CAPTURING THE THIRD DIMENSION

JRT LANE
UNIVERSITY OF PRETORIA

PO BOX 17314
GROENKLOOF
PRETORIA
0027
REPUBLIC OF SOUTH AFRICA
Tel. No. 27 012 361 4853
Jlane_ff@Yahoo.com
THIS IS A REGISTERED PATENT

Abstract: *Capturing 3-D interaction in a non-intrusive and natural way presents some obstacles. A digital camera has some appealing attributes to overcome such problems despite the fact that pictures taken by a camera have no inherent depth information. Several computer vision techniques exist to obtain 3-D information from pictures. Some of these methods tend to be expensive while others are not fully 3-D. A novel solution to the problem has been found which has numerous advantages over existing methods. This alternative 3-D vision system can be used to achieve 3-D interaction in a seamless and non-intrusive manner.*

Keywords: Interaction, non-intrusive, 3-D, computer vision system.

Computing Review Categories: I.3.1, I.3.6, I.4.0

CAPTURING THE THIRD DIMENSION:
ACHIEVING REAL-TIME 3-D NON INTRUSIVE INTERACTION USING A SINGLE CAMERA

1 INTRODUCTION

"For now we see in a mirror, darkly; but then, face to face; now I know in part, but then shall I know even as also I am known."

1 Cor 13 vs. 12. [1]

When using a virtual world, a user should not be hindered or restricted by the equipment that monitors the user's movements. Interaction should be seamless. Consider a virtual climbing wall. In the real world equivalent, light clothing that allows a certain freedom of movement is worn; e.g. a pair of shorts and a T-shirt. The whole body has to work together to pull off certain moves. In virtual reality (VR), to track the movement of every limb a full body suit would be required. This would ultimately restrict movement. A non-intrusive mechanism is needed to track such movements. Computer vision (using a camera and a computer to mimic human vision) provides a means of achieving interaction without getting in the way.

The problem with using a camera is that the pictures produced are only 2-D. Some cleverness is needed to calculate distances of objects from the camera. This is known as the 2-D to 3-D problem. Multiple cameras and other depth detection techniques have been used to overcome this problem. It should be noted that most 3-D vision systems tend to be expensive, both in computational and pecuniary terms.

A new 3-D computer vision system has been discovered which presents solutions to the above problems. This method provides a means of effectively obtaining 3-D information by making use of only a single camera and one or more mirrors. Determining the 3-D information can be done at the frame rate of the camera and for this reason the system can be used as an interaction device. This device will be non-intrusive.

It is this method that is discussed in this report. It is juxtaposed to other conventional methods. An application will be discussed as well as future uses and research. Finally conclusions are drawn.

2 SOME PROBLEMS

2.1 The 2-D to 3-D Problem:

A single image-capture device, such as a camera, captures two-dimensional pictures. This means that images captured have no depth information; i.e. it is difficult to determine the distance from one object to another or from an object to the camera, by simply looking at a picture. A problem therefore arises when attempting to determine 3-D information from such a 2-D image.

2.2 Miscellaneous Problems Encountered When Using Image Capture Devices for Interaction:
Image capture devices require large bandwidth and excessive processing ability to handle and analyze images. Furthermore it may be necessary to put restrictions on the background and lighting to make the task of image determination possible. These constraints are expressed by Smith in [7].

There are many ways of reducing the amount of processing needed, see [2, 3, 9]. One method of dealing with lighting and simplifying image processing is to produce a silhouette see [3]. This allows features to be extracted rapidly and easily. Van den Bergh presents a method of extracting the background of a picture to aid in the identification and tracking of an object [9]. Solutions to the problems in 2.2 are beyond the scope of this article. It should however be noted that many solutions and working examples exist that resolve or nullify these problems.

