Tracking the Motion of Box Jellyfish

Tobias Kjellberg Magnus Oskarsson Tobias Palmér Kalle Åström
Centre for Mathematical Sciences, Lund University, Sweden
{magnuso, tobiasp, kalle}@maths.lth.se

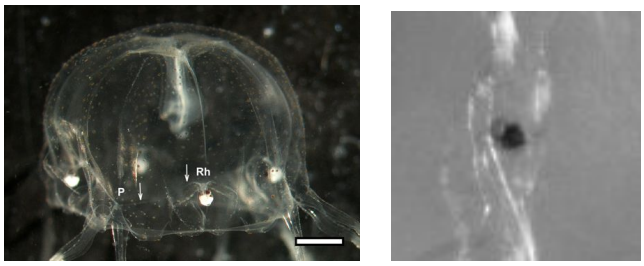


Fig. 1. Left: The box jellyfish *tripedalia cystophora* is only a couple of mm large and almost completely transparent. Right: A close-up of the rhopalia from one frame recorded in the experimental setup.

***Abstract*—In this paper we investigate a system for tracking the motion of box jellyfish *tripedalia cystophora* in a special test setup. The goal is to measure the motor response of the animal given certain visual stimuli. The approach is based on tracking the special sensory structures – the rhopalia – of the box jellyfish from high-speed video sequences. We have focused on a real-time system with simple building blocks in our system. However, using a combination of simple intensity based detection and model based tracking we achieve promising tracking results with up to 95% accuracy.**

I. INTRODUCTION

Box jellyfish, or cubomedusae, have very special visual systems, [1], [2]. The visual system is based on four identical sensory structures, which are called rhopalia. Each rhopalia consists of six different eyes: the upper and lower lens eyes, the pit eyes, and the slit eyes [3]–[8]. The lens eyes have image-forming optics and resemble vertebrate and cephalopod eyes [9]–[11].

The role of vision in box jellyfish is known to involve phototaxis, obstacle avoidance, control of swim-pulse rate and advanced visually guided behaviour [12]–[14]. Most known species of box jellyfish are found in shallow water habitats where obstacles are abundant [15]. Medusae of the studied species, *T. cystophora*, live between the prop roots in Caribbean mangrove swamps [16], [17]. They stay close to the surface [16] where they prey on a phototactic copepod that gathers in high densities in the light shafts formed by openings in the mangrove canopy. The medusae are not found in the open lagoons, where they risk starvation [12]. The overall goal of this project is to learn more about the visual system and the neural processes involved. One interesting problem is the study of the connection between visual stimuli presented to the animals and the motor response of the animal. However, tracking free-swimming jellyfish poses a very demanding tracking problem. Instead we look at a special experimental setup where the animals are tethered whilst they are submitted to different light stimuli. The goal is then to track how the animals direct themselves, i.e. how they would like to move. In order to do this, an initial goal which we discuss in this paper is how to track the four rhopalia. These appear as four dark discs, situated on the perimeter of the bell, see figure 2. There have been a number of previous studies on motor response from controlled visual stimuli, see e.g. [18], [19].

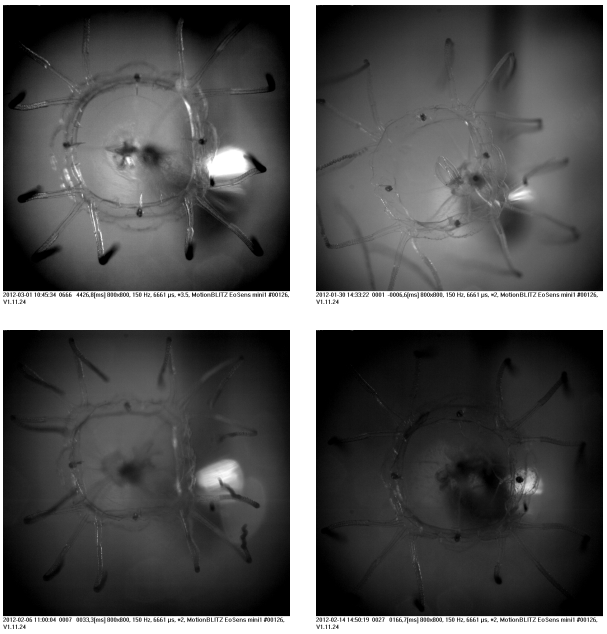


Fig. 2. Example input frames from a number of different sequences. Notice the high variance in lighting conditions. In some frames the rhopalia are barely discernible and in many frames there are structures that have an appearance very similar to the rhopalia.

