Autonomous Aerial Navigation and Tracking of Marine Animals

William Selby, Peter Corke, Daniela Rus

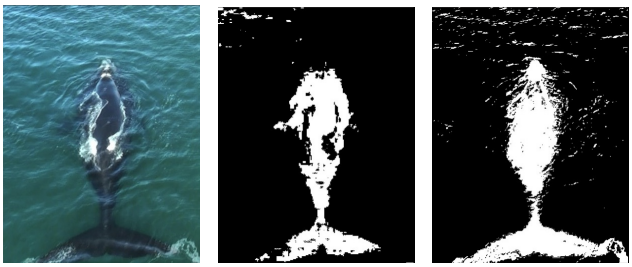
Abstract— **In this paper, we describe the development of an independent and on-board visual servoing system which allows a computationally impoverished aerial vehicle to autonomously identify and track a moving surface target. Our image segmentation and target identification algorithms were developed with the specific task of monitoring whales at sea but could be adapted for other targets. Observing whales is important for many marine biology tasks and is currently performed manually from the shore or from boats. We also present hardware experiments which demonstrate the capabilities of our algorithms for object identification and tracking that enable a flying vehicle to track a moving target.**

I. INTRODUCTION

We wish to develop robotic flying camera systems that can be used to automate data collection in support of biological and environmental studies. We are especially interested in applications where the data collection is over a large body of water and therefore difficult for humans to access.

In this paper we describe our work on autonomous visual servoing for the specific application of tracking whales. Observing whales is important for marine biology tasks such as taking census, determining family lineage, and general behavior observations. Currently, whales are observed manually using binoculars and cameras from the shore or from boats, and notes are made using pencil and paper. The process is error prone, non-quantitative and very labor intensive. Human-operated planes and helicopters are also used, but the data gathered this way is limited. Planes fly at high altitude, can not hover, and the data is limited in duration and precision. Helicopters can hover and fly closer to the sea surface, but they are noisy and effect the behavior of the whales being observed.

We used small hovering unmanned aerial vehicles such as the Ascending Technologies *Falcon 8* robot to assist in the data collection of whales. The robot is silent enough to fly close above the water's surface and not disturb the whales. The robot captures their natural behavior with images of unprecedented detail. Between August 20-25 2009, our team deployed a remote controlled *Falcon 8* robot over the sea at Peninsula Valdez, Argentina to collect data on Southern Right whales. We had several successful missions of approximately fifteen minutes each, during which the robot was piloted over groups of whales and video was recorded.



(a) Original Image (b) Hue Segmentation (c) Sat. Segmentation

Fig. 1. Image Thresholding

Motivated by this data, our goal was to create an autonomous flying robot that relies on vision-based position estimates and only on-board processing to navigate. The robot is a light weight platform such as the Ascending Technologies *Hummingbird* or *Pelican*[1] fitted with a liftable but computationally impoverished processor. Our robot differs from other reported systems that use additional sensors such as GPS or laser scanners or perform off-board processing of the video stream. We describe an autonomous aerial robot system that uses fast, robust but simple computer vision algorithms to identify and track targets in natural environments. The algorithms are designed for targets that move against a fairly uniform background such as a whale moving on the surface of the sea. It could also be applied to other biological monitoring problems such as tracking animals moving in a meadow.

The resulting whale tracking algorithm performs object recognition using a computationally-cheap pixel-level classifier and domain knowledge. We find that image hue and saturation values are suitable invariants to segment whales from other elements in the scene and can be used to estimate the centroid of the moving target. This is complicated by the roll and pitch motion of the quadcopter which is necessary for the under-actuated vehicle to move in the horizontal plane.

