2013 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEPT. 22–25, 2013, SOUTHAMPTON, UK

VISUAL HULL RECONSTRUCTION FOR AUTOMATED PRIMATE BEHAVIOR OBSERVATION

Nastaran Ghadar, Xikang Zhang, Kang Li, Deniz Erdogmus

CSL, Northeastern University

*Guillaume Thibault, Alireza Bayestehtashk, Izhak Shafran Kris Coleman, Kathleen A. Grant**

Oregon National Primate Research Center, OHSU

ABSTRACT

The study of social animal interactions is used as means for understanding animal behavior and biology. In this work, we describe a computerized method that utilizes 3D visual hull reconstruction to identify and localize rhesus macaques in their social groups. There are three major steps in this study. First, we collect experimental data from four synchronized cameras at different locations and angels in a cage containing five rhesus macaques. Second, by using computer vision algorithms, we detect and identify animals using 2D observations that were provided from the previous step. This provides essential quantitative data for animal behavior research. Finally, by applying visual hull reconstruction algorithm, we automatically build a 3D model for each rhesus macaques on every frame. The results of this work can be used for tracking these animals in their cage, and furthermore it can be used for activity recognition of social interactions of rhesus macaques. The method we developed in this paper, shows promising results that are accurate, yet runs in a timely manner ; this makes this algorithm suitable for large datasets and we can use it for future high-level recognition tasks.

Index Terms— Visual hull reconstruction, social animals, object detection, background substraction

1. INTRODUCTION

Having an automated system to be able to observe and model the social activities and animal behaviors gives us the ability to recognize and understand the behaviors of social animals. Creating such mechanism has a wide range of applications. The ability to better protect animals which are in danger, using the animal models for health research, helping zoos for creating a better environment for animals are some of its applications. Another important aspect of such system is in animal biology research.

Our objective is to show how computer vision, machine learning research in general and 2D object detection, 3D reconstruction in particular can accelerate the rate and quality of research in behavior of social animals. Most of studies done on animal behavior have one major limitation and that is the slow rate of processing and analyzing collected data. Surveillance of social behavior is either processed by a human expert who has been trained to recognize different behaviors or by videotaping the animals and later coding the videotapes. As for the former method, apart from the fact that an expert should be available all the time, he also needs to pay attention to all the behaviors happening at the same time; therefore it is pretty common that some behaviors be missed or mislabeled. And the later method is very time and space consuming. Hence, having an automated system that does both the surveillance and behavior recognition could be a lot of help in understanding compound social behaviors with no need of human observation.

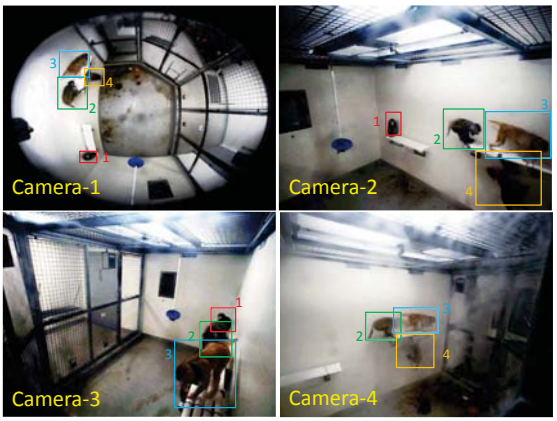

Fig. 1. Collected observation video dataset from four views. All cameras are synchronized. Primates are identified through colored collars around their necks.

In this work, we develop a general framework for detecting, localizing, and reconstructing images of social animals in 3D observation environment. Many 2D and 3D vision techniques are explored and tailored for the social animal research. We think that an automated all-directional observa-

*This work is supported by NSF under grant BCS-1027724. We acknowledge the help of John Hunt and Murat Akcakaya in this work.

978-1-4799-1180-6/13/$31.00

tion system is necessary for any practical application, because both recognizing high-level behaviors and kinematic modeling of animals require a system that is robust to object rotations, translations, occlusions, and depth variations. Without a 3D observation, these difficulties can not be addressed effectively.

Our primary contribution in the animal behavior research study is that we provide 3D observations from 2D videos. The major steps to get 3D observations from 2D videos are:

- **Primate detection and identification in 2D**: We train a primate detector and a background model from a small portion of labeled video data. Then these two techniques are integrated to generate 2D shapes of detected primates.
- **Multiview environment and calibration**: Once the silhouettes are gathered from multiple 2D views, we can use calibrated multiple cameras to project synchronized 2D observations into 3D space.
- **3D visual hull reconstruction**: Based on those projections and calibrated 3D environment, we can reconstruct 3D primates based on shape-from-silhouettes techniques.

