# CHAPTER 6. SINGLE-LENS MULTI-OCULAR STEREOVISION

The knowledge of single-lens trinocular stereovision presented in the previous chapter is extended and generalized to build a *single-lens multi-ocular* stereovision using prisms that have similar pyramid-like structure but have an arbitrary number (≥3) of faces, called a multi-face filter.

The term *multi-ocular* and another term *multi-view* which often have the same implications frequently appear in recent literatures. These two terms are normally used to describe a set of or a series of images captured from the same scene. These images may be acquired simultaneously or consecutively at different rates, such as, video rate vs. standalone shots. Multi-ocular or multi-view images normally provide more comprehensive information on the environment and have attracted a great amount of interest. Works carried out involving multi-view and multi-ocular images cover a very wide range, that includes stereovision, object detection, reconstruction and recognition, token tracking in image sequences, motion analysis, intelligent surveillance, etc. Many examples of the research works about multi-ocular or multi-view images can be found in [56]-[67].

A *multi-camera* system is generally needed if the multi-view or multi-ocular images are required to be captured simultaneously. However the camera setup, calibration and synchronization of such a multi-camera system are usually more difficult and complicated than typical single camera or two camera vision system. Some discussions on multi-camera calibration can be found in works reported in [68]-[70].

This chapter presents the analysis and the implementation of a single-lens multi-ocular stereovision system. This system is able to capture three or more different views of the same scene simultaneously using only one real camera with the aid from a multi-face filter. It combines the advantages of single-lens stereovision and multi-ocular stereovision. Dynamic scene image capturing or video rate image capturing is not a problem for this system too.

Each image captured by this single-lens system can be divided into three or more sub-images and these sub-images can be taken as the images taken by three or more virtual cameras which are created by the multi-face filter. Two approaches used by the previous trinocular system are also applied here to analyze this multi-ocular system with necessary modifications: the first based on calibration technique and the second one is based on geometrical analysis of ray sketching. The geometrical based approach attracts greater interest because of its advantage of a simpler implementation: it does not require the usual complicated calibration process but one simple field point test to determine the whole system once the system is fixed and pin-hole camera model is used. Experiments are conducted to test the feasibility of both approaches.

Developing such a single-lens multi-ocular stereovision may help to solve some problems of a multi-camera system to a certain extent. No work of single-lens multi-ocular simultaneous stereovision systems using similar method is reported before and this design of single-lens multi-ocular stereovision system using multi-face filters should be a novel design according to the author's knowledge. One design which has similar function as ours is the work by Park, et. al [71]. They presented one depth extraction system using one lens array and one CCD camera. Their design is very interesting and can capture many elementary images (or sub-

images) for stereovision; however, as the lens array has too many (13X13) elementary lenses, the CCD camera needs to capture the sub-images using multiple shoots, and also the elementary images need to be modified (or rectified) before used for stereo, hence it can be seen our design has better features in the aspect of higher sub-image resolutions, concurrent image capturing which is important for the applications in the dynamic scenes, and direct image utilization for stereovision without modification.

Part of the content of this chapter has been published in [72], and one journal paper [73] together with part of the content in previous chapter has been drafted.

## 6.1 Virtual Camera Generation

Firstly with reference to the 3F filter used in Chapter 5, a multi-face filter is defined as a transparent prism which has a number of planar faces (≥ 3) inclined around an axis of symmetry to form a pyramid and this axis of symmetry is normal to the back plane of prism and passes through the back plane center. A 3F filter can be seen in Figure 5.1 and Figure A. 6. Graphical illustrations of filters with 4 and 5 faces are given in the following figures:

If a multi-face filter is vertically positioned in front of a CCD camera as shown in Figure 5.1, the image plane of this camera will capture multiple different views of the same scene behind the filter simultaneously. These sub-images can be taken as the images captured by multiple virtual cameras which are generated by the multi-face filter. One sample image captured by a system using four-face filter is given in Figure 6.2, from which the obvious differences among the four sub-images caused by different view angles and view scopes of the virtual cameras can be observed. It is assumed that each virtual camera consists of one unique optical center

and one "planar" image plane. The challenge is to determine the properties of these virtual cameras such as their focal lengths, positions and orientations so that the disparity information on the sub-images can be exploited to perform depth recovery like a stereovision system. As these sub-images are captured simultaneously, this system should theoretically possess the advantages of a typical multi-ocular stereovision system including its special properties on epipolar constraints, which provide a significant advantage in correspondence determination.

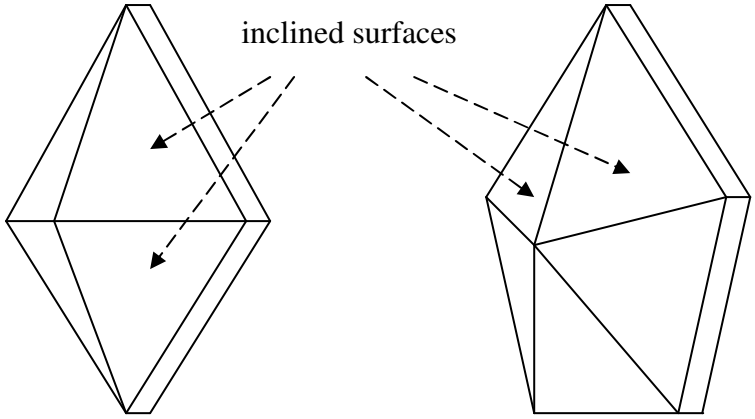


**Figure 6.1 Symbolic illustrations of multi-face filters with 4 and 5 faces**

Just like the virtual camera model used for single-lens binocular and single-lens trinocular stereovision systems described in the two previous chapters, it is assumed that the Field of View (FOV) of each virtual camera is constrained by two boundary lines (see Figure 5.4): one boundary line is the optical axis of the virtual camera which can be determined by back-extending the refracted ray that is aligned with real camera optical axis; and another FOV boundary line of the virtual camera can be determined by back-extending the refracted ray that is aligned with real camera FOV boundary line(s). And the optical center of virtual camera can be found

at the intersection between these two FOV boundary lines. Thus the generation of virtual camera(s) can be determined either by calibration or by geometrical analysis of ray sketching. The detailed determination process of virtual cameras for the trinocular system that is discussed in the previous chapter can be applied to this multi-ocular stereovision with minor modifications as the same principle is used to explain the virtual camera generation. The principle actually consists of two steps, the first step is the determination of individual virtual camera either by calibration or by geometrical analysis of ray sketching; and the second step is exploitation of stereovision information embedded in the sub-images. In the next few sections we will show that both the calibration based approach and geometrical analysis based approach used in previous single-lens trinocular system can be modified easily to cater for the virtual camera determination of virtual cameras of this multi-ocular system. The different number of virtual cameras relative to the trinocular stereo system results in different number of coordinate systems having different orientations and positions, different mapping between virtual camera image coordinates and real image plane coordinates, and also different disparity-depth recovery equations. All these are discussed in the following sections.

The basic requirements for building this system are repeated here:

1) the image plane of the CCD camera in use has consistent properties;

2) the multi-face filter is exactly symmetrical with respect to all of its apex edges;

3) the back plane of the multi-face filter is positioned parallel to the real camera image plane, and

4) the projection of the multi-face filter vertex on the camera image plane is located at the camera principle point and the projection of one apex edge of the filter on the image plane bisects the camera image plane equally and vertically.

If the above requirements are satisfied, the camera optical axis passes through multi-face filter vertex, the virtual cameras will then have identical properties and are symmetrically located with respect to the real camera optical axis. Thus the analysis of any one virtual camera would be sufficient as the results can be transposed to other virtual cameras theoretically.

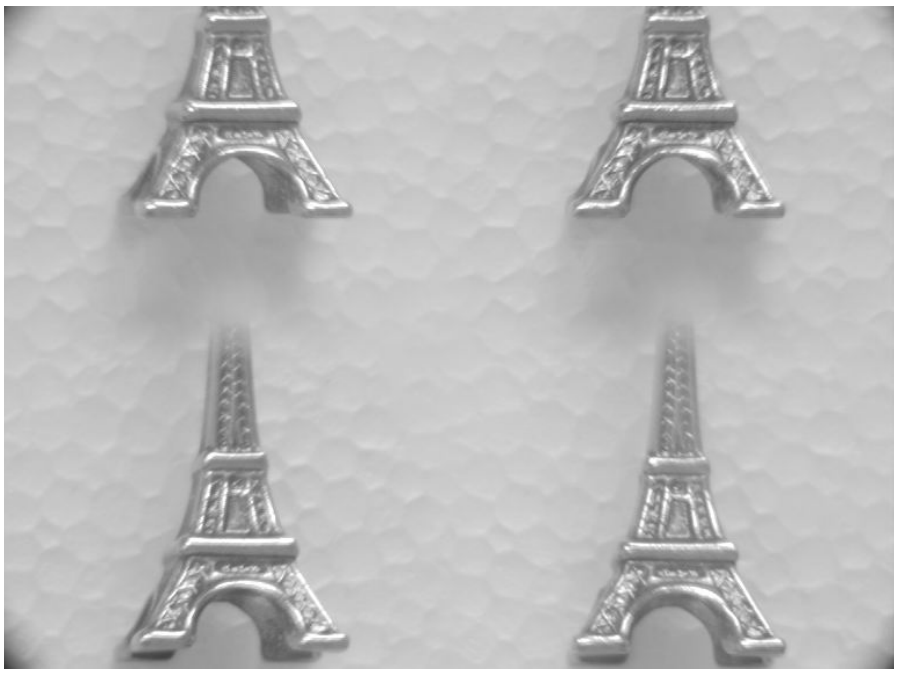

**Figure 6.2 One image captured by the single-lens multi-ocular system (4 faces)**

### 6.1.1 Determining the Virtual Cameras by Calibration

The calibration technique used to calibrate the virtual cameras of trinocular system in Chapter 5 can be also used for the multi-ocular system, with slight modifications. The same camera model is used and various coordinate systems can

be created on the virtual cameras analogously. The latter include the distorted virtual camera 2D image coordinate systems ($X_{d,i}$, $Y_{d,i}$), where $i = 1, 2, \ldots, n$, and $n$ is the total number of faces of the filter used; undistorted virtual camera 2D image coordinate systems ($X_{u,i}$, $Y_{u,i}$) and the 3D Virtual Camera Coordinate System located on the virtual camera optical center. ($X_{d,i}$, $Y_{d,i}$) can be linked to the computer image coordinates ($X_f$, $Y_f$) via:

$$X_{d,i} = (C_x - X_f) \bullet dx', \qquad Y_{d,i} = (C_Y - Y_f) \bullet dy', \tag{6.1}$$

where $dx'$ and $dy'$ are the pixel size of the computer sampled images and can be obtained by actual CCD pixel size times its resolution and then divided by computer sampled image resolution in both $x$ and $y$ directions. Hence the calibration of virtual cameras becomes possible. Each virtual camera can be calibrated one-by-one using the information provided by its correspondent sub-image captured by real camera image plane, from which the whole system can be fully described.

This system is now ready to perform depth recovery using the similar technique used for the trinocular system presented previously.

From the coordinate system set up for camera calibration the following equations can be obtained:

$$
\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = R_1 \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + T_1;
$$

..

...

$$
\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = R_i \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + T_i;
$$

..

