Topic 10: Stereo Imaging

10.1 Introduction

The lecture covers the basic background of stereo imaging and how it can be used to extract depth information from images. Image detected by two-dimensional camera contains no *depth* information. However in many system we need depth information, for example in automated map making, robotic vision and target tracking. There are a range of schemes to extract depth information, these being:

1. **Active Measurement:** Range of *pulse-echo* techniques to measure distance to a point, for example radar, ultrasound, laser pulse or laser line scan. The most common of these schemes is the laser line scan where a three dimenional, solid, object in rotated while been scanned by a laser beam. This obtains exetrnal shape of the object, and thus it three-dimensional shape. These are not really imaging in the normal sense.

2. **Stereo Imaging:** Use two, or more spatially seperated cameras to form images from different directions. The depth information is then extracted from the *differences*.

3. **Holography:** Active optical system that records full three-dimensional object information. Very difficult to extract the information, and its not really useful as an depth analysis scheme.

Of these three only *Stereo Imaging* uses conventional cameras and image processing to extract the information, and this is the only technique we will consider in detail.

10.2 Basic Stereo Imaging Scheme

In stereo imaging we use *two* spatially seperated cameras to image a three-dimensional object as shown in figure 1, where in the simplest case the optical axis of the two cameras are parallel being separated by a distance S.

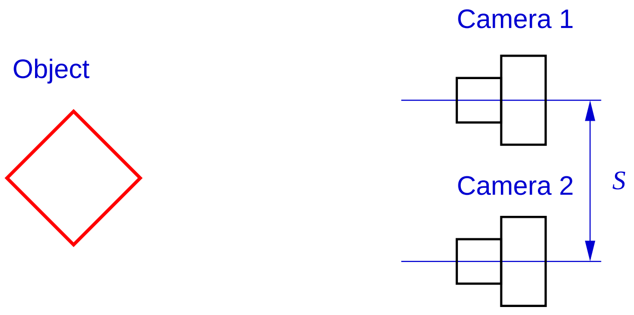

Figure 1: Basic layout for parallel stereo imaging

From the two camera we get a *different view* of the same object with the typically for a cube shown in figure 2. So we get shifts in the *vertical lines* only, from which we can extract depth information. This system is equivalent to the human visual system where we have two eyes separated by between $60 \rightarrow 70\,\mathrm{mm}$ so when we view a three dimensional object we see two slightly different views.

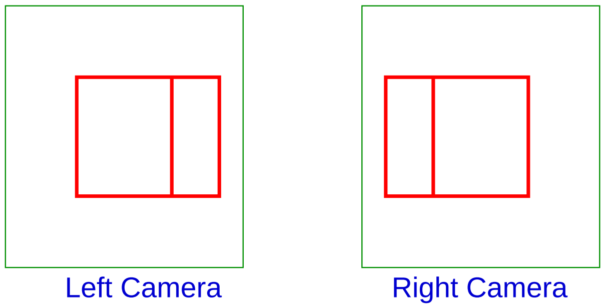

Figure 2: Two stereo images of a cube.

Consider the system in more deatail as shown in figure 3 where we assume the optical axis of the two camera is parallel and separated by distance S. Now consider the image of a point P which is located a distance X_0 from the axis of one camera and a distance X_1 from the optical axis of the other.

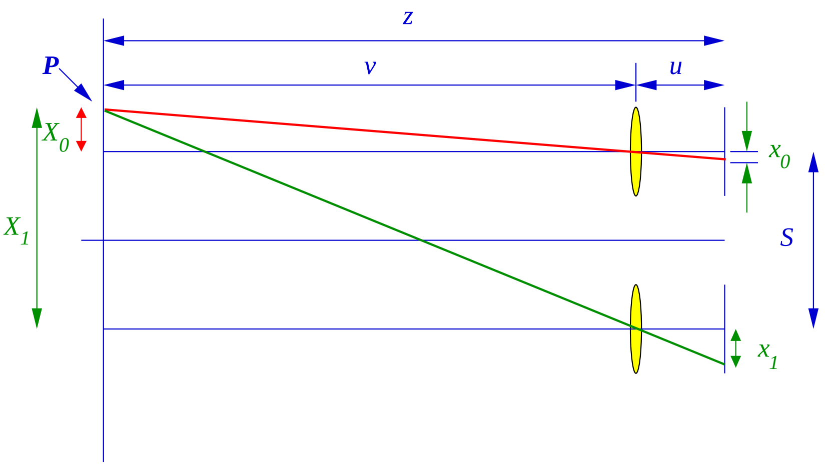

Figure 3: Details of parallel geometry stereo system.

