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 $\mathbf{w}\mathbf{x} = \mathbf{s}$, 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

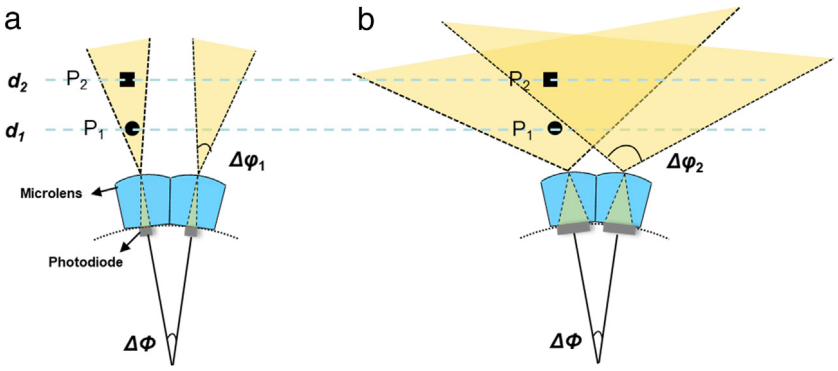


Fig. 1. Structures and fields of view of (a) Conventional compound eye with $\Delta\varphi_1 \leq \Delta\phi$ and (b) Proposed COMPU-EYE system with $\Delta\varphi_2 \gg \Delta\phi$.

ommatidia viewing the object and the area of the object that is visible by the ommatidia depend upon the object distance. The variation of these quantities with object distance is used for depth estimation in the proposed method.

Here, we give an example of the variation for different object distances. In Fig. 2, the measurement matrices and corresponding number of nonzero elements per column are shown, in which a compound eye consists of 5 × 5 ommatidia with a radius of 6.92 mm, focal length of micro lens of 1.35 mm, $\Delta\phi = 12°$, and $\Delta\varphi = 30°$. Three objects are located at $d_1 = 2$ mm, $d_2 = 20$ mm, and $d_3 = 40$ mm from the compound eye. The object plane is composed of 12 × 12 pixels with a uniform distribution. As the object moves away from the compound eye, the areas of the ommatidial receptive fields and the overlap between them both increase. Accordingly, as shown in Fig. 2(a), the number of nonzero elements in the measurement matrix increases with object distance. In Fig. 2(b), the number of nonzero elements per column in the measurement matrices varies with respect to the object distances, implying that each pixel is uniquely observed by a different set of ommatidia with different perspectives. Thus, a unique measurement matrix is generated with respect to object distance. By using the relationship between the unique measurement matrix and the object distance, we propose the following method for estimating object distances.

2.3. System model for depth estimation

First, we set the range of interest $R = [d_{\min}, d_{\max}]$, where $d_{\min}$ and $d_{\max}$ are the minimum and maximum distances, respectively. The range of interest can be application-specific; for example, it can be 10–25 mm for endoscopic applications [24]. For DSP, we assume that the object distance can be sampled as a set of discrete distances $\mathbf{d} = \{d_1, d_2, \ldots, d_L\}$ within the range of interest, where L is the number of distance elements. In this paper, we consider uniform discrete distances within the range of interest. The depth resolution $\Delta d = (d_{\max} - d_{\min})/L$ depends on the number of distance elements and depth range of interest. According to the predetermined $\mathbf{d}$, a measurement matrix $\mathbf{A}_{d_l}$ for $l \in \{1, 2, \ldots, L\}$ can be obtained from the structure of the compound eye and the object located a distance d_l away from the compound eye. By concatenating L measurement matrices, a dictionary matrix $\mathbf{A} \in \mathbb{R}^{M\times(L\cdot N)}$ can be formed as $\mathbf{A} = \left[\mathbf{A}_{d_1}\mathbf{A}_{d_2} \ldots \mathbf{A}_{d_L}\right]$. Then, the linear representation of $\mathbf{y}$ in Eq. (1) can be rewritten in terms of all possible measurement matrices as

$$\mathbf{y} = \sum_{i=1}^{L} \mathbf{A}_{d_i}\mathbf{x}_i = \mathbf{A}\mathbf{X}, \tag{4}$$

where $\mathbf{X} = \left[\mathbf{x}_1^T, \ldots, \mathbf{x}_L^T\right]^T = \left[x_{1,1}, \ldots, x_{1,N}, \ldots, x_{L,1}, \ldots, x_{L,N}\right]^T \in \mathbb{R}^{(L\cdot N)\times 1}$. When an object is located at a certain distance in the set $\mathbf{d}$, a valid observation $\mathbf{y}$ can be sufficiently represented by a linear combination of the columns from the corresponding measurement matrix. For example, when the object distance matches the lth measurement matrix, the linear equation becomes

$$\mathbf{y} = \mathbf{A}\mathbf{X}_0, \tag{5}$$

where $\mathbf{X}_0 = \left[0, \ldots, 0, x_{l,1}, \ldots, x_{l,N}, 0, \ldots, 0\right]^T$ is a sparse coefficient vector whose entries are zero except for those associated with the lth measurement matrix. $\mathbf{x}$ can be sparsely represented as $\mathbf{x} = \mathbf{w}^T\mathbf{s}$. Similarly, $\mathbf{X}$ can be sparsely represented as $\mathbf{X} = \mathbf{W}^T\mathbf{S}$. Here, $\mathbf{S}$ is an $L \cdot N \times 1$ sparse vector and $\mathbf{W}$ is a block diagonal matrix containing L instances of $\mathbf{w}$, that is, $\mathbf{W} = diag(\underbrace{\mathbf{w}, \ldots, \mathbf{w}}_{L}) \in \mathbb{R}^{(L\cdot N)\times(L\cdot N)}$, where $diag(\cdot)$ represents a diagonal matrix. By using $\mathbf{S}$, Eq. (4) becomes