3 CONVENTIONAL SOLUTIONS FOR SOLVING THE 2D TO 3D PROBLEM

The depth of an object in a scene can be determined by making use of multiple cameras in a setup that mimics human vision. Usually these methods use two cameras placed at a fixed distance from each other [8]. In this way stereo vision is achieved and 3-D image reconstruction can take place.

Single camera methods of estimating depth are also available. These methods accomplish depth detection by means of analyzing shaded regions or by searching for hints, within the image, that give away the required depth information; e.g. searching for distortions in lines. A method, implemented at Massachusetts Institute of Technology, determined depth by examining a scene under different lighting conditions [8].

4 PROBLEMS WITH CONVENTIONAL METHODS

In implementing stereo vision techniques two or more cameras are needed. For each camera additional image conversion equipment is necessary, i.e. multiple analog to digital converters are needed. Furthermore capturing the extra images will require extra memory. Cameras and video equipment do not come cheap. For each of the added cameras additional bandwidth will be required. These methods are also particularly complex and require special software. Added complexity and bandwidth result in higher processing costs. This has a detrimental effect on the use of these methods in certain real-time applications, which require rapid interaction. Stereo computer vision techniques are extremely expensive and complex. For these reasons they are used only in applications where their use is critical. Hence these methods are not in common use.

Depth detecting methods via other schemes (single camera methods) require additional processing to be done to analyze the data and scene being studied, to extract the depth hints. The extra processing adds to the already high image processing costs. This may tend to be a problem in applications that require a very high frame rate to cater for rapid interaction. These techniques do not fully capture three-dimensionality. They add certain depth and positional information to a 2-D scene. In essence these methods are "2.5-D" [8]. That is, the dimensionality lies somewhere between 2-D and 3-D.

5 A NOVEL SOLUTION TO THE PROBLEM

The 2-D to 3-D problem exists because a picture lacks depth information. If that, depth information, which is lacking can somehow be inserted into the picture then the problem will be solved. This is the idea behind this method.

A camera can accurately capture some 2-D positional information of an object. By placing a mirror inside the camera's view area, it is possible to capture additional 2-D information about the object's position. By combining this information it is possible to solve the problem; i.e. to find an objects depth or 3-D information.

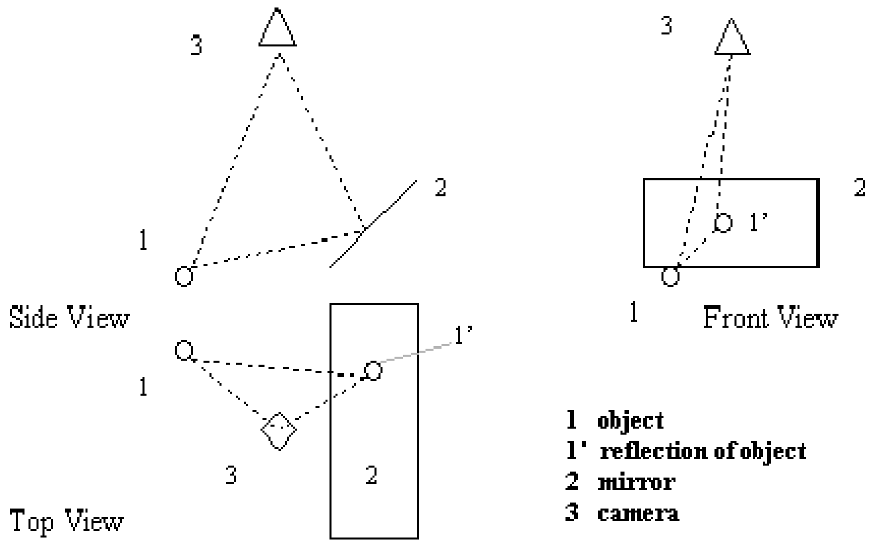

Figure 1. Side view, front view and top view of the camera/mirror setup.

