DESIGN OF A SINGLE-LENS STEREO CAMERA SYSTEM

*Ardeshir Goshtasby** and *William A. Gruver***

A camera system is designed that can obtain stereo information in a single shot and through a single lens. A single image obtained by this camera is equivalent to two images obtained by two perfectly aligned cameras with exactly the same optical properties. In this system, stereo is achieved by viewing the reflections of a scene in two mirrors that have a common axis. The angle between the mirrors functions like the baseline length in a two-camera system. When the mirrors are fixed to the camera in such a way that the common axis of the mirrors appears vertical in obtained images, corresponding points fall on the same scan line in an image, thus facilitating the correspondence process.

1. INTRODUCTION

Depth perception by stereo disparity has been extensively studied in computer vision[(1-4)]. Stereo disparity is a powerful cue which without the presence of other monocular and binocular cues, can measure depth[5]. Depth perception via stereo disparity is a passive method that does not require any special lighting or scanner to acquire the images. This method may be used to determine depths of points in indoor as well as outdoor scenes, and depths of points that are inches or miles away from the viewer.

A major step in stereo depth perception is the determination of correspondence between points in the images. This step is known as the *correspondence process*, and many algorithms for it have been developed[(6-14)]. Another major step in stereo depth perception is the computation of depth values from point correspondences [(15-18)]. The objective of this paper is neither to propose a new correspondence algorithm nor to implement a new method that determines depth values from the correspondences, but rather, to propose a new camera system that can obtain stereo images free of unnecessary geometric and intensity differences, thus facilitating the correspondence process. We will first determine the desirable properties of an ideal stereo camera system and then design a system that provides those properties.

A correspondence algorithm can produce more reliable matches if the underlying images have smaller intensity and geometric differences. Some geometric differences between stereo images are unavoidable. Actually, it is the local geometric difference between stereo images that results in the perception of depth. Image geometric difference due to rotation of one camera with respect to the other, however, does not have anything to do with the 3-D structure of a scene and only confuses the correspondence process. If the scene has Lambertian surfaces, there should be no difference in the intensities of corresponding points. Differences in the optical properties of two cameras, however, cause intensity differences between stereo images. These unnecessary geometric and intensity differences should be removed from stereo images, or reduced as much as possible, to increase the correspondence reliability and improve the accuracy of computed depth values.

When stereo images are obtained by two cameras, it is possible that the focal lengths and zoom levels of the cameras are slightly different. The two cameras may have lenses that do not have exactly the same optical properties. The optical axes of the cameras may not be parallel and normal to their baseline, and the two cameras may not be oriented in the same direction. These camera differences contribute to the geometric and intensity differences between stereo images, making the correspondence process less reliable. It is desirable for a camera system to obtain a pair of stereo images in a single shot and through a single lens so that unwanted geometric and intensity differences between the images can be avoided. A single camera with multiple shots will remove some of the unwanted image differences, but if images of a dynamic scene are required, since the scene can change during the time the images are obtained, positions of corresponding points will no longer relate to the depths of points.

* Ardeshir Goshtasby is with the Department of Electrical Engineering and Computer Science, The University of Illinois at Chicago, Chicago, Illinois 60680-4348, Email: ardeshir@uicbert.eecs.uic.edu

** William A. Gruver is with the Center for Robotics and Manufacturing Systems, University of Kentucky, Lexington, KY 40506-0108, Email: WILLIAM.A.GRUVER@ukwang.uky.edu

Stereo images obtained from two cameras that are not aligned require larger search areas to establish image correspondences. Larger search areas, however, increase the probability of mismatches. Therefore, it is desirable that the cameras are aligned so that corresponding points fall on the same scan line in the images. Then, given a point in one image, its correspondence can be determined in the other image by searching along the same scan line.

The distance between the camera lens centers, known as the ***baseline length***, determines the coarseness of the computed depth values. It is desirable to have a camera system whose baseline length can be varied so that when viewing nearby objects a small baseline length is used, and when viewing far away objects a larger baseline length is used.