Our contribution is a system that uses an effective but computationally modest scene classifier to track whales moving on the water surface using on-board computation. Extensive results are presented for scene classification. We analyzed over 3,500 frames representing Southern Right whales with a resulting 98.99% recall. A state estimator

This work was supported by the MAST project (549969), the MURI SWARMS project (W911NF-05-1-0219), the MURI SMARTS project (N00014-09-1-1051), and the Boeing Company

W. Selby is a student in the Mechanical Engineering Department, Massachusetts Institute of Technology, United States selbywc@mit.edu

P. Corke is with the School of Engineering Systems, Queensland University of Technology, Australia peter.corke@qut.edu.au

D. Rus is with the Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology, United States rus@csail.mit.edu

[1]We used the Hummingbird for indoor experiments and the Pelican for outdoor work.

determines the relative pose of the target and compensates for the roll and pitch motion of the quadcopter which induces large apparent target motion. We also present experimental results from a realistic indoor test where the aerial vehicle autonomously tracks moving whales projected onto the floor of the laboratory.

A. Related Work

Computer vision techniques can be applied to identify and track marine animals to aid marine biologists [1], [2], [3]. Our work builds on an important body of research in visual servoing [4], autonomous navigation for quadrotor robots, and object identification and tracking.

A good review of vision-based control of aerial vehicles is found in [5]. Typically aerial navigation systems involve the use of a combination of GPS and INS sensors [6], [7], [8] to stabilize the vehicle. Our design fuses the relative target position, estimated using vision, with IMU measurements to create a system independent of an external localization source. This makes our system capable of operating without GPS updates for extended periods of time by utilizing a local target centric reference frame. Existing quadrotor systems use sensors such as laser scanners [9] to provide accurate pose estimation but the sensors are heavy and a power burden. SLAM techniques are used to estimate position. SLAM is computationally expensive and may require prior knowledge of the environment [10]. In [11], visual SLAM was used to navigate a helicopter in a known indoor environment using optical flow techniques. For this application SLAM is not appropriate since it computationally expensive and mapping is not required. SLAM would also be difficult given the relatively featureless and dynamic ocean environment.

Most vision-only quadrotor systems require off-board computation since the platform is too small to carry a computer capable of executing the necessary algorithms at a fast enough frequency [12], [13], [14], [11], [15]. Off-board computation requires images and control commands to be transmitted wirelessly and these are subject to interference and dropouts. On-board computation avoids these complications as well as potential time delays. Additionally, many of the vision only systems including [14] and [16] have no means of estimating altitude. Corke [17] described control of a small helicopter using on-board processing with inertial sensors and a stereo camera which yields height information. Our method of altitude estimation is accurate at low altitudes and can estimate the six degree of freedom pose using only information from the on-board IMU and a single camera.

Object identification has been a very active area of computer vision research. Most object identification methods are based on a known model described in terms of point features or edge configuration, color or texture [18], [19], [20], [21]. If the cameras are fixed and the background is constant, foreground extraction techniques can be used as well as motion detection to identify targets moving in the image [22], [23], [18]. If labeled information is available, classifiers can be trained on sample data to learn a model of the object.

Our system uses computationally inexpensive target identification algorithms, and therefore high frequency, to estimate the relative distance from the quadrotor to the target. The images are able to be processed on-board a low payload aerial vehicle which avoids latencies and errors due to wireless image transfer and avoids reliance on external computation. The resulting target position estimates are fused with IMU measurements to stabilize a highly dynamic and underactuated system in a local reference frame independent of an external localization source such as GPS or a motion capture system.

B. Outline

This paper is organized as follows. Section II describes a fast method for object identification and tracking using vision. Section III evaluates this method using the whale footage collected in 2009. Section IV discusses autonomous tracking results.

II. Image Segmentation and Object Tracking

We require a solution that is robust but also computationally efficient enough to run on our small on-board computer at a high frequency. Techniques such as graph cuts [24], [25] and MSER [26] are too expensive for this purpose. Instead our object recognition algorithm characterizes the target in the hue (H) and saturation (S) plane and then identifies the target in subsequent frames based on this characterization.

A. Target Color Model

The model used to represent the target is a two-dimensional histogram. First, an initial frame containing the target similar to Figure 1(a) is captured and converted from RGB color space to HSV color space. A two-dimensional histogram is computed which describes the probability distribution of the HS pair values within the image.

We currently require the user to interactively select minimum and maximum threshold values for the H and S images using sliding trackbars in order to identify the sections of the image corresponding to the whales. We denote these values as H_{min}, H_{max}, S_{min}, and S_{max}.

