3D SHAPE ACQUISITION OF FLYING BATS IN HIGH-SPEED MOTION USING A STEREOVISION SENSOR

Keywords: High-speed video, 3D dynamic shape, stereo vision, motion, tracking.

Abstract: 3D shape acquisition of fast-moving objects is an emerging area with many potential applications. This paper presents a novel application of 3D acquisition for studying the dynamic external morphology of live bats in flight. The 3D acquisition technique is based on binocular stereo vision. Two high-speed (500 fps) calibrated machine vision cameras are employed to capture intensity images from the bats simultaneously, and 3D shape information of the bats are expected to be derived from the stereo video recording. Since the high-speed stereo vision system and the bat dynamic morphology study application are both novel, it was unknown to which extent the system could perform 3D acquisition of the bats’ shapes. We carried out experiments to evaluate the performance of the system using artificial objects in various controlled conditions, and the knowledge gained helped us deploy the system in the on-site data acquisition. Our preliminary analysis of the real data demonstrates the feasibility of gathering 3D dynamic measurements on bats’ bodies from a few selected feature points and the possibility of recovering dense 3D shapes of bat heads from the stereo video data acquired. Issues are revealed in the 3D shape recovery, most notably related to motion blur and occlusion.

1 INTRODUCTION

A relatively unexplored area of computer vision is the acquisition of 3D shapes from very fast-moving objects, mainly due to the constraints on the speed of imaging hardware. With the recent advance in video recording technology, nowadays 2D images can be acquired at the rate of thousands frames per second (fps) at affordable cost (Pendley, 2003). An increasing demand is to capture 3D shape information of objects at high speed, since such information can help answer various questions, e.g., how the body of a professional sportsman moves to perform a highly accurate and yet sophisticated task. Extensive study has been conducted to obtain 3D shape information from 2D images. While methods using monocular visual cues such as shape from shading (Woodham, 1980), shape from texture (Witkin, 1983), shape from motion (Ullman, 1979), etc., suffer from being unable to achieve a robust metric-based 3D shape acquisition, methods incorporating information from multiple cameras, e.g. stereovision are believed to be able to produce more reliable 3D reconstruction (Trucco and Verri, 1998).

In the past, stereovision suffered from the problem of establishing point correspondence (Scharstein and Szeliski, 2002). Active vision methods have been adopted to reduce ambiguity in solving the correspondence problem. However, an active vision system requires more complicated hardware and a more complex calibration procedure, and is often bottlenecked by its speed (Zhang and Huang, 2006). When observing moving objects, speed is a valuable parameter since higher speed allows more details of the object under motion to be captured. To acquire 3D information from objects at high speed, stereovision is a sensible choice since it can fully take use of the speed of modern digital cameras. Although the problem of stereo correspondence still exists, research has shown it is possible to achieve a good quality 3D recording using stereovision in a controlled environment where the object surface is well illuminated and the object texture is visible to the cameras (Siebert and Urquhart, 1994). In fact the increasingly-improving spatial resolution in digital imaging permits to capture more texture details, which reduces the ambiguity in stereo matching and makes stereovision more technically sound for dynamic 3D shape acquisition.

This paper presents a novel application of 3D shape acquisition using a high-speed stereovision system. The objective is to obtain 3D shape measurements of bats in flight when performing prey-hunting tasks. It’s well known that bats are the nature’s experts in using sonar to locate and identify objects from environment (echolocation). However, it is largely unknown how bats exploit echolocation effectively and efficiently, perhaps due to the large...
diversity of bat species (bats account for 20 percent of all mammal species (Tudge, 2000)) and the highly complex behaviours bats developed (Fenton, 2001). For acoustic engineers, knowledge of bat echolocation could help them build advanced sonar systems. 3D shape information of bats is believed to contain some clue to these questions.

This paper describes the progress in the application. The configuration of the stereovision sensor and the related performance evaluation are described in Section 2. Details of the sensor deployment in real data acquisition are given in Section 3. Section 4 reports the preliminary analysis of the data we collected from live bats on-site. Some conclusions are drawn and future work is discussed in Section 5.