...

$$
\begin{bmatrix} x_n \\ y_n \\ z_n \end{bmatrix} = R_n \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + T_n,
$$

**(6.2)**

where $i$ is the integer from 1 to $n$ ($n$ is total number of faces of the filter used), and

$$
R_i \equiv \begin{bmatrix} r_{i,1} & r_{i,2} & r_{i,3} \\ r_{i,4} & r_{i,5} & r_{i,6} \\ r_{i,7} & r_{i,8} & r_{i,9} \end{bmatrix}, \qquad T_i = \begin{bmatrix} T_{i,x} \\ T_{i,y} \\ T_{i,z} \end{bmatrix}.
$$

Each virtual camera should use the same world coordinates for the proceeding equation to be true.

$R_i$, $T_i$ and $f_i$ can all be obtained from calibration.

Also from the definition of the calibration coordinate setup, the following equations can be obtained:

$$x_1 = \frac{X_{u1}}{f_1} z_1 \qquad y_1 = \frac{Y_{u1}}{f_1} z_1;$$

..

...

$$x_i = \frac{X_{ui}}{f_i} z_i \qquad y_i = \frac{Y_{ui}}{f_i} z_i;$$

..

...

$$x_n = \frac{X_{un}}{f_n} z_n \qquad y_n = \frac{Y_{un}}{f_n} z_n.$$

**(6.3)**

Hence one set of linear equations can be obtained:

$$\frac{X_{u1}}{f_1} z_1 = r_{1,1} x_w + r_{1,2} y_w + r_{1,3} z_w + T_{1,x},$$
$$\frac{Y_{u1}}{f_1} z_1 = r_{1,4} x_w + r_{1,5} y_w + r_{1,6} z_w + T_{1,y},$$
$$z_1 = r_{1,7} x_w + r_{1,8} y_w + r_{1,9} z_w + T_{1,z};$$

..

...

$$\frac{X_{ui}}{f_i} z_i = r_{i,1} x_w + r_{i,2} y_w + r_{i,3} z_w + T_{i,x},$$
$$\frac{Y_{ui}}{f_i} z_i = r_{i,4} x_w + r_{i,5} y_w + r_{i,6} z_w + T_{i,y},$$
$$z_i = r_{i,7} x_w + r_{i,8} y_w + r_{i,9} z_w + T_{i,z};$$

..

...

$$\frac{X_{un}}{f_n} z_n = r_{n,1} x_w + r_{n,2} y_w + r_{n,3} z_w + T_{n,x},$$
$$\frac{Y_{un}}{f_n} z_n = r_{n,4} x_w + r_{n,5} y_w + r_{n,6} z_w + T_{n,y},$$
$$z_n = r_{n,7} x_w + r_{n,8} y_w + r_{n,9} z_w + T_{n,z}.$$

**(6.4)**

Reconstruction proceeding equations:

$$Ac = B$$

**(6.5)**

with:

$$A = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & -\frac{X_{u1}}{f_1} & 0 & .. & 0 \\ r_{1,4} & r_{1,5} & r_{1,6} & -\frac{Y_{u1}}{f_1} & 0 & .. & 0 \\ r_{1,7} & r_{1,8} & r_{1,9} & -1 & 0 & .. & 0 \\ r_{2,1} & r_{2,2} & r_{2,3} & 0 & -\frac{X_{u2}}{f_2} & .. & 0 \\ r_{2,4} & r_{2,5} & r_{2,6} & 0 & -\frac{Y_{u2}}{f_2} & .. & 0 \\ r_{2,7} & r_{2,8} & r_{2,9} & 0 & -1 & .. & 0 \\ .. & .. & .. & .. & .. & .. & .. \\ .. & .. & .. & .. & .. & .. & .. \\ .. & .. & .. & .. & .. & .. & .. \\ r_{n,1} & r_{n,2} & r_{n,3} & 0 & 0 & .. & -\frac{X_{un}}{f_n} \\ r_{n,4} & r_{n,5} & r_{n,6} & 0 & 0 & .. & -\frac{Y_{un}}{f_n} \\ r_{n,7} & r_{ni,8} & r_{n,9} & 0 & 0 & .. & -1 \end{bmatrix},$$

$$c = \begin{bmatrix} x_w & y_w & z_w & z_1 & z_2 & .. & z_n \end{bmatrix}^T,$$

and

$$B = \begin{bmatrix} -T_{1,x} & -T_{1,y} & -T_{1,z} & -T_{2,x} & -T_{2,y} & -T_{2,z} .. & ... & .... & -T_{n,x} & -T_{n,y} & -T_{n,z} \end{bmatrix}.$$

The least square solution is $c = (A^T A)^{-1} A^T B$.

**(6.6)**

All the elements in $A$ and $B$ can be obtained either from calibration or pixel reading from the image captured. Once the $z_i$'s in $c$ are found, the distance $Z_i$'s between real camera optical centers and the point of interest can be determined, and the average of $Z_i$'s can be used in the later experiments. The redundant information (as any two virtual cameras are enough for stereo) is handled by least square method, and the condition number appearing in equation (6.6) would not pose a problem when

calculating the matrix inverse (explained in section 5.1.1). The system is ready for depth recovery.

### 6.1.2 Determining the Virtual Camera by Geometrical Analysis of Ray Sketching

As has been explained at the beginning of this chapter, the same principle of determining the virtual camera used for the trinocular system can be applied to the multi-ocular system. Hence the geometrical analysis based approach used to determine the virtual cameras of single-lens trinocular system is also used to explain this multi-ocular system with some modifications. The differences are mainly caused by different size and geometry the multi-face filters, which are not very difficult to handle. Pin-hole camera is still used to approximate the virtual cameras.

For this geometrical analysis based approach, the assumptions made are: the real camera is not calibrated; the size and resolution of camera CCD chip are known; the computer sampled image resolution is known; geometry of the multi-face filter and also its relative position with respect to the real camera are known. The ray sketching in Figure 5.4 is still usable to understand this approach. This is because the determination of the individual virtual camera does relate to the angle between the inclined plane and the back plane of the prism, and does not relate to span angle or the area of each inclined plane which is the only change involved by using the filter of different number of faces.

A *quick review* on the description made about this approach in Chapter 5 is: find a point $P$ on the real camera image plane which defines one FOV boundary line of a virtual camera (its choice depends on how the effective range of the real camera image plane is defined) such that the line jointing point $P$ and focal point $F$ intersects

with the line $O''D$ (the line which bisects triangle $O''AC$) at point $M$, and this ray $PM$ after two refractions on filter surfaces becomes ray $NL$ (point $N$ is on plane $A'B'C'$) and goes into the view zone behind the filter. If this ray $NL$ defines the boundary of the captured scene or the interested boundary within one sub-region on real camera image plane, then it also defines the view boundary of the virtual camera that is correspondent to this sub-region.

Next, we look at ray $KO''$, where point $K$ is the camera image plane center and point $O''$ is the filter vertex, this ray becomes ray $JS$ (point $J$ is on plane $A'B'C'$) after two refractions. As this ray $KO''$ defines the real camera optical axis, then ray $JS$ defines the virtual camera optical axis according to the description about virtual camera model in section 5.1. By back-extending the ray $NL$ and $JS$, their intersection can be found, which is the optical center $F'$ of the virtual camera. This intersection always exists as ray $NL$ and $JS$ are located in a same plane.

This describes the basic ideas on how the virtual cameras are determined via geometrical analysis of ray sketching. As the symmetry is assumed for each virtual camera, the determined position and orientation of any virtual camera can be transposed to other virtual cameras easily via coordinate rotation skills.

Similar camera and image coordinate systems can be built on virtual cameras as what have been done with the calibration based approach, except that:

1) in the camera model, the optical centers of virtual camera are positioned behind the image plane for more accurate description on virtual camera generation;

2) the 2D computer image coordinate systems are rotated with respect to their $z$-axes such that their $x$-axes bisect the correspondent sub-regions on real camera image plane of each virtual camera for easier analysis. Hence 2D camera image coordinates can be linked to the computer image coordinates $(X_f, Y_f)$, $i = 1, 2, .., n$, via:

$$X_{d,i} = (X_f - C_x) \bullet \sin(\frac{180°}{n}(2i-1)) \bullet dx' + (Y_f - C_y) \bullet \cos(\frac{180°}{n}(2i-1)) \bullet dy'$$
$$Y_{d,i} = (X_f - C_x) \bullet \cos(\frac{180°}{n}(2i-1)) \bullet dx' + (C_Y - Y_f) \bullet \sin(\frac{180°}{n}(2i-1)) \bullet dy'$$

**(6.7)**

The field point testing process is still needed to determine the focal lengths of virtual cameras and real camera before the system is ready to perform depth recovery. The depth recovery equations are given as below:

The coordinate setup gives

$$\begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = {}^{w}R_1 \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} + {}^{w}T_1;$$

..

...

$$\begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = {}^{w}R_i \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} + {}^{w}T_i;$$

..

...

$$\begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = {}^{w}R_n \begin{bmatrix} x_n \\ y_n \\ z_n \end{bmatrix} + {}^{w}T_n,$$

**(6.8)**

where ${}^{w}R_i \equiv \begin{bmatrix} {}^{w}r_{i,1} & {}^{w}r_{i,2} & {}^{w}r_{i,3} \\ {}^{w}r_{i,4} & {}^{w}r_{i,5} & {}^{w}r_{i,6} \\ {}^{w}r_{i,7} & {}^{w}r_{i,8} & r_{i,9} \end{bmatrix}$ $\quad$ ${}^{w}T_i = \begin{bmatrix} {}^{w}T_{i,x} \\ {}^{w}T_{i,y} \\ {}^{w}T_{i,z} \end{bmatrix}$.

Please note that the same world coordinates (same origin and orientation) is used for each virtual camera coordinates to refer to.

From the geometrical analysis ${}^{w}R_i$, ${}^{w}T_i$ can be obtained. Also $f_i$ are taken as equal and can be obtained from a field point test.

It is also known that:

$$x_1 = -\frac{X_{u1}}{f_1} z_1, \qquad y_1 = -\frac{Y_{u1}}{f_1} z_1;$$

..

...

$$x_i = -\frac{X_{ui}}{f_i} z_i \qquad y_i = -\frac{Y_{ui}}{f_i} z_i;$$

..

...

$$x_n = -\frac{X_{un}}{f_n} z_n, \qquad y_n = -\frac{Y_{un}}{f_n} z_n.$$

**(6.9)**

Thus a set of linear equations can be obtained:

$$x_w = -{}^w r_{1,1} \frac{X_{u1}}{f_1} z_1 - {}^w r_{1,2} \frac{Y_{u1}}{f_1} z_1 + {}^w r_{1,3} z_1 + {}^w T_{1,x},$$

$$y_w = -{}^w r_{1,4} \frac{X_{u1}}{f_1} z_1 - {}^w r_{1,5} \frac{Y_{u1}}{f_1} z_1 + {}^w r_{1,6} z_1 + {}^w T_{1,y},$$

$$z_w = -{}^w r_{1,7} \frac{X_{u1}}{f_1} z_1 - {}^w r_{1,8} \frac{Y_{u1}}{f_1} z_1 + {}^w r_{1,9} z_1 + {}^w T_{1,z};$$

..

