Height estimation for an autonomous helicopter

Peter Corke, Pavan Sikka and Jonathan Roberts
CSIRO Division of Manufacturing Technology
Pinjarra Hills, AUSTRALIA 4069.
http://www.cat.csiro.au/cmst/automation

Abstract

This paper describes our solution to the problem of estimating the height of an autonomous helicopter above ground. Height is a critical variable for a height control loop as required for the task of hovering. Two sensing approaches are investigated: ultrasonic and stereo imaging, which have complementary characteristics with respect to height.

1 Introduction

The CSIRO Automation group has recently started development toward an autonomous scale-model helicopter, see Figure 1. Our first target is to achieve autonomous hovering, and we plan to demonstrate this before August of this year. One component of the hover control task is height regulation. If the helicopter is trimmed then height is controlled by one or both of collective pitch and throttle inputs. The vertical thrust on the helicopter is

$$T_R \propto N^2 C_L$$

where N is the blade rotational rate and C_L is the collective pitch. The lift dynamics can be approximated by a double integrator system.

This paper is concerned with the problem of estimating the height of the vehicle above the ground which is an essential input to any height regulation loop. We are currently investigating two approaches: ultrasonic for very close range (less than 2m) and stereo imaging for greater heights.

Our previous work has concentrated on dense area-based stereo vision[Banks, 1999], but for this work we have focussed on feature-based techniques. These provide the raw information for height estimation as well as attitude rate estimation. The disparity between corresponding image features in the left and right cameras can be used to provide sparse range information which is sufficient for height control purposes. In addition the features can be tracked in time[Roberts, 1994][cite Jon thesis] to provide information about the vehicle's change in attitude.

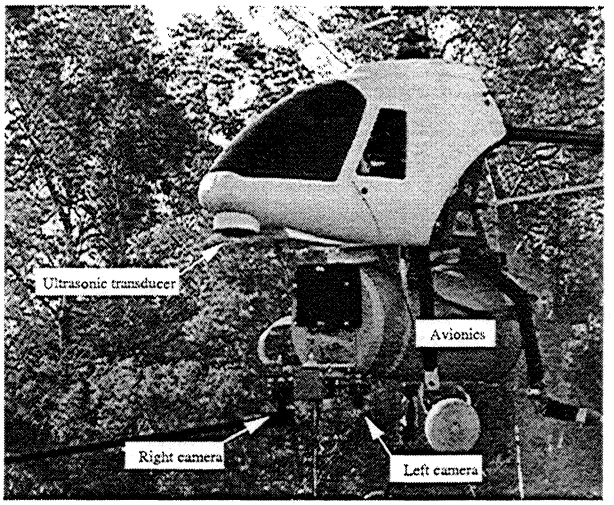

Figure 1: The CMST scale-model helicopter in flight.

2 Stereo-based height estimation

Stereo imaging has many desirable characteristics for this application. It can achieve a high sample rate of 60Hz (NTSC field rate), is totally passive, and gives height above the surface rather than height with respect to some arbitrary datum (as does GPS). Its disadvantages are computational complexity, and failure at night[1]. In our experiments the helicopter is flying over a grassy area that provides sufficient irregular texture for stereo matching. Our goal is to achieve height estimation without artificial landmarks or visual targets.

2.1 The camera system

The ST1 stereo head, shown in Figure 2, comprises two 1/4 inch CMOS camera modules mounted on a printed circuit board that allows various baselines. Currently we are using the maximum baseline of 160mm.

The two cameras are tightly synchronized and line multiplexed into an NTSC format composite video signal. Thus

[1] Infra-red cameras could be used but size and cost preclude their use with this vehicle.

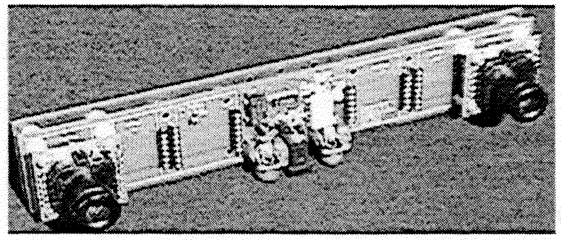

Figure 2: The small stereo camera head used on the helicopter in a downward looking configuration.