Algorithm 1 shows the process used to identify the target. Example outputs for the H and S planes are shown in Figures 1(b) and 1(c) respectively, with input from the image shown in Figure 1(a).

Using the threshold values selected by the user, the two-dimensional histogram M is modified so that the values of bins outside the threshold values are set to zero. The resulting histogram, M_t represents the target that the user has identified as a discrete probability distribution in hue-saturation space.

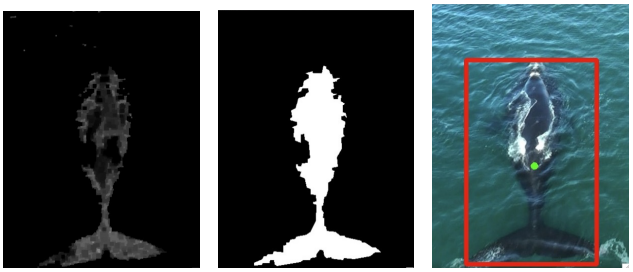
B. Target Identification and Localization

After the user defines the threshold values for target identification, the model histogram M_t is used to find all pixels in future frames which have a high probability of being the target. This process is shown in Algorithm 2. For each consecutive frame, the image is converted to HSV color

Algorithm 1 Target Color Model

Input: Image with target, two-dimensional histogram M
Output: Thresholded H and S plane image and back-projection image

Convert image to HSV color space
Display original H and S plane images
Calculate M from HS image
while Target not completely segmented **do**
    Threshold H and S images to get H_{min}, H_{max}, S_{min}, and S_{max}
    Threshold M to compute M_t
    Calculate back-projection using M_t
    Display back-projection image
    Display thresholded H and S plane images
end while



(a) Back-projection (b) Region Output (c) Vision System Output

Fig. 2. Target Identification

space. The H and S values of each pixel are back-projected through the histogram to form a whale probability image, for example Figure 2(a).

A greyscale morphological opening operation with a 3×3 structuring element is used to remove small false positive regions in the back-projection image which are often noise. Next, a greyscale morphological closing operation with a 3×3 square kernel is used to join together positive regions which are close in proximity and to fill in small gaps.

After the back-projection image has been filtered, contiguous groups of non zero-value pixels are identified as binary regions. We assume the intended target is large in the image. We eliminate regions with small dimensions based on a user defined minimum area. This procedure is effective for instances when the target object is large in the frame. A sample output of the region image is shown in Figure 2(b).

Processing continues until there are no regions remain or until a user defined maximum number of regions has been reached. The center of mass and an axis-aligned bounding box around the region are plotted to identify the target to the user as shown in Figure 2(c).

There are some situations where the user may wish to track separate objects at the same time. By relying on one distinct target histogram, it is not possible to identify and track targets with significantly different H and S values. To overcome this obstacle, the classifier code supports the creation of multiple target histograms. If a new object comes into view, the user has the ability to capture a frame and create a separate target histogram for the new object using the process in Algorithm 1. Each time the user creates a new target histogram, the target histogram and corresponding threshold values are stored in a vector including the previous target histograms and their threshold values. Algorithm 2 is then run on each target histogram separately. Each target is identified with a bounding box and center of mass marker.

Algorithm 2 Object Identification

Input: Image, target histogram M_t
Output: Image with target(s) identified

Convert image to HSV
Calculate back-projection image using M_t
Perform morphological operations on back-projection
Identify pixel groups as regions
Identify largest region height and width
if $\text{Perimeter}_{Region} < \text{Perimeter}_{Max}$ **then**
    Remove region
end if
if $\text{Width}_{Region} < \text{Width}_{Max}$ AND $\text{Height}_{Region} < \text{Height}_{Max}$ **then**
    Remove region
end if
Calculate center of mass
Plot center and bounding box and display output image

Once the appropriate objects have been identified in the image, a vector that points to the average location of the objects from the image center is calculated and optionally displayed. For images with only one target, the vector points directly to the center of mass of the target. However, when there are multiple target in the frame, this vector is computed in two separate ways. The first method takes each target's x and y center of mass coordinates and computes the average center of mass location

$$(x_e, y_e) = \Big(\frac{1}{n}\sum_{i=1}^{n} x_i, \frac{1}{n}\sum_{i=1}^{n} y_i\Big) \tag{1}$$

where (x_e, y_e) is an error vector representing the average location of the identified objects, x_i and y_i represent the center of mass coordinates for each object, and n is the total number of objects found in the image.

