



ChiRoPing Deliverable D1.1.2

Evaluation of performance of Computer Vision methods for bat head shape measurement

Principal Author: UEdin
Contributors:
Dissemination: PU

Abstract: To acquire visual evidence of the bat behaviour, the project is using a novel, custom designed 500 frame/second 3D scanner. To get best performance, and to have some idea of what that performance is, we have been doing extensive performance analysis. This report summarises the results of that analysis, and how this relates to the ability to collect usable bat data. The main conclusion is that the sensor can give sub-millimetre accurate depth data over a 20 cm (typically 50 frames) flight range when set up properly, but additional work is needed to improve the spatial resolution. Several interesting observations have already been made from the data collected in Denmark.

Deliverable due: Month 12

1 Analysis of the high speed 3D video capture system

1.1 Problem Background

As described in the proposal, the main goal of the visual scanning portion of the ChiRoPing project is to acquire shape and other visual information, when bats scan for and capture prey. As the project has proceeded, the information that has the highest priority is:

- Evidence of the bat behaviour during target acquisition and capture
- Measurements of components of the bat (eg. mouth opening and orientation, wing angle versus time).
- Measurement of the bat head shape and pose over time, for input into the acoustic modelling portion of the project.

This information is to be acquired for 4 species, 3 trawling bats (*Macrophyllum macrophyllum*, *Noctilio leporinus*, *Myotis daubentonii*) and 1 gleaning bat (*Micronycteris microtis*).

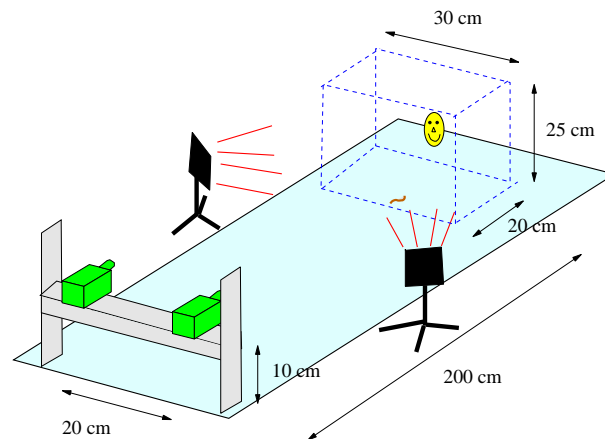
The bulk of the visual information is to be obtained by use of a novel high speed dense stereo acquisition system, custom produced for the project by Dimensional Imaging (DI). This equipment, seen here



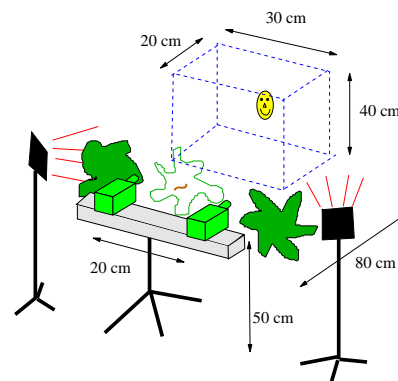
consists of 2 500 frame per second light sensitive cameras, 2 associated PCs, capture cards and framestores, plus 2 infrared light panels. The 2 cameras and capture cards are synchronised, resulting in a set of synchronised images. The extended core memory on the 2 PCs allowed approximately 1000 frames or 2 seconds of data to be captured at this data rate. After data capture, DI's proprietary 3D shape recovery algorithm produces a dense 3D image, with an (x,y,z) value at every pixel.

This is a novel (and unique in the world) piece of equipment, to be used in a restricted capture environment and time slot. We believed that it was necessary to do a performance characterisation of this device, so we could properly design the experiments before the two scheduled capture sessions. This report summarises the results of the equipment evaluation, some preliminary results from the November 2008 Denmark capture trip, and then discusses the future plans for project image analysis.