Figure 1 shows a top, front and side view of a camera a mirror and an object. The object, (1) can be seen by the camera,(3) directly. It can also be seen indirectly by the camera in the mirror,(2); i.e. the camera can see the reflection of the object, (1') in the mirror.

The frames (individual pictures) taken by the camera therefore contain two different views of the object. This is illustrated in figure 2. Once an object and its reflection have been located in the image the calculation of the relative 3-D co-ordinates can take place.

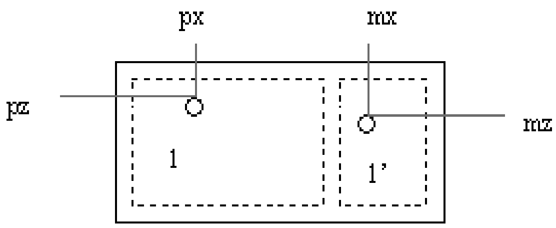

Figure 2. An image obtained by the camera.

The mirror acts as a picture taken by a second camera. Part of the camera's view area needs to be sacrificed to view the mirror. Therefore this method can be considered to be a form of multiplexing since a tradeoff exists between the size of the view area of the camera and the ability of a camera to capture 3-D using a mirror.

The region of interaction will determine the size, orientation and placement of the mirror relative to the camera. The area of interaction to be captured in 3-D must be in the intersection of the camera's field of view and the mirror's field of view. This is illustrated in Figure 3. Striped region X, is that area in which an object can be moved and have its 3-D co-ordinates calculated. This is the region in which an object and its reflection can be seen simultaneously by the camera.

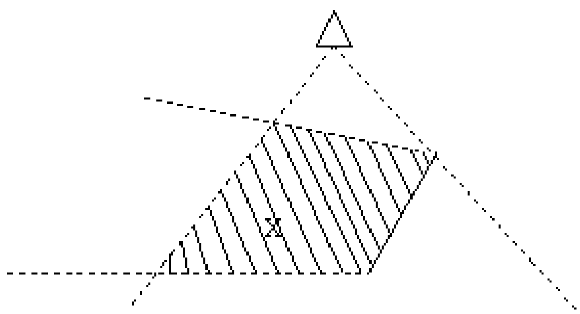

Figure 3. Side view of interaction area.

In addition to calculating the positional information of an object the physical dimensions of an object (height, width and breadth) may also be determined by making use of this method.

It should be noted that mirrors have been used successfully in applications for accurately positioning objects. A very specific example of this is Michelson's interferometer, which has been used to establish high precision length standards [10]. It can thus be expected that this method can be used for precision measurements of objects and determining their exact positions in 3-D.

6 CALCULATING THE 3D CO-ORDINATES

6.1 Calibration

Before calculations can begin the system needs to be calibrated. This includes firstly taking into account the region of interaction to be monitored. A mirror needs to be placed and oriented inside the camera's view area, so that the camera can view both the mirror and the interaction region. The mirror must be correctly oriented, so that the camera can view the reflection of an object in the interaction region. The orientations of both the camera and the mirror need to be taken into account as well as the position of the mirror relative to the camera. The resolution of the camera also needs to be considered. Hereafter accurate calculation of the 3-D position of an object can be done.

