

# Vision System for Wing Beat Analysis of Bats in the Wild

Mikhail Breslav<sup>1</sup>, Nathan W. Fuller<sup>2</sup>, and Margrit Betke<sup>1</sup>

<sup>1</sup> Image and Video Computing Group, Department of Computer Science

<sup>2</sup> Center for Ecology and Conservation Biology, Department of Biology, Boston University

## Abstract

*Bats are among a variety of animals that researchers are analyzing with computer-vision methods. The focus of our work is on using shape analysis to estimate the wing beat of individual bats. Our proposed system works with thermal infrared video of bats flying in their natural environment where field conditions make it challenging to record high-quality data. We present wing beat estimates for 20 different bats during their emergence from a cave.*

## 1. Introduction

There is growing interest in the automated analysis of videos of insects, birds, and a host of other flying animals using computer-vision methods. Bats are of interest to researchers in biology and engineering [3]. To date, vision-based techniques have been applied to tracking bats [11], and analyzing their kinematics [2, 8], behaviors [6], and flight trajectories [10]. We propose here a method to extract the wing beat of individual bats by analyzing their shape. Having an estimate of the wing beat frequency of a bat species may improve our ability to design algorithms for detection and tracking of bats in video data, much like periodicity estimation has done for video analysis of pedestrians [4, 5].

Existing vision-based approaches that estimate the wing beat of individual bats typically do so by placing the bats in laboratory spaces and setting up visible-light cameras with high frame rates to record the flying bats in close proximity. Our system can extract the wing beat of individual bats as they fly in their natural environments. Field conditions are more challenging than laboratory conditions because we cannot guide the direction of flight of the bats, our infrared cameras are relatively far from the bats yielding less pixels per bat, and bats stay in the field of view for a short amount of time ( $< 1.5$  s).

The video data used in this work show Brazilian free-tailed bats (*Tadarida brasiliensis*) as they emerge from a cave in Texas, recorded with a thermal infrared camera. Our FLIR SC8000 camera records 14-bit video at a resolution of  $1024 \times 1024$  pixels with a frame rate of 131.5 frames per second. A representative frame from our infrared data is seen in Figure 1 (top) and segmented bats are shown underneath (Figure 1 (bottom)).

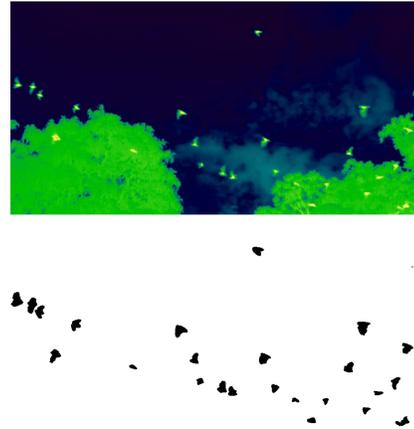


Figure 1: Frame of infrared video showing Brazilian free-tailed bats emerging from a cave (top) and the same frame segmented (bottom). An average bat shown here has a projected area of  $30 \times 30$  pixels.

## 2. Approach

An overview of our system, which is made up of three stages, is shown in Figure 2. The main goal of the first stage is to extract the sequence of shapes generated by an individual bat as it is observed flying through the field of view of the camera in our video data. We use the word “shape” to mean the binary connected component associated with a bat as produced by our segmentation algorithm. The sequence of shapes combined with the 2D trajectory of the bat (obtained by a 2D tracker) forms a signal which we refer to as a “shape-time signal.” Example shape-time signals are shown in Figure 3b. The other component of the first stage of our system involves manually choosing shapes that serve as prototypes for unique 3D bat poses.

The second stage is the main part of our system which operates on individual shape-time signals. The goal is to assign a label corresponding to a discrete pose to each shape in the shape-time signal. To assign the best label to a given shape, the algorithm computes a distance between the shape and each prototype shape. To make these assignments more robust to noise, our system estimates where the wings of the bat are in relation to the body. The end result is that each shape in the shape-time signal is assigned a set of scores indicating