To help set some context, here are sketches of the data capture scenario for the water trawling bats:



and the leaf gleaning bat:



In the case of the water trawling bats, we see a long pool of water, with the cameras (and some of the acoustic measurement equipment) at one end. The bat generally flies lengthwise along the pool, capturing a mealworm that has been placed at the optimal sensing distance (1-2 metres) from the cameras. The IR lights provide the extra illumination needed by the high-speed cameras without disturbing the bats. The dotted box shows the zone within which good quality capture data is obtainable. This scenario was tested in the Denmark trip and was shown to be effective.

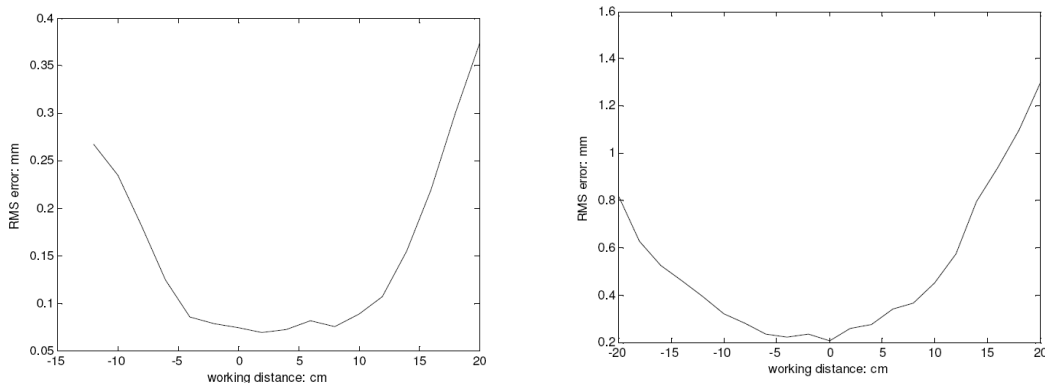
In the case of the leaf gleaning bat, we see the cameras placed between foliage, upon which an insect has been placed for capture by the bat. We will not be able to test this scenario until the Panama expedition, so we are uncertain about various questions, like how far the bat hovers from the leaves during scanning, which will affect where we place the cameras. As well as the configuration shown, the cameras could also be below the foliage looking upwards, or be placed in a vertical orientation either adjacent to the foliage or at the side.

1.2 Acquisition Questions, Methods and Results

The use of the sensor for the capture of dynamic bat behaviour required the answer to several questions:

1. What 3D shape noise levels are possible? What is the size of the capture window?

We placed a well-textured planar surface at various distances from the focus point, and measured the RMS on the residuals from the 3D plane fitting. The results when using 50mm focal length lenses and the target at 80 cm are shown on the left, with the results from 75mm lenses and the target at 200 cm on the right.



What we observe is a zone about 20 cm deep for both lenses where there is excellent 3D shape recovery (about 0.1 mm and 0.25 mm respectively).

2. Setup: what camera baseline should be used, and what aperture should be used?

We measured the RMS error levels while varying the camera baseline separation and concluded that 195-260 mm separation worked best for the 50mm lenses and 260-390 mm separation worked best for the 75mm lenses.

We measured the RMS error levels while varying the camera aperture and concluded that there was not much effect on the accuracy, but there was a smaller zone of good accuracy for the recovered stereo when using F2.8. Apertures smaller than F5.6 allowed too little light for stereo recovery.

3. What was the spatial and temporal distribution of noise?

We computed the RMS noise levels for the well-textured planar surface locally (as contrasted with globally across the whole plane in the previous experiments). We concluded that there was no observable variation of the noise across the field of view.

We correlated the noise residual vectors between consecutive images and observed high correlation on well textured surfaces. We concluded that the correlation arises because the sensor is measuring actual shape variation on the plane, and therefore one would expect to observe a similar pattern of residuals in consecutive images. When texture was weaker, which means that the stereo algorithm does not produce as good evidence, the correlation decreased significantly.

4. What are the minimum resolvable spatial features?

We did 2 sets of experiments, one where two thin threads were observed with varying separation between them. The second experiment used a shallow planar target, where the separation between the edges of the target was varied. The issue was how close two features could be placed before they were measured as a single feature, which would give information about the minimum observable spatial resolution of the features. The threads represent transitory features of several depths, and the edges represented shape feature transitions.