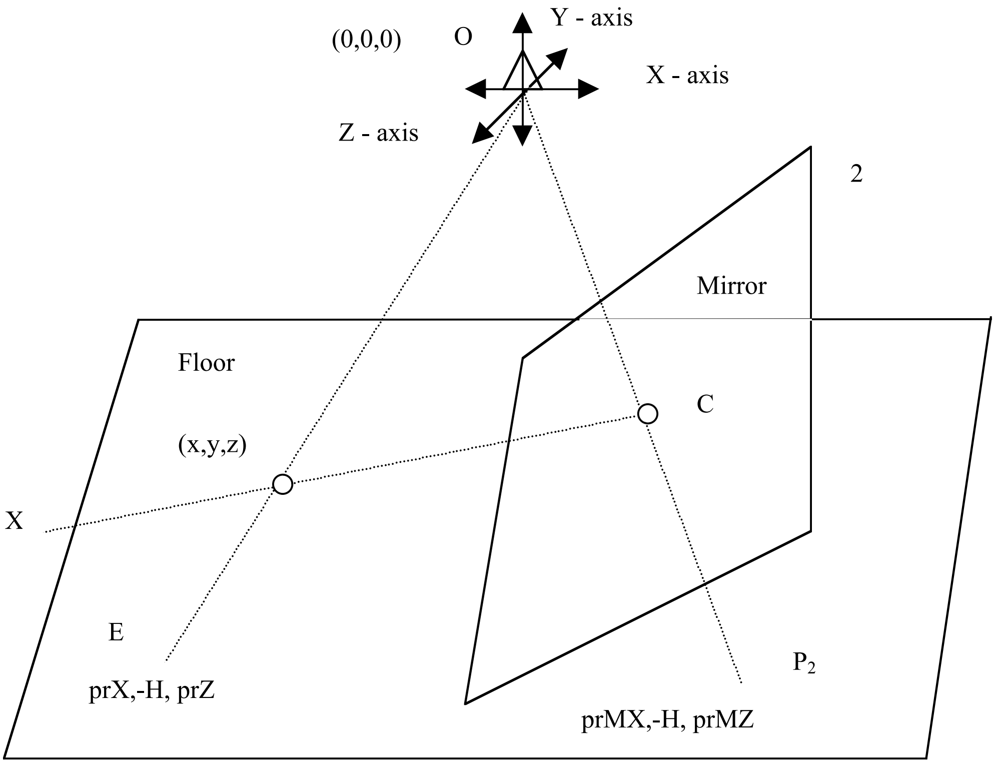

Figure 4. Setup of a camera and mirror for calculations.

The step that precedes 3-D determination is the identification and location of the object and its reflection in the picture. Interaction applications require this step to be done in real-time. Fortunately there are applications and methods of identifying and locating objects at frame rate [2, 5, 9].

After the above image analysis, two co-ordinate pairs are obtained. These are the two points (px, pz) and (mx, my) as seen in figure 2, where the object and its reflection were located in the picture.

6.2.1 Method of Calculation:

i. Since the picture taken by a camera has a perspective view, one can calculate the projection of these co-ordinates onto some back plane, in this case the floor. Thus (px, pz) becomes (prX, -H, prZ) and (mx, mz) becomes (prMX, -H, prMZ); see figure 4.

ii. Since the mirror is a flat surface it can simply be represented mathematically by the equation of a plane.

iii. The point C (the relative 3D co-ordinate of where the object is seen in the mirror) can be calculated by finding the point of intersection of the line OP_2 and the plane [4].

iv. A second point now needs to be found on the line CX. This may be done, by taking into account that the angle of incidence equals the angle of reflection. From this the line equation of CX may be determined.

v. The point of intersection of the lines CX and OE is the relative 3D position of the object. The final data obtained from these calculations yields the x, y, and z co-ordinates of an object relative to the set of axes selected.

6.2.2 A simpler but less accurate approach may be used if the camera is positioned above the interaction area and the mirror placed in front of the interaction area, as in figures above. (px, pz) may be used as an approximation of the X and Y co-ordinates of the object. The value mx can be used to approximate the depth. Thus a rough estimate of the 3-D co-ordinates of an object are (px, mx, pz).

The velocity and acceleration of an object can be determined by observing the displacement of an object over time.

7 ADVANTAGES OF USING THE SINGLE CAMERA AND MIRRORS APPROACH

The proposed technique has the following advantages over conventional methods:

1. In terms of money, it is less expensive than stereo vision techniques. A single set of recording equipment is necessary and one or more mirrors are needed. Generally mirrors are cheaper than cameras.
2. It is fast and simple. After vision processing has been done on an image, some simple calculation is necessary to determine the relative 3-D co-ordinate data.
3. The method can be as accurate as required. Furthermore, by using enough mirrors the full three-dimensionality of a scene or object can be obtained.
4. A single camera uses a fraction of the bandwidth that conventional multiple camera 3-D vision systems use.

