Abstract - Why can the compound eye of an insect capture the prey so exactly and quickly? From the biological aspect, the compound eye is excellent at detecting motion. Based on the computer vision aspect, limited studies regarding this issue exist. It has been verified that a trinocular visual system incorporating a third CCD into a traditional binocular is very helpful to solve translational motion. Extending from this concept, a spherical compound-like eye of superposition type for pan-tilt rotational motion is established. It can be concluded that the more the number of ommatidium in an insect has, the higher the detection capability of capturing prey becomes. In this paper, the secret of compound-eye of insects is explored through the point of view of computer vision.

I. INTRODUCTION AND MOTIVATION

The configuration of the compound eye of insects always attracts many researchers’ attentions. Recently, the biologically inspired visual studies have flourished with a boom in the microlens technology. The development of image acquisition systems based on the framework of the compound eye has also progressed more quickly ever than before. The fabrication of micro compound eye has been reported in the literature with a gradual orientation towards its commercial applications. Those well-known commercial applications include the TOMBO compound eye proposed by Tanita et al. [1], and the hand held plenoptic camera by Ng et al. [2]. The former is a multiple-imaging system with a post-digital processing unit that can provide a compact hardware configuration as well as processing flexibility. The latter is similar to the Adelson and Wang’s plenoptic camera [3], but with two fewer lenses, which significantly shortens the optical path and results in a portable camera. In addition, some related publications contain one photoreceptor per view direction [4,5], a miniaturized imaging system [6], the electronic compound eye [7], the curved gradient index lenses [8], an artificial ommatidia [9], and a silicon-based digital retina [10]. All of them belong to the same category of the image acquisition systems with the framework of the compound eye.

Why can the compound eye of insects capture the prey so accurately and quickly? This interesting topic has not been completely answered. The biologists believe that it is because of the flicker effect [11]. An object moves across the visual field, ommatidia are progressively turned on and off during which the bees measure the distance by the motion of images received by their eyes as they fly [11-16]. Therefore, many researchers have put considerable effort in building images viewed from a compound eye and reconstructing the environmental image from those image patterns. Most of those researches are limited to static images. However, the emphasis of this paper focuses on the dynamic vision of the compound eye. In order to achieve motion recovery for visual servo, ego-motion estimation needs to be investigated first. Neumann et al. applied plenoptic video geometry to construct motion equations, and optimized the error function to acquire motion parameters [8]. Tisse used off-the-shelf micro-optical elements to develop self-motion estimation [10]. Nevertheless, they did not discuss this topic, and the noise interference problem was not seriously examined. Recently, Lin and Cheng [17] presented a trinocular approach for recovery of the translational motion. It has been indicated that the third camera not only provides more image information, but also significantly improves both efficiency and accuracy for estimation of motion parameters. Extending from this concept, the recovery of the rotational motion with the configuration of compound eye will be examined.

II. THE COMPOUND-LIKE EYE IN COMPUTER VISION

According to the mosaic theory, there are two basic types of compound eyes, apposition and superposition [18]. Due to the construction between these two types being clearly different, the former acquires the image from ommatidium just a small part of overall object, and exploit each ommatidium to make up a complete, ambiguous image; while the latter can acquire a whole image through adjusting ommatidia. Each ommatidium can itself receive an ambiguous image, and every image will be different based on different positions from its arrangement. Strictly speaking, superposition referred to here is neural superposition [19, 20]. These two types mentioned above are based on the ecological aspect, which can help us to realize how to produce images from compound eyes very clearly. But based on the computer vision (CV) aspect, which configuration should we adopt? The compound-like eye using in CV proposed first by Aloimonos [20] can be divided into two types, one is planar compound-like eye, and the other is spherical compound-like eye. The spherical compound-like eye as shown in Fig. 1 will be studied in this paper.
How to form the configuration of spherical compound-like eye in computer vision? First the configuration of the spherical compound-like ommatidium needs to be defined. Suppose each ommatidium can look at an object. Based on this situation, a number of CCD cameras treated as ommatidium are arranged on the surface of the sphere. There will be a fixed horizontal distance between adjacent ommatidia. In order to distinguish the apposition type compound-like eye, this specific arrangement has to be clearly defined to be the superposition type. The image viewed by each ommatidium will have a similar and ambiguous outlook of an object. But each ommatidium can generate its image that depends on its different arrangement location. So the images generated by the superposition-type spherical compound-like eyes (SSCE) will be different but similar and vague, which are similar to blurred patterns viewed by ommatidia of an insect.