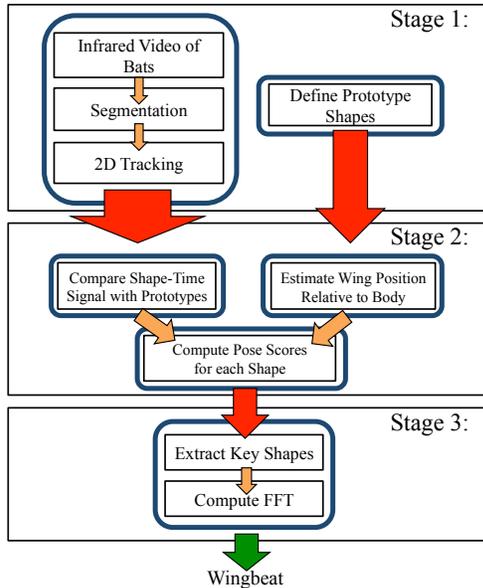


Figure 2: Overview of our 3 stage system. The goal of stage 1 is to generate shape-time signals and prototype shapes. Stage 2 takes as input a shape-time signal and uses shape analysis to assign scores to each shape indicating how well they match different poses. Stage 3 processes these scores and determines which shapes play a key role in estimating the wing beat. Then, time intervals, associated with the repetition of key shapes, are used by the FFT implementation to estimate wing beat.

how well the shape matches each discrete pose.

The third stage processes the scores and defines which shapes in the shape-time signal are key for the estimation of wing beat. Then, time intervals, which are associated with the repetition of key shapes, are passed along to a Fast Fourier Transform (FFT) implementation that yields the final wing beat estimate. Stages 2 and 3 must be executed once for each bat (shape-time signal).

## 2.1 Segmentation and Tracking

Segmentation of bats from infrared video is performed by modeling the distribution of background intensities at each pixel as a Gaussian. The mean and variance of the Gaussian are updated over time and any intensity value outside some fixed number of standard deviations from the mean (15) is considered a bat. Morphological operations are used to fill holes and delete single pixel components. A sample segmentation is shown in Figure 1 (bottom). Using segmented video frames, a Kalman filter tracks the 2D image coordinates of bats and at the same time saves their shapes for further analysis. The collection of shapes stored for an individual bat is the shape-time signal we defined earlier.

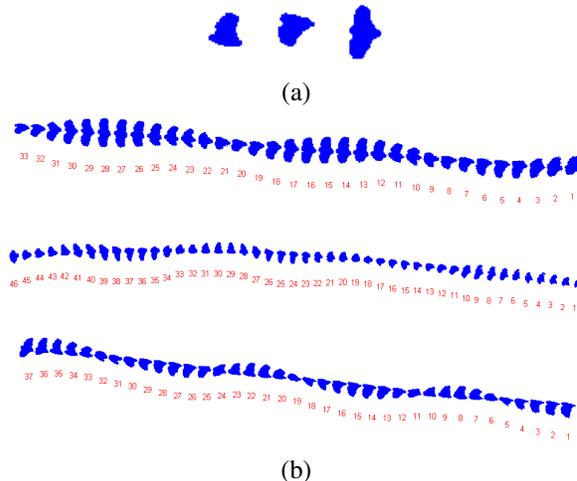


Figure 3: (a) From left to right, three prototype shapes that represent three unique 3D poses: wings above the body nearly completing the upstroke (‘up’), wings below the body nearly completing the downstroke (‘down’), and wings in a neutral position, level with the body, and spread out (‘neutral’). All three shapes are of bats headed in a direction perpendicular to the optical axis of the camera. (b) Three examples of shape-time signals. For visualization purposes, the projections of each bat are spaced out horizontally so the full shape is seen. Each bat is flying from right to left in the field of view of the camera and the numbers underneath a bat indicate the frame number in which it was imaged.

Three example shape-time signals are shown in Figure 3b. In this work, we do not deal with occlusions, so only bats that stay unoccluded are used in our analysis.

## 2.2 Prototype Shapes and Intuition

We motivate defining prototype shapes by observing that some 2D projected shapes of bats relay more information about the 3D pose of a bat than others. We make the assumption that, for a fixed camera setup, certain 2D projected shapes map to unique 3D poses. Shapes that satisfy this property will be defined as the prototype shapes because they model unique 3D poses. Now it follows that the repeated occurrence of a prototype shape in a shape-time signal is equivalent to the repeated occurrence of a unique 3D pose. Since each shape has a time stamp, based on the frame it came from, the periodicity of a 3D pose can be estimated. Using this central idea, our system can compute the wing beat for individual bats. In our work, we manually selected prototype shapes from already acquired automatic segmentations of bats. As an example, three prototype shapes are shown in Figure 3a.