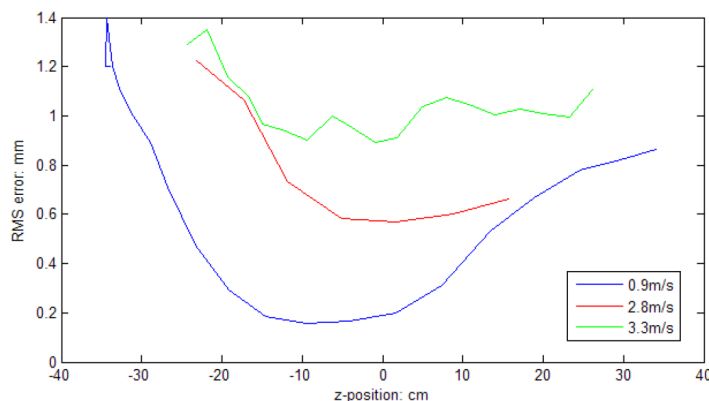
For the threads, the minimum separation varied between 10 and 27 pixels, depending on the thread size and vertical or horizontal presentation. In scene dimensions (50 mm lenses, 130 mm baseline, 80 cm target distance), this corresponded to 2-5 mm. The

minimum allowable separation of the edges of the planar target varied between 6 and 11 pixels, depending on the target width and vertical or horizontal presentation, which corresponded to 1-2 mm.

5. What range of target speeds are possible to be captured?

We dropped or swung a well-textured ball in the 3 coordinate directions relative to the camera, and measure the RMS of the residuals to a sphere fit. A sphere was used because this avoided having to solve the model registration problem when surface fitting. Different performances were expected in the 3 directions because of the different interactions between the motion blur and the 3D reconstruction algorithms. We moved the ball at different speeds to assess how the reconstruction error varied and what range of target speeds is feasible.

With vertical motion (ie. motion blur is perpendicular to the stereo baseline), the RMS error varied from 0.5 mm (3.4 m/s) to 0.75 mm (5.0 m/s). With horizontal motion (ie. motion blur is parallel to the stereo baseline), the RMS error varied from 0.4 mm (0.3 m/s) to 1.4 mm (2.8 m/s - ie. much more sensitive than the vertical motion). Finally, with motion towards the camera (ie. radial motion blur), the RMS varied somewhat complexly from 0.2 mm at 0.9 m/s to about 1.0 mm at 3.3 m/s. The figure below shows the results. For low speeds, we hypothesise that the depth of field blurring is the dominant factor, hence the 20 cm range within which low RMS is measured. At high speeds, motion blur dominates, largely independent of the 'sweet spot'.



1.3 Observations and Implications for Data Capture

Based on the characterisation experiments, we have a good understanding of the sensor and are able to capture usable data; in fact it was used in the Denmark data capture trip. Several nice captures were obtained, which will be on the project web site shortly. Altogether, we captured only 2-3 good runs where the bat attacks the prey, plus several cases where the bat launched itself from the trainer's hand.

One important issue arose during the Denmark trip, concerning the DI capture trigger. Initially, the trigger was a pre-trigger, whereby the user would start the 2 second capture process. Unfortunately, it was hard to predict when the bats would attack the prey, and the resetting of the trigger mechanism after a false alarm meant that we often missed the actual attack. This was one cause of the limited number of runs acquired in Denmark.