The second method computes a weighted average location based on the bounding box area of each target so that objects that appear larger in the frame are emphasized in this calculation

$$(x_e, y_e) = \Big(\sum_{i=1}^{n} a_i x_i, \sum_{i=1}^{n} a_i y_i\Big) \tag{2}$$

where each object's x and y center of mass coordinates are multiplied by the pixel area a_i of the bounding box enclosing the object and $\sum_{i=1}^{n} a_i = 1$ and $a_i > 0$.

III. Whale Classification Results

We implemented the object tracking algorithms described above and tested them in two ways: (1) on existing video

TABLE I
VISION SYSTEM RESULTS: OVERHEAD PERSPECTIVE

	Minimum	Average	Maximum	Stand. Dev.
False Negative	0.00%	3.94%	14.43%	6.12%
False Positive	0.00%	39.59%	193.81%	77.13%
Splitting	0.00%	4.66%	13.85%	5.28%
Using Pixel Value Logic				
False Negative	0.00%	1.73%	8.46%	3.33%
False Positive	0.00%	1.63%	5.73%	2.24%
Splitting	0.07%	4.51%	9.90 %	3.55%

footage taken of whales from a quadcopter and (2) operating on-board a quadrotor robot tracking a target in real time in the laboratory. Our goal was to determine if Algorithm 2 can be used for robust real time identification of whales in an image sequence collected by a robot. This section presents results which show that the algorithm is effective for our application and we present experimental hardware results utilizing this algorithm in the next section.

Our data set was the video footage our team recorded using a manually piloted aerial vehicle. From this data set, six clips were extracted resulting in 3,537 frames.

Some clips contained multiple identifiable whales. Other clips contained whales that appeared from outside the frame or were occluded by another whale. The clips also varied in duration as well as environmental conditions such as sea state and light level and direction.

The classification results were manually inspected to calculate the performance of the algorithm based on the following metrics: the rate of false positives, false negatives, and region splitting. Region splitting occurs when one object is represented by multiple regions. For each frame, any region identified by the classification system but did not contain a whale was counted as false positive. Each whale in the frame not identified was counted as a false negative. When a whale was identified with multiple bounding boxes, this was counted as an instance of splitting. The total number of false positive, false negative, and splitting instances were totaled and divided by the total number of frames in the clip to calculate the percentages shown in Table I. Percentages over 100 meant that the clip had more than one of that particular type of instance per frame. While it is possible to manually update the classifier during the clips to improve performance, this was not done. The classifier system was initialized only once using the first frame in the clip.

Six clips were analyzed with a mixture of our aerial vehicle footage from Argentina. The view angle from the robot resulted in a large distance between the camera and the whale. This distance reduced significant lighting variations in the image which resulted in more consistent HS values for both the whales and the background. The only noticeable adverse effect of this viewpoint was caused by the wakes of the whales. When the whales were on the water's surface, their wake's HS characteristics were very similar to the HS characteristics of the whales. The statistics for all clips with an overhead view are shown in Table I.

In particular, waves, wakes, and the beach caused high numbers of false-positives. The size difference between a large group of whales and a single whale in the image made size-based filtering ineffective. However, pixel value heuristics which estimated the percentage of *wave*, *wake*, and *beach* pixels in a bounding box vastly improved the classifier performance as shown in the bottom half of Table I. The HS characteristics of *wave*, *wake*, and *beach* pixels were estimated before the experiment was run. To determine which pixels were possibly *wave*, *wake*, and *beach* regions, HS classifications were used. The user isolated the *wave*, *wake*, and *beach* regions independently by manipulating the HS minimum and maximum values in order to create HS characterizations for these regions. As the experiment ran, the classifier identified candidate target regions. Once a target was identified by the classifier, the pixels in the bounding box were analyzed to confirm that the region was a whale and not a false positive. Using the HS classifications defined previously for *wave*, *wake*, and *beach* pixels, the percentage of *wave*, *wake*, and *beach* pixels in the bounding box were calculated. If a bounding box had more than a user defined percentage of *wave*, *wake*, and *beach* pixels, the region was assumed to be a false positive and ignored.