$$\mathbf{y} = \mathbf{A}\mathbf{W}^T\mathbf{S} = \mathbf{B}\mathbf{S}, \tag{6}$$

where $\mathbf{S} = \left[\mathbf{s}_1^T, \ldots, \mathbf{s}_L^T\right]^T = \left[s_{1,1}, \ldots, s_{1,N}, \ldots, s_{L,1}, \ldots, s_{L,N}\right]^T \in \mathbb{R}^{(L\cdot N)\times 1}$ and $\mathbf{B} = \mathbf{A}\mathbf{W}^T$. As does Eq. (3), l_1 norm minimization provides a sparse vector $\hat{\mathbf{S}}$:

$$\hat{\mathbf{S}} = \arg\min_{\mathbf{S}} \|\mathbf{S}\|_1 \text{ subject to } \left\|\mathbf{y} - \mathbf{A}\mathbf{W}^T\mathbf{S}\right\| \leq \eta, \tag{7}$$

where η is a small constant.

3. Depth estimation method

After $\hat{\mathbf{S}}$ has been obtained from Eq. (7), the problem of estimating object distances can be reformulated as a classification problem whose objective is to find the distances at which the object has the highest probability of being located. Because the compound eye imaging system can be sparsely represented in Eq. (6) and the measurement matrices are uniquely generated with respect to object distances, sparse-representation-based classification (SRC) can be used to estimate object locations. SRC has been widely studied, and its accuracy has been demonstrated in many applications including face recognition [25] and brain computer interface systems [26]. SRC usually finds the most compact representation of a test sample, where the representation is expressed as a linear combination of columns in an overcomplete dictionary matrix, and then, it determines a class that contributes most to represent the test sample [27]. In this paper, we use SRC to estimate the depths of multiple objects. Unlike conventional SRC, the observed signal in this depth estimation framework is superposed with respect to the number of objects. Therefore, our problem is defined as a multiclass classification problem. We first describe an SRC-based depth estimation algorithm in the compound eye imaging system; we then propose an iterative depth estimation method that updates dictionaries in a coarse-to-fine manner.

We first specify a classification rule by using sparse signal reconstruction. As l_1 norm minimization provides a sparse solution for Eq. (7), most of the nonzero components in $\hat{\mathbf{S}}$ reside in the class in which the object exists with high probability. One of the classification rules is

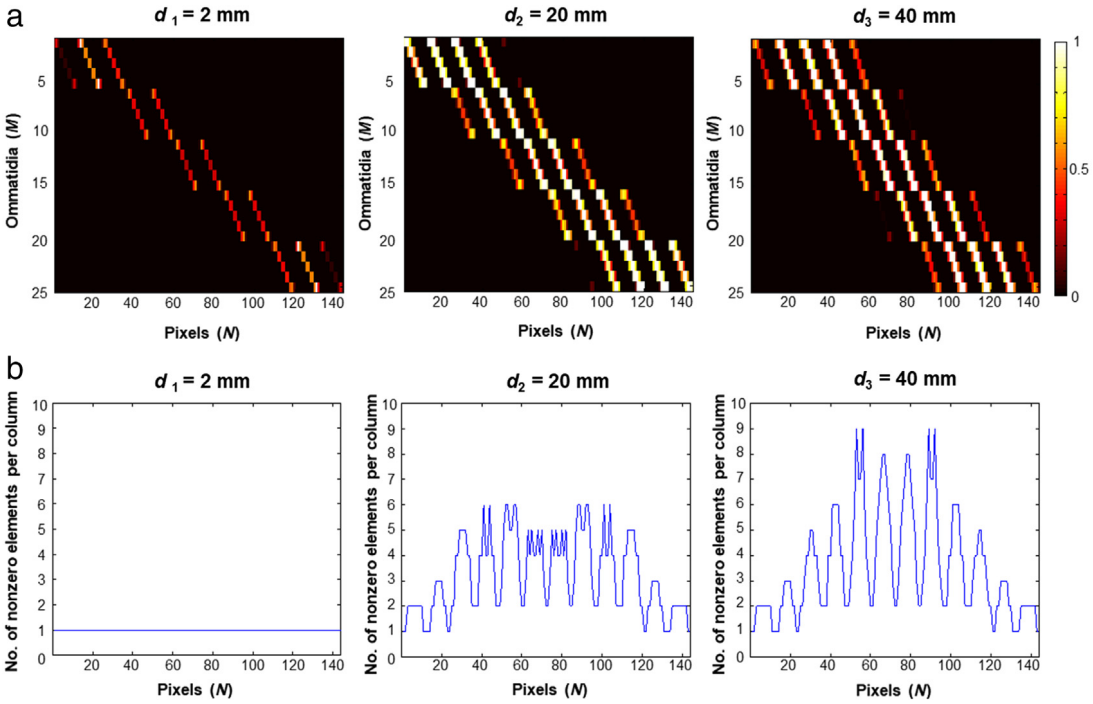


Fig. 2. (a) Measurement matrices and (b) their number of nonzero elements per column for $d_1 = 2$ mm, $d_2 = 20$ mm, and $d_3 = 40$ mm, where $M = 5 \times 5$, $N = 12 \times 12$, $\Delta\phi = 12°$, and $\Delta\varphi = 30°$.

