TreeScanV and Frame Mosaicing

Thesis submitted as partial fulfilment of the requirements of the

Masterate Degree in Information Engineering

By: Farshad Nourozi
February 1997

Department of Production Technology
Massey University

ABSTRACT

In 1993 the Department of Production Technology carried out a feasibility study of applying digital imaging technology in the pre-harvest inventory assessment for the forestry industry. Consequently a scanning mechanism was developed to capture a series of overlapping images along the stem of a tree. These overlapping images needed to be registered and combined to form a single long and thin high resolution image of the tree.

This report describes different methods of finding the overlaps between the consecutive images. Algorithms developed here fall into two broad categories: Spatial Domain and Frequency Domain feature matching. Comparison of different algorithms is made and advantages and disadvantages of each one are discussed. Finally a robust algorithm is developed which combines the strengths of the other algorithms.

ACKNOWLEDGMENTS

If it were not for the encouragement of my dear wife, Fariba, I would not have started the project. Moreover, if it were not for her sacrificial support during the course of this project, I could not have finished it by now. My sincere thanks to her.

I would also like to thank all the people in the Department of Production Technology who helped me in one way or another in developing this system. In particular I would like to thank my supervisors Professor Bob Hodgson and Dr. Ralph Pugmire for providing direction in this development.

I would also like to thank Swidbert who was working on a similar project for making his work available to the image processing community on the internet.

TABLE OF CONTENTS

1- INTRODUCTION

A two dimensional grey scale digital image consists of a matrix of numbers often represented by 8 bits each. Each number represents one picture element or pixel, each pixel has one of 256 possible grey levels. The resolution depth of the image is defined by the number of bits in a pixel or the grey scale quantisation. The spatial resolution of the image is defined by the number of pixels in each row and column of the image matrix.

When a digital image of an object is used for measurement of the dimensions of the object, one of the factors that limits the resolution of the measurement is the spatial resolution of the image. As a rule of thumb, in simple image measurements one pixels is required per smallest unit of measurement. However there are interpolation techniques which allow measurement to sub-pixel resolution. These techniques rely on the fact that each pixel, although does not show every detail of the underlaying object, it has mixed and integrated those fine details into a lump. Therefore some information about the underlaying details may be extracted from the lump .

In 1993 the Department of Production Technology carried out a feasibility study for Tasman Forestry to determine possible ways of capturing high resolution images of standing trees in forests for tree feature measurement. The required accuracy on the measurements from the stem of the tree was ±1cm. A typical tree has a stem diameter of about a meter and a height of about 40 meters. At that time it was decided that an image resolution of about 500×8000 was needed to easily provide the required accuracy.

Among other possible methods, using a small video tape camera to capture a series of images with some overlap from sections of the tree (TreeScanV) was proposed. To obtain a high resolution image, these individual low resolution images had to be combined to form a suitably high resolution image (See Figure 1.1).

This method was implemented in two phases. One involved the design and construction of a scanning system to capture a series of overlapping images and transfering them to a personal computer. This work was carried out in 1995 and the details of its implementation are included in appendix 1. Phase two of the implementation involved analysing the low resolution images obtained in phase

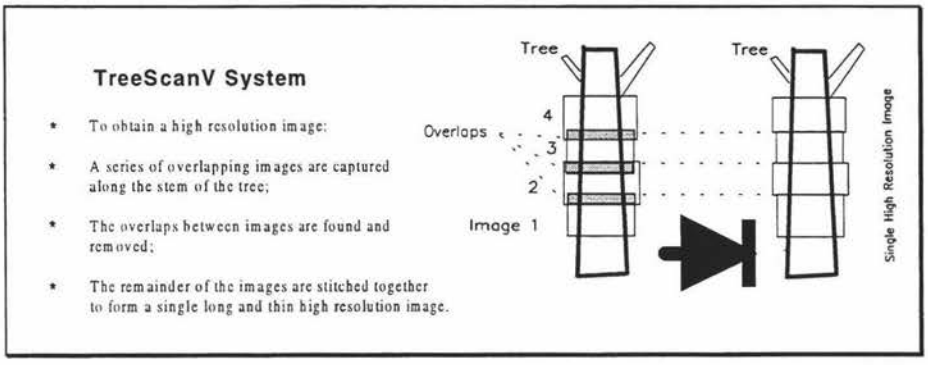

Figure 1.1- TreeScan

one and producing a single high resolution image. This report describes phase two of the implementation.

2- SYSTEM OVERVIEW

Considering the top level representation of the system as shown in figure 2.1, the objective of the system is to process the input images in order to produce the output as outlined in the following paragraphs:

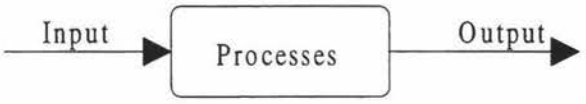

Figure 2.1- System Top Level

2.1- The Input:

- A stack of maximum 20 images. A stack is defined as a series of 2D images represented as a 3D structure. The third dimension identifies the serial location of a 2D image in the structure. Each image is an 8-bit grey scale image with resolution of 768×512. Two consecutive images have a minimum vertical overlap of 50 pixels (10%) and a maximum of 300 pixels (60%). There is a horizontal shift in both left and right direction up to a maximum of 50 pixels in either direction. A detailed account of the characteristics of the input images to the system is discussed in the next section.