If the poistions of the image of P in the two cameras is x_0 and x_1 respectively, then from the above diagram we have that,

$$\frac{x_0}{u} = \frac{X_0}{v} \quad \text{and} \quad \frac{x_1}{u} = \frac{X_1}{v}$$

where u is the distance from the lens to the detector plane, which we assume is the same if both cameras. We therefore have that the locations of the two images of P are at

$$x_0 = \frac{uX_0}{v} \quad \text{and} \quad x_1 = \frac{uX_1}{v}$$

Now from the geometry we have that, we know that the optical axis of the two cameras is seperated by S, so if their two axis are parallel, then

$$X_1 = X_0 + S$$

so that by substitution we have that

$$x_1 = x_0 + \frac{uS}{v}$$

where v is the horizontal distance from P to the two camera lenses. We can now solve from v to give,

$$v = \frac{uS}{(x_1 - x_0)} = \frac{uS}{\Delta x}$$

where Δx is the difference between the image of P in the two images. Now for a lens of focal length f we have the position of the object and image planes given by the Gaussian lens formula of.

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$$

However in most practical systems we have a *distance* object so that for a *short* focal length lens, then $v \gg u$. Therefore we can take the approximation that $u \approx f$, being that the lens to detector distance is simply given by the focal length of the lens. If we apply this approximation then we have that

$$v \approx \frac{fS}{\Delta x}$$

and also noting that z the distance from the image plane to object plane is given by $z = u + v$, then we get a final expression for the distance from the image plane to the point P given by

$$z = f\left(1 + \frac{S}{\Delta x}\right)$$

So given the focal length of the lenses, which we can assume we know, the separation, again known, then the depth z of point P can be measured from Δx which is the displacement imaged point between the two images. To provided we can locate the image of the same point in the two images we can simply calculate it *depth* from the system geometry.

10.2.1 Error in Depth Measure

Before simply using this system we do have to consider the errors introduced and how they effect the process. In any imaging system we will only be able to locate the image of P to a particular accuracy, typically given by the spatial sampling of the image. So if we take an error of δx in the measurement of Δx, therefore we measure,

$$\Delta x = \Delta x_0 \pm \delta x$$

so if we define as the ideal distance from the object oint P to the lens given by

$$v_0 = \frac{uS}{\Delta x_0}$$

then, due to the error in Δx we will have an an error in the depth measure of δv, so the measurement of depth we will obtains is $v = v_0 \pm \delta v$. We can then find δv is terms of δx by Taylor expansion, to be

$$\delta v = \frac{uS}{\Delta x_0^2} \delta x$$

so substituting for Δx_0, we get that

$$\delta v = \frac{v_0^2}{uS} \delta x$$

which shows that for a fixed S and δx then the error in the depth measurement rises as the *square* of distance from the camera. So to get good depth resolution you need a large S so large camera separation, but also for distant object we expect a poor depth resolution.

10.2.2 Example System

Consider a typical, practical example of using two normal video quality CCD cemera for stereo imaging. For a CCD camera the error δx is given by the size of the one CCD sensor, typically about 20μm is a reasonable camera. The orher parameters are typicall.

Sensor Size (δx)	20μm
Focal length f	25mm (typical)
Separation	100mm

If we put in the numbers for various approximate distances we find that.

$$\begin{aligned} v_0 \approx 1\text{m} \quad &\Rightarrow \quad \delta v \approx 8\text{mm} \\ v_0 \approx 10\text{m} \quad &\Rightarrow \quad \delta v \approx 800\text{mm} \end{aligned}$$

so *close* objects up to about 1 m we are obtaining a reasonable depth resolution of better that 1% error, but as the distance increases to 10 m the error rises to almost 10%, which is rather poor. Very similar to the human visual system, stereo vision useful up to about 10m, beyond that we use *size* and *perspective* to estimate distance.

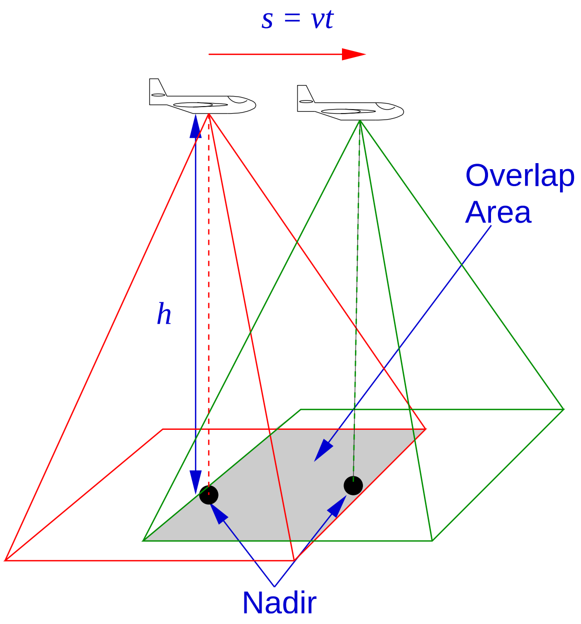

Figure 4: Stereo photography from a moving aircraft.