In summary, we have evaluated over 3,500 frames of real world footage containing several types of whales in various poses. Our results show 98.9% recall for the Southern Right whale footage we collected in 2009. This high success rate motivates our desire to automate the whale identification and tracking task using a small, visually guided aerial robot which we describe next.

IV. AUTONOMOUS VISUAL SERVOING

In this section we describe an autonomous on-board visual navigation and tracking system for an Ascending Technologies *Hummingbird* [27] quadrotor vehicle to support the whale tracking application. Due to the limited payload of the robot, we are restricted to a computationally impoverished SBC such as a Fit-PC2. In this section we describe the hardware and software of this robot system. Initially, highly accurate feedback from a motion capture system was used to stabilize the vehicle while the vision system indicated the desired motion within the plane. Next, the motion capture feedback was replaced by estimated state measurements from an Extended Kalman Filter (EKF) and the motion capture data served as ground truth for performance evaluations. The EKF estimated the quadrotor relative pose using vision and IMU data

A. System Architecture

The vision system was run on the vehicle using a 2.0 GHz Intel Atom processor (Fit-PC2) with a Point Grey Firefly MV USB camera. The camera had a resolution of 640×480 pixels which was down sampled to 320×240 pixels to reduce computational cost. Since the intended target is relatively large in the image this does not result in significant loss of precision about the target's relative position. The full system combined for a total payload of 535 g, well above the recommended maximum payload of 200 g for this

Fig. 3. Experimental Hardware

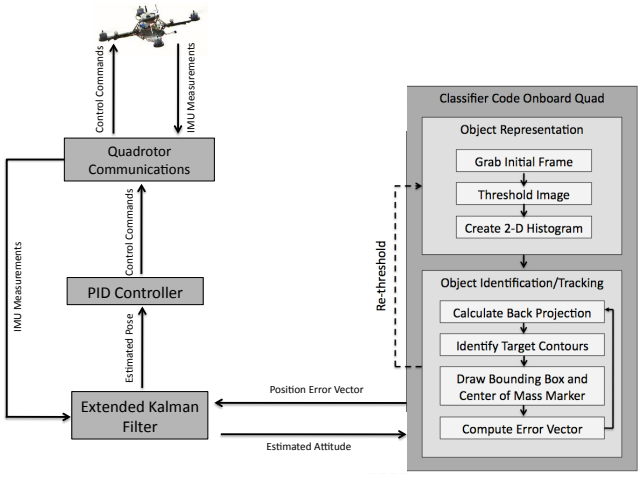


Fig. 4. Experimental Software Architecture

platform, but our experiments show that the system remains sufficiently manoeuvrable. The complete aerial system is shown in Figure 3.

A laptop was used to remotely run the commands on the Fit-PC2 via a wireless SSH connection. The software architecture is described in Figure 4. Several modules were run simultaneously on the Fit-PC2 and message routing between modules as well as data logging was handled by LCM [28].

The vision system module used Algorithm 2 to identify the target in the image frame. The controller module utilized four independent PID controllers to compute the roll, pitch, yaw and thrust commands. It received the pose of the robot either from the motion capture module or from the EKF module. The low-level on-board autopilot software computed the individual motor commands.

The quadrotor communication module was used to interface with the quadrotor's Autopilot board. This module received IMU measurements and sent control commands via a USB to serial connection.

The EKF module, which is discussed in more detail in [29], received the vision system output and IMU readings and estimated the target-relative pose of the quadrotor which was used to calculate control commands.

We demonstrated whale tracking by projecting clips of robot-captured video of whales onto the floor of the lab. Using the Vicon motion capture system for state feedback, the system was able to identify the whales projected onto the floor and remain stably positioned over the group. A sample run of this experiment can be seen in the accompanying video. Shadows from the robot were avoided by projecting the video at a sharp angle and adjusting the optics to project a rectangular frame.

