PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Real time lobster posture estimation for behavior research

Sheng Yan, Jo Arve Alfredsen

Sheng Yan, Jo Arve Alfredsen, "Real time lobster posture estimation for behavior research," Proc. SPIE 10225, Eighth International Conference on Graphic and Image Processing (ICGIP 2016), 102250F (8 February 2017); doi: 10.1117/12.2266430

SPIE.

Event: Eighth International Conference on Graphic and Image Processing, 2016, Tokyo, Japan

Real Time Lobster Posture Estimation For Behavior Research

Sheng Yan Jo Arve Alfredsen

Norwegian University of Science and Technology

ABSTRACT

In animal behavior research, the main task of observing the behavior of an animal is usually done manually. The measurement of the trajectory of an animal and its real-time posture description is often omitted due to the lack of automatic computer vision tools. Even though there are many publications for pose estimation, few are efficient enough to apply in real-time or can be used without the machine learning algorithm to train a classifier from mass samples. In this paper, we propose a novel strategy for the real-time lobster posture estimation to overcome those difficulties. In our proposed algorithm, we use the Gaussian mixture model (GMM) for lobster segmentation. Then the posture estimation is based on the distance transform and skeleton calculated from the segmentation. We tested the algorithm on a serials lobster videos in different size and lighting conditions. The results show that our proposed algorithm is efficient and robust under various conditions.

Keywords: Animal behavior, lobster, skeleton, posture estimation

1. INTRODUCTION

The attempt to design and fine tune the computer vision algorithms to assist the animal behavior research is quite few. However, the measurement of the trajectory of the animal and its quantitative posture description are demanding in animal behavior research. Many researchers just manually determine the type of the behavior and only record the span of time for their research [1]. The useful data, such as the posture along the trajectory is omitted. The main contribution of our lobster posture estimation algorithm is that it can be a strong tool to boost the existing behavior research approaches and makes the automatic identification of the type of behavior pattern possible, which subsequently could be dramatically decrease the manually work for the animal behavior researcher. Furthermore, the quantitative posture description of the lobster in each frame could open the new possibility to apply the newly developed data analysis and machine learning algorithms to the animal behavior research.

To fulfill such task, our algorithm is tailored to meet the requirement in three aspects. First is the simplicity for setup. Many algorithms are using supervised or unsupervised learning algorithm to determine each parts of the animal for further processing. The work for collecting representative learning data itself is very time consuming. On contrary, our algorithm just take one sample of animal color or predefined color. Second is the effectiveness. It is meaningless if the video recording for an animal is ten minutes but need hours to compute. Our algorithm can work in real-time. Finally, our algorithm is robust for the varying lighting condition and the various size of lobsters in different biological phases.

Due to the vast applications, the object detection and posture estimation has drawn much attention in the realm of computer vision recently. The whole recognition system can be divided into two parts. The first part is the texture or feature description. Among the abundant publications, the Normalized Histogram of Oriented Gradient (HOG) descriptors [2] is outperforms many other descriptors in object recognition tasks for its simplicity and the tolerance of the variation of the object. However, the orientation is always assumed to be up-right which limits the application to animal behavior research where the object can be in any orientation. More recently, the work of [3] shows high speed for detection up to 100 frames per second but it is only specified for the pedestrian detection.

The other part of the detection system is supervised learning algorithm. To recognize a specific object such as human pedestrian, the animal or the vehicle, the positive images contain the training object and negative images without the objects are fed to the learning algorithm such as Adaboost [4], support vector machine , randomized decision trees and forests [5] or Neural network.

Eighth International Conference on Graphic and Image Processing (ICGIP 2016), edited by Tuan D. Pham, Vit Vozenilek, Zhu Zeng, Proc. of SPIE Vol. 10225, 102250F · © 2017 SPIE · CCC code: 0277-786X/17/$18 · doi: 10.1117/12.2266430

However, all of the algorithms discussed above involve the procedure of sample data collecting and learning. It is tedious, time consuming but skilled work to collect all the learning samples with manually posture labels which limits its usage in the animal behavior research. Therefore, our proposed algorithm comes into play to address the problem to make the posture estimation simple and effective.