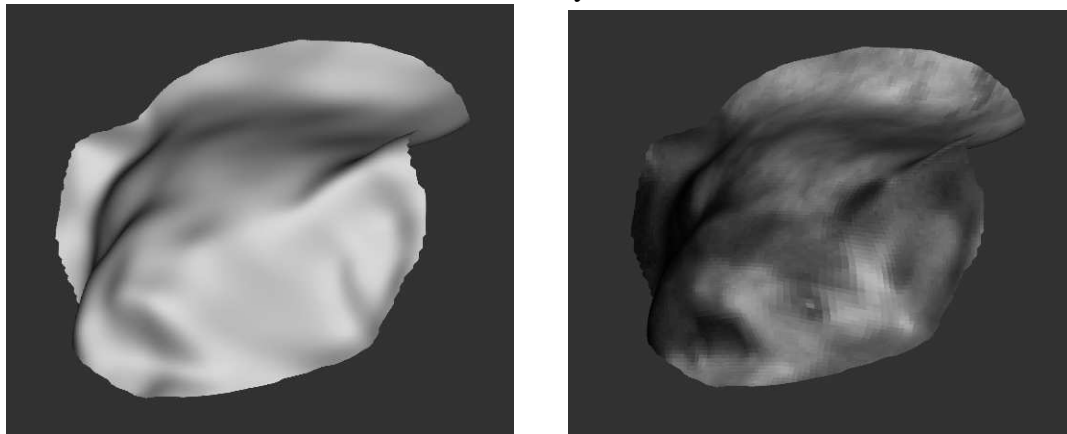
DI kindly modified their software to allow post-triggering, whereby the system would continuously capture images, overwriting previous ones, until a 'stop' signal occurred. We believe that this will allow a much greater number of captures from the Panama trip.

We were worried that the limitation of about 3 m/s in bat movement towards the camera system might limit the ability to capture 3D data. However, the Denmark capture trip showed that the *Myotis* bats were travelling at 1-2 m/s during capture, although they are capable of higher speeds. This suggests that we might cope with the speeds of the larger *Noctilio* trawling bat in Panama, if it also slows down during prey capture.

We were also worried about the narrow 20 cm capture window depth. However, the Denmark data showed that we could capture about 50 consecutive frames of good data as the bat moved through the window before it was either too blurry or out of the IR lighting range.

The Denmark trip has resulted in several interesting observations already about the *Myotis* bat, namely regarding the use of the tail membrane for lift and stabilisation when launching, there appears to be a head-dip just before prey capture, and the opening of the bat mouth seems to be reasonably constant during echolocation. The audio track has been synchronised to the video by use of a random-pulse timing light.

Stereo depth has been recovered from some of the Denmark captures, an example of which is here without and with the overlaid intensity data:



Because of the head size, and the 10-20 pixel scale of resolvable features, we can see that the 3D shape details are currently limited. Over the next year, we will have a student project that will investigate how to achieve resolvable features at about the 4-5 pixel range, which should give more 3D detail. The detail shown above was also improved by manually isolating the bat's head before the stereo calculation, which reduced the complexity arising from interference with the other parts of image. We will be looking at semi-automatic bat head extraction.

The analysis of this new type of sensor has been presented at a workshop and has been submitted for review as a journal article.

1.4 Future Plans

An important project goal is to obtain measurements of properties that are useful to the biologists, acoustic scientists and roboticists. So far, the information of main interest seems to be:

1. Mouth opening versus time

2. Head direction versus time
3. Wing and tail positions correlated with each other and with mouth opening
4. Ear tilt
5. Head shape (for use in the HRTF-based acoustic modelling).

All but the last item can be obtained by tracking various landmarks through time and then computing the 3D positions. Some semi-automated tools should make acquiring these landmarks easier, from the few thousand images we expect (only about 300 from the Denmark trip).

The head shape is harder because of the lower resolution, the blurring and shape noise. We are planning a model-based shape fitting approach to extracting each bat head from the 3D video, as summarised in the diagram below. From the micro-CT bat head scanning we have high-detail 3D mesh models available for the bats. We will extend these to allow jaw, mouth, ear and nose-leaf motion (top left 2 boxes). When working with a specific bat, we propose to acquire a lower resolution static shape capture while the bat is held (top center task). This will allow the initial model to be deformed to the specific bat's shape ("Specific ...Model"). This model will then be fitted to each video frame's data ("Mesh Model Fitting"). The plan is to then exploit the fact that the bat's shape is only changing slowly when observed at 500 fps, to improve resolution and reduce noise variance. Finally, the parameters of interest to the biologists and roboticists can be extracted from the fitted model.