III. PAN-TILT EGO-ROTATIONAL MOTION FOR ONE CCD

For the purpose of analysis, a spherical compound-like eye can be modelled as a pan-tilt compound-like eye system. Assume total rotational angles for both pan and tilt are less than 180°. Based on this specific configuration, the normal 3D rotational motion for one CCD can be transferred to 2D rotational motion. After acquiring pan-tilt ego-rotational motion for one CCD, the superposition image of spherical compound-like eyes will be discussed in the next section.

To simulate compound-like eyes for an ommatidium looking at the environment, assume an object is moving relative to the CCD platform, and define that the origin of CCD platform is located at the optical center of CCD as well as the rotation center of the platform. When a single rigid object is moving, two things need to be considered:
1. Using normal 3D rotational motion, pan-tilt rotational motion and image transformation can be established.
2. Based on the observed image of the CCD, the ego-rotational angle of the CCD can be resolved.

A. Pan-Tilt Rotation and Image Transformation

Given a world coordinate system, a rotation R applied to a 3D point \( P = (X, Y, Z)^T \) is accomplished through a displacement: \( P \to P' \). A normal 3D rotation about the arbitrary axis through the origin of the coordinate system can be described by successive rotations \( \psi, \theta \) and \( \phi \) about its \( Z, Y \), and \( X \) axes, respectively. Then the transformation \( M \) for the arbitrary rigid motion in 3D space is thus given by

\[
M: P' = R \cdot P = R_x(\psi)R_y(\theta)R_z(\phi)P
\]  

where each individual rotation matrix is expressed by

\[
R_x(\phi) = \begin{bmatrix}
1 & 0 & 0 \\
0 & \cos \phi & -\sin \phi \\
0 & \sin \phi & \cos \phi \\
\end{bmatrix}, \quad R_y(\theta) = \begin{bmatrix}
\cos \theta & 0 & \sin \theta \\
0 & 1 & 0 \\
-\sin \theta & 0 & \cos \theta \\
\end{bmatrix}
\]

\[
R_z(\psi) = \begin{bmatrix}
\cos \psi & -\sin \psi & 0 \\
\sin \psi & \cos \psi & 0 \\
0 & 0 & 1 \\
\end{bmatrix}
\]

It should be noted that the rotational operations do not commute.

For the composition of SSCE, each CCD (ommatidium) will match up with the cadence of pan-tilt motion and be put on the surface of the sphere. This formation manner can build a pattern similar to Fig. 1. Therefore, the rotational motion of SSCE basically is a simplification of a pure 3D rotational motion, just set the rotation of \( Z \) axis being 0, and consider only the motion behavior in \( X \) and \( Y \) axes.

First, a 3D point \( P \) is moved to a new location by rotating the platform about the \( X \) axis by an angle \( \phi \) and about the \( Y \) axis by an angle \( \theta \), given as \( P \to P' \). And after matrix multiplication, \( R \) is given by

\[
R = R(\theta)R(\phi) = \begin{bmatrix}
\cos \theta & \sin \theta \sin \phi & \sin \theta \cos \phi \\
0 & \cos \phi & -\sin \phi \\
-\sin \theta & \cos \theta \sin \phi & \cos \theta \cos \phi \\
\end{bmatrix}
\]

Under perspective imaging, a point \( P = (X, Y, Z)^T \) in 3D space is projected onto a location in the image plane \( p = (x, y)^T \), with

\[
x = \frac{fX}{Z}, \quad y = \frac{fY}{Z}
\]

where \( f \) is the focal length of the CCD.

Following the order of pan-tilt rotational motion \( R(\theta)R(\phi) \), then the rotational motion transformation of the 3D point can be viewed as moving an image point \( p = (x, y) \) to a corresponding image point \( p' = (x', y') \) based on a 2D image rotation mapping:
3D space point: \( R(\theta)R(\phi): P(X, Y, Z) \rightarrow P'(X', Y', Z') \)

2D image point: \( r(\theta)r(\phi): p = (x, y) \rightarrow p' = (x', y') \)

Using (2) and (3), therefore, the image point \( p \) of \( P \) moves to \( p' \) described by

\[
\begin{bmatrix}
  x' \\
  y'
\end{bmatrix} = f \begin{bmatrix}
  x \cos \theta + y \sin \theta \sin \phi + f \sin \theta \cos \phi \\
  y \cos \phi - f \sin \phi
\end{bmatrix}
\]

\( i(-x \sin \theta + y \cos \theta \sin \phi + f \cos \theta \cos \phi) \)

(4)

Note that the movement of the 3D space point has been changed into image point. (4) depicts the 2D image rotational transformation based on a pan-tilt rotational motion to a 3D point with 3D rotational motion. It means that there is no 3D space variable contained in this transformation. It should be, therefore, emphasized that this image transformation does not require any information about the scene when the CCD rotates around its lens center.