The equipment needed to implement this method is common, e.g. a PC, a web-cam and the bathroom mirror (kids ask your parents first). This implies that methods of 3-D interaction and vision systems can be made more available to the public.

It should be noted that many of the advantages of computer vision techniques and those of certain interaction devices can be inherited by this method. Following is a list of advantages this method can offer to interaction tasks and environments:

1. The interaction device being used does not restrict movement.
2. A more natural type of interaction can be achieved; e.g. selecting can be done simply by the user pointing at the object of choice, “That one!”. This can be especially advantageous in virtual reality environments.
3. A camera out of reach is less prone to accidents and wear and tear than other interaction devices.
4. The naturalness of interaction that can be achieved using this method should contribute to the ease of use of an application. Since the method of interaction will be intuitive to a specific application, learning to interact and use an application should take place quickly.

In computer vision techniques the accuracy and ability to obtain extra information can be used to assist and improve object identification. In the image understanding phase of the computer vision process determining object characteristics can be aided by the added views obtained from the mirrors.

8 LIMITATIONS AND DISADVANTAGES OF THE PROPOSED METHOD

The following disadvantages and limitations are foreseen:

1. The interaction area is limited by the view area of the camera and the size of the mirror. Applications where a large interaction area (perhaps volume is a better term than area since this is in 3D) is needed will have to find creative ways of extending the method.
2. To set up a scene for interaction, care must be taken to set up the mirrors and camera so that the required interaction can clearly be seen; e.g. if a person moves his one arm over the other the interaction should not be lost, especially if it affects the application.
3. The angles and positions of the mirror and camera need to be taken into account. These must be incorporated into the calculations whether they are done via software or hardware.
4. Interaction can be captured and monitored only as fast as the camera’s frame rate. The camera’s speed needs to be high to capture fast real-time interaction. High-speed cameras are expensive.
5. Mirrors are fragile.

9 FUTURE USES

Some of the fields that this method can be applied to are:

- AI: computer vision, machine vision and robot vision.
- Graphics: virtual reality and interaction. A single camera has already been used to implement a 2-D interaction device [6]. By means of this method 3-D interaction devices may be created using a single camera and a mirror.

The following are fields that may use VR to aid and improve current techniques via simulation, using this alternative 3-D vision system.

- Education, teaching and co-ordination. A student may enter and interact in certain environments that they can normally only be taught about.
  Example: A student may climb aboard a Spanish galleon and flare the sails in a history simulator.
- Music, teaching and co-ordination.
  Example: The virtual drums project will allow a person to play a virtual drum kit.
- Medicine.
  Example: A doctor may practice examining a patient.
  3D-information from X-rays or sonar may be determined via this method.
- Simulators.
  Example: Piloting a plane.
- Mining/Geology
  Example: Diamond cutting.

There are a multitude of uses and applications for this 3-D vision system which are beyond the scope of this article.

10 AN APPLICATION: VIRUTUAL DRUMS

Virtual reality is an implementation of a computer-generated environment. The interaction in and with the environment should be realistic and natural to the real world equivalent. Virtual Drums attempts to create a 3D interactive environment for playing drums. Playing a drum kit requires a great deal of interaction. All four of the drummer’s limbs can be moving at any instance. Furthermore the movement of the drumsticks can be quite fast at times. To create this virtual drum kit in a seamless way a method of monitoring the above-mentioned movements is needed.

By making use of the new method it will be possible to use a single camera and mirror to achieve the interaction aspect of the project. The user (drummer) will be able to play the virtual drum kit in an intuitive and unrestricted way; i.e. as if playing on a real drum kit. Ultimately interaction is accomplished in a seamless manner.