...

$$x_w = -{}^w r_{i,1} \frac{X_{ui}}{f_i} z_i - {}^w r_{i,2} \frac{Y_{ui}}{f_i} z_i + {}^w r_{i,3} z_i + {}^w T_{i,x},$$

$$y_w = -{}^w r_{i,4} \frac{X_{ui}}{f_i} z_i - {}^w r_{i,5} \frac{Y_{ui}}{f_i} z_i + {}^w r_{i,6} z_i + {}^w T_{i,y},$$

$$z_w = -{}^w r_{i,7} \frac{X_{ui}}{f_i} z_i - {}^w r_{i,8} \frac{Y_{ui}}{f_i} z_i + {}^w r_{i,9} z_i + {}^w T_{i,z};$$

..

...

$$x_w = -{}^w r_{n,1} \frac{X_{un}}{f_n} z_n - {}^w r_{n,2} \frac{Y_{u2}}{f_2} z_n + {}^w r_{n,3} z_n + {}^w T_{n,x},$$

$$y_w = -{}^w r_{n,4} \frac{X_{un}}{f_n} z_n - {}^w r_{n,5} \frac{Y_{u2}}{f_2} z_n + {}^w r_{n,6} z_n + {}^w T_{n,y},$$

$$z_w = -{}^w r_{n,7} \frac{X_{un}}{f_n} z_n - {}^w r_{n,8} \frac{Y_{u2}}{f_2} z_n + {}^w r_{n,9} z_n + {}^w T_{n,z}.$$

**(6.10)**

Manipulating the proceeding equations yields:

$$A'c' = B'.$$

**(6.11)**

with:

$$A' = \begin{bmatrix}
1 & 0 & 0 & {}^{w}r_{1,1}\frac{X_{u1}}{f_1} + {}^{w}r_{1,2}\frac{Y_{u1}}{f_1} - {}^{w}r_{1,3} & 0 & .. & 0 \\
0 & 1 & 0 & {}^{w}r_{1,4}\frac{X_{u1}}{f_1} + {}^{w}r_{1,5}\frac{Y_{u1}}{f_1} - {}^{w}r_{1,6} & 0 & .. & 0 \\
0 & 0 & 1 & {}^{w}r_{1,7}\frac{X_{u1}}{f_1} + {}^{w}r_{1,8}\frac{Y_{u1,}}{f_{1,}} - {}^{w}r_{1,9} & 0 & .. & 0 \\
1 & 0 & 0 & 0 & {}^{w}r_{2,1}\frac{X_{u2}}{f_2} + {}^{w}r_{2,2}\frac{Y_{u2}}{f_2} - {}^{w}r_{2,3} & .. & 0 \\
0 & 1 & 0 & 0 & {}^{w}r_{2,4}\frac{X_{u2}}{f_2} + {}^{w}r_{2,5}\frac{Y_{u2}}{f_2} - {}^{w}r_{2,6} & .. & 0 \\
0 & 0 & 1 & 0 & {}^{w}r_{2,7}\frac{X_{u2}}{f_2} + {}^{w}r_{2,8}\frac{Y_{u2}}{f_2} - {}^{w}r_{2,9} & .. & 0 \\
.. & .. & .. & .. & .. & .. & .. \\
.. & .. & .. & .. & .. & .. & .. \\
.. & .. & .. & .. & .. & .. & .. \\
1 & 0 & 0 & 0 & 0 & .. & {}^{w}r_{n,1}\frac{X_{un}}{f_n} + {}^{w}r_{n,2}\frac{Y_{un}}{f_n} - {}^{w}r_{n,3} \\
0 & 1 & 0 & 0 & 0 & .. & {}^{w}r_{ni,4}\frac{X_{un}}{f_n} + {}^{w}r_{n,5}\frac{Y_{un}}{f_n} - {}^{w}r_{n,6} \\
0 & 0 & 1 & 0 & 0 & .. & {}^{w}r_{ni,7}\frac{X_{un}}{f_n} + {}^{w}r_{n,8}\frac{Y_{un}}{f_n} - {}^{w}r_{n,9}
\end{bmatrix}$$

$$c' = \begin{bmatrix} x_w & y_w & z_w & z_1 & z_2 & .. & z_n \end{bmatrix}^T,$$

and

$$B' = \begin{bmatrix} {}^{w}T_{1,x} & {}^{w}T_{1,y} & {}^{w}T_{1,z} & {}^{w}T_{2,x} & {}^{w}T_{2,y} & {}^{w}T_{2,z} .. & ... & .... & {}^{w}T_{n,x} & {}^{w}T_{n,y} & {}^{w}T_{n,z} \end{bmatrix},$$

and the least square solution is $c' = (A'^{T}\ A')^{-1} A'^{T}\ B'$

**(6.12)**

All the elements in $A'$ and $B'$ can be obtained either from geometrical analysis of ray sketching or reading from the image captured. Once $z_i$'s are found, the distance $Z_i$'s between the real camera optical center and the point of interest can be determined, and the average of $Z_i$'s is used in experiment. And the condition number of $(A'^{T}\ A')$ is not a problem when finding its inverse.

As was observed from the trinocular system, the mathematics involved in this geometrical based approach may not be simpler than the calibration based approach, however using this approach a complicated calibration procedure, including the camera calibration software and hardware preparation and calibration operation itself can be avoided, and instead an alignment between the n-face filter and real camera and also a procedure of field point testing only need to be considered. Hence this approach gives a much simpler system implementation process.

## 6.2 Experiment and Discussion

Similar experimentation technique and devices used for the single-lens trinocular system (described in section 5.2) can be applied to this single-lens multi-ocular system with necessary modifications. The experimentation is designed and conducted to test the feasibility and accuracy of both the approaches that are used to model this virtual stereovision system. It is still be divided into three main steps. The first step is to calibrate the real camera to get its properties; and second step is to determine the virtual cameras either via calibration or via the geometrical analysis based approach which includes one field point testing; and the third step is the depth recovery test. One image captured during the calibration of virtual camera is shown in Figure 6.3 and one image captured during depth recovery is shown in Figure 6.2. In the experiment, the correspondence searching ends at pixel level and does not go into sub-pixels. The redundancy caused by the extra virtual camera during depth recovery (as any two virtual cameras would be enough for stereo) is handled by the least square method as shown in the depth recovery equation in the previous section (equation (6.6) and (6.12)).

The generalized multi-ocular theory is firstly tested for depth recovery using a 3F filter under the same conditions described in Chapter 5. The result obtained by using the theory of the generalized multi-ocular system is identical with the result obtained by using the theory of trinocular system: it is shown that for the depth ranged from 0.9m to 1.5m the geometrical analysis based approach can give an absolute depth recovery error of about 1% in average using a typical setup in the experiment (see equations (6.11) and (6.12), $n$ = 3), while the calibration based approach can give an error of about 3% under the same condition (see equations (6.5) and (6.6), $n$ = 3).

Subsequently, experiment was carried out with a filter with four faces. The only 4-face filter available for the experiment has a diameter of 49.8mm, and its $l$ is 35.3mm, $a$ is 25.33mm, $t$ is 3.87mm and $h$ is 5.3mm. The way used to define the geometrical properties of a prism given in Appendix C can also be applied here. Experiment shows that for the depth ranged from 0.5m to 0.75m the geometrical analysis based approach can give an absolute depth recovery error of about 2% on average using this prism in the experiment (see equations (6.11) and (6.12), $n$ = 4), while the calibration based approach can give an error about 6% (see equations (6.5) and (6.6), $n$ = 4). Table 6.1 gives some detailed depth recovery result ($\lambda$ = 45mm, here $\lambda$ is the smallest distance between optical centers of any two virtual cameras).

According to the results we can observe that both approaches can determine the system. The main constraint encountered in our experiments was the availability of the filters. The geometry of the four face filter used can roughly satisfy the requirements but it only gives baselines of about 35mm to 45mm in the experiment, which obviously limits the precision of depth recovery. In addition, the view zone of each virtual camera is also limited, which affects the accuracy of calibration and

hence the accuracy of the calibration based approach. Efforts are now being spent to acquire better filters for further testing.

In normal trinocular stereovision systems, the relative locations among the cameras affect the positions of epipolar lines on the camera image planes and hence the feasibility of using the epipolar constraints to reduce the correspondence searching effort. For example, if all the three cameras are positioned such that their optical axis are approximately on one plane and hence their baselines form very small angles with respect to each other, then their epipolar lines will have too many arbitral intersections and hence too many hypothesized correspondence points, because the resolution the CCD pixels is not infinite. A discussion on this problem has been given by [45], which suggests that the orthogonal epipolar lines (and hence orthogonal camera baselines) are the most optimal for the correspondence searching to get the fewest potential candidates in trinocular stereovision system when using the rule of epipolar constraints. This discussion is obviously also applicable in the case of the multi-ocular stereovision system. For the single-lens trinocular stereovision system described in this thesis the angles between any two baselines connecting any virtual cameras is about 60°, and for the single-lens four-virtual-camera stereovision described here the angles between any two virtual camera baselines are about 45° or 90° (the virtual camera image planes are taken as approximately coplanar), which are reasonably close to the optimal camera geometry to utilize the epipolar constraints according to the comments given in [45]. For any other single-lens multi-ocular stereovision system of this type with different number of virtual cameras, the angles between any two virtual camera baselines can be easily and approximately inferred and hence whether the optimal camera geometry is satisfied for application of epipolar constraints in correspondence searching can be easily known.

Another issue concerns the irregularity of multi-face filters geometry. Some filters may not completely satisfy the requirements made on the symmetry of multi-face filters. But once all the inclined faces are still intersecting at one point, both approaches can still be applied to model and determine this system. There is no change required for the calibration based approach as virtual cameras are to be calibrated one by one; for the geometrical analysis of ray sketching based approach, since the symmetry no longer holds, the analysis of the ray sketching of virtual camera now also needs to accomplished one by one according to the different geometry of each inclined filter face. These ideas can also be applied to deal with the imperfect positioning between the filter and CCD camera.