2 SYSTEM EVALUATION

The sensor we employed comprises two Mikrotron™ high-speed monochrome video cameras, two infrared (IR) light panels and two processing computers. The cameras have maximum resolution 1280x1024 and maximum speed of 1000 fps. There is a trade-off between resolution and speed, since higher speed and higher resolution cannot be achieved together within the budget. The monochrome image modality is chosen to reduce data capacity therefore allowing higher frame rate in data acquisition given the same system bandwidth.

The cameras are connected to the frame grabbers running on master-slave mode on the computers, which allows synchronized stereo video capture up to 500 fps with recording duration of 2 seconds. The IR lights are used to illuminate the acquisition scene. The IR wavelength was carefully selected to overlap the visibility spectrum of the cameras and illuminate the scene without disturbing the bats. The computers are paired to receive, store and process raw intensity images captured by the stereo cameras. They share buffers so that the data can be processed in parallel. However, due to the huge data capacity, real-time processing is not possible and data has to be analyzed after acquisition.

Two groups (water trawling and insect gleaning) of bat species were to be studied. For the water trawling bats, 3D shape data ought to be acquired when the bats are about 1.2–2 meters away from the cameras, which allows a reasonable take-off distance after the bats perform prey capture. For the insect-gleaning bats, as they can hover over the prey for a short while, we would like to capture 3D data when the bats are in position as close as possible to the cameras and yet the field of view of the stereovision sensor is big enough to observe 3D morphology of bats in manoeuvre.

To suit the above capturing scenarios, two sets of lenses were purchased: a pair of Fujinon CF50HA-1 (focus length 50mm) and a pair of Fujinon CF75HA-1 (focus length 75mm). Both sets fit the camera mount and resolution requirements. The CF50HA-1 lenses have a focusing range of [0.4m,∞], and the CF75HA-1 lenses have a focusing range [0.9m,∞] and a larger magnification factor than CF50HA-1. With the two sets of lenses, the stereovision sensor is able to observe bats in sensible precision at distances in the range [0.5m, 2m] in a window about 30cm(wide)x40cm(high), as shown in Sections 3.

Because of the novelty of both the sensor and the application, we carried out experiments to evaluate the performance of the sensor before the real data acquisition took place. First we examined the depth range in which 3D measurements of an object can be validly made by the sensor (working range). We used a planar surface as a ground truth object. The surface is rigid and textured. The stereo images of the planar surface were processed using D13D™ stereo photogrammetry software [Khambay, 2008] to generate 3D images of the object. Each 3D image was fitted with a 3D plane, and the RMS (Root Mean Square) value of the fitting residuals was calculated.

![Figure 1: RMS errors of measuring a 3D plane around working distance.](image-url)
We placed the planar surface at a number of depths sampled around a chosen position where the sensor was focused (the distance between the chosen distance and camera baseline is the working distance of the stereovision sensor). At each position, a 3D image of the planar surface was acquired by the sensor and a RMS error was calculated. Fig. 1 illustrates the RMS errors obtained from all test positions around working distances 80cm and 200cm. The working distances 80cm and 200cm were selected as we believe they were the most likely working distances for real data acquisition. It can be seen that the RMS errors in both Fig.1(a) and Fig.1(b) exhibit clearly basin shapes around the chosen working distances. The basin shapes are consistent with what we expect as the cameras have their depth of field. It appears that there is about 10-15cm bottom in each basin where the RMS errors are rather small. Beyond the range of these bottoms, 3D measurements can still be obtained though with larger errors. If we set the upper error threshold as 2 times of the minimum RMS errors, we can obtain about 20cm of working range in either test in Fig.1. In real data acquisition, the object will be of course different from the planar surface, but the RMS error curves in Fig.1 can indicate to some extent the quality of 3D measurements the sensor can achieve at those working distances.

Figure 2: Spatial coherence: RMS errors of fitting a plane to 30x30 in a 3D image.