Detection helps us towards tracking rhesus macaques in a video sequence and the 3D visual hall reconstruction gives us information regarding shape and relative distance of these species from each other. These two components are essential for activity recognition since some social behaviors can be extracted from relative distance of species such as dominance, or aggression; and some behaviors are extracted from the 3D shape of the monkeys, such as locomotion or grooming.

2. RELATED WORK

In this section, we provide a selected review of closely related work. After briefly describing behavior research of primates relevant to this work and current measurements, we summarize the state-of-the-art algorithms from computer vision for tackling computational challenges in our specific aims.

Primate Behavior Research: Behaviors, especially social interactions, among primates are widely studied. There are some common categories for these social behaviors [1], some of these categories are as follows: A nonviolent demonstration between a pair of primates; a series of social connections between primates which can be interpreted as some sort of relation between them, e.g. dominant/subordinate; and social hierarchy, which goes one step further from social relations between a pair and gives more general information about their groups , e.g. which primate is the leader/follower.

Computer Vision based Animal Behavior Analysis: In computer vision community, many studies utilize videos of animals as testing dataset for developing new algorithms, especially for tracking or behavior recognition.

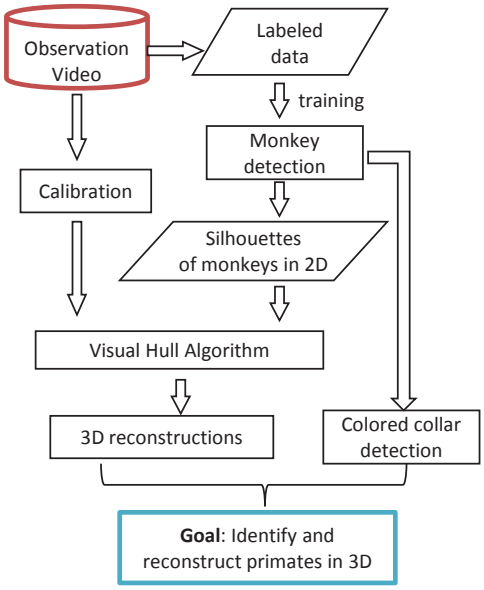

Fig. 2. Framework.

Khan et al. [2] developed a system that can automatically generate the three dimensional trajectory of primates in an outdoor environment. Their purpose is to evaluate the navigational abilities of non-human primates. Their system extracts primate kinematic features such as path length, speed, and other variables impossible for an unaided observer to note. From trajectories, they computed and validated a path length measurement and proposed a method for automatic behavior detection. Also, their system is used to examine the sex differences in spatial navigation of rhesus primates. Chaumont et al. [3] proposed a computerized method and a software called MiceProfiler, that uses geometrical primitives to model and track social interactions in mice. Their system monitors a comprehensive repertoire of behavioral states and temporal evolution, which is utilized for identifying the key events that trigger social contact. Balch et al. [4] proposed an automated labeling system for studying social insect behaviors, which is inspired by multirobot systems, speech recognition and computer vision research. Their ultimate goal is to automatically create executable models of animal behavior. Several algorithms are developed for tracking, recognizing, and learning models of social animal behavior.

3. METHODOLOGY

3.1. Primate detection and identification in 2D

3.1.1. Primate detection

The goal is to detect and localize primates in 2D image frames from videos of each view.There are two main phases in this module: training phase and detection phase.

- **Training phase**: For each of the four views, we manually labeled 1K instances of monkeys and randomly sampled 2K negative non-monkey patches with similar size. and created a fixed-resolution (60×60) normalized training image data set, in total 12K annotated images. Then we encoded images into feature spaces. Here we selected Histogram of Oriented Gradients (HOG)[5] and Average RGB(aRGB) as the feature descriptors. Popular HOG descriptor aims at extracting object local appearance and shape information, which can be described by the distribution of intensity gradients over different edge directions. aRGB was mainly used to supplement the color information of objects. For each view, based on the labeled data, we trained a binary classifier or detector by using linear SVM [6]. For each detector and parameter combination a preliminary detector is trained and tested again on the training set. Using all the false positives obtained, the detector is re-trained. This re-training significantly improved the performance of each detector.