2.2- The Processes:

- Finding vertical overlaps and horizontal shifts between consecutive images;
- Removing overlaps and shifting images for alignment and then stitching images together to get a single composite high resolution image;
- Indicating the validity of the results. The system has to be fail safe. If the input images are significantly distorted or there is no real match between frames the process has to be able to recognise that and prompt the user in an appropriate way.

2.3- The Output:

- A single composite high resolution image. The resolution of this final image depends on the overlaps and shifts of individual images that were used to construct the high resolution image. Smaller overlaps and shifts produce a higher resolution image output;
- A log file containing the following information:

a) Vertical overlaps and horizontal shifts information to be used by measurement software for calibration.

b) Validation of the results.

3- INPUT IMAGES

The input images to the system are primarily snap shots of sections of trees captured and digitised by a frame grabber card. They are originally recorded on a video tape and are taken in forest conditions. In the following paragraphs some important characteristics of the input images are discussed.

3.1- Light Variation

The lighting conditions in the forest vary greatly depending on the density of the forest, the time of the day and weather conditions. For a tree that is 40 meters high, the light level at the bottom of the tree where the camera is looking at the floor of the forest could be a few hundred times less than the top of the tree where the camera is essentially looking into the sky. The images at the bottom of the tree have relatively low contrast, but in a dense forest, due to the diffused nature of lighting at the floor of the forest show much detail. This is not the case at the boundaries of the forest. Here the camera may look into the light outside of the forest which is far more intense than the light level inside the forest. This results in a backlit situation where the captured image is of high contrast but much of the detail disappears. The images captured at the top of the tree are worse than those captured at the boundaries of the forest and the backlit situation is more severe at the top of the tree.

A full stack of images is taken over a time span of approximately 20 seconds. Sometimes the light level changes within this time due to moving clouds etc. However this variation is less significant than the change of light level due to the height at which different sections of tree being imaged. The image capture system is designed to deal with the change in light level at each section by allowing the camera to automatically adjust the aperture and integration time combination. This improves the quality of images. However, the camera can not compensate for the backlighting condition.

Figure 3.1 shows a typical input image to the system from the bottom and the top of a tree.

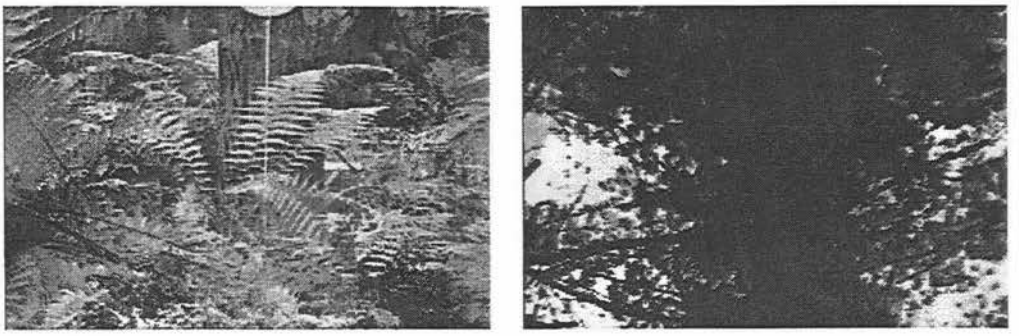

Figure 3.1- Typical input image from bottom and top of a tree

3.2- Video Timing

Another important characteristic of input images is the distortion caused by inconsistency of the video timing signals and the sampling process of the analog video from the tape. To understand the problem, a brief discussion of the timing format of the analog video signal and the sampling process performed by the frame grabber is beneficial.

The analog video signal produced by replaying the tape is a raster image in PAL format. When this signal is displayed on a monitor, a raster image is produced by scanning an electron beam across and down the face of the CRT in precisely positioned horizontal lines. The beam travels from left to right, and from top to bottom. The beam starts at the top left corner of the display and traces the first horizontal line by scanning across to the upper right corner. When it reaches the upper right corner of the screen, the beam skips downward and is repositioned at the left of the screen. The beam traces another line and this process is repeated until the beam reaches the last line at the bottom of the screen. The beam is then reset to the top of the screen and this process is repeated.

The period of time when the beam is tracing a horizontal line is called the *Active Video* portion of the signal. The period of time when it is reset from the right end of one line to the left edge of the next line is called *Horizontal Blanking Period.* The period of time when the beam is reset from the end of the last line at the bottom of the screen to the beginning of the first line at the top of the screen is called *Vertical Blanking Period.*

The action of beam is controlled by the horizontal and vertical sync pulses. These pulses are recorded on the tape along with the Active Video and define the timing and the duration of periods when the beam is repositioned.