II. EXPERIMENTAL SETUP

In this study we used in total 33 animals, with sizes ranging from 0.43 to 0.89 cm. The animals were tethered by the top of the bell during the experiments, using a glass pipette with gentle suction, and placed in a Plexiglas tank with inside dimensions of $5 \times 5 \times 5$ cm. The vertical walls of the tank were covered with diffusing paper and a neutral density filter. Each vertical wall was illuminated from the outside by four blue-green LEDs. The diffuser was used to make a plane light source, while the neutral density filter was used to increase the contrast between lit and dark panels and switching one or more panels off was used as the behavioural trigger. The colour of the LEDs matched the maximum spectral sensitivity of the animals and had a peak emission at 500 nm. During the experiments a box was placed over the set-up in order to eliminate visual cues coming from outside. Image sequences

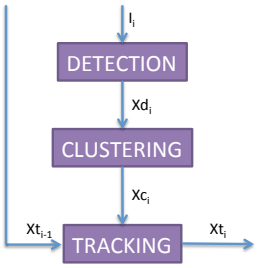


Fig. 3. An overview of the system. For each input frame I_i we run the detection algorithm. This produces a number of tentative points Xd_i. This set of points is then sent to the clustering algorithm, which then outputs a smaller number of refined positions Xc_i. These points are fed into the tracking algorithm, alongside the four point positions from the previous frame, Xt_{i-1}. The final output is then the four detected points Xt_i.

were recorded with a high-speed camera operated at 150 frames per second. The dataset consists of 15 video sequences, each with around 100 greyscale frames. Each greyscale frame has a resolution of 800 $\times$ 864 pixels. Depending on how the jellyfish is set up, the light and shadows form differently which will make the video sequences different from each other, see figure 2. Even though great measure has been done in order to minimize artefacts in the film sequences the difference between them can be quite large. Some of the film sequences are brighter, making it easier to find the rhopalia while some are darker and thus making it hard to distinguish the rhopalia from the background. The tethering of the animal is also visible in all sequences. This tether-shadow causes problem when the jellyfish is contracting and the rhopalia is moving over the tether-shadow. Since the box jellyfish is moving in every video sequence some parts of the jellyfish are moving in and out of focus. The physical nature of the jellyfish, i.e. being transparent, also affects the appearance and causes refraction on the surroundings.

The rhopalia in each frame in each video sequence has been manually annotated in order to perform evaluation and testing of algorithms.

III. SYSTEM OVERVIEW

In this section we will describe our system. Since the focus is on real-time, the chosen building blocks are quite simple in nature, especially the first steps. In figure 3 an overview of the system is shown. We have divided it into three parts. For every frame i, a first detection step – which only uses the local intensity distribution around a point – produces a large number of detections, Xd_i. These points are then clustered into a set of points Xc_i, in order to remove multiple detections, as well as for improved positional accuracy in the detected points. Finally these clustered positions are sent to the tracking step, which also gets as input the previous frame's four detected points, Xt_{i-1}. The output is the four detected points Xt_i. In the following subsections we will describe the different steps in more detail.