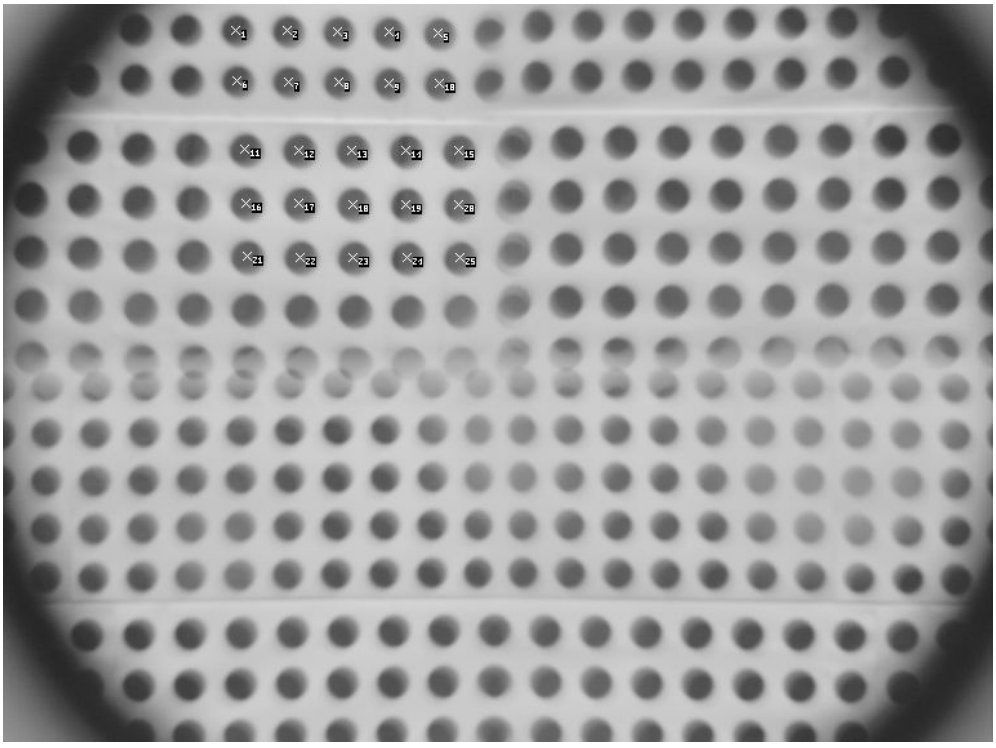

**Figure 6.3 Calibration of virtual cameras (4 faces filter used)**

| **Actual Depth** (mm) | **Correspondence Pixel Triplet** (in the order of left, bottom and right subsections of computer screen) | **Recovered Depth** (mm, Cali based Approach) | **Absolute Error in Percentage (%)** | **Recovered Depth** (mm, Geo Analysis based Approach) | **Absolute Error in Percentage (%)** |
|---|---|---|---|---|---|
| 500 | (223,27) (218,430) (624,438) (628,30) | 425.15 | 14.97 | 484.02 | 3.20 |
| | (124,54) (118,457) (527,464) (530,58) | 454.61 | 9.08 | 485.14 | 2.97 |
| | (203,60) (198,461) (605,469) (608,61) | 433.06 | 13.39 | 483.89 | 3.22 |
| 550 | (230,32) (223,460) (656,468) (660,33) | 499.29 | 9.22 | 546.87 | 0.57 |
| | (140,55) (135,484) (568,491) (570,59) | 513.3 | 6.67 | 546.1 | 0.71 |
| | (211,59) (205,486) (638,495) (641,62) | 500.17 | 9.06 | 545.21 | 0.87 |
| 600 | (209,43) (203,489) (655,499) (658,46) | 567.78 | 5.37 | 602.41 | 0.40 |
| | (127,63) (120,510) (572,517) (574,67) | 576.15 | 3.98 | 599.5 | 0.08 |
| | (192,67) (187,515) (637,523) (639,72) | 566.58 | 5.57 | 599.53 | 0.08 |
| 650 | (182,21) (178,481) (643,491) (647,24) | 631.79 | 2.80 | 653.22 | 0.50 |
| | (105,39) (99,504) (568,510) (570,45) | 643.66 | 0.98 | 660.07 | 1.55 |
| | (167,44) (161,505) (627,515) (631,47) | 635.42 | 2.24 | 655.33 | 0.82 |
| 700 | (190,36) (185,509) (666,519) (669,37) | 709.48 | 1.35 | 714.01 | 2.00 |
| | (119,51) (111,528) (593,535) (596,54) | 710.05 | 1.44 | 716.57 | 2.37 |
| | (175,56) (168,530) (650,541) (653,58) | 710.29 | 1.47 | 716.2 | 2.31 |
| 750 | (187,19) (181,504) (674,515) (678,20) | 789.85 | 5.31 | 774.14 | 3.22 |
| | (121,34) (113,522) (607,531) (610,38) | 776.89 | 3.59 | 774.69 | 3.29 |
| | (174,38) (167,523) (661,535) (663,41) | 778.76 | 3.83 | 771.56 | 2.87 |
| AVG | | | 5.57 | | 1.72 |

**Table 6.1 Recovered depth by multi-ouclar stereovision, 4 face filter, λ=45mm**

## 6.3 Summary

This ends the presentation on a single-lens multi-ocular stereovision system using multi-face filter. One image acquired by this system can be split into three or more sub-images and these sub-images can be taken as the images captured by multiple virtual cameras. As what have been discussed in the binocular and trinocular single-lens stereovision system presented in the two previous chapters, two different approaches were used to determine this system are given: one is based on calibration technique and the other one is based on geometrical analysis of ray sketching. The latter method does not require complex calibration and has a much

more simplified implementation process. Experiments were conducted to show the effectiveness of both approaches.

Compared to the single-lens binocular and trinocular system reported in the previous chapters, this modeling of the system is obviously more complex. It does, still possess the advantage in correspondence search of a stereovision as the two previous systems. This multi-ocular system can theoretically capture more stereo information, which may lead to better depth recovery accuracy. However, each virtual camera would have even less view zone since now four or more virtual camera share only one CCD matrix. For this system, we have tested our theoretical development with a setup which is able to capture four views of the same scene. A fair comparison of its performance with a binocular or trinocular system is again quite difficult due to the limitation in hardware and the acquisition of a suitable prism for the purpose. Nevertheless, our experiment results, though limited, are believed to be adequate to prove and demonstrate our idea.

# CHAPTER 7. CONCLUSION

A single-lens multi-ocular stereovision system using pyramid-like glass filters has been developed, analyzed, implemented and tested. A systematic investigation on this system of the types: binocular, trinocular and multi-ocular systems, has been carried out. Each image captured by this system can be split into two, three or multiple sub-images and these sub-images can be considered as the images simultaneously captured by the virtual cameras generated by the prism. The determination and hence the modeling of these virtual cameras is the key challenge in the study. Two different approaches were presented; one is based on a traditional camera calibration technique and the one is based on the geometrical analysis of ray sketching. The latter provides a method with a much simpler implementation requirement and better accuracy.

The involved work started with a literature review, followed by the system design and theoretical analysis. Subsequently software and hardware design, acquisition and implementation were carried out. With the system fully setup, the next natural step concentrates on experiment design and implementation and analysis of the results. The main constraints encountered during the study is the limited hardware that was available and affordable, in particular, the available filters did not have the required quality and property.

The system setup did present some problems in the calibration process; in particular, the filters found for the trinocular system and multi-ocular system (a four face system tested) do not create large enough view zones and the accuracy in the results of calibration suffers. This is further complicated by the less than ideal precision in the mechanical fixtures and the calibration board used. In addition, the

acquired filters do not provide large virtual camera baselines, which also limit the accuracy of depth recovery.

Nevertheless, the experiments have been carried out with reasonable accuracy. The final results showed that the two approaches developed to model and represent this system are both effective and sufficient to enable the system to work as binocular, trinocular and multi-ocular stereovision systems. The results, we believe are sufficient, within the allowable experimental errors, are adequate to verify the main ideas presented in this thesis.

As the stereo image pairs, triplets and sets are captured simultaneously, this system will have the typical advantages of binocular, trinocular and multi-ocular stereovision system, and in particular, the epipolar properties provide significant advantage in correspondence searching for the trinocular and multi-ocular systems. The system also has many other advantages, including:

1) Low cost. The use of one filter/prism to replace one or more cameras will significantly reduce the cost of building a multi-camera stereovision system;

2) Space saving;

3) More disparity information captured by one camera;

4) Fewer system parameters and easy implementation, especially for the approach of determining the system using geometrical analysis of ray sketching, which does not need calibration hardware and software at all, and;

5) No inter-camera synchronization needed since only one camera is used.

However, this system also have two disadvantages: firstly, all the virtual cameras created by this system share one CCD matrix, and hence each virtual camera virtually has fewer CCD pixels to represent a captured image; secondly, the common

view zone and baseline between virtual cameras are constrained by the filter size and shape, and hence this system possibly can only find its main application domain in close-range stereovision.

Finally, because of its many advantages, it can be expected that this deign will have good application potentials, such as, close-range 3D information recovery, indoor robot navigation / object detection, small size hand-hold stereovision system for dynamic scene, and economic 3D feature checker in industries, etc.

# CHAPTER 8. FUTURE WORK

Due to the limitation of time and hardware available, the different software of the system are not completely integrated, such as, the operations of calibration and depth recovery are running in different computer programs. It is recommended to integrate the different parts and different operational procedures of this system into one single computer program, so that this system is more like a product.

It is essential to acquire more precise devices so that more precise experiments can be carried out to further validate and investigate the system. Using current devices the experimentation can only basically prove our design and idea, and cannot support further and more thorough study. The limitations mainly have two aspects: one is on the mechanical stand of the system and in particular, the accuracy of calibration patterns provided by the calibration boards. The calibration boards need to be improved for better accuracy since they have caused noticeable calibration errors, especially when the virtual camera view zones created by filters are small and are not able to capture a large number of calibration patterns; another one is that on the filters currently in use can only provide very limited virtual camera baselines and view zones. Glass filters made with specified geometry according to our requirement are really costly (the mass-produced filters should not be as expensive). As a result, this system can only perform basic depth recovery in a close range and is not able to render good 3D reconstruction.

We recommend after acquiring better system devices based on our demand we should carry out further study on the error analysis and system performance, in particular, for the approach based on geometrical analysis of ray sketching. One intricate problem which might affect the accuracy of system accuracy is that: when the position and orientation of the virtual camera model is determined, different

selection of the interested area on the CCD plane, which also defines the boundary of the view scope of each virtual camera, will cause slight shift of the virtual camera positions and orientations and also the mapping of the interested points from the real camera image plane to the virtual camera plane. Some testing has been done on this point but no significant results were observed with the current experimental setup. Using a more precise setup could reveal more in this aspect.

Another area that may worth further study is the effects of the geometry of the filters, such as, how the angle between inclined plane and back plane of the prism will affect the effectiveness of the virtual camera modeling using the approach based on geometrical analysis.

# BIBLIOGRAPHY


[1] D. Marr, *Vision - A Computational Investigation into the Human Representation and Processing of Visual Information*, Freeman, San Francisco, 1982.

[2] L. Grewe and A. Kak, *Handbook of Pattern Recognition and Image Processing: Computer Vision*, pp. 239-317, Academic Press Inc, 1994.

[3] R. Gonzalez and R. Woods, *Digital Image Processing*, pp. 51-71, Addison Wesley, 1993.

[4] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, 1998.

[5] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, Cambridge, Massachusetts: MIT Press, 1993.

[6] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 1st ed., Cambridge, UK: Cambridge University Press, 2000.

[7] M. Sonka, V. Hlavac and R. Boyle, *Image Processing, Analysis and Machine Vision*, Thomson Asia, 2002.

[8] Y. Nishimoto and Y. Shirai, *A feature-based stereo model using small disparities*, in Proc. Conference on Computer Vision and Pattern Recognition (CVPR), pp. 192–196, 1987.

[9] W. Teoh and X. D. Zhang, *An inexpensive stereoscopic vision system for robots*, in Proc. Int. Conf. Robotics, pp. 186–189, 1984.

[10] A. Goshtasby and W. A. Gruver, *Design of a single-lens stereo camera system*, Pattern Recognition, vol. 26, pp. 923–936, 1993.

[11] M. Inaba, T. Hara and H. Inoue, *A stereo viewer based on a single camera with view-control mechanisms*, in Proc. of 1993 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, Vol. 3, pp. 1857-1865, 1993.

[12] S. Nene and S. Nayar, *Stereo with mirrors*, in Proc. Int. Conf. Computer Vision (ICCV), pp. 1087–1094, 1998.