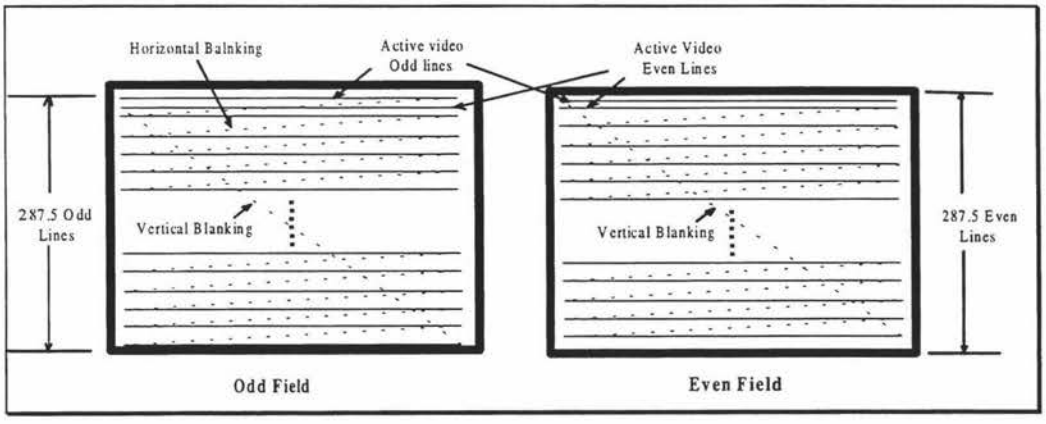

Figure 3.2.a- One Video Frame

A *frame time* refers to the time it takes to trace all the horizontal lines including all blanking intervals. The PAL signal contains 25 full *frames* each second. PAL *frames* are *interlaced* format signals. This means that each frame is divided into 2 separate *fields* of video information typically called *even* and *odd* fields (See Figure 3.2.a).

The *even field* contains all the even numbered lines in a frame and the *odd field* contains all the odd numbered lines. Interlacing involves displaying one field at a time. At the start of each frame, even lines are scanned one at a time starting with line 0 during the first field time which takes 1/50 of a second. When the last even line is scanned the beam is positioned to the beginning of the first odd line (line 1). Then during the second field time all the odd lines are scanned in between the even lines, thus the term interlacing. This interlacing cuts down on the visually perceptible flickering of an image changing every 1/25 of a second.

If the object and the camera are moving relative to each other then due to the interlaced format of the video the image in the even and odd fields will be misaligned. The amount of misalignment is directly proportional to the relative speed of the movement between the camera and the object. To avoid this misalignment problem the scanning platform stops at each section of the tree while the camera is recording frames from that section.

Each PAL *field* consist of 287.5 horizontal line periods which contain active video followed by a 25 horizontal line time *vertical blanking* interval. This interval contains, among other things a series of *vertical sync pulses*. So each PAL *frame* consist of 287.5 + 25 + 287.5 + 25 = 625 horizontal line times[2]. One horizontal line of video signal as observed at the output of a video camera is shown in Figure 3.2.

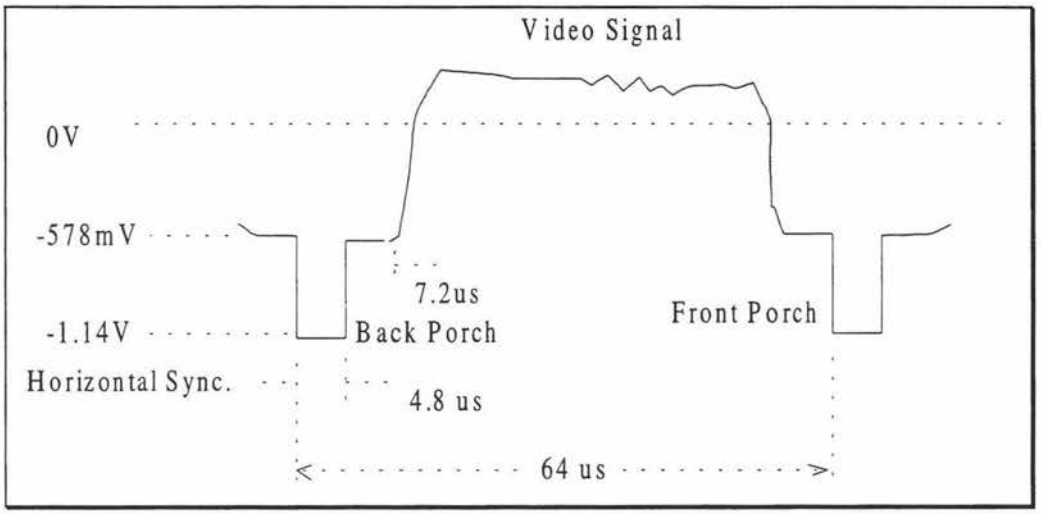

Figure 3.2- One horizontal line of video signal

The nominal horizontal line time is 64µs. However video signals may contain a lot of noise, making the determination of the sync edges unreliable. Particularly when the analog video signal is recorded on a magnetic tape and reproduced using a VCR, the noise on the video signal and variation in the horizontal line time is increased. For VCR, instantaneous line to line variations are up to ± 100ns. Line variations between the beginning and end of a *field* are up to ±5µs. When VCRs are in special modes such as fast forward or still picture, the time between horizontal sync signals may vary by up to ±20% from nominal [3].

Most modern VCRs perform head switching at the field boundaries, usually somewhere between the end of active video and the start of vertical sync. When head switching occurs, one video signal (for field n) is replaced by another video signal (for field n+1) which has an unknown phase offset from the first video signal. There maybe upto a ±1/2 line variation in vertical timing each field. As a result, longer than normal horizontal and vertical syncs may be generated. [3]

Apart from the noise in the video signals generated by the VCR, since all the timing information is record on a thin tape, any physical stretch to the tape will

increase the length of the tape and therefore the period of sync pulses. Any speed variation of the motor spinning the reading head of the VCR and speed variation in the motor driving the tape mechanism will also cause a variation of the period of the sync pulses.