- **Detection phase**: The basic detection sliding window sizes we set are [60 60], [60 100], [100 60], [100 100]. Then we use 10-level pyramid images to detect instances of multiple scales. The scale step size is 1.05. The same HOG and aRGB features are extracted over the windows at all locations and linear SVM classifier is run to decide if an instance is a primate or not. Finally, multiple detections are fused with non-maximum suppression. Methods that extract features such as Scale-invariant Feature Transform (SIFT) or Speeded Up Robust Features (SURF) cannot be used in this study since the extracted features using these methods cannot classify the primates and background with the current cage setup.

3.1.2. Extract silhouettes

We want to reconstruct and locate 3D pose (visual hulls) of primates in the observation environment. The basic idea is to use 2D silhouettes information from synchronized multiple views reconstructing 3D shape of the objects. These types of methods in the literature are called shape from silhouettes [7, 8, 9]. So a good extraction of silhouettes in 2D images is critical for later 3D reconstruction. Two type of background substraction techniques are used here to handle static and motive cases of primates respectively. Since there are multiple moving objects (primates) under observation, to have separate silhouettes for each primate we need to take advantage of both primate detection and background substraction techniques. In our approach, we only consider background substraction within each detected bounding boxes.

For background subtraction we created a static model for the background in each view based on some images of the empty cage. To obtain the foreground, for each frame we normalized the created static background image with that image and then subtracted it to obtain the foreground. Finally we binarized the result just to have the primates as foreground and everything else as background.

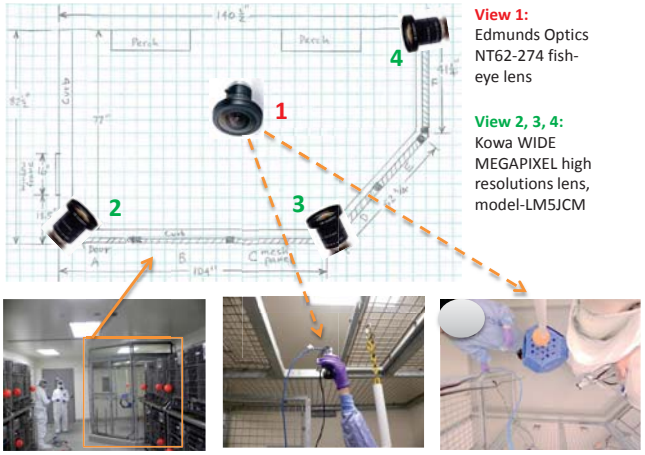

Fig. 3. Environment set up and lens installation.

3.2. Multiview environment and calibration

In order to determine the visual hull corresponding to a set of primate silhouettes, the cameras that produced the images must be calibrated. This means that the intrinsic camera parameters (such as focal length, principal point) and the pose must be (at least approximately) known. So camera calibration is another necessary step in building our 3D vision assisted observation environment. We use four cameras from different views as a quantitative sensor to recover 3D quantitative measures about the observed scene from 2D images. For our study, from a calibrated camera we can measure how far a primate is from the camera, or the height of the primate, etc. Here we briefly introduce the calibration algorithm we applied in our system and some specifications about the environment.

Figure 3 shows the experimental environment of our project. We are using 3 of the Kowa lenses (LM5JC1M 2/3", focal length 5mm, f/2.8) for the wall cameras and the Edmunds fish eye lens (Optics NT62-274, focal length 1.8mm, f/1.4) for the ceiling camera. The calibration algorithm we used is very similar to [10] which will estimate the intrinsic parameters, including focal length, principal point, skew coefficient, and distortions, and extrinsic parameters including rotations and translations. Figure 4 illustrates our calibration process.

3.3. 3D visual hull reconstruction

If camera intrinsic and extrinsic parameters are known from calibration, then the visual hull [11, 12, 13] can be computed