Stereo cameras have to be calibrated to relate the coordinates of corresponding points in stereo images with the depths of points in a scene. It is desirable for the camera to have a simple geometry so that, by knowing the positions of a few points in the scene and their correspondences in the images, it could be calibrated.

To summarize, a desirable stereo camera system should have the following properties:

1. It should obtain stereo images in a single shot and through a single lens.
2. Its baseline length should be adjustable.
3. Its geometry should allow easy calibration of the camera.
4. It should not introduce unnecessary geometric and intensity differences between images in a stereo pair.
5. It should produce images with corresponding points lying on the same scan line.

In the following, after reviewing past works, a stereo camera system is designed that provides these properties.

2. PAST WORKS

Stereo images are typically obtained either by displacing a single camera in the scene[8] or using two cameras mounted on a platform separated by a small distance[7]. Although these set-ups acquire stereo images in a simple manner, they rarely provide the desired properties mentioned in the previous section. Unnecessary geometric and intensity differences between stereo images make the correspondence process unreliable. By devising a more accurate calibration process and using higher quality cameras, the unnecessary geometric and intensity differences between the images can be reduced. Some image differences, however, will still exist.

Nishimoto and Shirai[19] propose a single-lens camera system that can obtain stereo images. In this system, a glass plate is placed in front of the camera, and images are obtained with the plate at two different rotational positions, see Fig. 1a. When the glass plate is rotated, the optical axis of the camera shifts slightly, simulating two cameras with parallel optical axes. The obtained stereo images have very small disparities making the point correspondences easy. However, from the small disparities only coarse depth values can be obtained. This camera system requires two shots from a scene and, therefore, should be used only in static environments. Otherwise, the scene will change during the time the images are obtained, and point correspondences will no longer correspond to the depth values.

Teoh and Zhang[20] describe a single-lens camera system with the geometry shown in Fig. 1b. Two mirrors, fixed to the body of the camera, make a 45° angle with the optical axis of the camera. A third mirror which can rotate is placed directly in front of the lens. The rotating mirror is made parallel to one of the fixed mirrors and an image is obtained. Then, it is made parallel to the other fixed mirror and another image is obtained. Here also, although a single camera is used, the result is the same as using two cameras with parallel optical axes. Note that since two shots of a scene are required, the camera should only be used in static scenes.

Both of these methods considerably reduce unnecessary geometric and intensity differences between stereo images. But the cameras have parts that should be rotated when obtaining a pair of stereo images. Exact rotation of the parts is a major design issue in these systems, and since two shots of a scene are required, the scene under study must not change during the time the images are obtained. In the following, a camera system is designed that can obtain stereo images in a single shot and through a single lens. Since the two images of a stereo pair are obtained at the same time, this camera can be used in dynamic scenes.

3. A SIMPLE STEREO CAMERA GEOMETRY

To derive the geometry of the desired stereo camera, let's start with the geometry of a single pin-hole camera as shown in Fig. 2. Suppose point **p** with coordinates (x, y) is the image of point **P** in the scene with coordinates (X, Y, Z). Also, suppose **O** is the camera's lens center, **h** is the camera's piercing point with coordinates (x_0, y_0) (this is the point where the principal ray OZ intersects the xy-plane), f is the focal length of the camera, XYZ is the world coordinate system with origin at **O**, and XY-plane is parallel to the image plane. From the perspective rule we have