Parameter	Value
Baseline	160mm
Pixel width	14.4um
Pixel height	13.8um
Focal length	6.3mm
$\alpha_x f$	438
$\alpha_y f$	457

Table 1: Camera and stereo head parameters.

each video field contains half-vertical resolution images, 320 pixels wide and 120 pixels high, from each camera.

The lenses are 12mm diameter screw fit type with a nominal focal length of 6.3mm. The lenses are made of plastic and have fairly poor optical quality — barrel distortion and poor modulation transfer function are clearly evident in the images we obtain.

The essential camera parameters are given in Table 1. Since the two cameras are multiplexed on a line-by-line basis the effective pixel height is twice that of the CMOS sensor itself.

Mechanically the stereo head is manufactured so that scan lines are parallel and vertically aligned to reasonable precision. The mounting configuration also ensures that the optical axes are approximately parallel. Another PCB, parallel to the main one and connected via nylon spacers stiffens the structure.

2.2 Calibration

For parallel camera geometry, which this stereo head approximates, the horizontal disparity is given by

$$d = \frac{\alpha_x f b}{r} \tag{1}$$

where α_x is the pixel pitch (pixels per metre), f is the focal length of the lens, b is the baseline and r is the range of the point of interest. The stereo camera head for the helicopter must be calibrated in order to establish the relationship between disparity (that the stereo vision system estimates) and range. It is also necessary to know in advance the disparity range that must be searched so as to minimize computation.

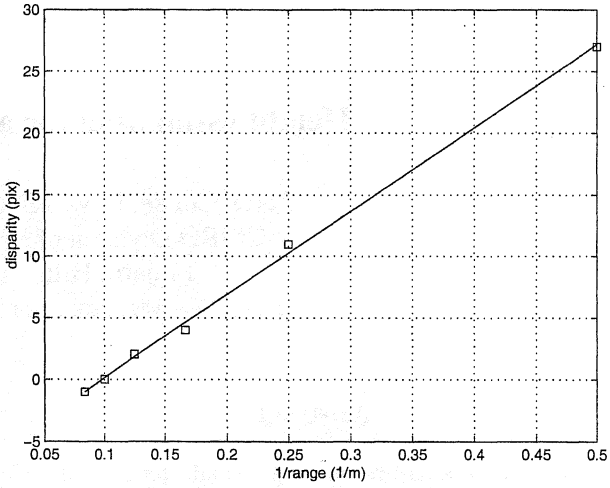

Figure 3: Disparity versus inverse range.

The derived parameters $\alpha_x f$ and $\alpha_y f$ given in Table 1 are the lumped scale parameters that appear in the projection equations. The computed values computed in the table agree well with simple calibration tests which give values of $\alpha_x f = 445$ and $\alpha_y f = 480$, which are consistant with a slightly higher focal length, approximately 6.5mm.

An experiment was conducted to verify the disparity relationship of (1). A dark rod was moved away from the camera in 2m steps upto 12m and the disparity in the images was measured. Figure 3 shows measured disparity plotted against inverse distance which is approximately linear. The empirical relationship is

$$d = \frac{67.7}{r} - 6.65 \tag{2}$$

which indicates that the cameras are slightly verged and the horopter is at 10.2m — beyond that range disparity is positive.

For our target working range of 2m to 20m the disparity would lie in the range -4 to 27 pixels, or a 32 pixel disparity search range.

Substituting data from Table 1 into (1) for ideal parallel geometry we obtain

$$d = \frac{70.1}{r}$$

which is quite similar to that obtained experimentally.

2.3 Architecture for visual processing

An overall view of the proposed system is given in Figure 4.