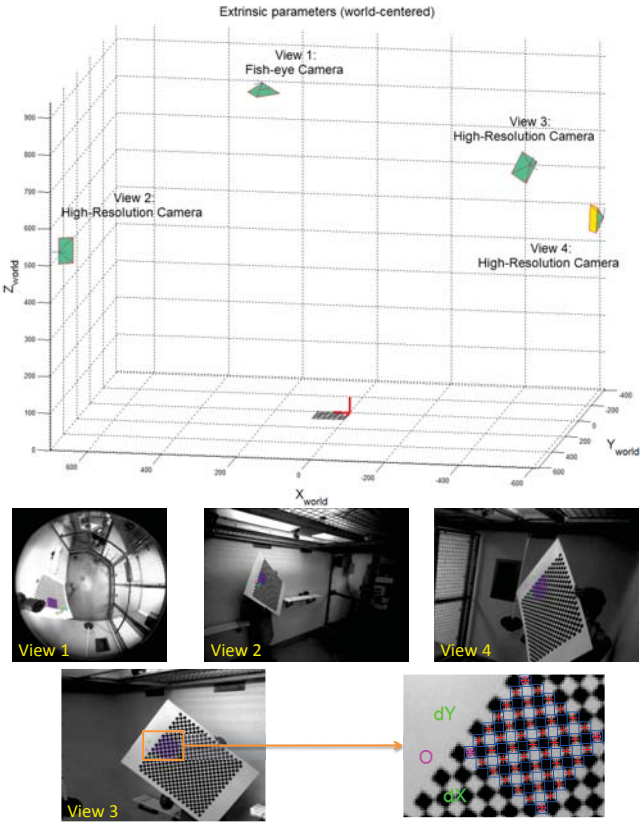

Fig. 4. Calibration process. A checkerboard of size $16.8'' \times 24''$ is used for calibration. The top figure shows the 3D locations of each camera.

by intersecting the visual cones corresponding to silhouettes captured from multiple views. The visual hull of a 3D object S is the maximal volume consistent with silhouettes of S. A formal definition of Visual Hull (VH) is first introduced by Laurentini [11] as following:

> "The visual hull $VH(S, R)$ of an object S relative to a viewing region R is a region of E^3 such that, for each point $P \in VH(S, R)$ and each viewpoint $V \in R$, the half-line starting at V and passing through P contains at least a point of S."[11]

Directly building visual hulls by intersecting the visual cones is very difficult in practice due to the curved and irregular surface of objects, which results in a complex geometrical representation for its cones. Therefore approximation methods are required. Polyhedral shape based approach [12] and volume based approach [14] are normally used for this purpose. We adopt the latter approach for its efficiency. Algorithm 1 shows a pseudo code of the approach.

Algorithm 1 Visual hulls approximation from K views

1. Divide the 3D space of interest into $N \times N \times N$ discrete voxels $v_n, n = 1, .., N^3$. voxels
2. Initialize all the N^3 voxels as object voxels
3. For $n = 1$ to N^3 {
   — For $k = 1$ to K {
   —— Project v_n into the k^{th} image plane by the projection function P^k;
   —— If the projected area $P^k(v_n)$ lies completely outside S^k, then classify v_n as non-object voxels;
   — }
   }
4. The visual hull VH is approximated by the union of all the object voxels.

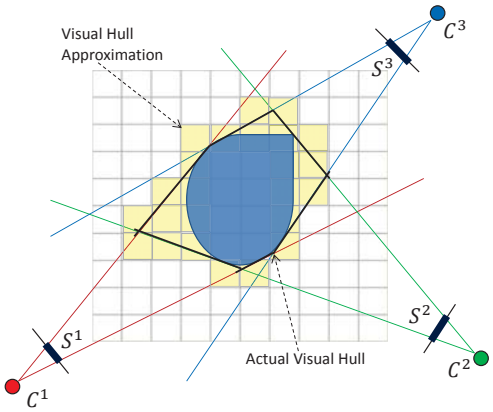

Fig. 5. 2D example of the visual hull approximation algorithm. C^1, C^2, C^3 are different views with corresponding silhouettes S^1, S^2, S^3. The yellow area is the approximation of the visual hull; the area enclosed by black lines is the actual visual hull; and the blue shape in the center is the object.

4. EXPERIMENTAL EVALUATION

4.1. 2D primate detection results

Figure 6, Figure 7, and Table 1 show our results on 2D primate detection. The challenge of detection comes from multiple factors. Firstly, due to the settings of the environment, the illumination varied in different locations, furthermore, it may change from time to time, too. So we cannot simply rely on background subtraction or illumination-sensitive features. Secondly, although the primates wear collars of different color, these are easily occluded when they move, or become indistinguishable when the illumination is low. Therefore, the main feature we used to detect primates is HOG.