$$\frac{X}{Z} = \frac{x - x_0}{f} \tag{1}$$

$$\frac{Y}{Z} = \frac{y - y_0}{f}. \tag{2}$$

Now, suppose two such cameras are positioned horizontally and separated by $2B$ with their optical axes parallel and normal to the baseline (the line connecting the two cameras lens centers), as shown in Fig. 3. Suppose $\mathbf{O}_L$ and $\mathbf{O}_R$ are the left and right camera lens centers, respectively, and the world coordinate system is midway between the two camera lens centers with the Z-axis parallel to the optical axes of the cameras. Therefore, in this set-up, the coordinates of the left and right camera lens centers are $(-B, 0, 0)$ and $(B, 0, 0)$, respectively. Also, suppose $\mathbf{h}_l$ and $\mathbf{h}_r$ are the left and right camera piercing points with coordinates (x_{0l}, y_{0l}) and (x_{0r}, y_{0r}), respectively; **P** with coordinates (X, Y, Z) is a point in the scene whose projections in the left and right image planes are $\mathbf{p}_l$ and $\mathbf{p}_r$ with coordinates (x_l, y_l) and (x_r, y_r), respectively; and both cameras have the same focal length f. Then, using relations (1) and (2), the relation between a point in the scene and its images in the left and right cameras can be, correspondingly, written as

$$\frac{X + B}{Z} = \frac{x_l - x_{0l}}{f} \tag{3}$$

$$\frac{Y}{Z} = \frac{y_l - y_{0l}}{f} \tag{4}$$

and

$$\frac{X - B}{Z} = \frac{x_r - x_{0r}}{f} \tag{5}$$

$$\frac{Y}{Z} = \frac{y_r - y_{0r}}{f}. \tag{6}$$

From relations (4) and (6) we find

$$y_l - y_{0l} = y_r - y_{0r} \tag{7}$$

and if $y_{0l} = y_{0r}$ then

$$y_l = y_r. \tag{8}$$

Relation (8) shows that if the piercing points of the two cameras appear at the same scan line in the images, corresponding points lie on the same scan line in the images. This is a desirable property which simplifies the correspondence process, and we would like the new camera system to have piercing points that fall on the same scan line.

From (3) and (5) we have

$$\frac{f}{Z}(X + B) + x_{0l} = x_l \tag{9}$$

$$\frac{f}{Z}(X - B) + x_{0r} = x_r. \tag{10}$$

The horizontal displacement between corresponding points in the images, known as the *disparity*, may be computed by subtracting relation (10) from (9):

$$x_l - x_r = \frac{2Bf}{Z} + (x_{0l} - x_{0r}). \tag{11}$$

From (11) the relation between the depth of a point and its image disparity can be determined:

$$Z = \frac{2Bf}{[(x_l - x_r) - (x_{0l} - x_{0r})]}. \tag{12}$$

From relation (12) we can conclude that with the geometry of Fig. 3, depth is inversely proportional to image disparity, and for a given depth, the larger the baseline length the larger the disparity. Therefore, baseline length may be increased to increase the accuracy of measured depth values. This is a desirable property, and we would like the new camera system to provide it through a simple mechanism.

Corresponding points in images obtained from two aligned cameras that are displaced horizontally and have exactly the same focal length and zoom level will fall on the same scan line in the images. When two cameras are used, a small rotational difference between the cameras will move image points, and make corresponding points fall on different scan lines, thus complicating the correspondence process.

If the cameras are separated horizontally with optical axes parallel and perpendicular to the baseline, then when two vertical mirrors are placed equidistant to the camera lens centers, as shown in Fig. 4a, the reflections of the cameras in the mirrors produce images that fall on top of each other. If the optical axes of the two cameras are not parallel and normal to the baseline, or if the cameras are slightly rotated with respect to each other, the two virtual cameras will not match, producing double images. One way to align the cameras, therefore, is to shift and rotate one of the cameras with respect to the other while viewing the reflections of the cameras in the mirrors from point O until a perfectly matched image is obtained. One can actually align the cameras in this manner.

Our objective in using the mirrors is not to align the cameras but rather to design a stereo camera system using the geometry of Fig. 4a. Imagine putting a real camera in the place of the perfectly matched virtual cameras. Also suppose the polished sides of the mirrors are reversed so that the images of the real camera in the two mirrors become the virtual cameras shown in Fig. 4.b. With this geometry, taking an image from the real camera with the mirrors present would be the same as taking two images from the virtual cameras without the mirrors. In this way, with a single shot a stereo image can be obtained that is equivalent to two images obtained from two perfectly aligned cameras. This is the basic principle behind the proposed stereo camera system. Since a single image obtained in the new camera system contains stereo information, we will call it a *stereo image*. The left half of a stereo image corresponds to the image obtained from the right virtual camera while the right half of the image corresponds to the image obtained from the left virtual camera. In the following, the relation between points in a scene and their correspondences in a stereo image is derived.