We also used the planar object to test spatial and temporal coherence of the sensor. The objective is to test whether or not there is any systematic distortion of performance of the sensor in spatial and temporal domains. The spatial coherence test was conducted as follows. A 3D image of the planar surface was partitioned into 30x30 patches. Each patch was fitted with a 3D plane and a RMS error was calculated. The RMS errors are shown in Fig.2, where pseudo color represents the RMS errors. It appears the RMS errors are randomly distributed on the planar surface, which suggests no obvious bias to a specific spatial location in the 3D measurements. In the temporal coherence test, 20 frames of 3D images were obtained continuously from the planar surface in a still pose. For each frame, the 3D data was fitted with a plane and a map of fitting residuals was calculated. For each pair of frames, a correlation value was calculated between the residual maps. We found that the residual maps are highly correlated, with all correlation values larger than 0.9. However, correlation value varies randomly in the range [0.9,1.0]. We hypothesized that the high correlation values are caused by the following two reasons: 1) the stereovision sensor is able to detect the small uneven shape details on the planar surface; 2) the stereovision software we used generates systematic errors associated with the texture of the planar surface. These possibilities are yet to be confirmed.

Figure 3: RMS errors of fitting a sphere to 3D images of a ball swung along three orthogonal directions. Speeds illustrated in color.
The above experiments so far evaluate the sensor in static scenes. We also conducted tests to reveal properties of the sensor related to moving objects. We used a well-textured ball as the ground-truth object. The object was swung over the sensor in three orthogonal directions (horizontal, vertical and depth). By fitting a 3D sphere to the 3D image of the ball, we can calculate the position of the centroid of ball, which can be then used to estimate the ball speed. The fitting residuals can be used to calculate RMS errors, which is an indicator of the level of noise in the 3D images. The RMS errors in relation to motion in the three directions are illustrated in Fig. 3. It can be seen that speed of the ball has a clear effect on the RMS errors. In all three directions the RMS errors increase with the speeds. However, if we compare the three directions, the horizontal (x- in Fig.3(a)) direction seems more sensitive to motion than the vertical (y- in Fig.3(b)) and the depth (z- in Fig.3(c)) directions, as the same magnitude of velocity generates the largest RMS errors in the horizontal direction. The depth direction is least sensitive to motion. At 0.9m/s, the RMS error curve exhibits a clear basin shape, which is a characteristic of static acquisition as shown in Fig.1. At 3m/s, the effect of motion can be clearly seen -- the RMS error curves are much flatter than the lower speeds. The characteristics of the sensor revealed in the above experiments are valuable. They assisted us to figure out a sensible deployment of the sensor for real acquisition of 3D data of bats in relevant capturing conditions. In the next section, we report our results in real data acquisition.

3 DATA ACQUISITION

The real data acquisition took place in two sites: Odense in Denmark and BCI (Barro Colorado Island) in Panama. Four species of bats were examined including one species of European bat (Myotis daubentonii) and three species of tropical bats (Micronycteris microtis, Macrophyllum macrophyllum, Noctilio leporinus). The species were chosen for study because they represent a large variety of body form and acoustic behaviour. However, since bat species have a vast diversity, it is unclear in the literature about the details of sizes, speeds, and body forms of the chosen species. Therefore we had to keep a reasonable degree of flexibility in our sensor configuration to accommodate the possible changes in data acquisition of the bats.

For the water trawling bat species, we estimated their flying speeds are about 2-5m/s according to the experience of bat biologists. Our sensor evaluation test shows that this range of speed may affect the 3D measurement from the sensor. We constrained the flying path of bats in the depth direction of the sensor, as the depth direction is the least sensitive direction to motion blur (Fig.3). This acquisition setup is illustrated in Fig.4. Baits (fishes or mealworms) were put in the water at regular places. The bats usually flow through the narrow water path to take the baits, as they need a distance to take off from the water surface. The bats were therefore trained to move in the depth direction of the stereovision sensor.

Figure 4: Sensor setup for water trawling bats.