A. Detection

The rhopalia appear as dark discs on a brighter background in the images that are quite consistent in size and appearance in the images. See figure 4 for a close-up view. For this reason we have tested a number of template based approaches for detection. The template is based on the assumption that we have a number of pixels near the rhopalia that are dark. Further outside the rhopalia we should have pixels that are brighter. Figure 4 shows some example templates that we have tried. For speed we have adopted quite sparse templates. Top row shows pixels that should be inside the rhopalia, and hence should be darker. Bottom row shows the pixels that are assumed to be outside the rhopalia. For each point $X_i(j)$ in the input image I_i, we can then define a number of inside and outside points, $\Omega_{in}(j)$ and $\Omega_{out}(j)$. Examples of these point sets can be seen in figure 4. We have looked at two types of measures, one absolute and one relative. For the absolute measure we define a threshold for the inner pixels, t_{in} and one for the outer pixels, t_{out}. We then count the number of inside and outside pixels that fulfil the constraints, i.e.

$$N_{abs}(j) = \sum \Gamma(\Omega_{in}(j) \leq t_{in}) + \sum \Gamma(\Omega_{out}(j) \geq t_{out}), \quad (1)$$

where $\Gamma(x) = 1$ if x is true and zero otherwise, and

$$Xd_i = \{X_i(j) | N_{abs}(j) > N_{det}\}, \quad (2)$$

where N_{det} is some bound.

For the relative measure we randomly compare n inside and outside pixels, and count how many of the inside pixels are darker than the outside pixels. So if we let $R(\Omega)$ denote a function that randomly chooses a point from the set Ω we have,

$$N_{rel}(j) = \sum_{k=1}^{n} \Gamma(R(\Omega_{in}(j)) < R(\Omega_{out}(j)), \quad (3)$$

and

$$Xd_i = \{X_i(j) | N_{rel}(j) > N_{det}\}. \quad (4)$$

We have evaluated our whole system in order to find templates that generate enough possible detections, i.e. that in most cases at least generates the four correct points, but that do not generate excessive amounts of false positives. In figure 5 a typical detection result is shown. Since there are many rhopalia-like structures in the images we get a large number of false positives, but these will be eliminated in the following clustering and tracking steps.

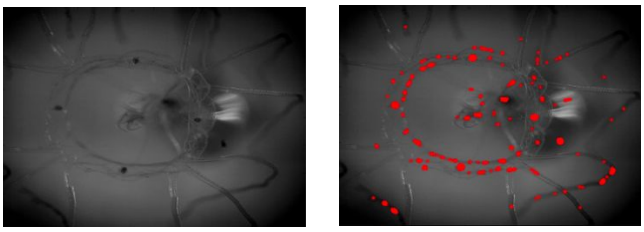


Fig. 5. A typical detection result is shown. Since there are many rhopalia-like structures in the images we get a large number of false positives, but these will be eliminated in the following clustering and tracking steps.

B. Clustering

We take a scale space approach for clustering the detections; by smoothing the input image I_i isotropically with a Gaussian kernel we get a low scale version I_{sm}. We then