4. THE NEW CAMERA GEOMETRY

Consider the camera geometry shown in Fig. 5. The real camera is positioned at point O′ facing the common axis of the mirrors. Suppose the reflections of the camera in the two mirrors are virtual cameras with centers at $\mathbf{O}_L$ and $\mathbf{O}_R$. Then, an image obtained from point O′ in the presence of the mirrors is the same as two images obtained from perfectly aligned cameras at points $\mathbf{O}_L$ and $\mathbf{O}_R$ without the mirrors. Suppose the distance of the real camera lens center to the mirrors' common axis is d and the angle between the mirrors is $(\pi - 2\alpha)$; then, the baseline length of the two virtual cameras is

$$2B = 2d \sin 2\alpha. \tag{13}$$

Using the geometry of Fig. 3 for a two-camera system, and assuming $X'Y'Z'$ is the new world coordinate system with the origin at the real camera lens center, and X', Y', and Z' axes are parallel to X, Y, and Z

axes, respectively, the relation between corresponding points in the XYZ and $X'Y'Z'$ coordinate systems may be written as

$$Z = Z' + d(1 + \cos 2\alpha); \quad X = X'; \quad Y = Y' \tag{14}$$

or

$$Z' = Z - d(1 + \cos 2\alpha); \quad X' = X; \quad Y' = Y. \tag{15}$$

Substituting relations (12) and (13) into (15), we obtain

$$Z' = \frac{2df \sin 2\alpha}{[(x_l - x_r) - (x_{0l} - x_{0r})]} - d(1 + \cos 2\alpha). \tag{16}$$

When an image is obtained from point **O'**, since the reflections of the principal rays of the two virtual cameras in the mirrors coincide, a single piercing point will be obtained. Thus, $x_{0l} = x_{0r}$ and relation (16) becomes

$$Z' = \frac{2df \sin 2\alpha}{(x_l - x_r)} - d(1 + \cos 2\alpha). \tag{17}$$

Note that since the image obtained by the real camera is the reflection of images obtained from the virtual cameras in the mirrors, image coordinates should increase from right to left. However, since in an image, coordinates increase from left to right, $(x_l - x_r)$ should be replaced by $n - (x_l' - x_r')$, where $(x_l' - x_r')$ is the distance between the corresponding points in the left and right halves of a stereo image, and n is a constant to be determined. n is the horizontal shift between the coordinates of the two image halves. Therefore, to determine the depth of a point in a stereo image obtained from the camera at point O', one has to determine the horizontal distance between corresponding points in the image, substitute it into

$$Z' = \frac{2df \sin 2\alpha}{n - (x_r' - x_l')} - d(1 + \cos 2\alpha), \tag{18}$$

and determine the value of Z'.

Parameters f and n can be computed by calibrating the camera using the coordinates of two points in the scene and their correspondences in the two image halves. Parameters d and α are either given or should be determined manually. Note that the units of measurement used in the scene and in the images are different. The ratio of the two units of measurements–let's denote it by u–will be a part of the focal length f in formula (18). The value of f obtained by calibrating the camera, therefore, would not be the actual focal length–let's denote it by F–but rather the product of the actual focal length and the ratio of the two units of measurement: $f = F \times u$. If F is known, then u and n become the two unknowns in relation (18), which can be determined by using the coordinates of two points in a scene and their correspondences in a stereo image.