The vision processing software runs under the on-board real-time operating system, LynxOS, and uses a custom streaming video driver for the framegrabber. LynxOS is based on UNIX System V and also supports some BSD APIs. LynxOS is fully compliant with POSIX.1, POSIX.1b (POSIX REALTIME) and POSIX.1c (PTHREADS) standards and

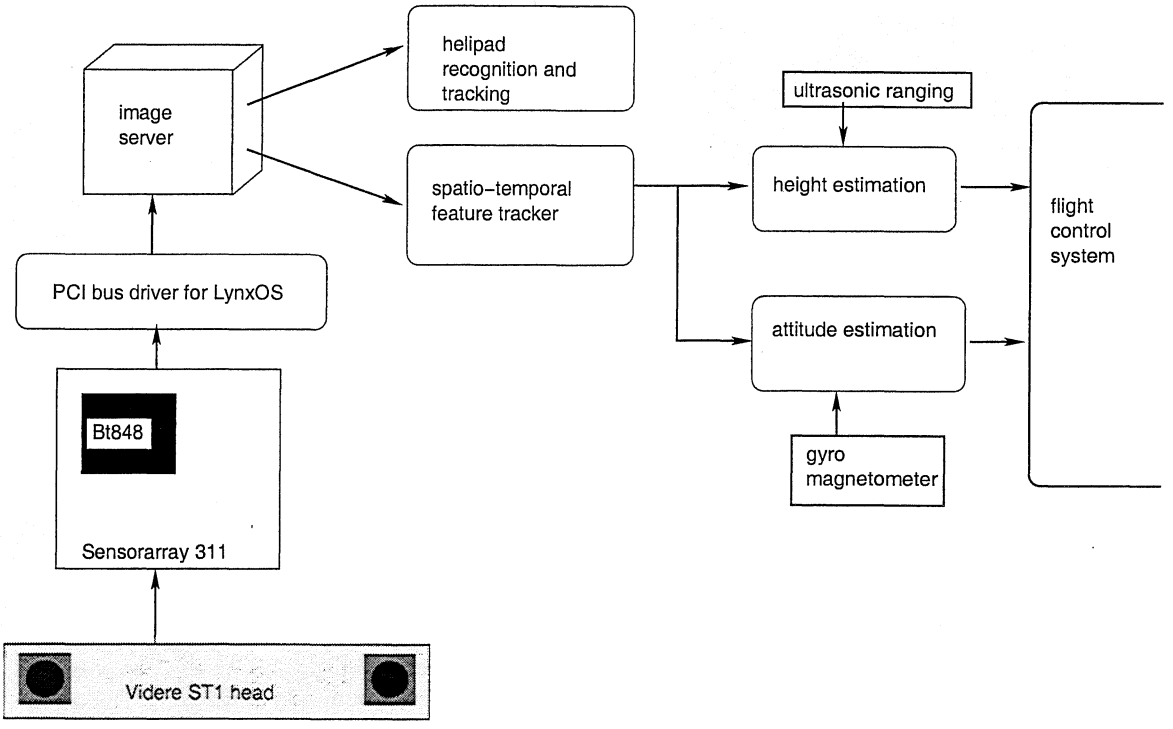

Figure 4: Overall system architecture.

that makes it very attractive for developing portable real-time applications.

Hardware

The helicopter control computer is based on the PC104/PC104+ bus, which is a stack-able architecture based on the ISA/PCI specifications. The framegrabber is a PC104+ module and unlike traditional framegrabbers that have large on-board frame buffers, this framegrabber has very little on-board memory (about 540 bytes) and writes images directly over the PC104+ (PCI) bus into system memory. This type of framegrabber is referred to as a FIFO framegrabber[Support, 2000] and has become feasible as system bus bandwidths have increased (132 MB/s for the PCI bus, for example). An advantage of this type of framegrabber is that since they capture an image directly into system memory, the CPU is freed from the task of image capture.

Software

LynxOS runs in protected mode and uses paging for memory management. This implies that user-provided buffer memory (contiguous in the user's code) is actually fragmented in the physical system memory. LynxOS provides mechanisms to translate virtual memory addresses to physical memory addresses and the framegrabber has the ability to capture an image into a series of small, fragmented buffers in physical memory. Therefore, the system provides the advantage of allowing an application to provide the memory for the framegrabber while still retaining all the advantages of memory protection.