B. Visual Target Tracking Using On-board Estimated Pose

We developed an algorithm for tracking whales that relies entirely on visual feedback. This algorithm is described in detail in [29]. The problem is considered a planar tracking problem. An EKF is used to estimate the pose of the quadrotor in $SE(2)$ with respect to the target. Roll and pitch cannot be independently controlled and altitude is controlled by a separate mechanism. This estimated pose was sent to the control module which computed commands to maneuver the quadrotor to the center of the target. The EKF was adapted extensively from [10], [30] and implemented using the KFilter library [**?**]. This filter combined position estimates from the target identification and tracking algorithms described in Section II at 10 Hz as well as attitude and acceleration information from the IMU at 30 Hz. The filter had to handle these asynchronous measurements and the inherent latencies in these measurements. While the IMU measurements were noisy and included a bias that drifted over time, the slower relative position estimates from the vision system were more accurate and without bias. The filter relied more heavily on the vision system position estimates and utilized the IMU measurements to estimate the quadrotor's target-relative pose between updates from the vision system. This improved the quality of the velocity estimate which is essential for stable and high-quality vehicle motion.

1) Indoor Experimental Results: We implemented the visual servo algorithm described in [29] on a *Hummingbird* quadrotor robot. The computation was performed entirely on-board using a FitPC. The robot was tasked to track whales in a movie clip projected on the floor of our lab. Figure 5 shows a sample configuration. In limited testing we observe that the system correctly tracks the pod of whales. In 44% of frames the whales are oversegmented, that is the whale regions are split resulting in more regions than there are whales in the frame. Since the system tracks the centroid of the regions according to (2) the overall tracking performance is not noticeably degraded.

2) Outdoor Experimental Results: For outdoor operation as envisaged for this application GPS will be available. GPS can be used to geotag the whale's location while also providing low-rate absolute position information for control. It can also be used to fly a grid pattern search for whales. The GPS navigation controller provided on-board the *Pelican* autopilot board is is able to keep the quadrotor within a 5 m radius of the set point when there is good GPS reception. When the quadrotor is placed in GPS mode, the most recent latitude, longitude, and height estimates are used as a desired set point for the GPS controller which attempts to hover

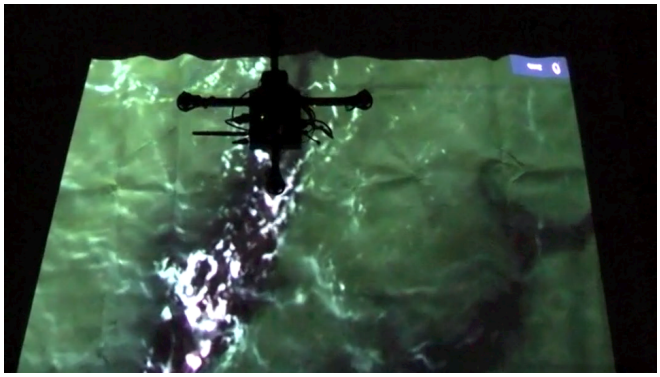


Fig. 5. Experimental snapshot of robot visually servoing to whales. A whale movie clip recorded in Argentina was projected on the lab floor and used to evaluate the visual tracking of whales.

around the desired set point. A simple proportional controller is used to calculate inputs to the GPS navigation controller from the attitude-compensated vision system output. This results in an updated GPS waypoint (latitude, longitude, and height) for the on-board GPS controller.

Experiments were conducted on a grassy field at the MIT campus. This field was flat and uniformly colored without any obstacles. The field was a circle about 50 meters in diameter. The desired target was a red wagon approximately 1.4×0.7 m which was initially static and then manoeuvred by hand. The wagon was pulled as close to 1 m/s as possible. The experiment was designed so that that human would also be identified as a target once in the frame. The vision system correctly identified the target in 91.7% of the frames. When the target is moved, the vision system creates a new GPS waypoint that keeps the quadrotor over the wagon.