## 2.3 Shape Comparison

Our system compares every shape in the input shape-time signal to every prototype shape using the shape

context descriptor [1] and the Hungarian algorithm [9]. The shape context descriptor is a log-polar histogram that is computed for points along the contour of a shape. The histogram at a point is computed by finding the relative distance and angle to other contour points. All the histograms taken together yield a shape descriptor that is invariant to translation, scale, and rotation (with some added work). Next, the Hungarian algorithm produces a correspondence between points on the first shape with those on the second. If one shape has more points than the other, “dummy points” are added to the smaller shape. The cost of matching point  $i$  in shape 1 to point  $j$  in shape 2 is the  $\chi^2$  distance between their histograms. If one of the points being matched is a dummy point, then a dummy cost can be used (e.g. 0.5). The final distance is obtained by summing the costs of all corresponding points.

## 2.4 Wing Position Relative to Body

The choice of using a rotationally invariant shape descriptor makes it difficult to disambiguate shapes that are similar under rotation such as ‘up’ and ‘down’ (Figure 3a). To resolve this, our system estimates where the wings of a bat are in relation to the position of its body. We observed that the 2D position of the bat changes smoothly across time, which allows us to approximate where the body of the bat is by fitting a polynomial to the shape-time signal. Figure 4 shows part of a shape-time signal with the fitted polynomial. To help differentiate ‘up’ from ‘down,’ four features are extracted from the shape: area  $A_a$  above the polynomial, area  $A_b$  below the polynomial, the furthest point  $D_a$  on the shape above the polynomial, and the furthest point  $D_b$  below the polynomial. Then the ratio  $W = (A_a D_a) / (A_b D_b)$  indicates whether the shape is more likely to be ‘up’ (large ratio), ‘down’ (small ratio), or ‘neutral’ (ratio  $\approx 1$ ).



Figure 4: A fifth-order polynomial (red) is fit to the shape-time signal to approximate the location of the body of the bat.

## 2.5 Pose Scores

Our system classifies each shape in a shape-time signal as one of the prototype shapes (representing unique 3D poses) or ‘neither’ (the label given when no prototype shape is a good match). For this classification shape distances (Section 2.3) are combined with the ratio  $W$  (Section 2.4) to assign a set of scores indicating how similar the shape is to each class. Shape distances are converted to scores by normalizing the distance to

each prototype and subtracting the scores from 1 to reflect similarity. A fixed score is given for ‘neither,’ and the scores are renormalized. The ratio  $W$  is converted to a set of scores by observing that a large ratio indicates a larger score for ‘up,’ and a smaller score for everything else. Similar reasoning is applied for small ratios, and ratios close to 1. The final set of scores is an average of these two component scores, and they sum to 1.

## 2.6 Wing Beat Estimation

The last stage of our system uses the previously computed scores to classify each shape in the shape-time signal, and subsequently produce a wing beat estimate. Each shape is assigned the class for which it has the highest score, and only shapes scoring high enough ( $\geq 0.3$  is a ‘confident’ score) are used for the wing beat estimate. Next, our system finds repetition of poses that are close together in time, which occurs because bats maintain the same general pose for several consecutive frames. To remove these redundant measurements, our system selects a key shape to represent the group. Once key shapes are extracted for the whole shape-time signal, they can be organized based on the pose they represent. Key shapes of the same pose, along with their time stamps, form a time signal which exactly specifies how frequently a pose repeats. All time signals, one for each pose, are summed and sent to an FFT implementation. The result is a frequency spectrum where the fundamental frequency is the wing beat estimate.

## 3. Preliminary Experimental Results

In our experiments, we tracked and analyzed the movement of 20 different bats from an infrared video sequence consisting of 1,000 frames (7.6 s). The shortest track lasted 26 frames and the longest 146 frames (0.19 s and 1.1 s respectively). The prototype shapes used are the same three from Figure 3a (‘up’, ‘down’, and ‘neutral’). For shape comparison we subsample the contour of the larger shape until the number of contour points is roughly equal to that of the smaller shape. Any remaining difference in the number of points is filled with dummy points. The shape context histogram contained 12 angle bins (uniformly across  $360^\circ$ ) and 5 distance bins (log-uniform from 0 to 2). The FFT is computed at 1,024 points, where the sampling rate is the frame rate of our cameras (131.5 frames per second). The fundamental frequency extracted from the frequency spectrum was taken to be the highest peak with a positive frequency immediately following 0 Hz. Using our system, we obtained wing beat estimates for 20 bats. We compared them with “ground truth” wing beat estimates obtained from the results of manually performed shape matching (Table 1).

Table 1: Automatic wing beat estimates for 20 bats and manually obtained wing beat estimates. Their mean  $\mu$  and standard deviation  $\sigma$  are also provided.