The LynxOS device driver for the framegrabber supports three modes of operation — demand, streaming and field-streaming. In the demand mode, the frame-grabber is setup for use and frame capture is triggered by the user. This mode is useful for snapping discrete images and for processing video at slow and varying rates.

In the streaming mode, the framegrabber is setup to capture images continually into a user-provided ring of image buffers. This makes it possible to capture images at frame rate. In the field-streaming mode, each field is considered an image. This allows for image capture at two times the frame rate.

Discussion

The two main issues associated with processing images at frame rate are bus bandwidth and disk bandwidth. The PCI bus has a theoretical maximum bandwidth of 132 MB/s. Full-size PAL color video at frame rate consumes about 44 MB/s. However, there is other activity on the system and depending on the number of devices using the system bus, this can be a significant issue. The disk bandwidth becomes an issue when video has to be written to disk at frame rate. The disk subsystem requires almost the same bandwidth as the framegrabber and that could prove to be too much for the system. For the hardware we use, we have found that it is possible to capture

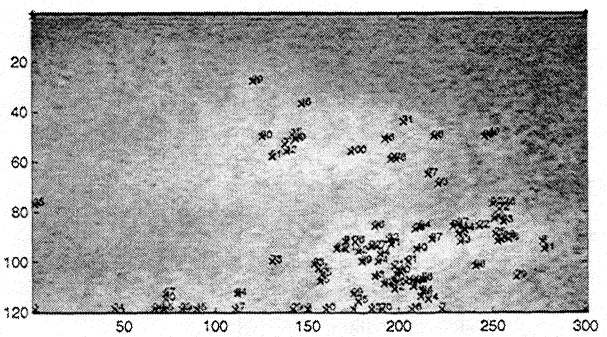

Figure 5: Left hand camera image from helicopter image sequence with corner features indicated.

monochrome 192x144 sized images to disk at frame rate.

2.4 Stereo vision

Despite our background with dense area-based stereo matching[Banks, 1999] it was unsuitable for this application. The non-parallel imaging geometry meant that the epipolar lines were not parallel, which given the large range over which we are matching the vertical disparity becomes significant. Projective rectification was investigated but is computationally too expensive for this application.

Since, at this stage, the information is required only for height control we do not need dense range data. A few points, sufficient to establish consensus, will be adequate. This led us toward feature-based methods, with the added advantage that the features can be temporally tracked to provide odometry and attitude information.

The main processing stages are as follows:

1. compute the cornerness function for every pixel in the left-hand image and select the N_{feat} strongest corners.
2. determine a region to search for the corresponding corner point in the right-hand image
3. search that region for a matching corner and compute disparity
4. repeat the steps for each corner in the left image
5. compute the median disparity and match strength information to determine the consensus disparity value.

2.5 Corner operators

Corner features are those which have high curvature in orthogonal directions. Many corner detectors have been proposed in the literature but we considered only two well known detectors: Harris[Charnley *et al.*, 1988] and SUSAN[Smith, 1992]. Roberts[Roberts, 1994] compared the temporal stability for outdoor applications and found that the Harris operator was superiour. Figure 5 shows a typical flight image and detected corner features.

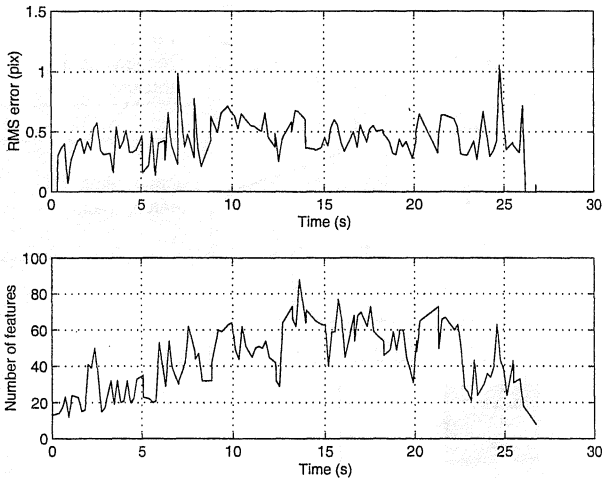

Figure 6: Matching statistics. RMS pixel error to epipolar geometry and number of corresponding feature points found.