[13] Z. Y. Zhang and H. T. Tsui, *3d reconstruction from a single view of an object and its image in a plane mirror*, in Proc. In Conf. Pattern Recognition, Vol. 2, pp. 1174-1176, 1998.

[14] Alexandre R.J. Francois, Gerard G. Medioni, Roman Waupotitsch, *Mirror symmetry ⇒ 2-view stereo grometry*, Image and Vision Computing, Vol. 2, pp. 137-143, 2003.

[15] J. Segen and S Kumar, Shadow Gestures: *3D hand pose estimation using a single camera*, in Proc. Computer Vision and Pattern Recognition, Vol. 1, pp. 479-485, 1999.

[16] Y. Nakazawa, T. Komatsu and T. Saito, *A simple cue-based method for camera calibration and 3-D shape measurement with a single moving camera*, in Proc. Int. Conf. on Image Processing, Vol. 2, pp. 293-296, 1996.

[17] Y. Suzuki and M. Takano, *Estimation of the position of a rigid body with a single camera*, in Proc. IECON'91, Int. Conf. on Industrial Electronics, Control and Instrumentation, Vol. 2, pp. 1309-1311, 1991.

[18] D. Moore and M. Hayes, *Tracking 3d position and orientation from 2D sequences using simple geometry*, in Conf. Record on Asilomar Conf. on Signals, Systems & Computers, Vol. 1, pp. 125-129, 1997.

[19] R. LeGrand and R. C. Luo, *Position estimation of selected targets*, in Proc. of Int. Conf. on Robotics and Automation, Vol. 2, pp. 1714-1719, 1996.

[20] E. H. Adelson and John Y. A. Wang, *Single lens stereo with a plenoptic camera*, IEEE Transactions on Pattern analysis and Machine Intelligence Vol. 14, pp. 99-106, 1992.

[21] J. Cardillo, M. A. Sid-Ahmed and J. J. Soltis, *3-D position sensing using a single camera approach*, in Proc. of the Midwest Symposium on Circuits and Systems, Vol. 1, pp. 325-328, 1990.

[22] Q. Z. Ye, S. H. Ong, X. Han，*A stereo vision system for the inspection of IC bonding wires*, International Journal of Imaging Systems and Technology, Volume 11, Issue 4 , Pages 254 – 262, 2000.

[23] D. H. Lee, I. S. Kweon and R. Cipolla, *A biprism-stereo camera system***,** in Proc. 1999 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, Vol. 1, pp. 82-87*,* 1999.

[24] D. H. Lee and I. S. Kweon, *A novel stereo camera system by a biprism***,** in Proc. 1999 IEEE Transactions on Robotics and Automation, Vol. 16, pp. 528-541, 2000.

[25] Lee. V. P, *Stereovision Using A Single CCD Camera*, Master Thesis, Department of Mechanical Engineering, NUS, 2001.

[26] Y. Xiao, *Stereovision Using Single CCD Camera*, Technical Report, Control and Mechantronics Laboratory, Department of Mechanical Engineering, NUS, 2000.

[27] L. C. Ng, *Stereo-vision Using Single CCD Camera*, Technical Report, Control and Mechantronics Laboratory, Department of Mechanical Engineering, NUS, 2001.

[28] C. W. Tan, *Stereo-vision Using Single CCD Camera*, Technical Report, Control and Mechantronics Laboratory, Department of Mechanical Engineering, NUS, 2003.

[29] K. B. Lim and Y. Xiao, *New understanding of single-lens stereovision using a biprism*, Proc. of IST/SPIE Symposium on Electronic Imaging, Vol. 5302, pp. 41-52, 2004.

[30] K. B. Lim and Y. Xiao, *Virtual stereovision system: new understanding on single-lens stereovision using a biprism*, J. Electron. Imaging 14, 043020, 2005.

[31] S. Ganapaphy, *Decomposition of transformation matrices for robot vision*, in Proc. of Int. Conf. on Robotics and Automation, pp. 130-139, 1984.

[32] T, M. Start, *Recovering the camera parameters from a transformation matrix*, in Proc. of DAPRA Image Understanding Workshop, pp. 264-271, 1984.

[33] Y. Yakimovsky, and R. Cunningham, *A system for extracting three-dimensional measurements from a stereo pair of TV cameras*, Computer Graphics and Image Processing, 7, pp. 195-210, 1978.

[34] O. Faugeras and G. Toscani, *The calibaration problem for stereo*, in Proc. of the Int. Conf. on Computer Vision and Pattern Recognition, pp. 15-20, 1986.

[35] O. Faugeras, Tuan Luong and S. Maybank, *Camera self-calibration: theory and experiments*, in Proc. of the 2nd European Conf. on Computer Vision, Vol. 588 of Lecture Notes in Computer Science, pp. 321-434, 1992.

[36] C. Huang and O. R. Mitchell, *Dynamic camera calibration*, in Proc. of Int. Symposium on Computer Vision, pp. 169-174, 1995.

[37] R. Y. Tsai, *A versatile camera calibration technique for high-accuracy 2D machine vision metrology using off-the-shelf TV cameras and lenses*, IEEE Journal of Robotics and Automation, Vol. RA-3, No. 4, pp. 323-344, 1987.

[38] M. Yachida, Y Kitamura and M. Kimachi, *Trinocular vision: new approach for correspondence problem*, Proc. of Eighth International Conference on Pattern Recognition, pp. 1041-1044, 1986.

[39] C. Hansen, N. Ayache and F. Lustman, *Towards real-time trinocular stereo*, Proc. of Second International Conference on Computer Vision, pp. 129-133, 1988.

[40] C. Stewart and C. Dyer, *The trinocular general support algorithm: a three-camera stereo algorithm for overcoming binocular matching errors*, Proc. of Second International Conference on Computer Vision, pp. 134-138, 1988.

[41] N. Ayache and C. Hanson, *Rectification of images for binocular and trinocular stereovision*, Proc. of International Conference on Pattern Recognition, Vol. 1, pp. 11-16, 1988.

[42] Y. Kitamura and M. Yachida, *Three-dimensional data acquisition by trinocular vision*, Advanced Robotics, Vol. 4, No. 1, pp. 29-42, 1990.

[43] U. Dhond and J. K. Aggarwal, *Binocular versus trinocular stereo*, Proc. of International Conference on Robotics and Automation, vol. 3, pp. 2045-2050, 1990.

[44] N. Ayache and F. Lustman, *Trinocular stereo vision for robotics*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 13, Issue 1, pp. 73-85, 1991.

[45] R. N. Chiou, C. H. Chen, K. C. Hung, and J. Y. Lee, *The optimal camera geometry and performance analysis of a trinocular vision system*, IEEE Trans. on System, Man, and Cybernetics, Vol. 25, No.8, pp. 1207-1220, 1995.

[46] V. G. S. Ernst, P. W. Sablik, J. Balendonck, Z. Houkes, P. P. L. Regtien, *A surface relief meter based on trinocular vision*, Proc. of International Conference on Image Processing and Its Applications, No. 410, pp. 281-285, 1995.

[47] J. Shen, P. Paillou, *Trinocular stereovision by generalized hough transform*, Pattern Recognition, Vol. 29, No. 10, pp. 1661-1672, 1996.

[48] T. Bouwmans, A. Bigand and J.P. Dubus, *A fuzzy model in trinocular vision matching*, Proc. of International Conference on Systems, Man, and Cybernetics, Vol. 4, pp. 2649-2652, 1996.

[49] P. Lehel, E. Hemayed and A. Farag, *Sensor planning for a trinocular active vision system*, Proc. of International Conference on Computer Vision and Pattern Recognition, Vo. 2, pp. 306-312, 1999.

[50] J. Mulligan and K. Daniilidis, *Trinocular stereo for non-parallel configurations*, Proc. of International Conference on Pattern Recognition, Vol. 1, pp. 567-570, 2000.

[51] S. Pollard, M. Pilu, S. Hayes and A. Lorusso, *View synthesis by trinocular edge matching and transfer*, Image and Vision Computing, Vol. 18, Issue 9, pp. 749-757, 2000.

[52] M. Agrawal and L. Davis, *Trinocular stereo using shortest paths and the ordering constraint*, Proc. of the IEEE Workshop on Stereo and Multi-Baseline Vision, pp. 3-9, 2001.

[53] S. Kurada and G. Rankin and K. Sridhar, *A trinocular vision system for close-range position sensing*, Optics & Laser Technology, Vol. 27, Issue 2, pp. 75-79, 1995.

[54] B. Ramsgaard, I. Balslev and J. Arnspang, *Mirror-based trinocular systems in robot-vision*, Proc. of 15th International Conference on Pattern Recognition, Vol. 4, pp. 499-502, 2000.

[55] Y. Xiao and K. B. Lim, *A Single-lens Trinocular Stereovision System Using a 3F Filter*, Proc. of IEEE Conference on Robotics, Automation and Mechatronics, pp. 396-400, 2004.

[56] A. A. Goshtasby, *Three-dimensional model construction from multiview range images: survey with new results*, Pattern Recognition, Vol. 31, Issue 11, pp. 1705-1714, 1998.

[57] D. M. Gavrila and L. S. Davis, *3-D model-based tracking of humans in action: a multi-view approach*, Proc. of International Conference on Computer Vision and Pattern Recognition, pp. 73-80, 1996.

[58] N. Grammalidis and M. G. Strintzis, *Disparity and occlusion estimation in multiocular systems and their coding for the communication of multiview image sequences*, Circuits and Systems for Video Technologies, Vol. 8, Issue 3, pp. 328-344, 1998.

[59] F. Lerasle, G. Rives and M. Dhome, *Tracking of human limbs by multiocular vision*, Computer Vision and Image Understanding, Vol. 75, Issue 3, pp. 229-246, 1999.

[60] S. Abbasi and F. Mokhtarian, *Automatic view selection in multi-view object recognition*, Proc. of International Conference on Pattern Recognition, Vol. 1, pp. 13-16, 2000.

[61] S. L. Dockstader and A. M. Tekalp, *Multiple camera tracking of interacting and occluded human motion*, Proc. of IEEE, Vol. 89, Issue 10, pp. 1441-1455, 2001.

[62] J. Black and T. Ellis, *Multi-camera image measurement and correspondence*, Measurement, Vol. 32, Issue 1, pp. 61-71, 2002.

[63] L. Zeinik-Manor and M. Irani, *Multiview constraints on homographies*, Pattern Analysis and Machine Intelligence, Vol. 24, Issue 2, pp. 214-223, 2002.

[64] Z. Q. Zhang, L. Zhu, S. Z. Li and H. J. Zhang, *Real-time multi-view face detection*, Proc. of International Conference on Automatic Face and Gesture Recognition, pp. 142-147, 2002.

[65] E. J. Gonzalez-Galvan, S. R. Cruz-Ramirez, M. J. Seelinger and J. J. Cervantes-Sanchez, *An efficient multi-camera, multi-target scheme for the three-dimensional control of robots using uncalibrated vision*, Robotics and Computer-Integrated Manufacturing, Vol. 19, Issue 5, pp. 387-400, 2003.