Figure 5: Stereo images of a water trawling bat (M. daubentonii) in hunting prey.

Figure 6: Sensor setup and for insect gleaning bats (M. microtis).
Once the bats were trained, the stereovision sensor was in place to acquire images of the bats performing tasks. The sensor was calibrated off-line with a fixed focusing point and working range. The distance between the sensor and the prey was adjusted to observe the different stages of bats performing the tasks. The height of the sensor above the water level was also adjusted accordingly to allow the bats take off properly after hunting the prey. The acoustic equipment to collect audio data of bats was set up behind the stereovision sensor as it works at longer distance and its size is much larger than the stereovision sensor.

An example of the stereo images collected is depicted in Fig. 5. A bat (M. daubentonii) was captured in the cameras just before taking the prey (mealworms). The light was focused around the worms, so the bat and the worms were properly illuminated while the background remained dark. The shape of the bat’s body parts including forehead, mouth, ears, and eyes can be seen in the images although with a noticeable amount of motion blur. We also set up a special light visible to the cameras (e.g., the circular dot in the left image in Fig.5), which flashed continuously in a random pattern to synchronize the audio and video data.

For the insect gleaning bats (M. microtis), object speed is not a major issue in 3D acquisition, since the bats usually reduce their speeds (in our measurements the speed was below 1m/s) to hover around the prey in search of a good hunting position. To obtain the best image quality, we placed the sensor at the closest point to the prey (Fig.6). The prey (insect) was placed on a leaf among a camouflage bush. The position of the prey on the leaf was adjustable to allow study of the bat’s searching behaviour before hunting the prey. The focusing point of the stereovision sensor was on the centre of the leaf. The orientation of the leaf was adjusted almost perpendicular to the depth direction of the sensor to avoid observing occlusion by the leaf. With this setup, the sensor was able to capture larger foreground objects.

The stereo data we collected from the four species of bats are summarized in Table 1. All together 147 valid data acquisition sessions were carried out in 12 working days. Among the species, N. leporinus has the most successful data acquisition sessions, and M. daubentonii has the least amount of valid data acquired. The uneven distribution of the successful data sessions is largely due to the complex behaviour of the bats and also due to how well the bats were trained. Whether the amount of data acquired from each species is sufficient or not is to be investigated.

<table>
<thead>
<tr>
<th>Species</th>
<th>M. dau</th>
<th>M. mac</th>
<th>M. mic</th>
<th>N. lep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Sessions</td>
<td>14</td>
<td>37</td>
<td>20</td>
<td>76</td>
</tr>
</tbody>
</table>

4 DATA ANALYSIS

The data from all four bat species are in the initial stage of processing and analysis. Due to the novelty of the sensor and the application, the total outcome the data could bring about is still under investigation. However, preliminary results already demonstrate that it is possible to derive 3D metric information of the bats from the recorded stereo data.

Our first endeavour was to make some 3D measurements on the bats’ bodies. A few feature points (landmarks) were manually placed in one (reference) frame of the left and right image sequences. Once the landmarks were selected in the reference frame, they were tracked across the remaining frames of the image sequences. The algorithm we applied for tracking the landmarks is based on optical flow (Ogale and Aloimonos, 2007). For each single landmark point on the current frame, a window of 20x20 centred at the landmark was selected and the motion vector field between the window and the image of next/last frame was calculated. The location of the landmark in the next/last frame was then estimated by incrementing the motion vector of the landmark to its position in the current frame. The calculation of the motion vector field allows us to maintain the continuity of pixels in the selected window. The choice of the 20x20 window size is from our experience in balancing data integrity and texture resolvability -- the pixels in the window are supposed to represent a continuous surface of the bat body and yet have sufficient texture to be distinguished from the surrounding regions. The tracking is currently running in a semi-automatic manner. In some frames where the scene has drastic changes, location prediction of the landmark points could be severely inaccurate, and then manual rectification of the locations was necessary.
Once the landmark points were extracted from the left and right image sequences, stereo triangulation was performed to calculate the corresponding 3D points, which can then be used to derive some measurements such as speeds, sizes, mouth openings of the bats, etc. Fig. 7 illustrates an example of such 3D measurements. The four corners of the mouth of the bat were measured in 64 consecutive frames. The recovered 3D positions of the mouth corners are depicted in Fig. 7, where the red lines represent the widths of the mouth and the blue lines represent the heights. The positions of the worms in Fig. 7 were also recovered, from which the water surface was calculated. The 3D positions of the bat mouth corners and the worms in Fig. 7 are displayed in a coordinate system in which the water surface is aligned with the x-y plane. Such an arrangement of coordinates renders the altitude of the bat in flight above the water level (z-coordinate) explicit. It is clearly seen that the bat approached the prey in a low and flattened path and then lifted above the water surface immediately after grabbing the prey.