5. PROPERTIES OF THE NEW CAMERA SYSTEM

From relation (18) we see that depth increases with the increase of the distance between corresponding points. This fact can be verified geometrically also from Fig. 6. Point **A** is closer to the viewer than point **B**. The images of points **A** and **B** in the stereo image would be the images of intersections of lines connecting the points to the centers of the virtual cameras with the mirrors. We see that points $\mathbf{A}_R$ and $\mathbf{A}_L$ fall inside the interval delimited by points $\mathbf{B}_R$ and $\mathbf{B}_L$, showing that the distance between images of point **A** is shorter than the distance between images of point **B**.

The field of vision of the camera is equal to 4α as shown in Fig. 7a. This field of vision is maximum when $4\alpha = (\pi - 2\alpha)$ or $\alpha = \frac{\pi}{6}$. When $\alpha > \frac{\pi}{6}$ the reflection of a scene in one mirror will appear in the other mirror causing multiple images of the scene in the same mirror. If the width of the mirrors is large enough, the field of vision decreases with a decrease in the value of α, as shown in Figs 7b and 7c. When α becomes zero, the field of vision of the camera becomes minimum and no point in the scene will be visible in both image halves.

Note that by increasing α, the field of vision of the camera increases and at the same time the distance between corresponding points increases. Since increased distance between corresponding points means more accurate measurement of depth, we see that by increasing α the same effect as increasing the baseline length in a two-camera system is obtained. This is an obvious result since by increasing α, the single-camera model simulates a two-camera model with a larger baseline length.

The field of vision of the camera decreases as the sizes of the mirrors decrease, as can be observed by comparing Figs 7b and 7d. Simple geometry shows that (see Fig. 7e) the width of each mirror should be at least $2d \sin \alpha$ for points in infinity to remain in the field of vision of the camera:

$$w \geq 2d \sin \alpha. \tag{19}$$

Or, for a fixed w, the angle between the mirrors should be at least equal to $(\pi - 2\alpha)$, where $\alpha = \sin^{-1}\left(\frac{w}{2d}\right)$, in order for points in infinity to be visible in both halves of a stereo image. We, therefore, see that α should be at most equal to the smaller of $\frac{\pi}{6}$ and $\sin^{-1}\left(\frac{w}{2d}\right)$. For instance, when $w = 2$ inches and $d = 4$ inches we should have $\alpha \leq 14°$, and when $w = 3$ inches and $d = 2$ inches we should have $\alpha \leq 30°$.

The mirrors should be installed in front of the camera such that the optical axis of the camera makes an angle either slightly smaller or larger than 90° with the common axis of the mirrors as shown in Fig. 8. Otherwise, the camera itself will appear in the obtained images. Note that rotating the camera about the X' axis (see Fig. 5) will not change any of the above formulas. Only the virtual cameras will be rotated by the same amount, while still keeping their optical axes parallel and normal to the baseline. To install the mirrors symmetric with respect to the optical axis of the camera, the mirrors should be adjusted such that the images of any point in the plane formed by the camera lens center and the common axis of the mirrors' appear equidistant to the mirrors' axis in the stereo image.

Except for the fact that the new camera system obtains stereo information in a single shot, it provides the same property as two perfectly aligned cameras with exactly the same optical properties. This camera eliminates inter-camera distortions, but intra-camera distortions still exist between stereo image halves which should be removed by a process known as the *decalibration*[21]. The decalibration process transforms an image obtained by a regular camera to one obtained by a pin-hole camera.

6. CONCLUSIONS

A camera system that can obtain stereo information in a single shot and through a single lens was described. This camera system has the following properties:

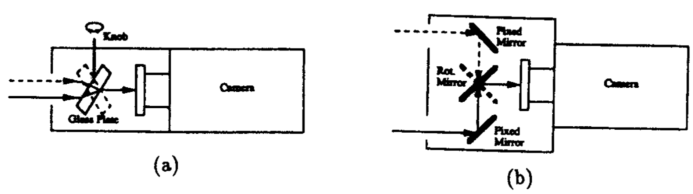
1. A single image obtained by this camera is equivalent to two images obtained from two perfectly aligned cameras with exactly the same optical properties.
2. Corresponding points lie on the same scan line in an image.
3. Inter-camera distortions no longer exist since a single camera is used to obtain the images. Intra-camera distortions may still exist which should be removed by a decalibration process[21].
4. The distance between corresponding points in a stereo image increases as the depth of the object point in the scene increases.
5. The angle between the mirrors should be larger than $\frac{2\pi}{3}$ to avoid reflection of one mirror into another and creation of multiple images of a scene in the same mirror.
6. In order that points in infinity become visible in both halves of a stereo image, it is required that the angle between the mirrors be larger than $(\pi - 2\alpha)$, where $\alpha = \sin^{-1}\left(\frac{w}{2d}\right)$, w is the width of the mirrors, and d is the distance of the camera lens center to the common axis of the mirrors.
7. The field of vision of the camera increases with the decrease of the angle between the mirrors.
8. Decreasing the angle between the mirrors has the same effect as increasing the baseline length in a two-camera system. Therefore, the angle between the mirrors may be varied to control the depth perception accuracy.

This camera system provides the desirable properties mentioned in the Introduction and is useful in applications where baseline length has to be varied to view nearby as well as far away objects.

REFERENCES

1. G. Sperling, Binocular vision: A physical and neural theory, *The American Journal of Psychology,* vol. 83, no. 4, 461–534 (1970).
2. S. T. Barnard and M. A. Fischler, Computational stereo, *ACM Computing Surveys,* vol. 14, no. 4, 553–572 (1982).
3. G. F. Poggio and T. Poggio, The analysis of stereopsis, *Ann. Rev. Neurosci,* vol. 7, 379–412 (1984).
4. M. R. M. Jenkin, A. D. Jepson, and J. K. Tsotsos, Techniques for disparity measurement, *Image Understanding,* vol. 53, no. 1, 14–30 (1991).
5. B. Julesz, Binocular depth perception without familiarity cues, *Science*, vol. 145, 356–362 (1964).
6. K. Mori, M. Kidode, H. Asada, An interactive prediction and correction method for automatic stereocomparison, *Computer Graphics and Image Processing*, vol. 2, 393–401 (1973).
7. M. Levine, D. A. O'Handley, and G. M. Yagi, Computer determination of depth maps, *Computer Graphics and Image Processing*, vol. 2, 131–150 (1973).
8. R. Nevatia, Depth measurement by motion stereo, *Computer Graphics and Image Processing*, vol. 5, 203–214 (1976).
9. D. Marr and T. Poggio, A computational theory of human stereo vision, *Proc. R. Soc. Lond.*, vol. B-204, 301–328 (1979).
10. S. T. Barnard and W. B. Thompson, Disparity analysis of images, *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 2, no. 4, 333–340 (1980).
11. K. Prazdny, Detection of Binocular Disparities, *Biol. Cybern.* vol. 52, 93–99 (1985).
12. N. Ayache and B. Faverjon, Efficient registration of stereo images by matching graph descriptions of edge segments, *Int. J. Computer Vision,* 107–131 (1987).
13. S. T. Barnard, Stochastic stereo matching over scale, *Int. J. Computer Vision,* vol. 3, 17–32 (1989).
14. D. J. Fleet, A. D. Jepson, and M. R. M. Jenkin, Phase-based disparity measurement, *Image Understanding,* vol. 53, no. 2, 198–210, 1991.
15. D. B. Gennery, Stereo-camera calibration, *Proc. Image Understanding Workshop,* 101–107 (1979).
16. O. D. Faugeras and G. Toscani, The calibration problem for stereo, *Proc. Computer Vision and Pattern Recognition*, 15–20 (1986).
17. H. Takahashi and F. Tomita, Self-calibration of stereo cameras, *Proc. Int. J. Conf. Computer Vision,* 123–128 (1988).
18. T. M. Strat, Recovering the camera parameters from a transformation matrix, *Proc. Image Understanding Workshop*, 264–271 (1984).
19. Y. Nishimoto and Y. Shirai, A feature-based stereo model using small disparities, *Proc. Computer Vision and Pattern Recognition,* 192–196 (1987).
20. W. Teoh and X. D. Zhang, An inexpensive stereoscopic vision system for robots, *Proc. Int. Conf. Robotics*, 186–189 (1984).
21. A. Goshtasby, Correction of image deformation from lens distortion using Bezier patches, *Computer Vision, Graphics, and Image Processing,* vol. 47, 385–394 (1989).