The virtual drums project should be completed by the end of this year, 2000.

11 FUTURE RESEARCH

The use of AI in this technique should be further researched. Especially in the areas of locating an object seen in a picture.

Flat mirrors and their use to this application have been discussed and considered in this report. The calculations for reflected light at plane surfaces are elementary. Spherical mirrors, although the calculations require a little more thought, may carry some added advantages. For instance a convex mirror has the property of focusing rays onto a single point. This could be useful in focusing all the rays of a large image into a camera. Lenses may also be used to accomplish this. Future research should be done in the use of lenses and different reflective surfaces.

The use of reflective surfaces, not just mirrors should be considered. It may be possible to implement the method using one way mirrors and semi reflective surfaces.

Calculations may be affected or changed to cater for objects that obscure others. Additional mirrors may be used to overcome this problem.

The view volume will be static as long as the mirrors and camera are static. Moving cameras and mirrors can be considered to broaden the application horizon to include high detailed interactions or those that cover large variable environments. By allowing paning, zooming and other fancy camera tricks, the above extensions could be implemented.

The future success of this method will dictate whether such research will be done and which fields will be influenced. Future research may open new avenues for the use of this method.

12 CONCLUSION

A new solution has been discovered for determining depth and 3-D information by means of a single image capture device and one or more reflective surfaces. This solution presents a simple and cost effective alternative to common methods for overcoming the 2-D to 3-D problem. These advantages should make 3-D interaction and vision techniques more available to the public.

The processing required to make the method effective can be done in real-time. This implies that this 3-D vision system may be extended for use in the field of interaction. The method has the added advantage that interaction can be done in a non-intrusive way. Since the method is new much research remains to be done. This research will play a significant role in the future success and use of this method.

13 REFERENCES

1. "HOLY BIBLE", The Scofield Reference Bible, pp. 1245, Oxford University Press, Inc. United States of America, 1967.
2. Andrew Blake and Michael Isard, "3D position, attitude and shape input using video tracking of hands and lips", pp. 1, Computer Graphics (Proc. Sigraph 94), ACM Press, New York, 1994.
3. James W. Davis and Aaron F. Bobick, "A Robust Human-Silhouette Extraction Technique for Interactive Virtual Environments", pp. 12-25, Nadia Magnenat-Thalmann, Daniel Thalmann (Eds.): CAPTECH'98, LNAI 1537, 1998.
4. Robert Ellis, Denny Gulick, "Calculus with Analytic Geometry", Fifth Edition, pp. 721-724, Saunders College Publishing, United States of America, 1978.
5. William T. Freeman, David B. Anderson, Paul A. Beardsley, Chris N. Dodge, Michal Roth, Craig D. Weissman, and William S. Yerazunis, Hiroshi Kage, Kauo Kyuma, Yasunari Miyake, Ken-ichi Tanaka, "Computer Vision for Interactive Computer Graphics", pp. 42, IEEE Computer Graphics I/O Devices, May/June 1998.
6. Foley, Van Dam, Feiner, Hughes, Phillips, "Introduction to Computer Graphics", pp. 297-319, Addison-Wesley, United States of America, 1997.
7. Joshua Smith, Tom White, Christopher Dodge, Joseph Paradiso, Neil Gershenfeld, "Electric Field Sensing For Graphical Interfaces", pp. 54, IEEE Computer Graphics I/O Devices, May/June 1998.
8. Efraim Turbam, "Expert Systems and Applied Artificial Intelligence", pp. 337-365, Prentice Hall, New Jersey, 1992.
9. Frans van den Bergh, "A Device-free Locator using Computer Vision Techniques", pp. 60-67 and 130. University of Pretoria, Pretoria, 1999.
10. Hugh D. Young, "University Physics", Eighth Edition, pp. 949, 976-1006, 1036-1037 Addison-Wesley, United States of America, 1992.