Id:	Wing Beat (Hz)		Id:	Wing Beat (Hz)	
	Autom.	Manual		Autom.	Manual
1	10.6	10.8	11	9.3	9.4
2	10.0	9.8	12	9.5	9.8
3	10.4	10.3	13	10.0	10.9
4	10.9	9.8	14	9.2	9.9
5	9.7	9.9	15	9.7	9.4
6	10.5	11.7	16	11.3	11.2
7	13.1	10.9	17	8.7	9.4
8	10.0	9.0	18	6.0	11.4
9	18.7	10.2	19	10.0	10.5
10	13.1	9.9	20	10.2	9.6
Autom.: $\mu = 10.5, \sigma = 2.4$			Man.: $\mu = 10.2, \sigma = 0.7$		

#### 4. Discussion and Conclusion

The main contribution of our paper is a method for isolating key shapes of bats in flight and using them to estimate wing beat frequencies. Our work relied on using the shape context descriptor along with the Hungarian algorithm to compare shapes. This choice necessitated that our system estimates the position of the wings of the bats relative to their bodies.

Sources of error in the current system include noisy segmentations, noisy correspondences from the Hungarian algorithm, and incorrect localization of the body of the bats. Wing beat estimates of bats 7, 9, 10, and 18 may be inaccurate due to too few observed wing beat cycles ( $\leq 2$ ).

To the best of our knowledge, our system is the first to estimate the wing beat of bats in the wild using computer-vision methods on thermal infrared video data. Our wing beat estimates differ from the manually obtained ones by 1.4 Hz on average and they fall within or are near the range of 10-15 Hz which is the wing beat frequency range reported in the biology literature for Brazilian free-tailed bats [7]. Future work includes: evaluation of system performance using different shape distance measures, learning which prototype shapes are best suited for a particular camera view and bat flight trajectory, and performing a quantitative analysis on the mapping between 2D shapes and 3D poses.

**Acknowledgments.** This material is based in part upon work supported by the National Science Foundation under Grant Nos. IIS-0910908 and IIS-0855065 and the Office of Naval Research under grant No. 024-205-1927-5. We would like to thank Diane H. Theriault, Zheng Wu, and Liliana Lo Presti for discussions on shape analysis.

#### References

- [1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):509–522, 2002.
- [2] A. J. Bergou, S. Swartz, K. Breuer, and G. Taubin. 3d reconstruction of bat flight kinematics from sparse multiple views. In *1st IEEE ICCV Workshop on Dynamic Shape Capture and Analysis*, pages 1618–1625, 2011.
- [3] M. Betke, D. E. Hirsh, N. C. Makris, G. F. McCracken, M. Procopio, N. I. Hristov, S. Tang, A. Bagchi, J. D. Reichard, J. W. Horn, S. Crampton, C. J. Cleveland, and T. H. Kunz. Thermal imaging reveals significantly smaller Brazilian free-tailed bat colonies than previously estimated. *Journal of Mammalogy*, 89(1):18–24, Feb. 2008.
- [4] A. Briassouli and N. Ahuja. Extraction and analysis of multiple periodic motions in video sequences. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(7):1244–1261, July 2007.
- [5] R. Cutler and L. Davis. View-based detection and analysis of periodic motion. In *International Conference on Pattern Recognition*, pages 1–4, 1998.
- [6] R. Fisher and Y. Xiao. Bat echolocation behavior from highspeed 3d video. In *Workshop on Visual Observation and Analysis of Animal and Insect Behavior, Istanbul, Turkey*, 2010. 4 pp.
- [7] R. C. Foehring and J. W. Hermanson. Morphology and histochemistry of flight muscles in free-tailed bats, *Tadarida brasiliensis*. *Journal of Mammalogy*, 65(3):pp. 388–394, 1984.
- [8] T. Y. Hubel, N. I. Hristov, S. M. Swartz, and K. S. Breuer. Changes in kinematics and aerodynamics over a range of speeds in *Tadarida brasiliensis*, the Brazilian free-tailed bat. *J. R. Soc. Interface*, 2012.
- [9] H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1–2):83–97, 1955.
- [10] D. H. Theriault, Z. Wu, N. I. Hristov, S. M. Swartz, K. S. Breuer, T. H. Kunz, and M. Betke. Reconstruction and analysis of 3D trajectories of Brazilian free-tailed bats in flight. In *Workshop on Visual Observation and Analysis of Animal and Insect Behavior, Istanbul, Turkey*, Aug. 2010. 4 pp.
- [11] Z. Wu, N. I. Hristov, T. L. Hedrick, T. H. Kunz, and M. Betke. Tracking a large number of objects from multiple views. In *Proceedings of the International Conference on Computer Vision (ICCV)*, Kyoto, Japan, September/October 2009. 8 pp.