Fig. 1.



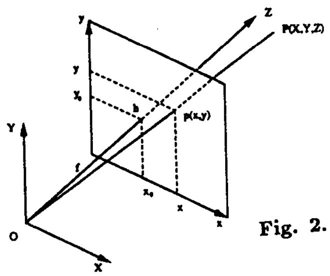


Fig. 2.

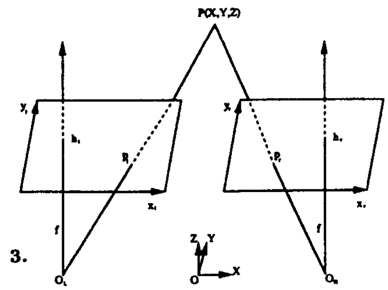


Fig. 3.

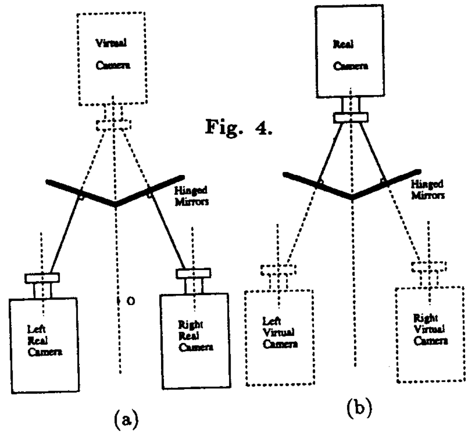


Fig. 4.

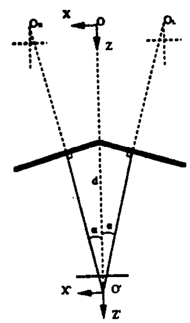


Fig. 5.

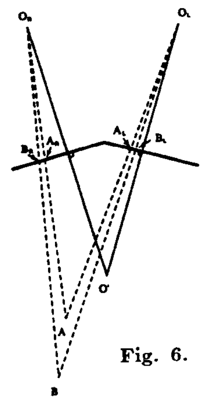


Fig. 6.

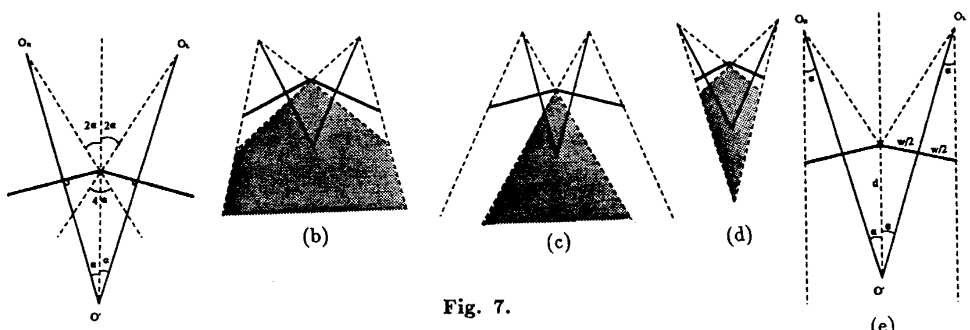


Fig. 7.

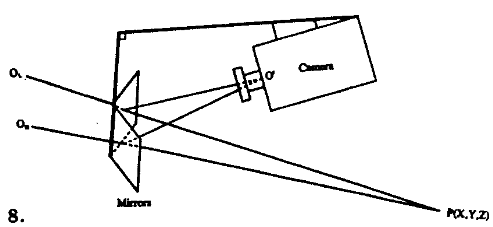


Fig. 8.