The sampling process performed by the frame grabber card is synchronised to horizontal and vertical sync pulses present in the video signal. Any error in the sync pulse periods in the analog video signal can introduce errors in the digitised video.

3.3- Overlapping Lines

The number of overlapping lines between consecutive images is dependant on the lens magnification (zoom), angle of movement between consecutive frames and the distance of the camera from the tree. To obtain a uniform resolution for all the images taken by the system regardless of the distance of the tree and the camera, the magnification of the lens is adjusted according to distance so that the field of view at the bottom of the tree includes about 2 meters by 1.5 meters. This provides a uniform resolution for all images since each frame consists of 768 x 512 pixels which are distributed uniformly over the field of view providing a spatial resolution of about:

2000 mm / 768 pixels = 2.6 mm / pixel.

Since the angle of movement between frames is fixed at 3.6°,the number of overlapping lines between consecutive images is directly proportional to the distance between the tree and the camera. Figure 3.3 shows this relationship diagrammatically. Table 3.1 presents some typical distances and the associated number of overlapping lines.

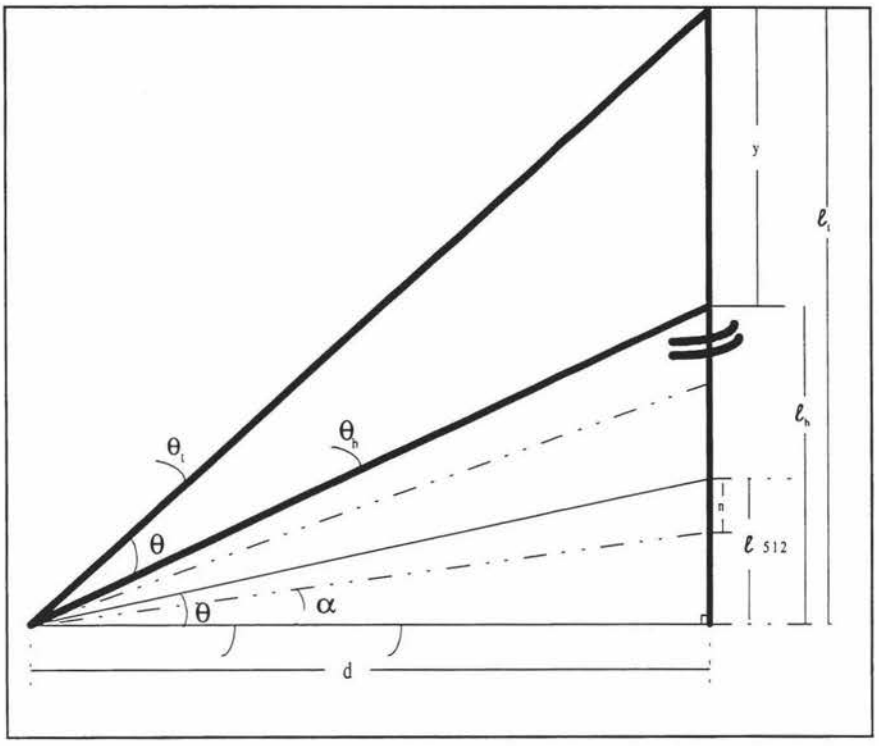

Figure 3.3- Distance and overlapping lines

As shown in figure 3.3:

$$\frac{Tan\theta}{Tan\alpha} = \frac{512}{512-n} \qquad \text{Equation 3.1}$$

For small angles of θ, α equation 3.1 approximates to:

$$\frac{\theta}{\alpha} = \frac{512}{512-n}$$

$$\Rightarrow n = \frac{\theta - \alpha}{\theta} \times 512 \qquad \text{Equation 3.2}$$

Table 3.1 is obtained using relationships in equation 3.2.

d (distance in meters)	θ (Vertical field of view in degrees)	n (No of overlapping lines in pixels)	α (angle of movement between frames)
10	**8.5°**	**295**	**3.6°**
15	**5.7°**	**188**	**3.6°**
20	**4.3°**	**83**	**3.6°**

Table 3.1- Overlapping lines variation due to varying distance

3.4- Image Resolution

Note that the field of view θ has been adjusted with the distance to keep the resolution of images constant. However, this image resolution is only constant in the first frame.

