Optics Communications 412 (2018) 178–185

Contents lists available at ScienceDirect

Optics Communications

journal homepage: www.elsevier.com/locate/optcom

Depth-estimation-enabled compound eyes

Woong-Bi Lee, Heung-No Lee *

Gwangju Institute of Science and Technology (GIST), School of Electrical Engineering and Computer Science, Gwangju, 61005, Republic of Korea

ARTICLE INFO

Keywords:
Computational imaging
Three-dimensional imaging
Compound eyes

ABSTRACT

Most animals that have compound eyes determine object distances by using monocular cues, especially motion parallax. In artificial compound eye imaging systems inspired by natural compound eyes, object depths are typically estimated by measuring optic flow; however, this requires mechanical movement of the compound eyes or additional acquisition time. In this paper, we propose a method for estimating object depths in a monocular compound eye imaging system based on the computational compound eye (COMPU-EYE) framework. In the COMPU-EYE system, acceptance angles are considerably larger than interommatidial angles, causing overlap between the ommatidial receptive fields. In the proposed depth estimation technique, the disparities between these receptive fields are used to determine object distances. We demonstrate that the proposed depth estimation technique can estimate the distances of multiple objects.

1. Introduction

Compound eyes, such as those of arthropods, have attracted widespread research interest owing to their unique features – such as wide fields of view (FOVs), excellent motion detection capability, and sensitivity to light intensity – that indicate their great potential for use in numerous applications, including unmanned aerial vehicles and endoscopic medical tools [1–4]. Recently, cameras inspired by compound eyes found in nature have been developed using curved optics and electronics [5,6] and discrete component integration at macroscopic levels [7].

Visual methods for depth estimation can be grouped into two main categories based on whether they use binocular or monocular cues [8]. Binocular cues are obtained from the minor disparities between the views of two eyes when the eyes are located close to one another and have overlapping views. These slightly different images of the same scene are sent to the brain and integrated into a single image containing depth information [9]. By contrast, monocular cues are obtained from two-dimensional images captured by a single eye; these cues include interposition, motion parallax, relative size and clarity, texture gradient, linear perspective, and light and shadow [8].

Some insects, such as praying mantids, that have binocular vision systems in the fronts of their heads use binocular cues to estimate target distances [9,10]. However, unlike humans' camera-like eyes that can focus on objects by changing the shapes or positions of their lenses, insects' compound eyes are inherently immobile and unable to focus owing to their structural limitations [8]. Thus, the binocular cues used for depth estimation in compound eyes are much less efficient, yielding images with low spatial resolutions and limited effective depth estimation ranges [11,12].

Insects can also estimate object distances using monocular cues. The motion parallax of objects in a visual scene that is caused by the relative motion between the observer and the objects yields information about object distances [8,13]. Specifically, nearby objects produce more apparent motion than distant ones. Insects' visual systems can easily detect the depths of objects that move independently of their surroundings by using motion parallax. For example, grasshoppers judge depths accurately by using the motion parallax generated by peering movements, that is, by moving their head from side to side [9], and bees measure distances by monitoring the apparent motion of an object relative to its surroundings [14].

Recently, artificial compound eyes that mimic natural compound eyes have been proposed. In these eyes, each ommatidium (individual imaging unit) has a limited acceptance angle, thus avoiding optical crosstalk among neighboring ommatidia [5–7,13]. In [6,13], object depths were estimated using monocular cues from optic flows (i.e., pattern of apparent motion) based on the phenomenon in which a closer object appears to move faster than a farther one. However, this method requires rotation or movement of the compound eye.

* Correspondence to: #C317, School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology (GIST), 123 Cheomdangwagi-ro, Buk-gu, Gwangju 61005, Republic of Korea.