10.3 Aerial Photography

One of the most common applications of stereo photography is aerial photography from a survey aircraft in level flight. Camera set to point *straight down*, with the centre of the field, given by the poin that the optical axis of the camera intersects the gound, called the *nadir*. We then taketwo images separated by time t, so if the aircraft speed with respect to the gounds is v then the camera seperation is $s = vt$ as shown in figure 4. Here the two images are taken by the *same* camera and the separation can be a large as required to obtain good depth information. Know cemara separation we can calculate height variation in overlap area and is used extensively in automatic map making.

This system looks simple but is *much* more difficult than expected since the and orientation of aircraft can change significantly between exposures. This results in the direction of the optical axis of the camera changing and thus the whole geometry of the imaging system changing. This can be componsated for but use of expensive self-leveling camera mounts and accurate tracking of the aircraft speed and angles which vastly complicates the system.

10.4 Satellite Stereo

Stereo imaging from a remote sensing satellite is a very pratical scheme where the two images detected from different orbits at different times as shown in figure 5. For a satellite the orbits are very stable and since there are no atmospheric effects, most of the problems associated with aircraft stereo do not occur and good quality stereo is very practical. On some satellite systems the imaging sensor can be directed to maximise the area of overlap. Particularly good stereo image are obtained from the SPOT system.

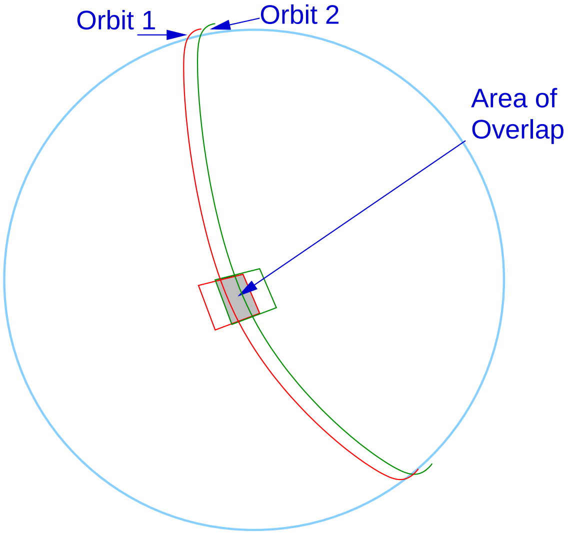

Figure 5: Stereo photography from a remote sensing staellite.

10.5 Converging Systems

In practical system the cameras as *not* parallel, but convergent to an average point in a plane at distance z_0 as shown in figure 6. We *can* solve for Δz, from the positions x_0 and x_1 in the two images. Expressions are much more difficult, but have the same form, and problems as the parallel case.

10.6 Extraction of Depth Information

Manually: Use human operator and optical viewing system. Used in most commercial map-making systems to trace contours.

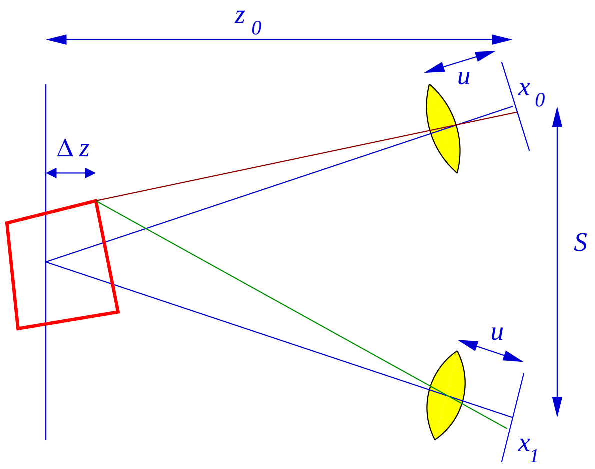

Figure 6: Geometry of a converging stereo system.

Box World: Simple geometric objects, such as found in most computer vision systems (automated inspection).

1. Enhance and detect *vertical* edges.
2. Search for corresponding edges in the two images.
3. Use difference to give depth-of-edge.

Works well for low noise images where edges are easy to find, problems are

- Missing edges in one image due to perspective change. (difficult to deal with).
- Missing edges due to noise (process images to form continuous edges).
- Confusion between edges. Problem if there are many edges, often solved by applying continuity rules between adjacent lines.
- Noise points. Try and remove before extracting depth information, but can severely upset analysis.

Some modern systems use *three* cameras in a triangle to resolve missing edge problems.

General Images: No simple, or general solutions. Range of possibilities based:

1. Identify and locate “known-objects” in each image (eg. cross-roads on a map). Works well for slow varying height information as found in aerial photography. (exact analogue of the human matching technique).
2. Region analysis of each image, and then match-up boarders of regions. Works well if you can break the image into coherent regions. (eg. in “box-world”).
3. Relaxation Labeling of edges or pixels. Hypothesis testing technique based on trying to form regions of common depth or slope within an image. AI scheme.