As the camera scans up the stem of the tree, the image resolution progressively decreases. This degradation of resolution is also a function of the distance between the camera and the tree. When there is a short distance between the camera and the tree, the resolution reduces at a higher rate than when the camera is further away from the tree. The perspective distortion is worse when the camera is close to the tree. Also note that the number of overlapping lines is a function of α, the angular movement of the scanning platform between consecutive frames. If the distance between the camera and the base of the tree is known, this angular movement can be adjusted based on the distance such that the number of overlapping lines remains the same for all the images regardless of the distance between the camera and the base of the tree. In the current design of the scanning platform this angle is fixed. In the following paragraphs the interdependency of horizontal and vertical resolution, distance of the camera and the base of the tree, and the height of the tree in the frame is further explored. As shown in figure 3.3:

$$\theta_t = \mathrm{Tan}^{-1}\left(\frac{\ell_t}{d}\right)$$

$$\theta_b = \theta_t - \theta$$

$$\ell_b = \mathrm{Tan}\theta_b \times d \qquad \textit{Equation} \quad 3.3$$

$$y = \ell_t - \ell_b$$

$$V_R = \frac{y}{512} \qquad \text{Vertival Resolution}$$

The horizontal field of view is adjusted so that about 1.5m of the tree is inside a frame at the bottom of the tree. Therefore the angle of view in the horizontal direction is:

$$\beta = 2 \times Tan^{-1}\left(\frac{1}{d}\right) \qquad \textit{Equation} \quad 3.4$$

The distance between the camera and the tree at the height of $\ell_b + (\ell_t - \ell_b)/2$ is:

$$d_h = \sqrt{(\ell b + (\ell t - \ell b)/2)^2 + d^2} \qquad \textit{Equation} \quad 3.5$$

Field of the view of the camera in the horizontal direction covers:

$$x = Tan\beta \times d_h \qquad \textit{Equation} \quad 3.6$$

Table 3.2 is calculated based on the relationships in Equations 3.3 through 3.6 and relates resolutions at different heights and different distances of the camera from the base of the tree:

Distance (m)	Height (m)	Theta (degrees)	Gamma (degrees)	Vertical Res(mm/pixel)	Horizontal Res(mm/pixel)
10	1	8.54	11.32	2.9	5.1
15	1	5.71	7.60	2.9	5.0
20	1	4.29	5.71	2.9	5.0
10	15	8.54	11.32	7.8	8.3
15	15	5.71	7.60	5.3	6.8
20	15	4.29	5.71	4.3	6.1
10	30	8.54	11.32	20.2	13.5
15	30	5.71	7.60	12.2	10.3
20	30	4.29	5.71	8.6	8.6

Table 3.2- Vertical and horizontal resolutions and field of views at different distances of camera from the base of the tree and at different heights up the tree

Figure 3.4 shows vertical resolution at different heights up the stem of the tree. It is assumed that the distance between the camera and the base of the tree is 15m. As can be seen from the graph, the slope of the line increases as the height increases. This relates to the increased perspective distortion observed high up the tree.

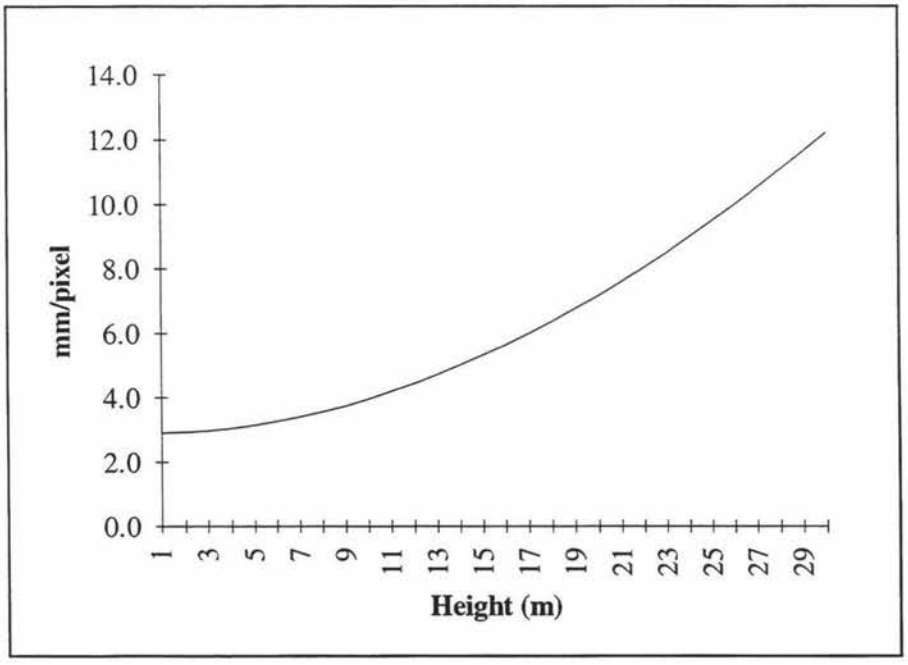

Figure 3.4- Vertical resolution at different heights when camera is positioned 15m from the base of the tree

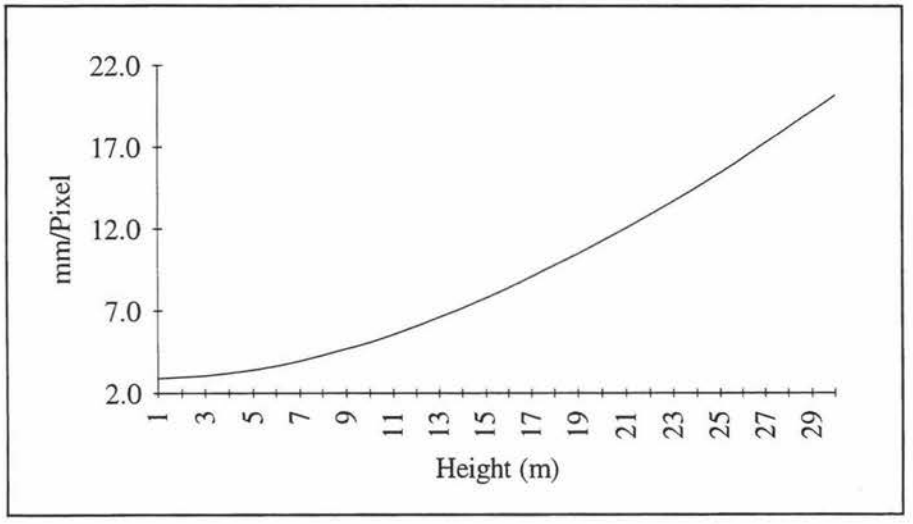

Figure 3.5- Vertical resolution at different heights when camera is positioned 10m from the base of the tree