These experiments validated our technical approach and proved that this system is capable of reaching our desired goal of an autonomous aerial platform that is capable of tracking a specific class of target. Out of 229 frames recorded, 7.86% of the frames had a false negative, 0.44% had a false positive, and 91.7% correctly identified the indented target. GPS inaccuracies caused the quadrotor to oscillate around the target as the quadrotor maneuvered. These experiments were done at low altitude which amplified the effects of these oscillations in the image. With a larger testing space and a larger target, the quadrotor can be flown at a high altitude which would minimize the effects of these errors.

V. CONCLUSIONS AND FUTURE WORKS

This research focused on identifying whales in an ocean environment. We have described computationally cheap vision algorithms that segment regions based on color as well as pixel value heuristics that eliminate image features such as waves, wakes, and beach areas. We compared the performance against human classification for many video clips of whales in the ocean, which gives quantitative performance data.

The vision system not only identified the intended target, but calculated a relative target position estimate which was used to control the quadrotor's position and enable the quadrotor to track the target. We have shown experimentally that the quadrotor is able to follow the target moving at constant velocity or maneuvering.

We have a significant program of future work planned before our next field-work campaign. Our ultimate goal is to field test this system in an ocean environment similar to where the initial video footage was captured. We are also exploring other means of classifying the intended target. Currently, our HS representation is a rectangle in HS space. Other methods which avoid this limitation will be explored to determine any perfomance gains.

VI. ACKNOWLEDGMENTS

This work was supported by the MAST project (549969), the MURI SWARMS project (W911NF-05-1-0219), the MURI SMARTS project (N00014-09-1-1051), and the Boeing Company. We are grateful for this support. We appreciate the technical support from Daniel Gurdan and Jan Stumpf of Ascending Technologies. We also thank Roger Payne, Mariano Sironi, and the Whale Conservation Institute for enabling the experiments to collect the whale video data. We would also like to thank Brian Julian, Daniel Soltero, and Abe Bacharach for their time and efforts with the quadrotor experiments.

REFERENCES

[1] A. Kreho, N. Kehtarnavaz, B. Araabi, G. Hillman, B. Wrsig, and D. Weller, "Assisting manual dolphin identification by computer extraction of dorsal ratio," *Annals of Biomedical Engineering*, vol. 27, no. 6, pp. 830–838, 1999. [Online]. Available: http://www.springerlink.com/content/mx8k54103337157p

[2] E. Ranguelova, M. J. Huiskes, and E. J. Pauwels, "Towards computer-assisted photo-identification of humpback whales," in *ICIP*, 2004, pp. 1727–1730.

[3] N. D. Kehtarnavaz, V. Peddigari, C. Chandan, W. Syed, G. R. Hillman, and B. Würsig, "Photo-identification of humpback and gray whales using affine moment invariants," in *SCIA*, 2003, pp. 109–116.

[4] S. Hutchinson, G. Hager, and P. Corke, "A tutorial on visual servo control," vol. 12, no. 5, pp. 651–670, Oct. 1996.

[5] A. Kurdila, M. Nechyba, R. Prazenica, W. Dahmen, P. Binev, R. DeVore, and R. Sharpley, "Vision-based control of micro-air-vehicles: progress and problems in estimation," in *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, vol. 2, Dec 2004, pp. 1635 – 1642 Vol.2.

[6] O. Amidi, "An autonomous vision-guided helicopter," Ph.D. dissertation, Robotics Institute, Carnegie Mellon University, 1996.

[7] M. Jun, S. Roumeliotis, and G. Sukhatme, "State estimation of an autonomous helicopter using kalman filtering," in *Intelligent Robots and Systems, 1999. IROS '99. Proceedings. 1999 IEEE/RSJ International Conference on*, vol. 3, 1999, pp. 1346 –1353 vol.3.

[8] J. J. Kehoe, R. S. Causey, M. Abdulrahim, and R. Lind, "Waypoint navigation for a micro air vehicle using vision-based attitude estimation," in *Proceedings of the 2005 AIAA Guidance, Navigation, and Control Conference*, 2005.

[9] R. He, S. Prentice, and N. Roy, "Planning in information space for a quadrotor helicopter in a gps-denied environment," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, May 2008, pp. 1814 –1820.

[10] N. R. A. Bachrach, R. He, "Autonomous flight in unknown indoor environments," *International Journal of Micro Air Vehicles*, vol. 1, no. 4, pp. 217–228, December 2009.