Table 1
SRC-based depth estimation algorithm.

Initial parameters: $\mathbf{y}, \mathbf{d} = \{d_1, d_2, \ldots, d_L\}, \mathbf{w}, \eta, \alpha$
Step 1: Set $\mathbf{A} = [\mathbf{A}_{d_1} \mathbf{A}_{d_2} \ldots \mathbf{A}_{d_L}]$ and $\mathbf{W} = diag(\underbrace{\mathbf{w}, \ldots, \mathbf{w}}_{L})$.
Step 2: Solve Eq. (7) from $\mathbf{y}$ given $\mathbf{A}$ and $\mathbf{W}$, and obtain $\hat{\mathbf{S}}$.
Step 3: Calculate the regularized residuals:
$r_l := \left\|\mathbf{y} - \mathbf{B}\delta_l\left(\hat{\mathbf{S}}\right)\right\|^2 / \left\|\delta_l\left(\hat{\mathbf{S}}\right)\right\|^2$ for $l = 1, \ldots, L$.
Step 4: Obtain the class of existence $I_e = \{l | r_l < \alpha\}$ and the estimated distance of the object $\hat{\mathbf{d}} = \{d_l | l \in I_e\}$.

to use the residuals [22]. For each class, we define its characteristic function $\delta_l : \mathbb{R}^{L \cdot N} \to \mathbb{R}^{L \cdot N}$ that selects the coefficients of $\hat{\mathbf{S}}$ associated with the lth class while nullifying the coefficients of other classes. Thus, for $\hat{\mathbf{S}} \in \mathbb{R}^{L \cdot N}$, $\delta\left(\hat{\mathbf{S}}\right) \in \mathbb{R}^{L \cdot N}$ is obtained by including the elements corresponding to the lth class and nulling all elements of $\hat{\mathbf{S}}$ from other classes. By using the characteristic function, we denote the regularized residuals as

$$r_l := \left\|\mathbf{y} - \mathbf{B}\delta_l\left(\hat{\mathbf{S}}\right)\right\|^2 / \left\|\delta_l\left(\hat{\mathbf{S}}\right)\right\|^2. \tag{8}$$

If the object is located at d_l, the r_l value is smaller than those at other distances. We denote I_e as a set of the indices of estimated distances at which the objects are expected to be located. With r_l for $l = 1, \ldots, L$, the classification rule is given by

$$I_e := \{l | r_l < \alpha\}, \tag{9}$$

where α is an arbitrary constant. A set of distances $\hat{\mathbf{d}}$ where the object is expected to be located can be determined by

$$\hat{\mathbf{d}} = \{d_l | l \in I_e\}. \tag{10}$$

Then, the images that only correspond to the estimated distances are reconstructed by solving $\hat{\mathbf{x}}_l = \mathbf{w}^T \hat{\mathbf{s}}_l$ for $l \in I_e$. The SRC-based depth estimation algorithm is summarized in Table 1.

Thus far, depth estimates have been obtained by finding locations in a dictionary, where the signals have small residuals. To improve the depth accuracy, the number of distance elements L must be increased in the form of the dictionary. However, the dictionary cannot include infinitely many possible distances owing to computational complexity and memory storage. To solve Eq. (7), $O(M \cdot N \cdot L)$ computations for every iteration and $O(M \cdot N \cdot L)$ storage are required; these are proportional to the number of distance elements [28]. We note that the l_1 norm minimization in Eq. (7) finds a sparse solution whose nonzero elements are most closely associated with the most correlated measurement matrix. By using the fact that the measurement matrices of neighboring distances are relatively more correlated than those of farther distances in dictionary matrix $\mathbf{A}$, we propose an iterative depth estimation method that is more efficient in terms of computational complexity and memory storage. Instead of universally searching for the object distances at once, we iteratively refine the set of distances in a coarse-to-fine manner [29]. The distances are investigated in detail only around regions where objects are expected to be present.

For iteration index i, we first choose a set of coarse distances within the range of interest $R^{(i)}$ as $\mathbf{d}_l^{(i)}$ for $l = 1, 2, \ldots, L_i$, at which the objects can potentially be located. The depth interval is $\Delta d^{(i)} = d_{l+1}^{(i)} - d_l^{(i)}$. Accordingly, $\mathbf{A}^{(i)}$ and $\mathbf{W}^{(i)}$ can be generated from the structure of the compound eye imaging system. The sparse signal $\hat{\mathbf{S}}$ is reconstructed by solving Eq. (7), and the estimate of distances $\hat{\mathbf{d}}^{(i)}$ can be obtained by solving Eq. (8)–(10). Then, the set of distances is updated by refining the range of interest and the depth interval. The range of interest is refined around the estimated distances, that is, $R^{(i+1)} = \left[\hat{d}_j^{(i)} - \Delta d^{(i)}/2, \hat{d}_j^{(i)} + \Delta d^{(i)}/2\right]$ for $j = 1, \ldots, \left|\hat{\mathbf{d}}^{(i)}\right|$, where $|\cdot|$ represents the cardinality. The depth interval is refined as $\Delta d^{(i+1)} = \Delta d^{(i)}/K$ for a positive real number $K > 1$. Then, the updated set of finer distances is

$$\mathbf{d}^{(i+1)} = \left\{\hat{d}_{j,k}^{(i+1)}\right\} \tag{11}$$

where $\hat{d}_{j,k}^{(i+1)} = \left(\hat{d}_j^{(i)} - \Delta d^{(i)}/2\right) + (k-1)\Delta d^{(i+1)}$ for $j = 1, \ldots, \left|\hat{\mathbf{d}}^{(i)}\right|$ and $k = 1, \ldots, \lceil K \rceil$. We repeat this process until the depth interval is sufficiently fine. The iterative coarse-to-fine depth estimation algorithm is summarized in Table 2.