Figure 3.5 shows vertical resolution at different heights up the stem of the tree when the camera is positioned 10m from the base of the tree. Comparing Figure 3.4 and Figure 3.5 shows a bigger slope in the resolution degradation line when the camera is closer to the base of the tree.

Figure 3.6 shows horizontal resolution at different heights up the stem of the tree where the camera is positioned 15m from the base of the tree. Comparing Figure 3.4 and Figure 3.6 shows that although the horizontal resolution of the image at the base of the tree is lower than the vertical resolution, at the top of the tree it is higher.

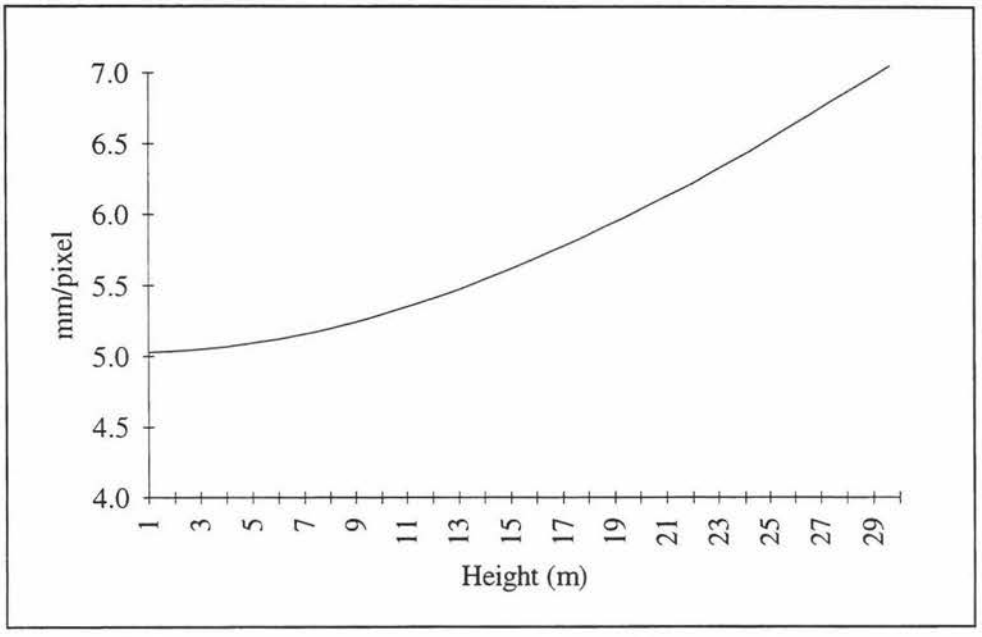

Figure 3.6- Horizontal resolution at different heights when camera is positioned 15m from the base of the tree

3.5- Perspective Distortion

This higher rate of resolution degradation in the vertical direction is the cause of the perspective distortion in the images. As the camera scans up the tree, the resolution both in the horizontal and the vertical direction decreases for two reasons:

- The distance between the camera and the section of the tree being imaged increases while the magnification is constant;
- The angle of view of the camera and the section of the tree being imaged changes.

This gives rise to the perspective distortion and also the different rates of degradation of the resolution in the horizontal and vertical directions. This effect is shown in Figure 3.7.