The main challenge to detect primates with HOG feature is the variable shape of the primate body. The reason that

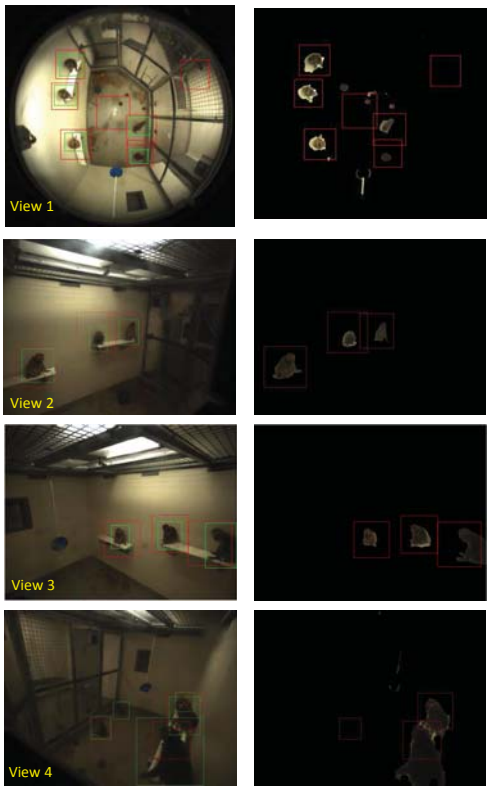

Fig. 6. Primate detection in 2D. In column one, green boxes are the ground truth; red boxes are the detection results. Column two shows the extracted silhouettes by background substraction over detected bounding boxes.

HOG can successfully detect pedestrians, for instance, is that the contours of all standing human beings look similar. The ratio between width and height is almost constant.. However, the contour of a crouching monkey is quite different from that of a jumping one.

For each view we trained a separate detector. We used a test video to evaluate the detector's performance. The results are shown in Table 1. TP stands for true positive, FP stands for false positive and FN stands for false negative. The PR curve in Figure 7 shows the relation between precision and recall rate with SVM threshold varied. From Figure 7, we can see that View 2 and View 3 are better than View 1 and View 4. It is reasonable because in View 2 and View 3 the background is simple and the monkeys are usually separated. In View 1, the background is strongly cluttered so there are many false positives. In View 4, the monkeys on the benches often occlude each other and the illuminance is low on the floor area, so it is difficult to locate monkeys and therefore many false negatives occur. Figure 6 is a good illustration for these points.

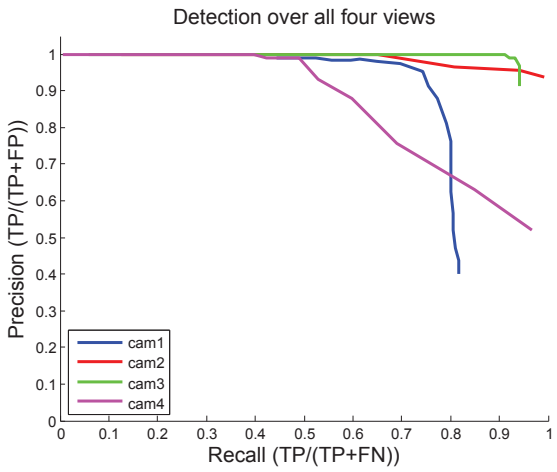

Fig. 7. PR-curve of 2D detection.

Table 1. 2D Primate Detection Results from 4 Views

Cameras	TP	FP	FN	Precision	Recall
View 1	180	79	45	0.70	0.80
View 2	98	29	37	0.77	0.73
View 3	93	9	9	0.91	0.91
View 4	129	11	77	0.92	0.63
Overall	**500**	**128**	**168**	**0.80**	**0.75**

4.2. 3D primate visual hull results

After detecting the primates, we used the detection results to reconstruct the 3D visual hulls of the primates in the cage. For each view, we have a detection log which gives us the bounding boxes around primates; combining the detection results and the foregrounds obtained from background subtraction technique, we can get a better estimate of the location and shape of primates in 2D. For each frame, we created a binarized image with primates as foreground and the rest as background in each view. Finally we used these images to create 3D visual hulls of the primates. However, since the number of cameras is limited, having a profound shape after 3D reconstruction is not possible; therefore the final shape of 3D primate reconstructions will not be accurate. On the other hand, it gives us the position of each primate at each frame, which later can be used for social behavior study of primates . Figure 9 illustrates this process for one primate in one frame. Due to lack of space, we cannot present a series of visual hull images.