Speeds of the bat were calculated from the 3D positions of the bat’s mouth corners. 3D coordinates of a landmark in two consecutive frames were differentiated to estimate the speed of the landmark. Fig. 8 illustrates the speeds of the mouth corners. It is evident that the noise level in speed measurement is high, which is due to the error amplification in differentiation. We calculated acceleration factors in the speed curves in Fig. 8 by applying first order linear regression to the curves, and we found three of the curves have consistent acceleration factors (-3.55 m/s$^2$, -3.95 m/s$^2$, -3.5 m/s$^2$), which indicates the bat was reducing its speed (with negative acceleration) to take the prey. The means of the speed curves are consistent as well, all falling in the range [2.8 ±0.03 m/s]. We also calculated the correlation values between the speed curves. As seen in Table 2, the correlation values are insignificant and random, which suggests the noise in speed measurement in Fig. 8 is independent of the bat’s real speed. Another observation about Fig. 8 is that the noise seems larger at the beginning and end than in the middle. This was confirmed by our calculation. We selected the frame in which the bat has the lowest altitude (frame 43). The data of this frame is supposed to have the best quality because the sensor was focused around the worm and the bat grabbed the worm at its lowest altitude. We calculated the standard deviations of the speeds of the bat mouth corners between frame 34 and 52 (frame 43 is the center of them), during which the bat traveled about 10 cm according to the mean speed of the bat we calculated. Not surprisingly, the standard deviations between frame 34 and 52 for all four mouth corners (1.1007 m/s, 1.0233 m/s,
0.8552m/s, 0.8553m/s for the left, right, top, bottom corners respectively) are significantly lower than those derived from the entire set of frames (1.2074m/s, 1.2170 m/s, 1.2338 m/s, 1.1728 m/s for the left, right, top, bottom corners respectively). The result confirms the ‘basin’ effect in the working range experiments (see Fig.1) in Section 2.

Table 2 Correlation factors between speed measurements from in Fig. 8.

<table>
<thead>
<tr>
<th>Correlation</th>
<th>Left</th>
<th>right</th>
<th>top</th>
<th>bottom</th>
</tr>
</thead>
<tbody>
<tr>
<td>left</td>
<td>1.0000</td>
<td>-0.2117</td>
<td>0.0638</td>
<td>-0.0335</td>
</tr>
<tr>
<td>right</td>
<td>-0.2117</td>
<td>1.0000</td>
<td>0.4241</td>
<td>-0.1859</td>
</tr>
<tr>
<td>top</td>
<td>0.0638</td>
<td>0.4241</td>
<td>1.0000</td>
<td>-0.3008</td>
</tr>
<tr>
<td>bottom</td>
<td>-0.0335</td>
<td>-0.1859</td>
<td>-0.3008</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Figure 9: 3D shape of M. daubentonii recovered stereo images.

Figure 10: 3D shape recovery: (a) M. microtis; (b) M. macrophyllum; (c) N. Leporinus.

