properties in catadioptric systems allow the best-focused image regions in multifocal images to be modeled by a series of neighboring concentric annuluses. The best-focused image regions can then be combined into a final overall well-focused image. As the model parameters are independent of the scene structure, once the parameters are determined, no additional model computation is needed for different scenes, which makes the process extremely fast.

This paper will benefit applications where a high-image resolution is desired over a large field of view. In particular, it will be useful for compact catadioptric systems and catadioptric systems that work with large apertures, which typically suffer from small DOF problems. Experiments on catadioptric images verify the model and the method. In practice, many off-the-shelf cameras are equipped with the focus-bracketing function, with which our method can be easily used by many existing catadioptric systems. Note that the focus-bracketing requires time for multiexposure, which might hinder the approach in dynamic environments. This problem would be alleviated by using newer cameras with their ever-increasing frame rate such as Casio EX-F1, which

can capture 60 f/s at 6 MP. The multifocal image acquisition can also achieve improved efficiency by using the techniques in this paper [13].

REFERENCES

[1] I. Bogdanova, X. Bresson, J. P. Thiran, and P. Vandergheynst, "Scale space analysis and active contours for omnidirectional images," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1888–1901, Jul. 2007.

[2] S. Baker and S. K. Nayar, "A theory of single-viewpoint catadioptric image formation," *Int. J. Comput. Vis.*, vol. 35, no. 2, pp. 1–22, 1999.

[3] X. Ying and Z. Hu, "Catadioptric camera calibration using geometric invariants," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 10, pp. 1260–1271, Oct. 2004.

[4] I. Tosic and P. Frossard, "Geometry-based distributed scene representation with omnidirectional vision sensors," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1033–1046, Jul. 2008.

[5] E. Hecht, *Optics*, 4th ed. Reading, MA: Addison-Wesley, 2002.

[6] R. Swaminathan, "Focus in catadioptric imaging systems," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–7.

[7] H. Ishiguro, "Development of low cost compact omni-directional vision sensors and their applications," in *Panoramic Vision: Sensors, Theory, and Applications*, R. Benosman and S. B. Kang, Eds. Heidelberg, Germany: Springer-Verlag, 2001, pp. 2–38.

[8] R. J. Pieper and A. Korpel, "Image processing for extended depth of field," *Appl. Opt.*, vol. 22, no. 10, pp. 1449–1453, 1983.

[9] D. G. Tieman, R. K. Murphey, J. T. Schmidt, and S. B. Tieman, "A computer-assisted video technique for preparing high resolution pictures and stereograms from thick specimens," *J. Neurosci. Methods*, vol. 17, no. 4, pp. 231–245, 1986.

[10] A. G. Valdecasas, D. Marshall, J. M. Becerra, and J. J. Terrero, "On the extended depth of focus algorithms for bright field microscopy," *Micron*, vol. 32, no. 6, pp. 559–569, 2001.

[11] K. Itoh, A. Hayashi, and Y. Ichioka, "Digitized optical microscopy with extended depth of field," *Appl. Opt.*, vol. 28, no. 16, pp. 3487–3493, 1989.

[12] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive digital photomontage," in *Proc. ACM SIGGRAPH*, vol. 23. 2004, pp. 294–302.

[13] S. W. Hasinoff and K. N. Kutulakos, "Light-efficient photography," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2203–2214, Nov. 2011.

[14] C. Kim, "Segmenting a low-depth-of-field image using morphological filters and region merging," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1503–1511, Oct. 2005.

[15] M. Antunes, M. Trachtenberg, G. Thomas, and T. Shoa, *All-in-Focus Imaging Using a Series of Images on Different Focal Planes* (Lecture Notes on Computer Science), vol. 3656. New York: Springer-Verlag, 2005, pp. 174–181.

[16] Z. Liu, K. Tsukada, K. Hanasaki, Y. K. Ho, and Y. P. Dai, "Image fusion by using steerable pyramid," *Pattern Recognit. Lett.*, vol. 22, no. 9, pp. 929–939, 2001.

[17] B. Forster, D. Van De Ville, J. Berent, D. Sage, and M. Unser, "Complex wavelets for extended depth-of-field: A new method for the fusion of multichannel microscopy images," *Microsc. Res. Tech.*, vol. 65, nos. 1–2, pp. 33–42, 2004.

[18] S. Li and B. Yang, "Multifocus image fusion using region segmentation and spatial frequency," *J. Image Vis. Comput.*, vol. 26, no. 7, pp. 971–979, 2008.

[19] F. Aguet, D. V. Ville, and M. Unser, "Model-based 2.5-D deconvolution for extended depth of field in bright field microscopy," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1144–1153, Jul. 2008.

[20] *Production List*. Accowle Company, Ltd., Tokyo, Japan [Online]. Available: http: http://www.accowle.com/english/products.html

[21] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 8, pp. 824–831, Aug. 1994.

[22] P. Favaro and S. Soatto, "A geometric approach to shape from defocus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 406–417, Mar. 2005.

[23] D. Malacara and Z. Malacara, *Handbook of Optical Design*, 2nd ed. New York: Marcel Dekker, 2004.

[24] D. Lanman, D. Crispell, M. Wachs, and G. Taubin, "Spherical catadioptric arrays: Construction, multi-view geometry, and calibration," in *Proc. Int. Symp. 3D Data Process., Vis., Trans.*, Jun. 2006, pp. 1–8.

Weiming Li received the Ph.D. degree in pattern recognition and intelligent systems from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2008.

He was with the Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong, from 2008 to 2011. He is currently a Researcher with the China Laboratory, Samsung Advanced Institute of Technology, Beijing. His current research interests include computer vision, computational imaging, and multimedia technology.

Youfu Li (SM'01) received the Ph.D. degree in robotics from Oxford University, Oxford, U.K., in 1993.

He was a Post-Doctoral Researcher with the Department of Computer Science, University of Wales, Aberystwyth, U.K., from 1993 to 1995. He joined the City University of Hong Kong, Kowloon, Hong Kong, in 1995. His current research interests include robot vision, visual tracking, robot sensing and sensor-based controls, mechatronics, and automation.

Dr. Li has served as an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING and the IEEE ROBOTICS AND AUTOMATION MAGAZINE.