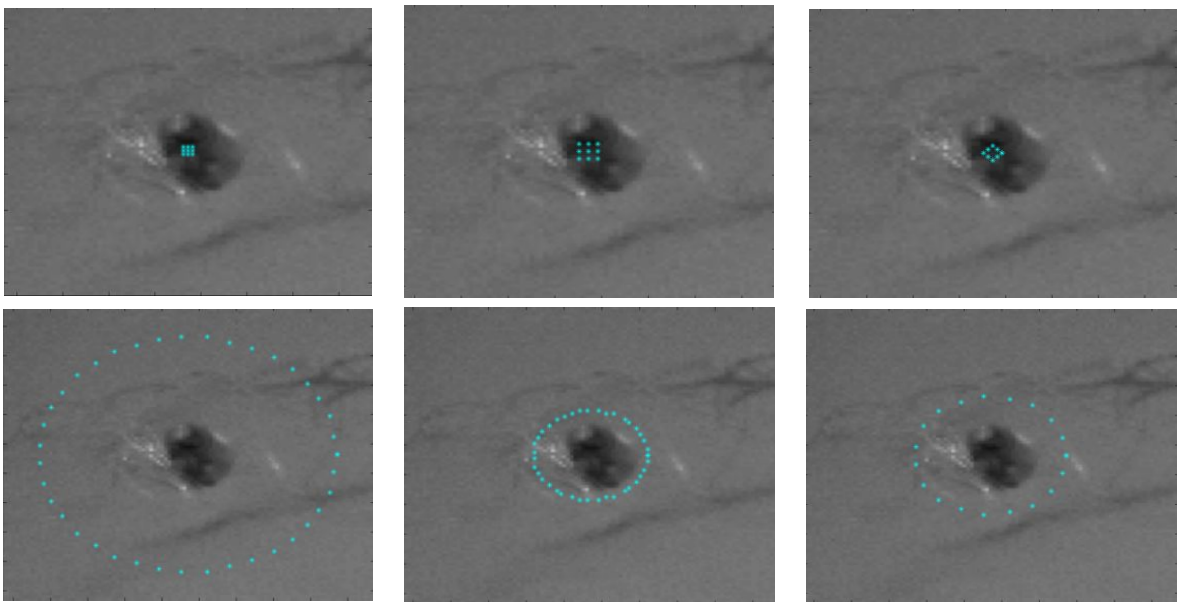


Fig. 4. A number of templates for the detection step. Top row shows pixels that should be inside the rhopalia, and hence should be darker. Bottom row shows the pixels that are assumed to be outside the rhopalia.

find all local minima X_{loc} of I_{sm}. The reason for this is the appearance of the rhopalia as dark spots in the images. We then calculate how many detections we have within a vicinity of each X_{loc},

$$N_{loc}(j) = \sum_{k=1}^{N_d} \Gamma(||X_{loc}(j) - Xd(k)||_2 < \epsilon_{cluster}), \quad (5)$$

and if there are a minimum number N_{min} of detections, then we add this local minimum to our clustered points Xc_i,

$$Xc_i = \{X_{loc}(j) | N_{loc}(j) \geq N_{min}\}. \quad (6)$$

This gives a fast, accurate and quite robust way of clustering the detections. See figure 6b for an example result of the clustering.

C. Tracking

For the tracking step we have looked at a number of simple algorithms. The input consists of the four points from the previous frame, Xt_{i-1} and a number of possible candidate points Xc_i. In our model we could also look at the scenario where the general statistical distribution of the four points is given in some form. Arguably the most simple tracking is just choosing as new points, the four closest points from the candidate points to the points from the previous frame, i.e.

$$Xt_i(j) = \arg \min_{Xc_i(k)} ||Xt_{i-1}(j) - Xc_i(k)||_2, \quad j = 1, \ldots, 4. \quad (7)$$

We have also looked at learning an affine shape model of the four points. In this case we can fix a canonical coordinate system by placing the first three points at $(0,0), (0,1)$ and $(1,0)$. The final points will then be placed at some point X_a. Using a large number of ground truth positions we can estimate the mean X_m and covariance matrix Σ for X_a. This gives us a way of finding the four points that statistically most resemble a four-point configuration, given affine deformations. For all possible subsets of four points Y_i of Xd_i we change coordinate system so that the first three points are at $(0,0), (0,1)$ and $(1,0)$ and the fourth point at Y_a. If there are n such subsets we find the best subset by

$$k_{opt} = \arg \min_k (Y_a(k) - X_m)^T \Sigma^{-1} (Y_a(k) - X_m), \quad (8)$$

and

$$Xt_i = Y_i(k_{opt}). \quad (9)$$

For both equation (7) and (8) we can choose $Xt_i = Xt_{i-1}$ if the optimal value is to large, i.e. if the best clustered points are too far away from the previous frame's points we choose the previous frame's point positions for the new frame. See figure 6c for an example result of the tracking.

We have not yet focused on the tracking, and more complex motion models are of course possible and quite easy to implement into the framework, such as e.g. Kalman filters [20] or Particle filters [21].