Table 2
Iterative depth estimation algorithm.

Initial parameters: $\mathbf{y}, R^{(1)}, \mathbf{d}^{(1)} = \left\{d_1^{(1)}, d_2^{(1)}, \ldots, d_{L_1}^{(1)}\right\}, \mathbf{w}, \eta, \alpha, i = 1$

Step 1: Set $\mathbf{A}^{(i)} = \left[\mathbf{A}_{d_1^{(i)}} \ \ldots \ \mathbf{A}_{d_{L_i}^{(i)}}\right]$ and $\mathbf{W}^{(i)} = diag(\underbrace{\mathbf{w}, \ldots, \mathbf{w}}_{L_i})$.

Step 2: Solve Eq. (7) from $\mathbf{y}$ given $\mathbf{A}^{(i)}$ and $\mathbf{W}$, and obtain $\hat{\mathbf{S}}$.

Step 3: Calculate the regularized residuals:

$$r_l := \left\|\mathbf{y} - \mathbf{B}\delta_l\left(\hat{\mathbf{S}}\right)\right\|^2 / \left\|\delta_l\left(\hat{\mathbf{S}}\right)\right\|^2 \text{ for } l = 1, \ldots, L.$$

Step 4: Obtain the set of indices of estimated distances $I_e^{(i)} = \{l | r_l < \alpha_i\}$.

Step 5: Update

$$R^{(i+1)} = \left[\hat{d}_j^{(i)} - \Delta d^{(i)}/2, \hat{d}_j^{(i)} + \Delta d^{(i)}/2\right],$$

$$\Delta d^{(i+1)} = \Delta d^{(i)}/K \text{ for } K > 1, \mathbf{d}^{(i+1)} = \left\{\hat{d}_{j,k}^{(i+1)}\right\} \text{ and } L_{i+1} = |\mathbf{d}^{(i+1)}|,$$

where $\hat{d}_{j,k}^{(i+1)} = \left(\hat{d}_j^{(i)} - \Delta d^{(i)}/2\right) + (k-1)\,\Delta d^{(i+1)}$

for $j = 1, \ldots, \left|\hat{\mathbf{d}}^{(i)}\right|$ and $k = 1, \ldots, \lceil K \rceil$.

Step 6: Set $i = i + 1$ and repeat from Step 1 until the depth resolution is sufficiently fine.

4. Results

To evaluate the performance of our depth estimation technique, we consider a hemispherical compound eye with a radius of 6.92 mm, where each ommatidium has a height of 1.35 mm [5]. The compound eye consists of an $M = 80 \times 80$ array of uniformly spaced ommatidia with $\Delta\phi = 1.8°$ and $\Delta\varphi = 45°$, such that $\Delta\varphi \gg \Delta\phi$. The 200×200 mm object consists of $N = 100 \times 100$ pixels. Thus, each measurement matrix has dimensions of 6400×10000. For the sparsifying basis $\mathbf{w}$, we use a db2 wavelet transform and a level of two. To solve Eq. (7), we use the fast and efficient alternating direction method [28].

First, we determine the depth estimation accuracy for the proposed compound eye. Because the measurement matrices corresponding to neighboring distances are more correlated with each other, we set a distance of 108 mm from the compound eye as the reference distance and compare with other distances by increasing the depth intervals. To evaluate the depth estimation accuracy with respect to the depth interval, we consider a sparse signal as an input, that is, $\mathbf{w} = \mathbf{I}$, where $\mathbf{I}$ represents an identity matrix. In each assessment, a sparse signal dimension of 10000×1 with 5%, 7.5%, and 10% of randomly located nonzero elements is used. The distance of the input signal is randomly chosen between the reference distance and the comparison distance. The tolerance α in Eq. (9) is set to be 0.1. This assessment is repeated 100 times. As seen in Fig. 3, as the object distances increase, the accuracy of the proposed depth estimation increases. For signals with 5% sparsity, if the depth intervals are larger than 0.3 mm, the proposed depth estimation works with more than 97% accuracy. For the l_1 norm minimization in Eq. (7), the reconstruction performance depends on the sparsity of the input signal, that is, low accuracy for the input signal with large sparsity. Thus, as the sparsity increases, the performance of the proposed depth estimation deteriorates as well.

The proposed COMPU-EYE imaging system used for evaluating the image reconstruction is shown in Fig. 4. The hemispherical compound eye observes an object consisting of four characters: G, i, S, and T. The characters are located at three different distances from the compound eye. G is 108 mm away from the compound eye, i and S are 109 mm away, and T is 112 mm away, as shown in Fig. 4(b). As shown in Fig. 4(c), the characters overlap one another, preventing the distance information from being inferred. The DSP technique introduced in Section 3 can be used to decompose each letter given its distance.

We demonstrate the performance of the proposed depth estimation method when the object distances are included in the set of potential distances in the dictionary. We assume that the depth range of interest is from 108 mm to 112 mm and the target depth resolution is 1 mm. Within the range of interest, the distances are uniformly sampled with 1-mm resolution, that is, $\mathbf{d} = \{108, 109, 110, 111, 112\}$. For depth estimation and object reconstruction, we construct a dictionary matrix $\mathbf{A} = \left[\mathbf{A}_{108}\ \mathbf{A}_{109}\ \mathbf{A}_{110}\ \mathbf{A}_{111}\ \mathbf{A}_{112}\right]$ in accordance with the potential distances.