We also tried to measure mouth openings for the bats. Fig.9 illustrates the distances between the top and bottom, the left and right mouth corners of the bat in different frames in Fig.7. It can be seen that the distances between the left and right mouth corners are quite stationary, indicating the unchanged width of the mouth during the frames. The distances between the top and bottom corners have a clear declining envelope, indicating that the mouth of the bat was closing during the frames. The mean value of mouth width measurements (Fig.9(b)) is 7mm and the standard deviation is 1.3mm. The noise level in the measurements is medium. Considering the bat speed (2.8m/s), the noise level is consistent with the result found during sensor evaluation tests in Fig.3(c) with respect to object motion in depth direction.

3D shape recovery from the stereo data was also investigated using the DI3D™ software, which has been successfully used to capture human faces in a number of applications (Windera, et. al., 2008). We considered three main issues which may affect the stereo software to obtain 3D shapes of bats such as: 1) lack of texture on bats’ bodies; 2) occlusion of body parts; 3) motion blur. Our experiments so far have shed some light on these concerns. Firstly, it was found that the texture of bats can be revealed under proper illumination. Fig.9(b) shows a bat head’s fine texture when the bat flow through the working range of the sensor on which the IR light was concentrated. With the texture visible to the cameras, the stereo software was able to recover part of the 3D geometry of the bat head as shown in Fig.9(a). Secondly, motion blur did take its toll on 3D shape recovery. For instance, the bat’s head in Fig.9(a) was smoothed with some fine shape details missing. Thirdly, the effect of occlusion was evident. For instance, the bat’s ears were squeezed into the head in Fig.9. In Fig.10 (a) and (b), noseleaves of the bats were largely missing. Despite these defects, the results show that it is possible to recover 3D shapes of bats using stereovision methods. Our next step is to develop more suitable stereo algorithms to compute 3D shapes from the data we collected which contain noticeable motion blur and evident object occlusion.

5 CONCLUSIONS

This paper reports an initial effort towards capturing 3D external morphology of bats in flight using a stereovision sensor. We discussed the data acquisition scenarios and evaluated the performance of the sensor accordingly. Stereo data were acquired from four species of live bats, and the preliminary analysis of the data confirmed that it is feasible to obtain 3D shape information of bats in flight for the chosen species using the stereovision method at 500fps. A number of issues were revealed in 3D shape recovery related to motion blur and occlusion, which helps identify the problems we should be still working on and revise the expectation of quality of 3D measurements we can draw from the stereo data.
The schedule for improvement was proposed. The first step is to build a landmarking tool, which aims to extract the positions of prominent landmarks of bats from stereo images. Template-based constraint will be used to improve the stability of the current point-based tracking algorithm. Some manual intervention may still be necessary due to the large uncertain properties of data, nevertheless, we expect to improve the portion of automatic landmarking since there are a large number (about 20,000) of frames of data to analyze. With data landmarked, a range of 3D information about bats will be calculated such as speed, size, mouth opening, head tilt angle, ear distance, etc., which is not only important to characterize the bats’ behaviour but also helps improve 3D shape recovery from the stereo images. For instance, the motion information can be used to de-convolve motion-blurred images, therefore revealing more missing details in the 3D shape recovery.

The next step is to investigate a ‘smarter’ stereovision algorithm for 3D shape recovery. Since consecutive frames of images preserve spatial continuity of the scene, the redundant information between the frames can be used to achieve higher spatial frequency. Our observation about the stereo data is that the head movements of bats are rather rigid and the non-rigid motion component is of low magnitude. Therefore integration of consecutive frames of stereo images (Zhang, 2003) should be a plausible way of retrieving more shape details of the bats’ heads.

Furthermore, a feature-area hybrid approach will be considered to improve the quality of stereo matching. Since landmarks will be extracted in the stereo images, they provide semantic information to stereo matching, which will improve the global optimality of stereo matching as well as the accuracy of matching around the landmarks (therefore allowing sharp features around the landmarks to be better revealed in 3D). For parts on bats bodies of vital acoustic importance such as ears, noseleafs, mouths, etc., specifically designed stereo matching taking into account information of both feature (e.g., ear contour) and texture will be considered.

REFERENCES