IV. EXPERIMENTAL RESULTS

In total we have in our test setup 1469 frames, and we have manually marked the true coordinates of the rhopalia in each frame. We have tested our system on the video sequences and compared it to ground truth in the following way. For the detection accuracy we count for each of the four rhopalia in the images if there are N or more detections within a circle with radius of 10 pixels around the true coordinate. We have used $N = 10$ and $N = 20$. For the cluster accuracy we have counted the percentage of rhopalia that have a cluster within a circle of 10 pixels around the correct coordinate, and likewise for the tracking accuracy we count the percentage of rhopalia with a tracked point within 10 pixels. We do this for all the 1469 frames, with four rhopalia in each frame. We have mainly tested different detection parameters and settings. In figure 7 the resulting accuracy percentages can be seen. We see that

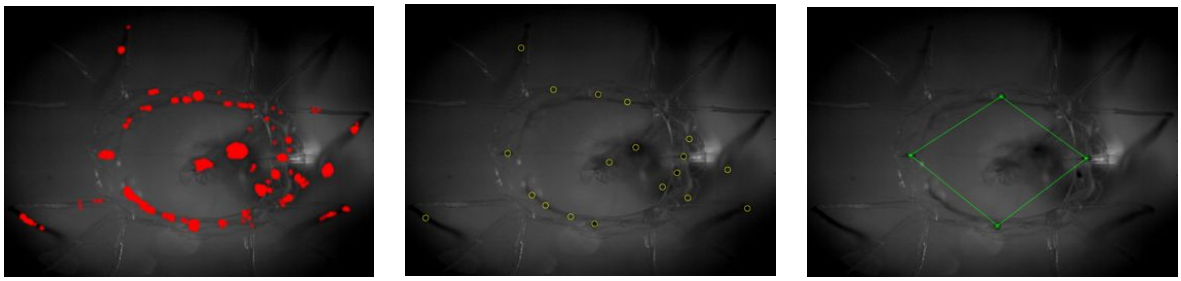
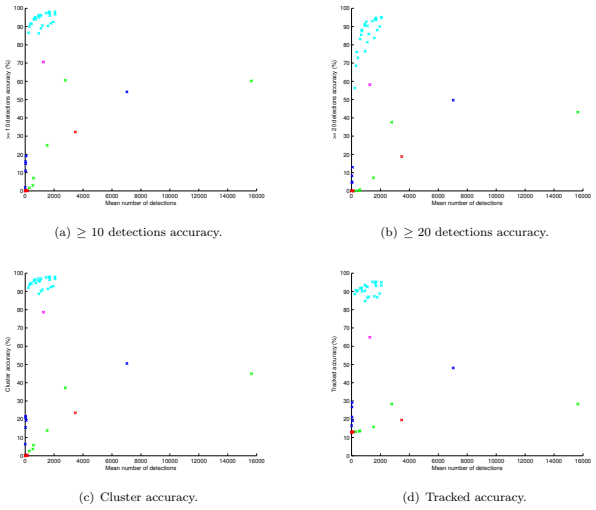


Fig. 6. The figure shows the output of the three steps of our system for an example input image. Left all the detected points are shown in red. These are fed into the clustering which outputs the resulting yellow clustered points in the middle. Finally the tracking outputs the four points to the right, depicted in green.



(a) $\geq$ 10 detections accuracy.

(b) $\geq$ 20 detections accuracy.

(c) Cluster accuracy.

(d) Tracked accuracy.

Fig. 7. Evaluation of a number of different detection parameters and settings. The accuracy for the different steps in the system is shown. The best performing systems have a tracking accuracy of 95%.

the best performing systems have a tracking accuracy of 95%.

V. Conclusions

We have in this paper investigated how a system for detection of the special eyes – the rhopalia – of box jellyfish can be constructed. We have shown that using a low-level detection method in combination with clustering and tracking we can get very good performance on a varying dataset. The system can be run in real-time. The basic idea is that the system should be used in order to learn more about the visual system and neural processes of box jellyfish. The next step would be to, from the tracking of the rhopalia, measure the motion of the whole bell of the jellyfish in order to measure the motor response of the animal given certain visual stimuli.

References

[1] E. W. Berger, "The histological structure of the eyes of cubomedusae," *Journal of Comparative Neurology*, vol. 8, no. 3, pp. 223–230, 1898.

[2] C. Claus, "Untersuchungen uber charybdea marsupialis," *Arb. Zool. Inst. Wien, Hft*, vol. 2, pp. 221–276, 1878.