**B. Ego-Rotational Motion of Pan and Tilt**

Given two image locations \( p_0(x_0, y_0) \) and \( p_1(x_1, y_1) \), they are the projections of a 3D point \( P \) at different time \( t_0 \) and \( t_1 \). When the image point \( p_0 \) moves onto \( p_1 \), assuming no translation occurs, what are the amount of rotational angles \( \phi \) and \( \theta \)?

It can be realized that (4) is a positive procedure for image rotation mapping. If we would like to compute the amount of rotations from a pair of observations, obviously it is an inverse problem. To resolve this issue, an intermediate point \( p_c(x_c, y_c) \), shown in Fig. 2, is assumed. Due to the horizontal rotation and vertical rotation are applied to the CCD separately, Prazdny [21] have proved that the points in the image move along a hyperbolic path. Similar to Burger and Bhanu [22], the first step is \( r(\phi) \), rotation around the X axis, moving the original point \( p_0(x_0, y_0) \) to an intermediate point \( p_c(x_c, y_c) \). The next step is \( r(\theta) \), moving the intermediate point \( p_c(x_c, y_c) \) to the final point \( p_1(x_1, y_1) \) by camera rotation around the Y axis. Therefore, the angles of ego-rotational motion of the pan and tilt can be obtained as

\[
\theta = \tan^{-1} \frac{x_c}{f} - \tan^{-1} \frac{y_0}{f} - \tan^{-1} \frac{x_0}{f} - x_c
\]

(5)

where the coordinates of the intersection point could be derived as

\[
x_c = f y_0 \left[ \frac{f^2 + x_0^2 + y_0^2}{f^2 + y_0^2 [f^2 + x_1^2] - (x_0 y_1)^2} \right]^{1/2}
\]

(6)

It should be noted that this is the ego-rotational motion model of one CCD if two successive images at two different times are applied.

**IV. EGO-ROTATIONAL MOTION OF SUPERPOSITION SPHERICAL COMPOUND-LIKE EYES**

Based on the images of SSCE generated from computer vision, the rotational motion using a SSCE can be resolved by the following procedures:

1. Each camera of the SSCE whose image is ambiguous is the same as ommatidium, and the fuzziness of all images is independent. The generation of the ambiguous image can be achieved by adding a random noise to an ideal image.

2. When the SSCE looks at an object, each ommatidium CCD will perceive a different profile in SSCE according to different locations of CCDs. In this manner, the compound-like eyes will observe a whole ambiguous image which is composed of many small, similar, independent, and ambiguous patterns.

3. When an object moves, the SSCE can detect this rotation using two whole images before and after the motion.

4. Generation of the image for each CCD camera is the same as with procedures from 1 to 3.

5. Using those two vague images that include the information of the rotation, the corresponding intersection point for each camera can be estimated.

6. Any single CCD camera is able to generate a pair of pan and tilt angles with its intersection point. However, it should be noted that the greater the total number of the camera is, the more the pan and tilt angles will be.

7. Taking the mean of all CCD camera’s ego-rotation angles, pan and tilt, unlike using standard least square in ego-translation [23], the ego-motion angles of pan and tilt of SSCE can be easily obtained.
In this manner, when the amount of ommatidium CCD camera is increasing, then the ego-motion angles of SSCE will grow. Validations of the SSCE being able to get the more exact ego-rotation angles even in the case of large noise will be explored.