[66] N. Ploskas, D. Simitopoulos, D. Tzovaras, G. A. Triantafyllidis and M. G. Strintzis, *Rigid and non rigid 3D motion estimation from multiview image sequences*, Signal Processing: Image Communication, Vol. 18, Issue 3, pp. 185-202, 2003.

[67] P. Remagnino, A. I. Shihah and G. A. Jones, *Distributed intelligence for multi-camera visual surveillance*, Pattern Recognition, Vol. 37, Issue 4, pp. 675-689, 2004.

[68] H. Maas, *Image sequence based automatic multi-camera system calibration techniques*, ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 54, Issue 5-6, pp. 352-359, 1999.

[69] F. Pedersini, A. Sarti and S. Tubaro, *Accurate and simple geometric calibration of multi-camera systems*, Signal Processing, Vol. 77, Issue 3, pp. 309-334, 1999.

[70] B. D. Olsen and A. Hoover, *Calibrating a camera network using a domino grid*, Pattern Recognition, Vol. 34, Issue 5, pp. 1105-1117, 2001.

[71] J. Park, S. Jung, H. Choi, Y. Kim, and B. Lee, *Depth extraction by use of a rectangular lens array and one-dimensional elemental image modification*, Appl. Opt. 43, 4882-4895, 2004.

[72] Y. Xiao, K. B. Lim and W. M. Yu, *A prism based single-lens multi-ocular stereo image capture system*, Proc. of IEEE Conference on Multi-Media Modeling (MMM), pp. 469-472, 2006.

[73] Y. Xiao and K. B. Lim, *A prism-based single-lens stereovision system - from trinocular to multi-ocular,* Image and Vision Computing, Submitted in 2005 (doing 2nd draft now).

# APPENDICES

## A. Epipolar Constraints

This section firstly gives a simple review on the concept of epipolar constraint based on [4] and then gives a short discussion on the considerations of its mathematical implementation, and ends with a review of the application of epipolar constraints in trinocular stereovision.

*Stereo Correspondence* or *Correspondence Searching* is always the key problem in stereovision system. Normal correspondence algorithm can basically be categorized into two groups: correlation based and feature based, besides various constraints, such as geometrical constraints and uniqueness constraint. Correlation based correspondence searching algorithm is built on the assumption that pixels (or a window enclosing certain amount of pixels) in correspondence have similar intensities; feature based correspondence searching algorithm basically uses set of points which have distinguishable geometrical characteristics, such as, lines and corners. Both algorithms need a kind of analysis called epipolar constraint to assist and improve their performance under most circumstances. The epipolar constraint method is a geometry-based constraint that enables vision system to narrow down the search scope for corresponding points from one whole image to one line on the image. Figure A. 1 shows the definitions of epipolar hole, epipolar plane, and epipolar line (Please note that pin-hole camera model (image plane – projection center) used in this diagram is different from the camera model (image plane – optical center) used in most previous formal text for convenience, but the theory described in this section can all be applied to all the camera modes with certain coordinate system modifications).

In Figure A. 1, two canonical camera models are constructed by image plane $I_L$ and $I_R$ and projection centers $C_L$ and $C_R$. Assuming that a point $P$ is located in the common view zone for both cameras, and its projection on the left image plane $I_L$ is point $P_L$, and that its projection on the right image plane $I_R$ is point $P_R$. Because the optical axis of these two cameras are not parallel, the line linking $C_L$ and $C_R$, or, the base line, always intersect with the left and right camera image planes at two points, $e_L$ and $e_R$ as shown in Figure A. 1, these two points are called *epipolar holes*. If the optical axes of these two cameras are parallel, then epipolar holes can be taken at the infinite distance on both ends of base line.

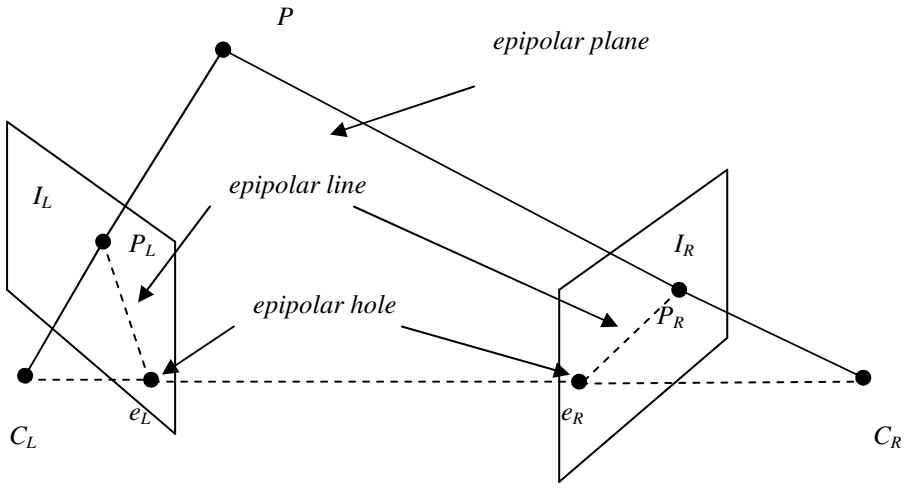


**Figure A. 1 Epipolar constraint**

Points $P$, $C_L$ and $C_R$ define a plane, and it must contain points $P_L$ and $P_R$. This plane is known as *epipolar plane*. However, in reality, the position of point $P$ are normally not known, instead only the projection of a point of interest on the image planes are known, for example, $P_L$ ($P_R$) from left (right) image plane $I_L$ ($I_R$). Thus, epipolar plane can also be defined by points $C_L$, $C_R$ and $P_L$ ($P_R$).

Next question is how to use this plane. Note that once the points $P_L$, $C_L$ and $C_R$ are known, one intersecting line between plane $P_LC_LC_R$ can always be found, the epipolar plane, and real camera image plane $I_R$, and vice-versa. This line $P_Re_R$ ($P_Le_L$) is called the *epipolar line*. The two corresponding points $P_L$ ($P_R$) are assumed to be located on this line. This is called *epipolar constraint*. Thus the searching scope of the corresponding points is narrowed down from one whole image to one line only, which will greatly reduce the search effort.

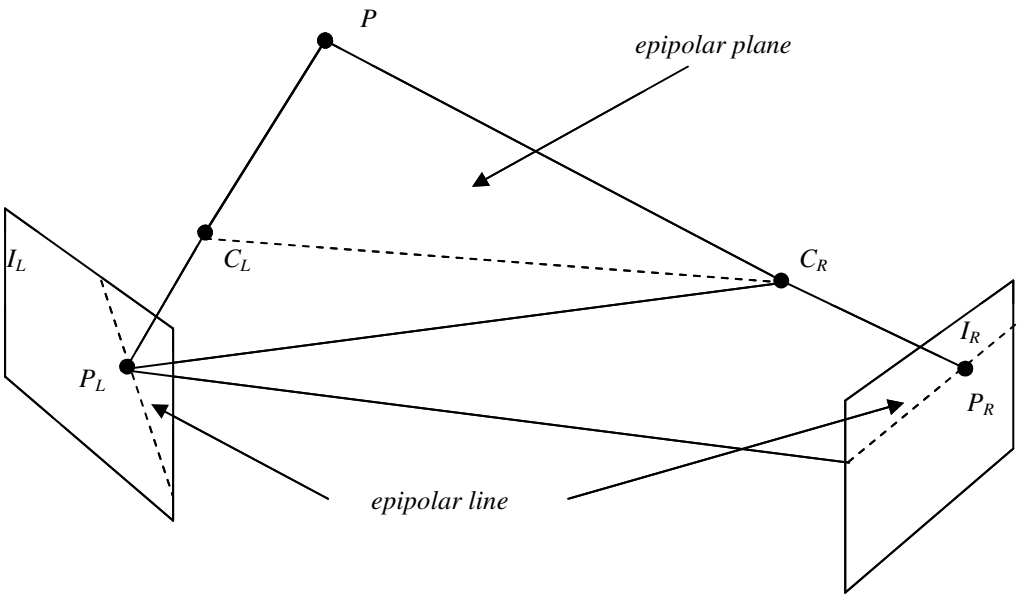


**Figure A. 2 Epipolar constraint (using different camera mode)**

In other camera models, same epipolar lines and epipolar plane can be constructed, as shown in Figure A. 2. Similarly, two canonical camera models are constructed by two image planes $I_L$ and $I_R$ and the corresponding optical centers $C_L$ and $C_R$. Assuming that point $P$ has a projection of the point $P_L$ on the left camera plane, and that of point $P_R$ on the right camera image plane, with the two optical centers $C_L$ and $C_R$ an epipolar plane will be formed. This plane will intersect the right

image plane $I_R$, to form an epipolar line, which can be used to reduce the searching effort in the corresponding point of $P_L$, i.e. $P_R$, and vice-versa.

Assuming that the coordinate values of point $P$, when referred to the world coordinate system defined on the left camera image plane is ${}^{L}C_L$ and that in the left image coordinate system (also defined on left camera image plane) is ${}^{L}c_L$, the relationship between ${}^{L}C_L$ and ${}^{L}c_L$ can be described by:

$$ {}^{L}c_L = \begin{bmatrix} M_L & 0 \end{bmatrix} \begin{bmatrix} {}^{L}C_L \\ 1 \end{bmatrix} = M_L \, {}^{L}C_L , $$

**(A. 1)**

where $M_L$ is defined as the calibration matrix describing the intrinsic parameters of this camera.

Similarly, ${}^{R}C_R$, ${}^{R}c_R$ and $M_R$ are defined in the same manner, and

$$ {}^{R}c_R = \begin{bmatrix} M_R & 0 \end{bmatrix} \begin{bmatrix} {}^{R}C_R \\ 1 \end{bmatrix} = M_R \, {}^{R}C_R . $$

**(A. 2)**

Assume that the relationship between the left and right world coordinate systems can be described by a translation $t$ and a rotation $r$ from the left coordinate system to that of the right, then:

$$ {}^{R}C_R = \left( r \, {}^{R}C_L - rt \right). $$

**(A. 3)**

Thus $c_R$ becomes:

$$ {}^{R}c_R = M_R ( r \, {}^{L}C_L - rt ) = \begin{bmatrix} M_R & -M_R rt \end{bmatrix} \begin{bmatrix} {}^{L}C_L \\ 1 \end{bmatrix}. $$

**(A. 4)**

If ${}^{L}C_L$, ${}^{L}C_R$ and $t$ are taken as free vectors as free vectors, expressed with respect to coordinate system of the left camera, as they are on the same plane, following equation can be obtained:

$${}^{L}C_L{}^{T}(t\times{}^{L}C_R)=0,$$

**(A. 5)**

Since ${}^{L}C_L = M^{-1}c_L$,

and ${}^{R}C_R = r\,{}^{L}C_R$, which gives ${}^{L}C_R = r^{-1}\,{}^{R}C_R$, hence

${}^{L}C_R = R^{-1}M_R^{-1}\,{}^{R}c_R$.