[3] J. S. Pearse and V. B. Pearse, "Vision of cubomedusan jellyfishes." *Science (New York, NY)*, vol. 199, no. 4327, pp. 458–458, 1978.

[4] G. Laska and M. Hündgen, "Morphologie und ultrastruktur der lichtsinnesorgane von tripedalia cystophora conant (cnidaria, cubozoa)," *Zool Jb Anat*, vol. 108, pp. 107–123, 1982.

[5] A. Garm, F. Andersson, and D.-E. Nilsson, "Unique structure and optics of the lesser eyes of the box jellyfish tripedalia cystophora," *Vision research*, vol. 48, no. 8, pp. 1061–1073, 2008.

[6] J. Piatigorsky and Z. Kozmik, "Cubozoan jellyfish: an evo/devo model for eyes and other sensory systems," *Int J Dev Biol*, vol. 48, no. 8-9, pp. 719–729, 2004.

[7] Z. Kozmik, S. K. Swamynathan, J. Ruzickova, K. Jonasova, V. Paces, C. Vlcek, and J. Piatigorsky, "Cubozoan crystallins: evidence for convergent evolution of pax regulatory sequences," *Evolution & development*, vol. 10, no. 1, pp. 52–61, 2008.

[8] F. S. Conant, *The Cubomedusæ: a memorial volume*. The Johns Hopkins press, 1898, vol. 4, no. 1.

[9] D.-E. Nilsson, L. Gislén, M. M. Coates, C. Skogh, and A. Garm, "Advanced optics in a jellyfish eye," *Nature*, vol. 435, no. 7039, pp. 201–205, 2005.

[10] V. J. Martin, "Photoreceptors of cubozoan jellyfish," in *Coelenterate Biology 2003*. Springer, 2004, pp. 135–144.

[11] M. Koyanagi, K. Takano, H. Tsukamoto, K. Ohtsu, F. Tokunaga, and A. Terakita, "Jellyfish vision starts with camp signaling mediated by opsin-gs cascade," *Proceedings of the National Academy of Sciences*, vol. 105, no. 40, pp. 15 576–15 580, 2008.

[12] E. Buskey, "Behavioral adaptations of the cubozoan medusa tripedalia cystophora for feeding on copepod (dioithona oculata) swarms," *Marine Biology*, vol. 142, no. 2, pp. 225–232, 2003.

[13] A. Garm and S. Mori, "Multiple photoreceptor systems control the swim pacemaker activity in box jellyfish," *The Journal of experimental biology*, vol. 212, no. 24, pp. 3951–3960, 2009.

[14] A. Garm, M. Oskarsson, and D.-E. Nilsson, "Box jellyfish use terrestrial visual cues for navigation," *Current Biology*, vol. 21, no. 9, pp. 798–803, 2011.

[15] M. M. Coates, "Visual ecology and functional morphology of cubozoa (cnidaria)," *Integrative and comparative biology*, vol. 43, no. 4, pp. 542–548, 2003.

[16] S. E. Stewart, "Field behavior of tripedalia cystophora (class cubozoa)," *Marine & Freshwater Behaviour & Phy*, vol. 27, no. 2-3, pp. 175–188, 1996.

[17] B. Werner, C. E. CUTRESS, and J. P. STUDEBAKER, "Life cycle of tripedalia cystophora conant (cubomedusae)," 1971.

[18] R. Petie, A. Garm, and D.-E. Nilsson, "Velarium control and visual steering in box jellyfish," *Journal of Comparative Physiology A*, vol. 199, no. 4, pp. 315–324, 2013.

[19] ——, "Contrast and rate of light intensity decrease control directional swimming in the box jellyfish tripedalia cystophora (cnidaria, cubomedusae)," *Hydrobiologia*, vol. 703, no. 1, pp. 69–77, 2013.

[20] G. Welch and G. Bishop, "An introduction to the kalman filter," 1995.

[21] N. J. Gordon, D. J. Salmond, and A. F. Smith, "Novel approach to nonlinear/non-gaussian bayesian state estimation," in *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, no. 2. IET, 1993, pp. 107–113.