5. CONCLUSION

In this paper, we presented a method to detect primates in the cage and build a 3D visual hull for each primate in the cage. Having this information, we were able to identify each pri-

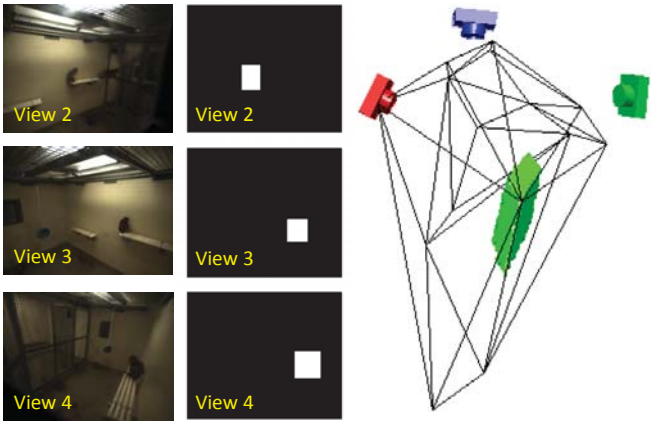

Fig. 8. 3D visual hull reconstruction result sample. Column one are the original images; Column two shows the binary images from 2D primate detection; Column three is the visual hull constructed from three views.

mate in 3D. Aside from originality of the problem, this study has a lot of challenges and we were able to overcome some in this work. The size of data is enormous and requires the most optimized techniques to reduce processing time. The environment is dark with a lot of illumination changes and colors of the collars are not visible or distinguishable in several time periods. There are many moving objects apart from the monkeys in the environment (people, swing, toys) For our future work, we will extend our detection to tracking primates in 2D and 3D,and further, using the 3D visual hull reconstruction in conjunction with 3D tracking results, we can extract activities and relative social behaviors of primates.

6. REFERENCES

[1] Robert A Hinde, *Growing Points Ethology*, CUP Archive, 1976.

[2] Zia Khan, Rebecca A Herman, Kim Wallen, and Tucker Balch, "An outdoor 3-d visual tracking system for the study of spatial navigation and memory in rhesus monkeys," *Behavior research methods*, vol. 37, no. 3, pp. 453–463, 2005.

[3] Fabrice de Chaumont, Renata Dos-Santos Coura, Pierre Serreau, Arnaud Cressant, Jonathan Chabout, Sylvie Granon, and Jean-Christophe Olivo-Marin, "Computerized video analysis of social interactions in mice," *Nature Methods*, vol. 9, no. 4, pp. 410–417, 2012.

[4] Tucker Balch, Frank Dellaert, Adam Feldman, Andrew Guillory, Charles L Isbell, Zia Khan, Stephen C Pratt, Andrew N Stein, and Hank Wilde, "How multirobot systems research will accelerate our understanding of social animal behavior," *Proceedings of the IEEE*, vol. 94, no. 7, pp. 1445–1463, 2006.

[5] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 1, pp. 886–893.

[6] Chih-Chung Chang and Chih-Jen Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 27, 2011.

[7] Wojciech Matusik, Chris Buehler, Leonard McMillan, and Steven J Gortler, "An efficient visual hull computation algorithm," Tech. Rep., MIT LCS Technical Memo 623, MIT Laboratory for Computer Science, Cambridge, MA 02141, 2002.

[8] Richard Szeliski, "Shape from rotation," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*. IEEE, 1991, pp. 625–631.

[9] Wojciech Matusik, Chris Buehler, Ramesh Raskar, Steven J Gortler, and Leonard McMillan, "Image-based visual hulls," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 2000, pp. 369–374.

[10] Janne Heikkila and Olli Silven, "A four-step camera calibration procedure with implicit image correction," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*. IEEE, 1997, pp. 1106–1112.

[11] Aldo Laurentini, "The visual hull concept for silhouette-based image understanding," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, no. 2, pp. 150–162, 1994.

[12] Jean-Sébastien Franco, Edmond Boyer, et al., "Exact polyhedral visual hulls," in *British Machine Vision Conference (BMVC'03)*, 2003, vol. 1, pp. 329–338.

[13] Keith Forbes, Anthon Voigt, Ndimi Bodika, et al., "Visual hulls from single uncalibrated snapshots using two planar mirrors," in *Proc. 15th Annual Symposium of the Pattern Recognition Association of South Africa*, 2004.

[14] Hiroshi Noborio, Shozo Fukuda, and Suguru Arimoto, "Construction of the octree approximating a three-dimensional object by using multiple views," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 10, no. 6, pp. 769–782, 1988.