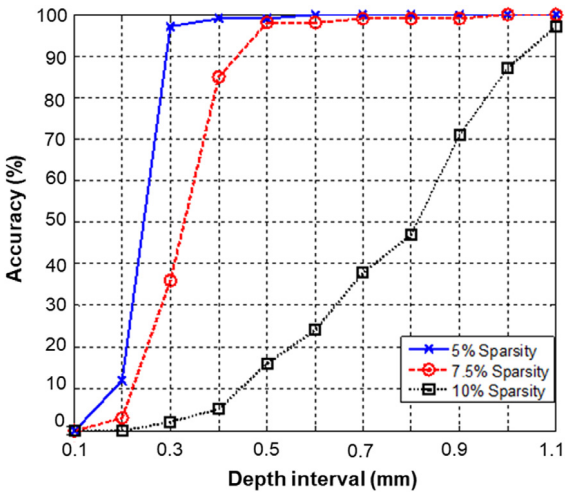


Fig. 3. Depth estimation accuracy (%) with respect to depth interval.

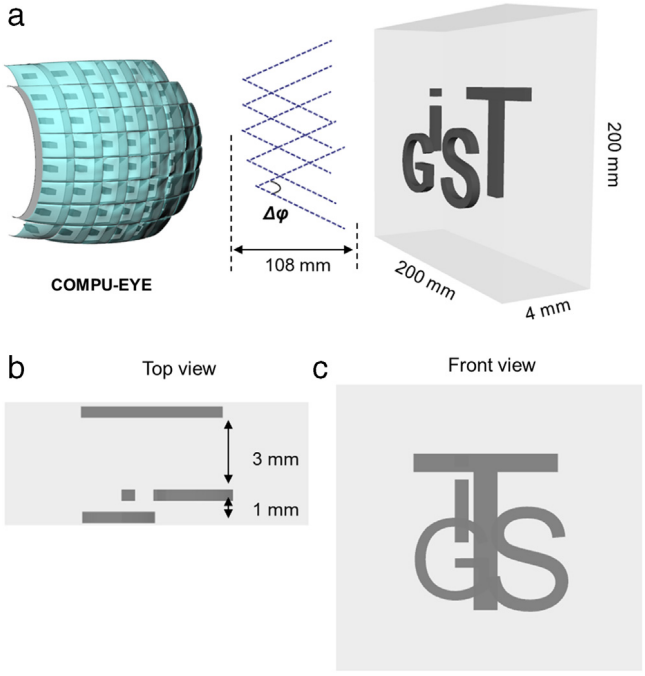


Fig. 4. Proposed COMPU-EYE imaging system: (a) Three-dimensional, (b) Top, and (c) Front views.

Given $\mathbf{A}$, we can solve Eq. (7) to obtain $\hat{\mathbf{S}}$ from $\mathbf{y}$. Then, $\hat{\mathbf{X}}$ can be obtained by calculating $\hat{\mathbf{X}} = \mathbf{W}^T\hat{\mathbf{S}}$. The reconstructed $\hat{\mathbf{S}}$ and $\hat{\mathbf{X}}$ are shown in Fig. 5(a) and (b), respectively. Owing to the sparse signal reconstruction, most of the nonzero signals in Fig. 5(a) are concentrated in the set of indices corresponding to distances of 108 mm, 109 mm, and 112 mm. We note that the reconstruction errors in Fig. 5(a) and (b) for 110 mm and 111 mm are caused by coherence among neighboring measurement matrices. As indicated in Fig. 5(c), the regularized residuals of the set of indices corresponding to distances of 108 mm, 109 mm, and 112 mm are smaller than those corresponding to the other distances. As a result, the index set of the estimated distances and the estimated distances of the objects are determined as $I_e = \{1, 2, 5\}$ and $\hat{\mathbf{d}} = \{108, 109, 112\}$, respectively. In Fig. 5(d), the reconstructed signals $\hat{\mathbf{x}}_l = \mathbf{w}^T\hat{\mathbf{s}}_l$ for $l \in \{1, 2, 5\}$ are represented as images. Note that the observation is highly distorted owing to the overlap among the ommatidial receptive fields.