[11] S. P. Soundararaj, A. K. Sujeeth, and A. Saxena, "Autonomous indoor helicopter flight using a single onboard camera," in *Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems*, ser. IROS'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 5307–5314. [Online]. Available: http://portal.acm.org/citation.cfm?id=1732643.1732910

[12] G. P. Tournier, M. Valenti, J. P. How, and E. Feron, "Estimation and control of a quadrotor vehicle using monocular vision and moire patterns," in *In AIAA Guidance, Navigation and Control Conference*. AIAA, 2006, pp. 2006–6711.

[13] E. Altug and C. Taylor, "Vision-based pose estimation and control of a model helicopter," in *Mechatronics, 2004. ICM '04. Proceedings of the IEEE International Conference on*, June 2004, pp. 316 – 321.

[14] A. Cherian, J. Andersh, V. Morellas, B. Mettler, and N. Papanikolopoulos, "Motion estimation of a miniature helicopter using a single onboard camera," in *American Control Conference (ACC), 2010*, July 2010, pp. 4456 –4461.

[15] M. Earl and R. D'Andrea, "Real-time attitude estimation techniques applied to a four rotor helicopter," in *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, vol. 4, Dec. 2004, pp. 3956 – 3961 Vol.4.

[16] A. Mkrtchyan, R. Schultz, and W. Semke, "Vision-based autopilot implementation using a quadrotor helicopter," in *AIAA Infotech@Aerospace Conference*, 2009.

[17] P. Corke, "An inertial and visual sensing system for a small autonomous helicopter," *J. Robotic Systems*, vol. 21, no. 2, pp. 43–51, Feb. 2004.

[18] G. Bradski, A. Kaehler, and V. Pisarevski, "Learning-based computer vision with intel's open source computer vision library," *Intel Technology Journal*, vol. 9, no. 2, pp. 119–130, May 2005.

[19] J. G. Allen, R. Y. D. Xu, and J. S. Jin, "Object tracking using camshift algorithm and multiple quantized feature spaces," in *VIP '05: Proceedings of the Pan-Sydney area workshop on Visual information processing*. Darlinghurst, Australia, Australia: Australian Computer Society, Inc., 2004, pp. 3–7.

[20] C.-F. Juang, W.-K. Sun, and G.-C. Chen, "Object detection by color histogram-based fuzzy classifier with support vector learning," *Neurocomputing*, vol. 72, no. 10-12, pp. 2464–2476, 2009.

[21] F. Porikli, "Integral histogram: A fast way to extract histograms in cartesian spaces," in *in Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2005, pp. 829–836.

[22] D. G. R. Bradski and A. Kaehler, *Learning opencv, 1st edition*. O'Reilly Media, Inc., 2008.

[23] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, p. 13, 2006.

[24] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 1222–1239, 2001.

[25] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," *ACM Transactions on Graphics*, vol. 23, pp. 309–314, 2004.

[26] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761 – 767, 2004, british Machine Vision Computing 2002. [Online]. Available: http://www.sciencedirect.com/science/article/B6V09-4CPM632-1/2/7e4b5f8aa5a4d6df0781ecf74dfff3c1

[27] D. Gurdan, J. Stumpf, M. Achtelik, K.-M. Doth, G. Hirzinger, and D. Rus, "Energy-efficient autonomous four-rotor flying robot controlled at 1 khz," in *Robotics and Automation, 2007 IEEE International Conference on*, April 2007, pp. 361 –366.

[28] A. S. Huang, E. Olson, and D. Moore, "Lcm: Lightweight communications and marshalling," in *Int. Conf. on Intelligent Robots and Systems (IROS)*, Taipei, Taiwan, Oct. 2010.

[29] W. Selby, P. Corke, and D. Rus, "Vision-based tracking for a quadcopter with onboard computation," in *Robotics and Automation, 2007 IEEE International Conference on*, May 2012, p. submitted.

[30] A. Bachrach, A. Winter, R. He, G. Hemann, S. Prentice, and N. Roy, "Range - robust autonomous navigation in gps-denied environments," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, May 2010, pp. 1096 –1097.