Then,

$$(M_L^{-1}\,{}^{L}c_L)^T(t\times r^{-1}M_R^{-1}\,{}^{R}c_R)=0$$

**(A. 6)**

If $t$ is $\begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$, then a skew symmetric matrix $s(t)$ can be used to replace the cross product operation of $t$, which is:

$$s(t)=\begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}.$$

**(A. 7)**

Finally next equation is obtained:

$${}^{L}c_L{}^{T}M_L^{-1}s(t)r^{-1}M_R^{-1}\,{}^{R}c_R=0,$$

**(A. 8)**

Let $F = M_L^{-1}s(t)r^{-1}M_R^{-1}$, $F$ is called the the *Fundamental Matrix* which relates the corresponding points in left and right views described in image coordinate systems (unit is in pixels), and let $E = s(t)r^{-1}$, $E$ is called the *Essential Matrix* which

relates the corresponding points in left and right views described in world coordinate systems (in physical units).

The concept of epipolar constraint can be used to simplify corresponding point searching; or in a reverse manner, if the two corresponding points are known already, epipolar constraint can be used to verify the correctness of system understanding or calibration. If the two optical centers of the two cameras are known, and one projection point (on either image plane) of any point of interest located in the common view zone of both cameras, then an epipolar plane can be determined and the epipolar line can be found also (on other image plane). The corresponding points must be located on this line, which provides a good way to test and verify the system presented in the formal text of this thesis.

To verify whether the hypothesized corresponding points are located on the epipolar line and hence satisfied the epipolar constraints, determining of this epipolar line mathematically on the camera image plane is needed. The following section provides the necessary mathematical procedures.

If three non-collinear points in space are known, which are ($x_1$, $y_1$, $z_1$), ($x_2$, $y_2$, $z_2$) and ($x_3$, $y_3$, $z_3$), then a plane determined by these three points is:

$$Ax + By + Cz = D,$$

**(A. 9)**

where,

$$A = y_1(z_2 - z_3) + y_2(z_3 - z_1) + y_3(z_1 - z_2)$$
$$B = z_1(x_2 - x_3) + z_2(x_3 - x_1) + z_3(x_1 - x_2)$$
$$C = x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2)$$
$$D = -(x_1(y_2 z_3 - y_3 z_2) + x_2(y_3 z_1 - y_1 z_3) + x_3(y_1 z_2 - y_2 z_1))$$

Another way of doing this is, assuming the positions of ${}^{w}P_L$, ${}^{w}C_L$ and ${}^{w}C_R$ are known, the normal of this epipolar plane is (the cross product of line ${}^{w}P_L{}^{w}C_L$ and line ${}^{w}P_L{}^{w}C_R$):

$$N_e = (\overrightarrow{{}^{w}P_L{}^{w}C_L}) \times (\overrightarrow{{}^{w}P_L{}^{w}C_R}),$$

**(A. 10)**

then this epipolar plane can be expressed in a standard way as:

$$N_e \bullet (p_e - a_e) = d_e,$$

**(A. 11)**

where $a_e$ is any point on the normal of this plane, and $p_e$ is any other point on this plane.

Assuming that the camera image plane is expressed as:

$$N_R \bullet (p_R - a_R) = d_R,$$

**(A. 12)**

where $a_R$ is any point on the normal of this plane, and $p_R$ is any other point on this plane.

Then the intersection line between the epipolar plane and the real camera image plane can be determined by:

$$p = k_e N_e + k_R N_R + k(N_e \times N_R),$$

**(A. 13)**

where

$$k_e = \frac{d_e(N_R \bullet N_R) - d_R(N_e - N_R)}{(N_e \bullet N_e)(N_R \bullet N_R) - (N_e \bullet N_R)^2},$$
$$k_e = \frac{d_r(N_e \bullet N_e) - d_e(N_e - N_R)}{(N_e \bullet N_e)(N_R \bullet N_R) - (N_e \bullet N_R)^2},$$

and $k$ is a parameter which changes for each corresponding point.

Thus epipolar line is determined mathematically by equation (A. 13).

If three cameras are used in a stereovision system, called trinocular stereovision, the epipolar constraint becomes very useful: it can help to narrow down the correspondence searching from a line to a point. Some earlier discussions regarding trinocular stereovision can be found in [38]-[44].

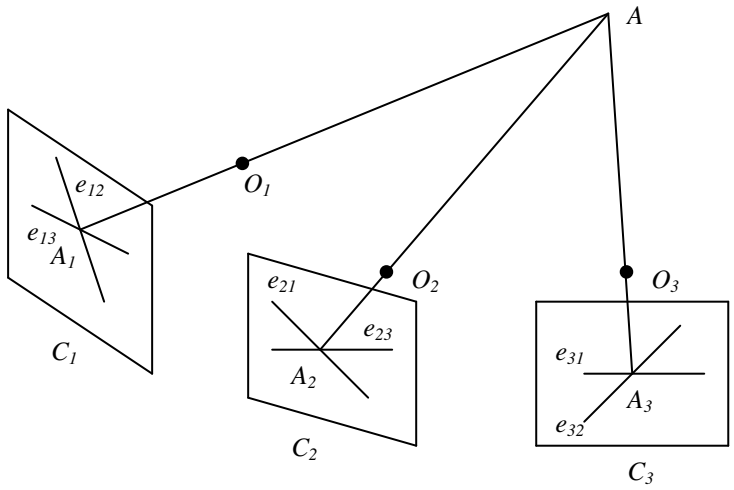


**Figure A. 3 Illustrations of Epipolar Constraints in Trinocular Stereovision**

See Figure A. 3, which shows a trinocular stereovision. The pinhole models of the three cameras are represented by three image plane $C_1$, $C_2$ and $C_3$ and the corresponding optical centers are $O_1$, $O_2$ and $O_3$. The projections of an object point $A$ on each camera mage plane are located at point $A_1$, $A_2$ and $A_3$. And the epipolar lines among these three image planes are drawn as shown. In a trinocular stereovision system, the hypothesized correspondence triplet is necessarily located at the intersection of epipolar lines. For example, if the correspondence points of $A_1$ are interested, on image plane $C_2$, the correspondence should be located on line $e_{21}$, and image plane $C_3$, the correspondence should be located on line $e_{31}$. Assume point $A_2$ is the hypothesized point on image plane $C_2$, and then on image plane $C_3$, the

hypothesized point should be located at also on the line $e_{32}$, then, the intersection point of line $e_{32}$ and $e_{31}$. Also, once $A_3$ is determined, the hypothesized point $A_2$ should also be located at the intersection of $e_{21}$ and $e_{23}$, which can be used to cross-check and validate hypothesis made on the correspondence. The idea is simple but very useful to reduce the effort the correspondence searching or validate the correctness of correspondence hypothesis.

## B. A Simple Calibration Technique

This section introduces a simple calibration technique to capture effective focal length of a CCD camera. Though it is theoretically more accurate and is able to find more system parameters, the calibration technique introduced in Chapter 3 is complex and difficult to implement, and any slight demerits on calibration patterns or disturbance introduced during calibration operation will cause errors. These errors might be the crucial factors that affect system performance, especially, the error on effective focal length, which directly affects depth recovery. This section will introduce a very simple calibration technique as a possible replacement of the existing calibration technique to determine or verify the effective focal length.

In a simple pin hole camera ray model (Figure A. 4), $A'B'$ stands for a line image projection on camera image plane, $A''B''$ stands for a real line, $O''$ is the midpoint of line $A''B''$and $O'$ is the midpoint of line $A'B'$. It is known that $OO'$ is vertical to $A'B'$. If it can be proved that $O'A'=O'B'$ when $O''A''=O''B''$, then $A'B'$ is parallel to line $A''B''$.

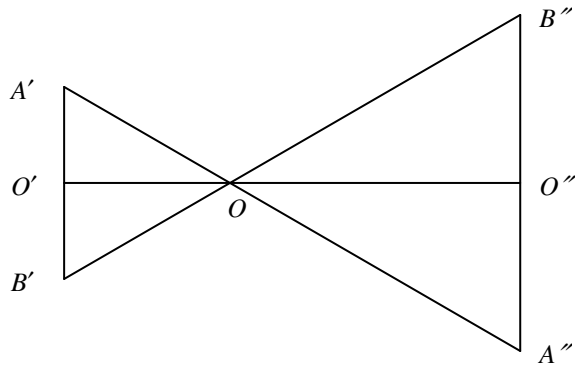


**Figure A. 4 A simple pin-hole camera model (side view)**

In three dimensional world (Figure A. 5), two crossing object line $C''D''$ and $A''B''$ project two image lines $C'D'$ and $A'B'$ from object plane $P''$ onto camera image plane $P'$. Using same reasoning described in the proceeding paragraph, it can be proved that $P'$ is parallel to $P''$ if $A'O' \leftrightarrows B'O'$ and $C'O' \leftrightarrows D'O'$, and when $B''O''=A''O''$ and $D''O''=C''O''$. Hence the absolute distance $OO''$ and focal length $O'O$ can be found using the properties of similar triangles.

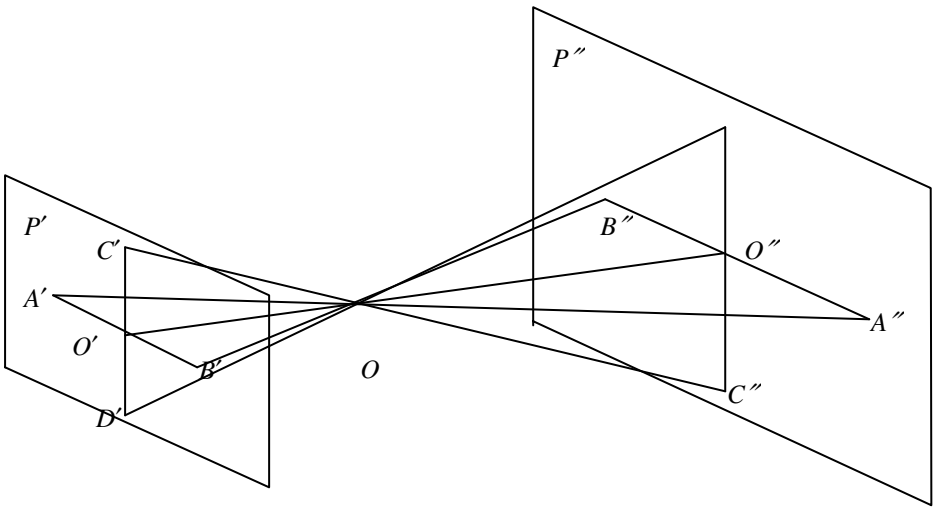


**Figure A. 5 A simple pin-hole camera model with two crossing object lines**

This calibration technique is simple and efficient in finding effective focal length.

The experimentation devices used to realize this calibration method are suggested to be: 1) CCD camera with known horizontal and vertical resolution and pixel size; 2) a calibration board with two crossing lines on a vertical plane and the length of each line is 0.5 meter. These two lines should divide each other equally; 3) another calibration board with the similar design but with different line length, e.g. 0.75 meter or 1 meter, for verification; 4) rulers.

And the experimentation procedures are suggested to be: 1) position CCD camera and make sure its optical axis is parallel to the work bench desk; 2) position and adjust the calibration board until the requirement described in algorithm is satisfied (i.e., $A'O' = B'O'$ and $C'O' = D'O'$). At this point, calibration board is parallel to the camera image plane. An extra verification procedure is that the four line segmentations should have the same length (not length in pixel numbers, but length in pixel numbers times pixel size. Pixels of CCD often have different dimensions in horizontal direction and vertical direction); 3) Measure the distance between camera lens center and calibration board. Use the knowledge of similar triangle to find effective focal length; 4) Use another calibration board to verify the result.