E-mail addresses: wblee@gist.ac.kr (W.-B. Lee), heungno@gist.ac.kr (H.-N. Lee).

https://doi.org/10.1016/j.optcom.2017.12.009
Received 22 September 2017; Received in revised form 19 November 2017; Accepted 3 December 2017
0030-4018/

In this paper, we propose a method for estimating object depths in a monocular compound eye imaging system based on the computational compound eye (COMPU-EYE) framework described in [15]. In COMPU-EYE, each ommatidium has a larger acceptance angle than its interommatidial angle, causing the ommatidial receptive fields to overlap significantly. As in binocular depth estimation methods, depth estimation in COMPU-EYE involves processing the multiple and slightly differing views received by the ommatidia by using a proposed digital signal processing (DSP) technique. Depth information can be estimated by using the dependences of the disparities between the ommatidial observations on object distance. We perform a numerical experiment to verify the effectiveness of the proposed method. In our experiment, we demonstrate that the proposed depth estimation technique can not only estimate the distances of multiple objects but also reconstruct object images with high resolution.

Depth estimation using the disparities between multiple subimages has been studied in multicamera systems such as integral imaging [16]. Integral imaging is a three-dimensional imaging and sensing system that uses an array of optical units. Each optical unit consists of a microlens and an array of photosensors, and it produces an elemental image. From multiple elemental images, a three-dimensional image is reconstructed optically or computationally [17]. In [18], an iterative reconstruction algorithm was proposed for improving image quality given distance information. A stereo matching method that used the spatial variations of parallax shifts in elemental images was proposed for depth estimation [2,19]. We note that multicamera setups are essentially different from our work. First, our structure can be considered a degraded integral imaging system with a single photosensor in each elemental image; this imitates the structure of apposition compound eyes found in nature. The number of sensors is thus reduced dramatically, and the sensors can be implemented in a fully hemispherical structure that provides a large FOV [5]. Some studies on integral imaging considered curved surfaces for realizing a large FOV [20]. However, with planar sensors, they require additional optical components like random phase masks; otherwise, mismatch occurs [20]. Second, three-dimensional information is highly compressed using a single photosensor per lens. Thus, more sophisticated reconstruction algorithms are required for imaging and depth estimation.

In Section 2, we describe the COMPU-EYE system model and the principle of depth estimation. In Section 3, we propose our depth estimation method, and in Section 4, we discuss the results. Finally, we present our conclusions in Section 5.

2. COMPU-EYE system model and depth estimation

2.1. COMPU-EYE system model

We consider an apposition compound eye imaging system in which a hemispherical eye observes a planar object. The hemispherical compound eye can be implemented by reformulating a stretchable set of a microlens and photodetector array [5]. As a result, this compound eye has a large FOV. This compound eye consists of a two-dimensional array of M ommatidia that are uniformly spaced with an interommatidial angle of $\Delta\phi$. As illustrated in Fig. 1(a), each ommatidium receives incident light within its acceptance angle $\Delta\varphi$. Based on the object's location, each observation at each ommatidium can be specified by a transfer function that describes the fraction of the input light that each ommatidium observes. We assume that the object is located a distance d (measured in millimeters) away from the compound eye and that the image to be reconstructed consists of N pixels that form an $N \times 1$ input vector $\mathbf{x} = \left[x_1, \ldots, x_N\right]^T$ in lexicographic order. Let y_i denote the output sample obtained by a photodetector at the ith ommatidium for $i \in \{1, 2, \ldots, M\}$. Through ray tracing analysis, y_i can be obtained using the linear equation $y_i = \boldsymbol{a}_{i,d}\mathbf{x}$, where $\boldsymbol{a}_{i,d}$ is a $1 \times N$ vector whose elements represent the visibility of the ith ommatidium at each of the N pixels of the object located at a distance of d [15]. Given the structure of the compound eye, specifically, the acceptance angles, interommatidial angles, and sizes of the compound eye and ommatidia, the receptive fields of ommatidia at a distance of d are determined. Each element of $\boldsymbol{a}_{i,d}$ is obtained by calculating the intersection area of the receptive field of the ith ommatidium and the jth pixel in the object for $j \in \{1, 2, \ldots, N\}$. The data acquisition model for M ommatidial observations can be expressed as a system of linear equations as follows:

$$\mathbf{y} = \mathbf{A}_d\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{y} = \left[y_1, \ldots, y_M\right]^T$ is a set of M output samples, $\mathbf{A}_d \in \mathbb{R}^{M\times N}$ denotes a measurement matrix whose ith row is $\boldsymbol{a}_{i,d}$, and $\mathbf{n}$ is an $M \times 1$ noise vector.