2.6 Geometric constraints

In order to find corresponding points we apply epipolar constraints as well as feature similarity matching. The fundamental matrix of the camera head is determined using the robust method of Zhang[Zhang *et al.*, 1994]. Given a point m_l in the left-hand image we know that the corresponding point is constrained to lie along the line

$$Fm_r$$

where F is the fundmental matrix, and the points are represented in homogeneous coordinates. The line constraint is satisfied when

$$m_l(Fm_r) = 0.$$

We use the epipolar line and known disparity search range to compute rectangular bounds for the corresponding point in the right image. The box is centered a distance $\hat{d}$ along the epipolar line, where $\hat{d}$ is either the expected disparity or the maximum disparity that is allowed. The height is such that epipolar line intersects the vertical sides of the box. The width is determined by the expected uncertainty in disparity. By limiting corner feature detection and matching to only a small region we greatly reduce computation time.

2.7 Matching similar corners

The first approach was based on earlier work by Roberts[Roberts, 1994] in temporal feature tracking. For each corner identified in the left image we compute all cornerss in the epipolar constrained search window in the right image. We then search for the corner that is closest to the epipolar line and has greatest similarity to the corner. Similarity is defined in terms of a 3-element feature vector

$$v = [I I_x I_y]$$

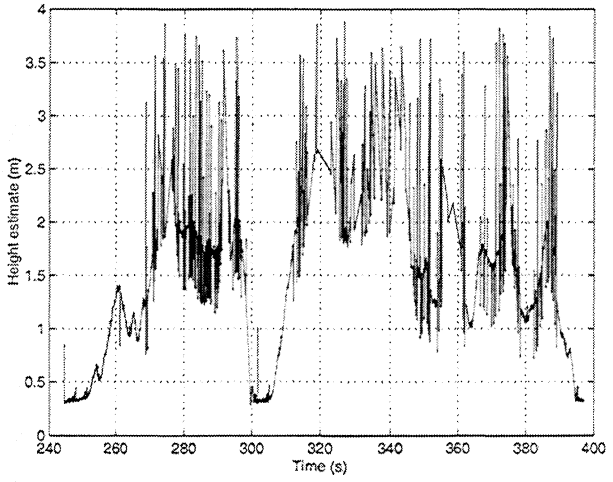

Figure 7: Raw height estimate from the ultrasonic transducer.

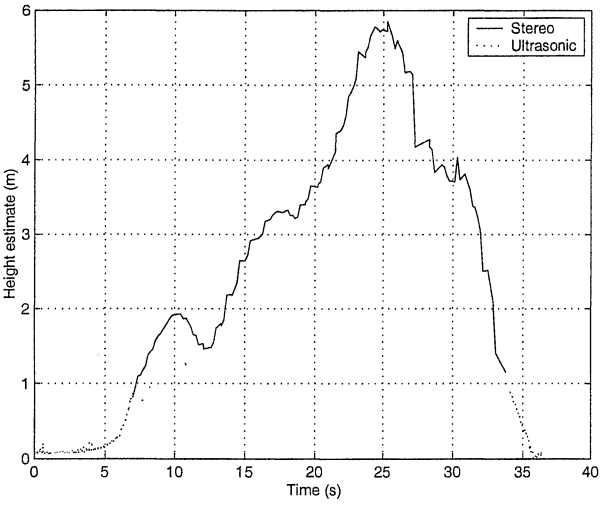

Figure 8: Combined height estimate from ultrasonic and stereo sensors with respect to the stereo camera.

which contains the intensity and the orthogonal gradients and is invarient to illumination level. The elements in the vector have been previously computed for the Harris corner detector. The vectors are compared using the GVM similarity measure

$$m = \frac{|v - w|}{\sqrt{|v|\,|w|}}$$

This measure has the virtue of being invariant to muliplicative intensity variation between the left and right image as occurs in practice.