V. EXPERIMENTAL RESULTS

In order to verify the performance of noise immunity for the SSCE, a synthesized cloud of fifty 3D points shown in Fig. 3 is chosen as the test object in the experiment. The contaminated image points by the noises before and after a movement are respectively modeled as:

\[
\begin{align*}
  x_i(i) + N_{x1}(i) \\
y_i(i) + N_{y1}(i)
\end{align*}
\]

\[
\begin{align*}
  x_2(i) + N_{x2}(i) \\
y_2(i) + N_{y2}(i)
\end{align*}
\]

where \(i\) indexes the image point, \((x(i), y(i))\) locates the ideal image point for the \(i\)-th point, and \(N(i)\) denotes a zero-mean Gaussian random noise at this position. The noise processes \(N_{x1}, N_{y1}, N_{x2}\) and \(N_{y2}\) are given by

\[
E[N^2_{x1}(i)] = E[N^2_{y1}(i)] = E[N^2_{x2}(i)] = E[N^2_{y2}(i)] = \sigma^2
\]

In other words, all image points are contaminated by a random noise with the same variance.

For convenience to subsequent analysis, we define the relative error in pan-tilt rotational angles as

\[
err = \frac{\sqrt{(\phi' - \phi)^2 + (\theta' - \theta)^2}}{\sqrt{\phi^2 + \theta^2}}
\]

where \((\phi', \theta')\) is the computed pan-tilt ego-rotational angle, and \((\phi, \theta) \neq 0\) is the actual pan-tilt rotational angle.

Based on the amount of movement angle, we use large and small rotation angles, respectively to validate, and defines small angle of rotation as: (Pan:1°, Tilt:-1°) whereas the large angle of rotation as: (Pan:-9°, Tilt:7°). The arrangements of the CCD in the single row SSCE were 1×3, 1×5, 1×7, 1×9, 1×11, 1×13, 1×15, and 1×25. All distances between adjacent CCD cameras were 5°. Each case was conducted for 300 trials. For small and large rotational motions, idealistically the relative errors of the single row SSCE are equal to zero.

A. Small Rotation Angle

Under different CCDs in a single row SSCE, the variance of noise changes from small to large, the performance of different CCDs in a single row SSCE using the presented algorithm can be compared. Before the validation, the different ambiguous image adding different levels of noise are demonstrated. The 1x25 single row SSCE under the small rotational motion: (Pan:1°, Tilt:-1°) was selected as an example. When the noise variance changes from 0.25 to 25, it appears that the interfered image is getting more ambiguous accompanied by the increasing noise variance.

Under the single row CCD from 1x1 to 1x25, during the small rotational motion by facing different levels of noise, the relative error of the single row SSCE can be calculated, and the results are listed in Table I.

<table>
<thead>
<tr>
<th>Vari.</th>
<th>1x1</th>
<th>1x3</th>
<th>1x5</th>
<th>1x7</th>
<th>1x9</th>
<th>1x11</th>
<th>1x13</th>
<th>1x15</th>
<th>1x25</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.93</td>
<td>0.58</td>
<td>0.42</td>
<td>0.34</td>
<td>0.31</td>
<td>0.29</td>
<td>0.23</td>
<td>0.26</td>
<td>0.16</td>
</tr>
<tr>
<td>4</td>
<td>1.96</td>
<td>1.12</td>
<td>0.84</td>
<td>0.71</td>
<td>0.59</td>
<td>0.53</td>
<td>0.44</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3.62</td>
<td>2.32</td>
<td>1.75</td>
<td>1.34</td>
<td>1.26</td>
<td>1.06</td>
<td>0.95</td>
<td>0.95</td>
<td>0.61</td>
</tr>
<tr>
<td>16</td>
<td>5.67</td>
<td>3.44</td>
<td>2.54</td>
<td>2.15</td>
<td>1.86</td>
<td>1.64</td>
<td>1.62</td>
<td>1.48</td>
<td>1.00</td>
</tr>
<tr>
<td>25</td>
<td>7.89</td>
<td>4.34</td>
<td>3.51</td>
<td>2.92</td>
<td>2.45</td>
<td>2.12</td>
<td>2.02</td>
<td>1.86</td>
<td>1.34</td>
</tr>
</tbody>
</table>

B. Large Rotation Angle

When the rotational motion is increased, the reliability of SSCE will increase, which means it can bear more intense noise. The noise variances were from 25 to 2500. The 1x25 single row SSCE under the large rotational motion: (Pan:-9°, Tilt:7°) for the noise variance varying from 25 to 2500 at two time instants (red and black) were displayed in Fig. 4.
Fig. 4 The interfered image of large rotational motion under the noise variance changes from 25 to 2500.

Comparing small and large rotation angles, the increasing noise variances made the image getting more and more ambiguous, and the interference is more severe. When the variance of the noise is above 400, obviously the interfered images shown in Fig. 4 (c)–(f) are all not easy to be distinguished from the original one. Even so, the presented approach still reaches very small relative errors for the large rotational motion as shown in Table II.