A signal is typically considered sparse if it can be represented with few nonzero elements. We note that any natural image can be represented as a sparse signal in a certain domain, such as by applying a wavelet, discrete cosine, or discrete Fourier transform [21]. That is, $\mathbf{x} = \mathbf{w}^T\mathbf{s}$ and $\mathbf{wx} = \mathbf{s}$, where $\mathbf{s}$ is a sparse $N \times 1$ vector and $\mathbf{w}$ is an $N \times N$ sparsifying matrix. By exploiting the sparse representation of $\mathbf{x}$, Eq. (1) can be expressed as

$$\mathbf{y} = \mathbf{A}_d\mathbf{w}^T\mathbf{s} + \mathbf{n}. \tag{2}$$

To obtain sufficiently high resolution, the number of pixels to be reconstructed is set to be larger than the number of ommatidia, that is, $N > M$. Then, Eq. (2) becomes an underdetermined system of linear equations. Given $\mathbf{A}_d$ and $\mathbf{y}$, $\mathbf{s}$ can be obtained by solving the following convex optimization problem [22]:

$$\hat{\mathbf{s}} = \min_{\mathbf{s}} |\mathbf{s}|_1 \text{ subject to } \left\|\mathbf{y} - \mathbf{A}_d\mathbf{w}^T\mathbf{s}\right\|_2 < \varepsilon, \tag{3}$$

where ε is a small constant. From $\hat{\mathbf{s}}$, the object image can be reconstructed by solving $\hat{\mathbf{x}} = \mathbf{w}^T\hat{\mathbf{s}}$.

2.2. Distance and measurement matrix

The COMPU-EYE imaging system proposed in [15] yields resolution improvements beyond the number of ommatidia owing to its use of large ommatidial acceptance angles in combination with a DSP technique. The large acceptance angles enable each pixel to be observed multiple times by multiple ommatidia with different perspectives. However, these ommatidial observations are severely distorted owing to the overlap in the ommatidial receptive fields. Given a measurement matrix, DSP can be used to reconstruct high-resolution images from the distorted observations by solving the underdetermined linear system in Eq. (1). The measurement matrix strongly depends on the object's properties, such as its distance. In [15], the object distance was assumed to be fixed and known, and the measurement matrix corresponding to this distance was given to the DSP system. However, assuming prior knowledge about object distances is impractical in reality. The reconstruction process works well only if the measurement matrix is correct; if an inappropriate measurement matrix is used, then the reconstructed image is severely distorted.

In the framework of COMPU-EYE imaging, we propose a new depth estimation method. In conventional compound eyes, $\Delta\varphi$ is designed to be smaller than or equal to $\Delta\phi$ to avoid aliasing [5,6,23]. As shown in Fig. 1(a), each ommatidium observes an independent section within $\Delta\varphi$. Consider two objects, P_1 and P_2, that are located at different distances from a compound eye. If the objects are observed by a single ommatidium in Fig. 1(a), their distances cannot be inferred. In contrast, the COMPU-EYE system has enlarged, overlapping ommatidial receptive fields, because $\Delta\varphi$ is much larger than $\Delta\phi$, as seen in Fig. 1(b). We note that a large acceptance angle can be realized by increasing the diameter of the photodetector, decreasing the focal length of the microlens, or using a material of higher refractive index for the microlens [15]. This configuration is shown in Fig. 1(b), in which object P_2 is observed by two ommatidia; thus, the compound eye can deduce that object P_2 is farther away than object P_1. When many ommatidia are present, the number of