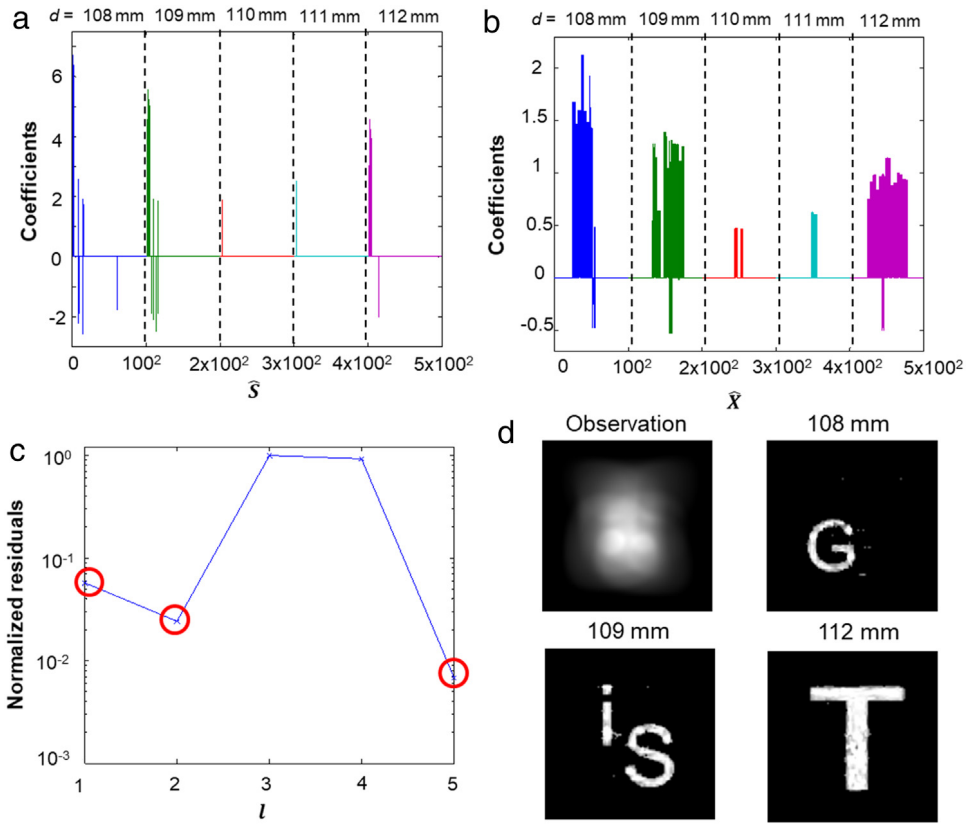


Fig. 5. (a) Reconstruction of $\hat{\mathbf{S}}$, (b) Reconstruction of $\hat{\mathbf{X}}$, (c) Normalized regularized residuals, and (d) Ommatidial observations and reconstructed images with respect to estimated distances.

The reconstructed characters at 108, 109, and 112 mm are clearly visible. This result indicates that COMPU-EYE achieves 1-mm depth resolution. We note that the reconstruction resolution is also improved by 1.56 times because 100 × 100-pixel images are reconstructed from 80 × 80-pixel ommatidial observations.

We now investigate the performance of the iterative depth estimation for the object shown in Fig. 4. We assume that the potential object locations are unknown and that the range of interest is from 100 mm to 120 mm, that is, $R^{(1)} = [100\,\text{mm}, 120\,\text{mm}]$. For the SRC-based depth estimation method in Table 1 to achieve a depth resolution of 1 mm, the dictionary requires 21 concatenated measurement matrices with dimensions of 6400 × 210000. The computational complexity of this task necessitates the use of the iterative depth estimation method described in Table 2. We first formulate a set of coarse distances $\mathbf{d}^{(1)} = \{100, 110, 120\}$ and $\mathbf{A} = [\mathbf{A}_{100}\ \mathbf{A}_{110}\ \mathbf{A}_{120}]$ correspondingly. The result of iterative depth estimation is shown in Fig. 6. At the 1st iteration, because the objects are located at around 110 mm, the residual value at 110 mm is the smallest. Thus, the object distance is estimated as 110 mm for $\alpha = 0.3$ at the 1st iteration. For detailed depth estimation, we further set a dictionary with finer distances at around 110 mm. The range of interest is refined as $R^{(2)} = [105, 115]$ and the depth interval, as $\Delta d^{(2)} = 3$ for $K = 3.33$. Then, the set of distances is updated as $\mathbf{d}^{(2)} = \{105, 108, 111, 114\}$. At the 2nd iteration, the residual values at distances of 108 mm and 111 mm are smaller than those at other distances. Thus, we estimate that the objects are located at around 108 mm and 111 mm for $\alpha = 0.2$. The range of interest is refined as $R^{(3)} = [107, 109] \cup [110, 112]$ and the depth interval as $\Delta d^{(3)} = 1$ for $K = 3$. Then, the set of distances is updated as $\mathbf{d}^{(3)} = \{107, 108, 109, 110, 111, 112\}$. At the 3rd iteration, the object distances are estimated as 108 mm, 109 mm, and 112 mm from the compound eye for $\alpha = 0.1$. As a result, the objects are efficiently reconstructed by using the proposed iterative depth estimation method.

Now, we aim to demonstrate depth estimation for an object with continuous depths. As a target, we consider a plane object that is slanted at 23° toward the compound eye and located 108 mm away from the compound eye. When the range of interest is from 108 mm to 111 mm, the object distance can be uniformly sampled as $\mathbf{d} = \{108, 109, 110, 111\}$ in Fig. 7(b). The proposed depth estimation method provides a depth map of the object with 1-mm depth resolution, as shown in Fig. 7(d). Consequently, an object with continuous depths can be well reconstructed by using the estimated distances, as seen in Fig. 7(c). In this manner, continuous depths can be estimated. We note that if we densely sample the range of distance, the depth map will be more accurate; however, there is a limit to the depth resolution, as seen in Fig. 3.

5. Conclusion

We have proposed a depth estimation method based on the COMPU-EYE imaging system, in which the ommatidial acceptance angle is much larger than the interommatidial angle. The ommatidial receptive fields overlap, and the disparities the between ommatidial observations vary with object distance. As a result, the uniqueness of the generated measurement matrix depends upon the object distance. In the proposed technique, the dependences of the disparities between the ommatidial observations and the measurement matrix uniqueness on object distance are used to estimate the depth. This work helps not only to estimate object distances but also to reconstruct objects with high resolution, and it is therefore essential for future development of the COMPU-EYE system.