Table II
Under the noise interference, the relative error of different CCDs in large rotational motion

<table>
<thead>
<tr>
<th>Var.</th>
<th>1x1</th>
<th>1x3</th>
<th>1x5</th>
<th>1x7</th>
<th>1x9</th>
<th>1x11</th>
<th>1x13</th>
<th>1x15</th>
<th>1x25</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>1.23</td>
<td>0.72</td>
<td>0.53</td>
<td>0.48</td>
<td>0.38</td>
<td>0.31</td>
<td>0.26</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>2.30</td>
<td>1.33</td>
<td>1.04</td>
<td>0.89</td>
<td>0.78</td>
<td>0.66</td>
<td>0.64</td>
<td>0.59</td>
<td>0.42</td>
</tr>
<tr>
<td>400</td>
<td>4.68</td>
<td>2.57</td>
<td>2.00</td>
<td>1.63</td>
<td>1.45</td>
<td>1.36</td>
<td>1.25</td>
<td>1.04</td>
<td>0.80</td>
</tr>
<tr>
<td>900</td>
<td>6.67</td>
<td>4.11</td>
<td>2.84</td>
<td>2.63</td>
<td>2.36</td>
<td>2.05</td>
<td>1.96</td>
<td>1.78</td>
<td>1.52</td>
</tr>
<tr>
<td>1600</td>
<td>9.51</td>
<td>5.09</td>
<td>4.03</td>
<td>3.46</td>
<td>2.99</td>
<td>2.04</td>
<td>2.03</td>
<td>2.22</td>
<td>1.72</td>
</tr>
<tr>
<td>2500</td>
<td>11.60</td>
<td>7.04</td>
<td>5.73</td>
<td>4.49</td>
<td>3.62</td>
<td>3.40</td>
<td>3.16</td>
<td>2.66</td>
<td>1.99</td>
</tr>
</tbody>
</table>

C. Discussion
Weng et al. [24] proposed that the motion displacement should be large to allow stable estimation of motion and structure. In the above two cases, when the motion displacement is small, the sensitivity to noise is high, while the bearing capability of noise-resistance is low. When the motion displacement is increasing, the reliability of performance will get improved and the noise-resistance will increase. Next, comparing Tables I and II with the noise variance 25, taking 1x25 for example, the relative error of large angle is 0.22, whereas the relative error of small angle is up to 1.74. From the above discussion, this situation has been confirmed.

In Tables I and II, no matter whether the rotation angle is small or large, also no matter how the noise levels are varying, when the number of CCD in a single row SSCE is increasing, the relative error in the rotational motion is getting reduced; and the more the number of CCDs, very obviously, the less the relative error. When the CCD number in the single row SSCE extends from 1x1 to 1x11, the relative error is down to 3 times as discussed above; when the CCD number is up to 1x25, the relative error is down to 5 times. For example, taking the noise
variance of 2500 in Table 2, and it is found that the relative error has been reduced from 11.66% down to 1.99%.

The dragonfly has nearly 30000 ommatidia in each eye, which makes sense because they hunt in flight, whereas butterflies and moths, which do not hunt in flight, only 12000 to 17000 ommatidia [12]. Due to the ommatidium increasing by more, the detection accuracy of compound eyes will be more and more enhanced, so that the compound eye is able to provide a more correct detection capability in 3D ego-motion. This phenomenon clearly corresponds to the above experimental results.

According to the above investigation, one of the reasons why the compound eye of the insect is able to help capturing its prey so exactly and quickly apparently is its sufficient large number of ommatidia. Based on sufficient multiple image patterns, powerful capability of noise-resistance for motion recovery can be accomplished without using any filters.

VI. CONCLUSION

The compound eyes of flying insect in the biological world are highly evolved organs. Although the images received from their view are ambiguous, they are still able to capture prey so exactly and quickly. Inspired by these insects, using pinhole image formation geometry to investigate the capturing behaviour of the SSCE to moving objects has been conducted. The concept of the SSCE configuration and the ego-rotation model of SSCE for pan-tilt motion are proposed. From the experimental results, it has been clearly found that if the number of ommatidium in an insect is getting more, under without using any filter, the higher the detection capability of capturing prey becomes.

ACKNOWLEDGMENT

This work was funded by National Science Council, Republic of China under contract NSC 93-2212-E-110-014.

REFERENCES