## C. Geometry Study of 3F Filter

In trinocular single-lens stereovision system, a 3F filter (tri-prism) is used to generate the three virtual cameras. The following discussion is a study of the geometry of a 3F filter, which provides some basic understanding of 3F filter geometrical structure. A 3F filter is shown in Figure A. 6 and Figure A. 7. Please note that in this diagram the round corners of the 3F filter used in the actual experimentations are trimmed to facilitate analysis. In this filter, the front plane profile is an equilateral triangle, and its side planes are perpendicular to the bottom plane, i.e,

$$\Delta ABC = \Delta A'B'C', \tag{A. 14}$$

and plane $ABC$ is parallel to plane $A'B'C'$.

In triangle $\Delta ABC$, $AB = BC = CA = l$, and $\angle ABC = \angle CAB = \angle BCA = 60°$.

Assuming that point $O$ is the center of triangle $\Delta ABC$, then,

$AO = BO = CO$.

Similarly, in triangle $\Delta A'B'C'$, $A'B' = B'C' = C'A' = l$,

and $\angle A'B'C' = \angle C'A'B' = \angle B'C'A' = 60°$.

Assuming that point $O'$ is the center of triangle $\Delta A'B'C'$, then,

$A'O' = B'O' = C'O'$.

If the apex is at point $O''$, then,

$AO'' = BO'' = CO'' = a$.

Points $O$, $O'$ and $O''$ are collinear.

In addition, line $O''O'$ is perpendicular to plane $ABC$ and plane $A'B'C'$.

Due to the geometry of the tri-prism, line $AA'$, line $BB'$ and line $CC'$ are parallel to line $O''O'$.

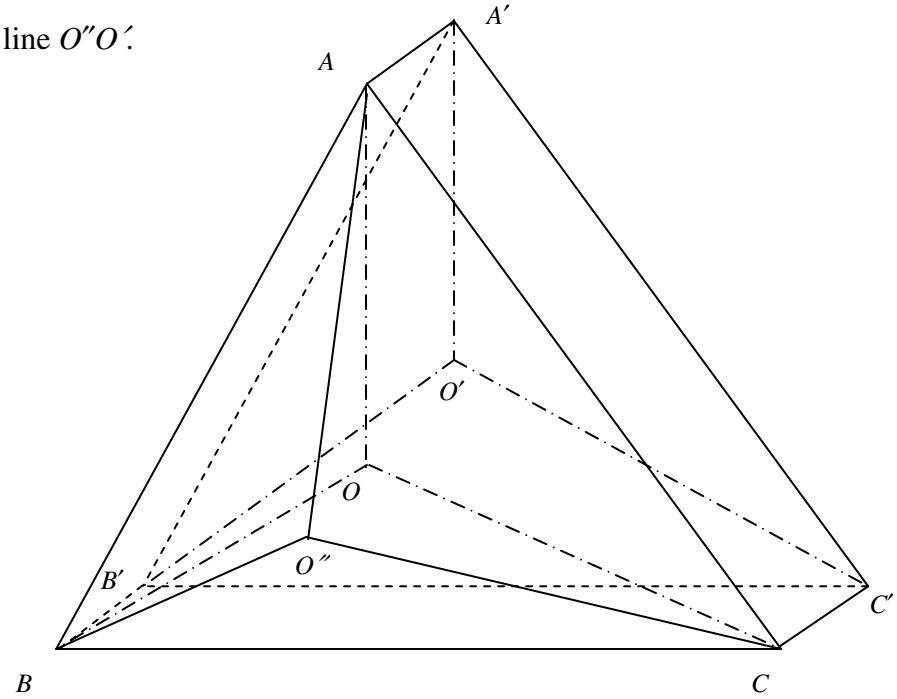


**Figure A. 6 Symbolic illustration of 3F filter structure**

Now it is needed to determine the important parameters of the 3F filter (see Figure A. 7): $\angle O''AC$, $\angle EO''F$, $\angle O''AO$, $\angle O''GO$, and $\angle AO''G$.

The procedures are outlined below.

To find $\angle O''AC$:

Let $\angle O''AC = \theta$,

Since $CO''^2 = AO''^2 + AC^2 - 2 \times AO'' \times AC \times cos\theta$, which is,

$$a^2 = a^2 + l^2 - 2 \times a \times l \times \cos\theta .$$

Hence $\theta = \cos^{-1}(\frac{l}{2a})$.

**(A. 15)**

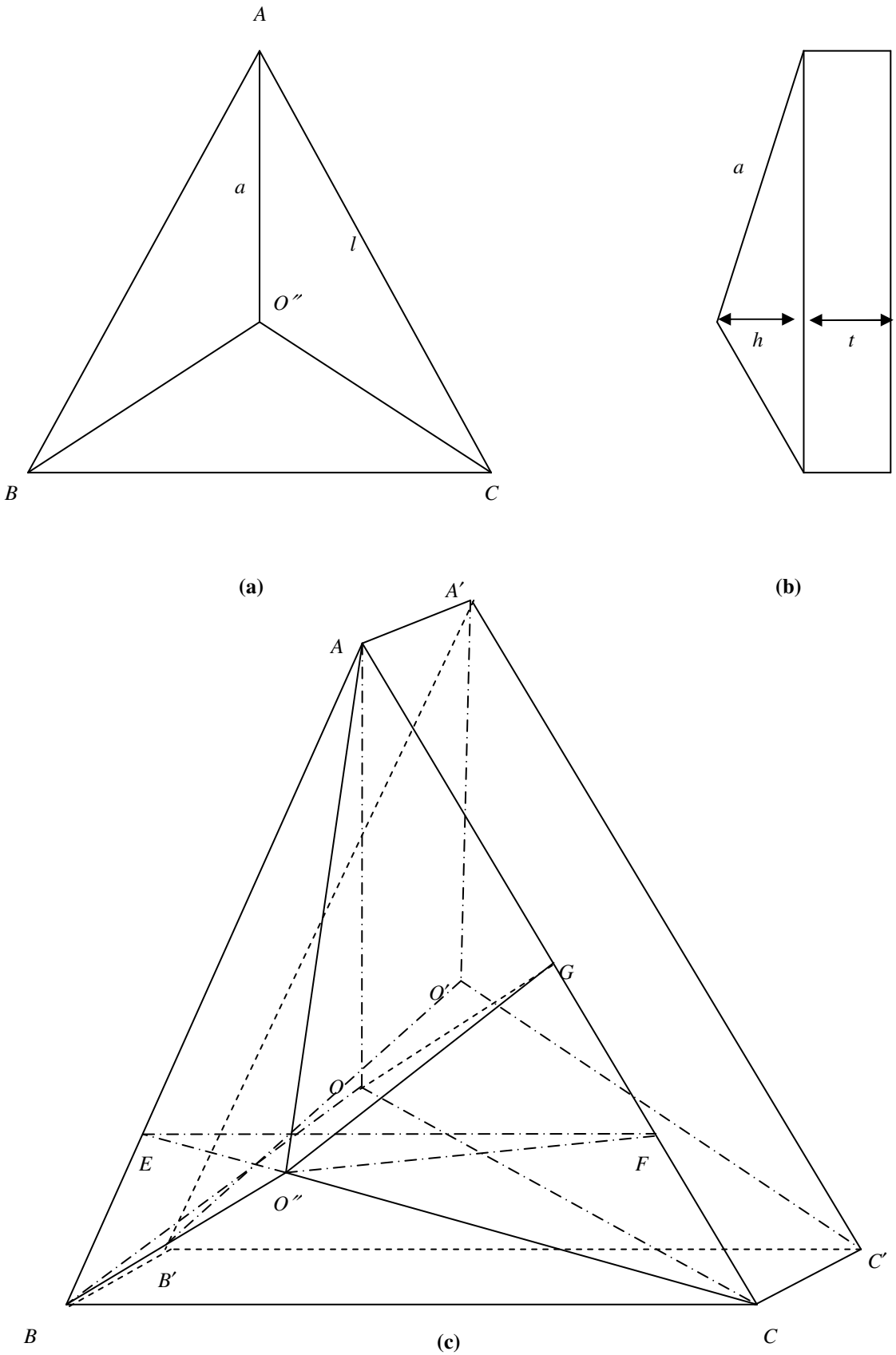


**Figure A. 7 3F filter 3D structure, with front and side view**

Then $\tan\theta = \dfrac{\sqrt{4a^2 - l^2}}{l}$.

**(A. 16)**

To find the length of $O''F$, where point $F$ is on line $AC$ and $FO'' \perp AO''$ (Figure A. 7),

$$O''F = \tan\theta \times AO'' = t\tan\theta \times \mathrm{a} = \frac{\sqrt{4a^2 - l^2}}{l} \times a .$$

**(A. 17)**

Also $O''E = O''F = \dfrac{\sqrt{4a^2 - l^2}}{l} \times a$, where point $E$ is on line $AB$ and $EO'' \perp AO''$.

To find the lengths of $AE$ and $AF$, where $AE = AF$,

$$AE = \sqrt{AO''^2 + O''E^2} = \sqrt{a^2 + \frac{4a^2 - l^2}{l^2}a^2} = \frac{2a^2}{l} = AF$$

**(A. 18)**

To find $EF$, since $\angle EAF = 60°$, $EF = AE$, following equation can be used:

$$EF^2 = AE^2 + AF^2 - 2 \times AE \times AF \times \cos 60° = AE^2$$
$$EF = \frac{2a^2}{l}$$

**(A. 19)**

To find $\angle EO''F$, following equation can be used

$$\cos\angle EO''F = \frac{O''E^2 + O''F^2 - EF^2}{2 \times O''E \times O''F} = \frac{2a^2 - l^2}{4a^2 - l^2}$$

**(A. 20)**

To find $\angle O''AO$, firstly,

$$AO = \frac{2}{3} \times \frac{\sqrt{3}}{2} \times l = \frac{\sqrt{3}}{3} l ,$$

and

$$h = \sqrt{AO''^2 - AO^2} = \sqrt{a^2 - \frac{1}{3}l^2},$$

$$\sin \angle O''AO = \frac{h}{a},$$

$$\text{So, } \angle O''AO = \sin^{-1}(\frac{h}{a}) = \sin^{-1}(\frac{\sqrt{a^2 - \frac{1}{3}l^2}}{a}).$$

**(A. 21)**

Let point *G* be the middle point of segment *AC* (see Figure A. 7). Obviously line $O''G \perp AC$ and $OG \perp AC$, which means $\angle O''GO$ is the angle between plane *OAC* (or plane *ABC*) and plane $O''AC$.

To find $\angle O''GO$,

$$\angle O''GO = \tan^{-1}(\frac{O''O}{OG}) = \tan^{-1}(\frac{h}{\frac{1}{3} \times \frac{\sqrt{3}}{2} \times l}) = \tan^{-1}(\frac{2\sqrt{3} \times \sqrt{a^2 - \frac{1}{3}l^2}}{l})$$

**(A. 22)**

This angle plays an important role when determining the position of virtual cameras as shown later.

Next to find $\angle AO''G$,

$\angle AO''G = 90° - \angle O''AC$, where $\angle O''AC$ is known from proceeding.

The gives a basic description on how to mathematically determine some important geometrical properties of a 3F filters. This knowledge can also be applied to any other multi-face filters that have similar structure after minor modification.