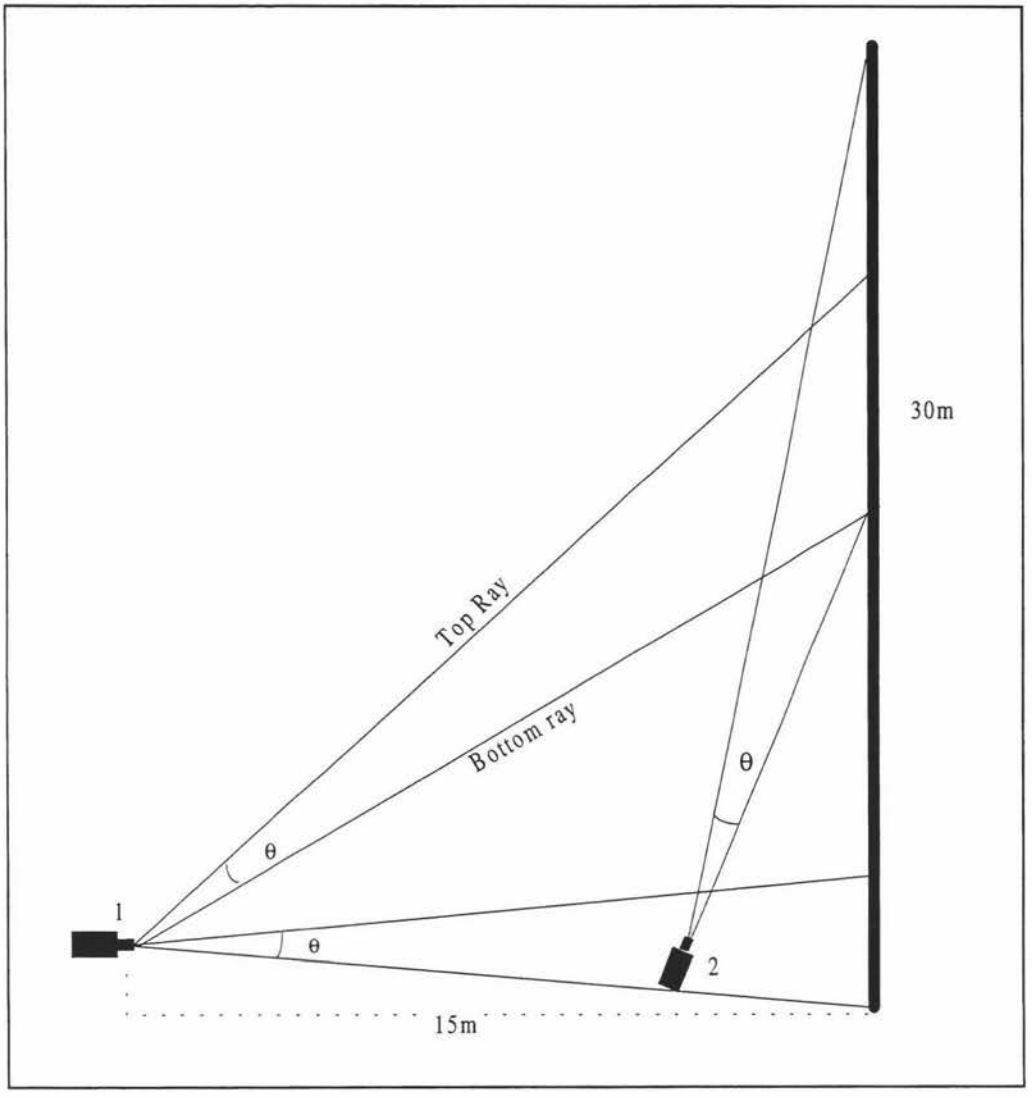

Figure 3.7- Different camera positions and angles

Figure 3.7 shows the vertical field of view of the camera at different distances from the tree as different heights of the tree are imaged. The magnification angle θ is the same in all situations. Note that in this position the top ray has to travel a longer distance to get to the camera than the bottom ray. This difference causes the perspective distortion. This effect becomes more severe as the angle of the field of view of the camera and the tree increases as shown in position 2 of the camera. Figures 3.8(a) and 3.8(b) shows images taken from positions 1 and 2 respectively.

Comparison of images at positions 1 and 2 shows that there is more perspective distortion when the camera is positioned closer to the subject and therefore the angle of the field of view is larger to provide the same physical dimensions of the subject in the view. These images show the top of a 30m high building

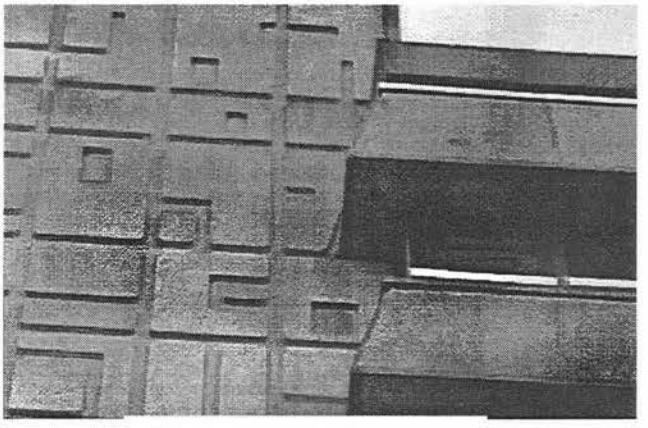

Figure 3.8(a)- Position 1

Figure 3.9 shows a frame at the base and one at the height of 30m up the stem of the tree. At the base of the tree, the field of the view of the camera includes 2m in the horizontal direction and 1.5m in the vertical direction. At 30m up the stem of the tree, the field of view of the camera includes 4.12m in the horizontal direction, a degradation in resolution of 4.12 / 2 = 2.06 times. However in the vertical direction, the field of view includes 6.25m which results in degradation in resolution of 6.25 / 1.5 = 4.2 times. This is about twice the degradation of resolution in the horizontal direction.

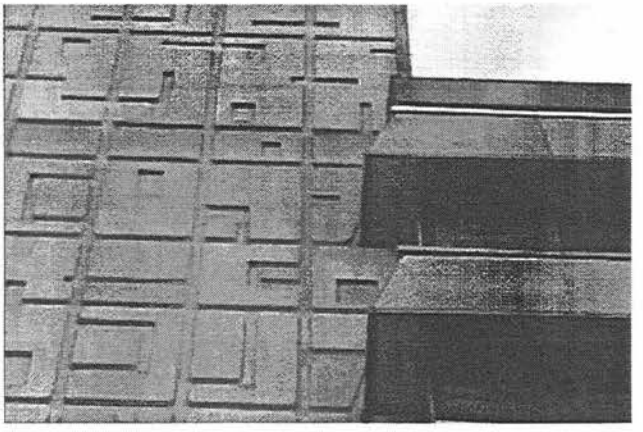

Figure 3.8(b)- Position 2

The degradation in the horizontal direction is due to the increased distance between the camera and the section of the tree being imaged since this distance is about twice the distance between the camera and the base of the tree. The degradation in resolution in the vertical direction is due to both the increased distance and the increased field of view of the camera.