Generally, disparity-based depth estimation methods have limitations for very distant objects because the disparities decrease [30]. By varying the acceptance angles of the ommatidia or arranging the ommatidia irregularly, the range of depth estimation can be extended

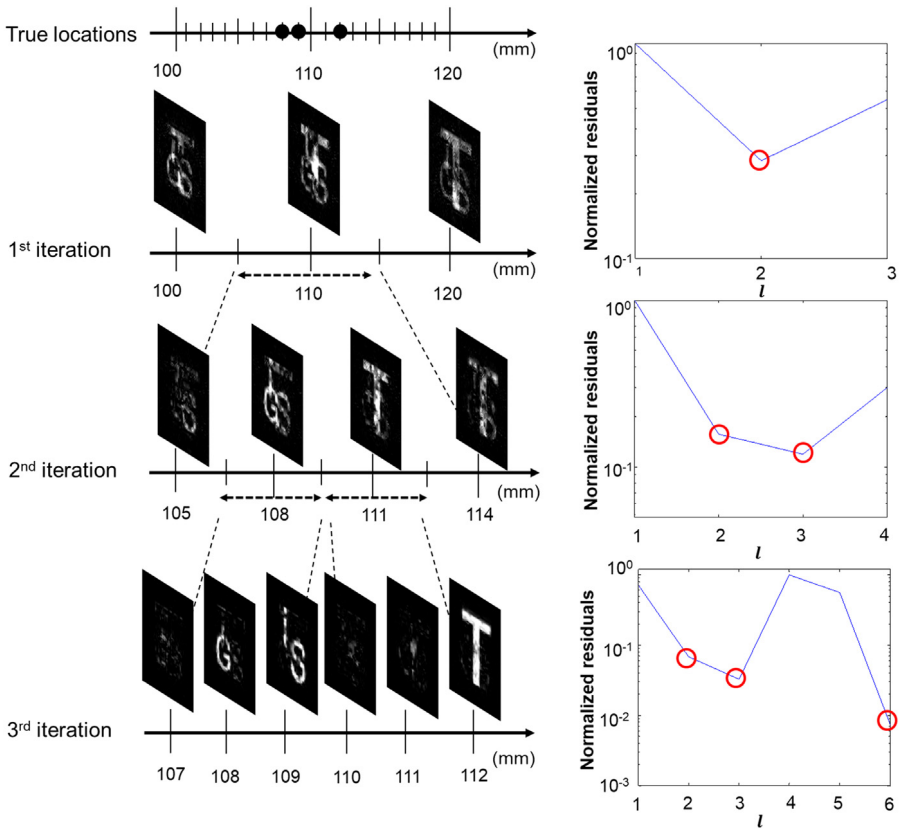
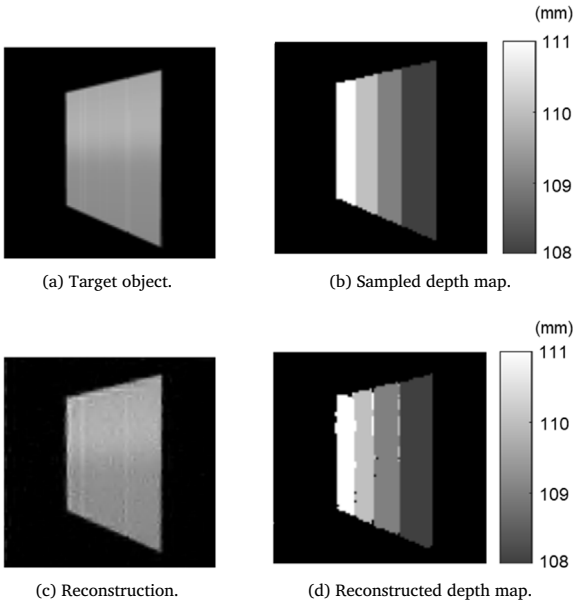


Fig. 6. An example of the iterative depth estimation method.



(a) Target object. (b) Sampled depth map.

(c) Reconstruction. (d) Reconstructed depth map.

Fig. 7. Depth estimation and object reconstruction for a slanted object.

adaptively, that is, large acceptance angles for small distances and small acceptance angles for large distances [31]. Our future work will focus on improving the depth resolution by designing COMPU-EYE to have high incoherence among measurement matrices with respect to object distances. Furthermore, we will improve the depth estimation performance by applying the l_0-norm based minimization to solve Eq. (6). This has been shown to provide better reconstruction performance than the l_1-norm minimization [32].

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIP) [NRF-2015R1A2A1A05001826].