All of these schemes work to a greater or lesser extend depending on the type of image and how much knowledge you have about it.

10.7 Extraction of Information in a Plane

Practical problem form computer vision Esprit project we were involved with.

Problem: Make a *road-sign* recognition system for in-vehicle use.

To aid the recognition process we want to image of the approaching road-sign of known size, with as little background *clutter* as possible. (We really want an edge detected road sign extracted from the image).

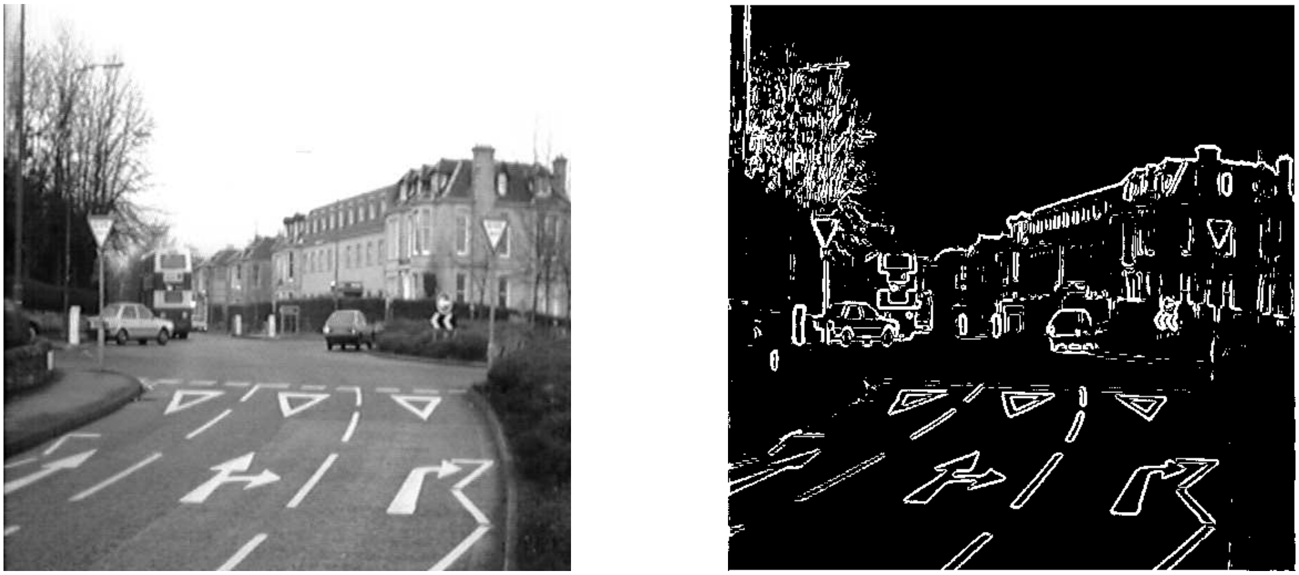

Figure 7: Road scene with two road signs and and binary threshold of edge image.

Figure 8: fig:junction

All road signs come in three standard sizes (EEC regulation), (assume single size at the moment). Want to use this to extract the sign. As car approached road we want to detect edges from one particular plane, so if the sign is in this plane, we will then detect the sign of the right size and ignore all other edge points in the image.

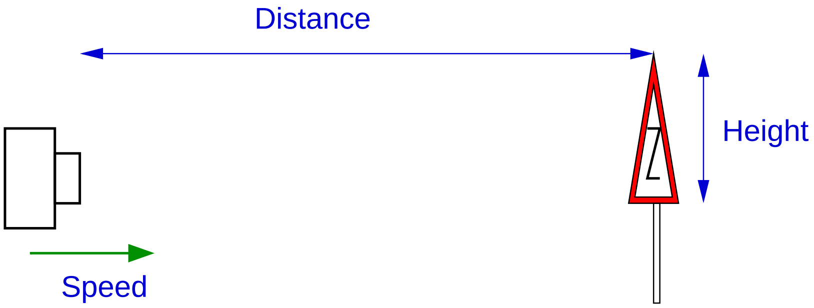

Figure 9: Geometry of imageing system from road sign recognition.

Scheme: Mount *two* cameras in the vehicle separated by a distance S.

1. Detect images from both cameras, and perform real time (simple) edge detection.

2. Threshold each to form binary edge images.
3. Reject all edge points *except* those that occur in both images with displacement Δx.
4. Use selected edge points to extract region(s) from input image.

All operations simple, (vertical edge detection, threshold, shift and logical `or`). Actually all can be done in analogue hardware.

Initial results form laboratory demonstrations looked hopeful, but it was never used on real system.