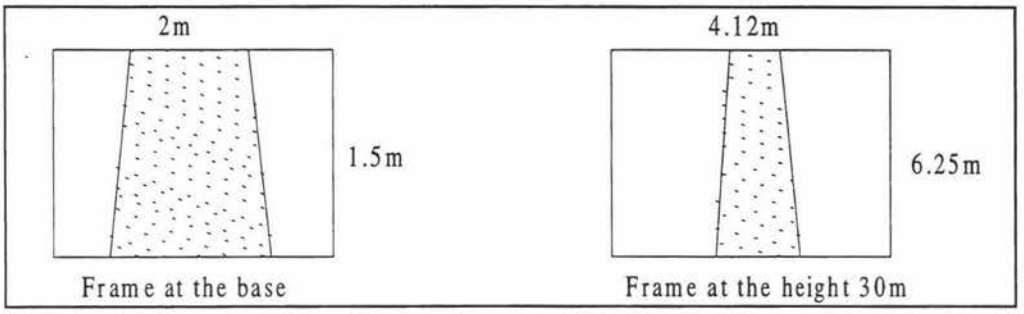

Figure 3.9- Frames at different heights, camera 15m from the base of the tree

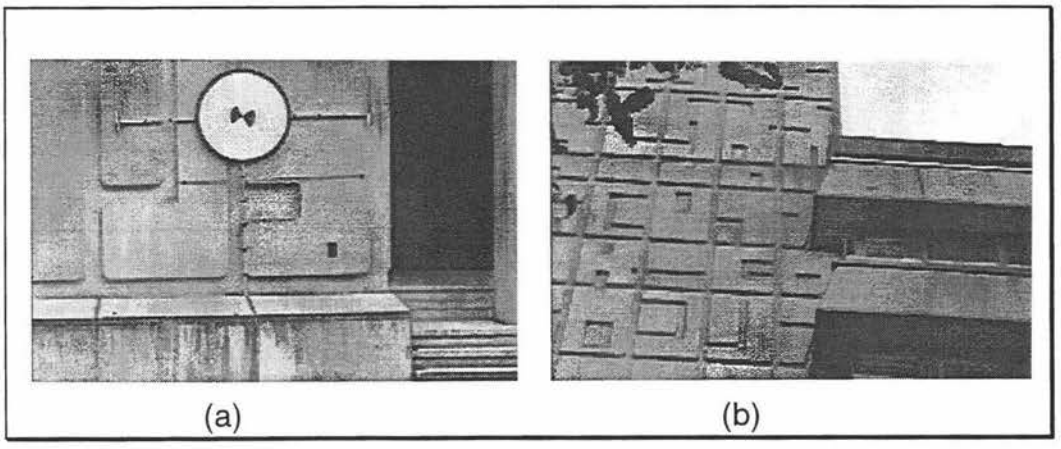

Figure 3.10- An image at the base of a building (a) and at the height of 30m (b)

Figure 3.10 shows an image of the base of a building and another at the height of 30 meters high up the same building. The horizontal distance between the camera and the building is 15.7m. Note the obvious perspective distortion in Figure 3.10.

Figure 3.11 shows two typical consecutive images input to the system.

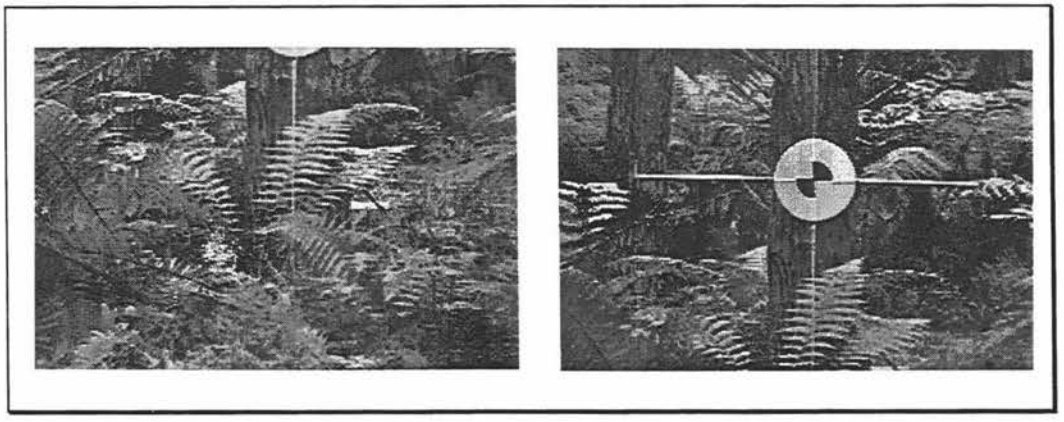

Figure 3.11- Typical consecutive input images