References

[1] A. Borst, J. Plett, Optical devices: Seeing the world through an insect's eyes, Nature 497 (2013) 47–48.
[2] K. Kagawa, K. Yamada, E. Tanaka, J. Tanida, A three-dimensional multifunctional compound-eye endoscopic system with extended depth of field, Electron. Comm. Jpn. (2012) 14–27.
[3] F. Expert, F. Ruffier, Flying over Uneven Moving Terrain Based on Optic-Flow Cues Without Any Need for Reference Frames or Accelerometers, Bioinspir. Biomim., IOP Publishing, 2015, pp. 1–25.
[4] A. Hassanfiroozi, Y.-P. Huang, B. Javidi, H.-P.D. Shieh, Hexagonal liquid crystal lens array for 3D endoscopy, Opt. Express 23 (2015) 971–981.
[5] Y.M. Song, Y. Xie, V. Malyarchuk, J. Xiao, I. Jung, K.J. Choi, Z. Liu, H. Park, C. Lu, R.H. Kim, R. Li, K.B. Crozier, Y. Huang, J.A. Rogers, Digital cameras with designs inspired by the arthropod eye, Nature 497 (2013) 95–99.
[6] D. Floreano, R. Pericet-Camara, S. Viollet, F. Ruffier, A. Brückner, R. Leitel, W. Buss, M. Menouni, F. Expert, R. Juston, M.K. Dobrzynski, G. L'Eplattenier, F. Recktenwald, H.A. Mallot, N. Franceschini, Miniature curved artificial compound eyes, Proc. Natl. Acad. Sci. 110 (2013) 9267–9272.
[7] O. Cogal, Y. Leblebici, An insect eye inspired miniaturized multi-camera system for endoscopic imaging, IEEE Trans. Biomed. Circuits Syst. 11 (2017) 212–224.
[8] M. Sarkar, Bioinspired Optical Imaging, Biologically Inspired Computer Vision, Wiley-VCH Verlag GmbH & Co., 2015, pp. 109–142. KGaA.
[9] M.F. Land, L. Chittka, R.F. Chapman, Vision, Structure and Function, Cambridge University Press, 2012, pp. 708–737.
[10] S. Rossel, Binocular vision in insects: How mantids solve the correspondence problem, Proc. Natl. Acad. Sci. U.S.A. (1996) 13229–13232.
[11] M.F. Land, Visual acuity in insects, Annu. Rev. Entomol. 42 (1997) 147–177.

[12] K. Kral, Behavioural–analytical studies of the role of head movements in depth perception in insects, birds and mammals, Behav. Process. (2003) 1–12.
[13] F. van Breugel, K. Morgansen, M.H. Dickinson, Monocular distance estimation from optic flow during active landing maneuvers, Bioinspir. Biomim. (2014) 025002–025010.
[14] H.E. Esch, J.E. Burns, Honeybees use optic flow to measure the distance of a food source, Naturwissenschaften (1995) 38–40.
[15] W.B. Lee, H. Jang, S. Park, Y.M. Song, H.N. Lee, COMPU-EYE: A high resolution computational compound eye, Opt. Express 24 (2016) 2013–2026.
[16] X. Xiao, B. Javidi, M. Martinez-Corral, A. Stern, Advances in three-dimensional integral imaging: Sensing, display, and applications [Invited], Appl. Opt. 52 (2013) 546–560.
[17] A. Stern, B. Javidi, Three-dimensional image sensing, visualization, and processing using integral imaging, Proc. IEEE 94 (2006) 591–607.
[18] L. Cao, J. Peter, Iterative reconstruction of projection images from a microlens-based optical detector, Opt. Express 19 (2011) 11932–11943.
[19] Y. Gao, W. Liu, P. Yang, B. Xu, Depth estimation based on adaptive support weight and SIFT for multi-lenslet cameras, 6th International Symposium on Advanced Optical Manufacturing and Testing Technologies (AOMATT 2012), SPIE2012, 2012, pp. 84190C–84194.
[20] Y. Yuan, X. Wang, X. Wu, J. Zhang, Y. Zhang, Improved resolution integral imaging using random aperture coding based on compressive sensing, Optik - Int. J. Light Electron Opt. 130 (2017) 413–421.
[21] E.J. Candes, T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Trans. Inform. Theory 52 (2006) 5406–5425.
[22] D.L. Donoho, M. Elad, V.N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, IEEE Trans. Inform. Theory (1995) 6–18.
[23] O. Cogal, Y. Leblebici, An insect eye inspired miniaturized multi-camera system for endoscopic imaging, IEEE Trans. Biomed. Circuits Syst. (2016) 1–13.
[24] K.W. Seo, D.-w. Lee, B.R. Min, A 3-D information acquisition algorithm for close range endoscopy, in: R. Magjarevic, J.H. Nagel (Eds.), World Congress on Medical Physics and Biomedical Engineering 2006: August 27–September 1, 2006 COEX Seoul, Korea Imaging the Future Medicine, Springer, Berlin, Heidelberg, 2007, pp. 2612–2615.
[25] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2009) 210–227.
[26] S. Younghak, L. Seungchan, L. Junho, L. Heung-No, Sparse representation-based classification scheme for motor imagery-based brain–computer interface systems, J. Neural Eng. 9 (2012) 056002.
[27] X. Song, Z. Liu, X. Yang, S. Gao, A new sparse representation-based classification algorithm using iterative class elimination, Neural Comput. Appl. 24 (2014) 1627–1637.
[28] J.F. Yang, Y. Zhang, Alternating direction algorithms for 1 problems in compressive sensing, SIAM J. Sci. Comput. (2011).
[29] D. Malioutov, M. Cetin, A.S. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Trans. Signal Process. 53 (2005) 3010–3022.
[30] R. Horisaki, S. Irie, Y. Ogura, J. Tanida, Three-dimensional information acquisition using a compound imaging system, Opt. Rev. 14 (2007) 347–350.
[31] H. Ryoichi, K. Keiichiro, N. Yoshizumi, T. Takashi, M. Yasuo, T. Jun, Irregular lens arrangement design to improve imaging performance of compound-eye imaging systems, Appl. Phys. Express 3 (2010) 022501.
[32] Y. Esmaeili Salehani, S. Gazor, I.-M. Kim, S. Yousefi, ℓ0-norm sparse hyperspectral unmixing using arctan smoothing, Remote Sens. 8 (2016) 187.