3 Ultrasonic height estimation

When close to the ground the stereo camera system is unable to focus and the blurred images are useless for stereo matching. Ultrasonics provide an alternative sensing modality in this height regime.

The ultrasonic sensor we have chosen is a Musto 1a unit which has a maximum range of 4m. The sensor has been adjusted so that the working distance range spans the output voltage range of 0 to 5V. The sensor emits 10 pings per second, and when it receives no return it outputs the maximum voltage.

The helicopter provides an acoustically noisy environment, both in terms of mechanically coupled vibration and "dirty" air for the ultrasonic pulses to travel in. We have attempted to shield the sensor from direct pressure waves from the main rotor by positioning it beneath the nose of the vehicle.

Figure 7 shows raw data from the sensor during a short flight. The minimum height estimate is 0.3m which is the height of the sensor above ground when the helicopter is on the ground. Useful estimates are obtained for heights below 1.5m, and good results below 1m. Above 1.5m the signal is extremely noisy which we believe is due to poor acoustic signal to noise ratio. Whether this is improved by isolating the sensor from the body of the helicopter has not yet been tested.

Nevertheless their is just sufficient overlap between the operating range of the stereo and ultrasonic sensors to provide some cross-checking.

4 Combined results

Figure 8 shows the results of combining raw ultrasonic data with stereo disparity data. We have plotted all ultrasonic range values below 1.2m and added the height differential between the ultrasonic and stereo sensor systems. The stereo disparity data is converted to range using (2).

On takeoff there is a very small amount of overlap and we have good stereo data to below 1m (disparity of 75 pixels). The landing occurred in a patch of long grass which led to very blurry images in which no reliable features could be found.

5 Conclusions and further work

We have shown that stereo and ultrasonic sensors can be used for height estimation for an autonomous helicopter. They are complementary in terms of the height regimes in which they operate. Both are small and lightweight sensors which are critical characteristics for this application. Feature-based stereo is computationally tractable with the onboard computing power, and provides sufficient information for the task.

The next steps in our project are:

1. "Close the loop" on helicopter height by end August.
2. Develop a new stereo-camera head to overcome the many limitations of the current ST-1 head, in particular exposure control.

3. Implement temporal feature tracking and use feature flow for visual gyroscope and odometry.
4. Visual-servo control of helicopter with respect to a landmark.

Acknowledgements

The authors would like to thank the rest of the CMST autonomous helicopter team: Leslie Overs, Graeme Winstanley, Stuart Wolfe, Reece McCasker, and Stuart Addinell.

References

[Banks, 1999] J. Banks. *Reliability Analysis of Transform-Based Stereo Matching Techniques, and a New Matching Constraint.* PhD thesis, Queensland University of Technology, 1999.

[Charnley *et al.*, 1988] D. Charnley, C. G. Harris, M. Pike, E. Sparks, and M. Stephens. The DROID 3D Vision System - algorithms for geometric integration. Technical Report 72/88/N488U, Plessey Research Roke Manor, December 1988.

[Roberts, 1994] J. M. Roberts. *Attentive Visual Tracking and Trajectory Estimation for Dynamic Scene Segmentation.* PhD thesis, University of Southampton, UK, dec 1994.

[Smith, 1992] S. M. Smith. A New Class of Corner Finder. In *Proceedings of the British Machine Vision Conference, Leeds*, pages 139–148, 1992.

[Support, 2000] Technical Support. Fifo frame grabbers : Application brief. Technical report, Sensoray Company Inc., Oregon, USA, 2000.

[Zhang *et al.*, 1994] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Technical Report 2273, INRIA, Sophia-Antipolis, May 1994.