2. ALGORITHM DESCRIPTION

Because the algorithm is aimed at applying to the laboratory environment, the distribution of input image pixels in RGB space from the video recording can be considered as invariant over time. The GMM model is calculated from the first image of the video and is applied to all the following images. Two following steps to extract the lobster pose are based on the segmented image. Details of the implementation are described in the following subsections.

2.1 GMM for lobster detection and segmentation

The input image I consists of n pixels $\{\boldsymbol{x_i} | \boldsymbol{x_i} \in \boldsymbol{R^3}\}_{i=1}^{n}$ which are presented in RGB color space. Unlike the model that is used to estimate image alpha [6], which consists of two GMMs, one for the background and the other one for the foreground, we model the pixels $\{\boldsymbol{x_i}\}$ in the whole image with only one GMM and alleviate the need for the initial estimated foreground and background as input. Therefore, the input image pixels are written as

$$\boldsymbol{Pr(x_i|\theta)} = \sum_{k=1}^{K} \boldsymbol{\lambda_k} \mathbf{Norm}_{\boldsymbol{x}}(\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k}) \tag{1}$$

where $\mathbf{Norm}_{\boldsymbol{x}}$ is the multivariate Gaussian with $\boldsymbol{x_i}$ as variable. $\boldsymbol{\mu_{1...K}}$ are the means of normal distribution and their corresponding covariance matrices $\boldsymbol{\Sigma_{1...K}}$ are taken to be full-covariance. Each $\boldsymbol{\lambda_k}$ is the weight coefficient and sums to one.

To fit the parameter $\boldsymbol{\theta} = \{\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k}, \boldsymbol{\lambda_k}\}_{k=1}^{K}$ from the whole image pixels $\{\boldsymbol{x_i}\}_{i=1}^{n}$, the expectation maximization algorithm is used [7]. After the parameter fitting, we could find which Gaussian distribution in GMM model is most suitable to present the lobster in RGB space by measuring the Mahalanobis distance from the selected color to the mean of each Gaussian distribution in GMM .

$$\boldsymbol{L} = \underset{\boldsymbol{k}}{\mathbf{argmin}} (\boldsymbol{x_l} - \boldsymbol{\mu_k})^{\boldsymbol{T}} \boldsymbol{\Sigma_k^{-1}} (\boldsymbol{x_l} - \boldsymbol{\mu_k}) \tag{2}$$

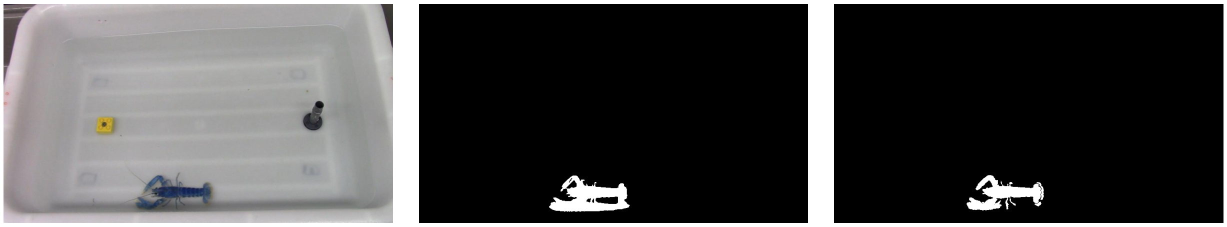
The $\boldsymbol{x_l}$ is predefined RGB color vector presenting the color of the whole body. This value could be easily obtained by sampling one pixel from the lobster body in the given image unless it is the outlier when considering the distribution of the body color of the lobster as a Gaussian distribution. $\boldsymbol{L}$ is the index of the Gaussian distribution within GMM which presents the lobster color distribution.

Once the model setup for RGB color space is finished, we could use this model for the given image and for the following image from the video stream obtained with the laboratory settings, at which the background and the object is under a very limited variation and we could safely regard the GMM model remains invariant over time. Assume that the segmentation represents the lobster consists of $\boldsymbol{m}$ points and denotes by $\boldsymbol{P} = \{\boldsymbol{p_i}\}_{i=1}^{m}$, we have

$$\boldsymbol{P} = \{\boldsymbol{p_i} | \underset{\boldsymbol{k}}{\mathbf{argmax}}\, \boldsymbol{\lambda_k} \mathbf{Norm}_{\boldsymbol{I(p_i)}}(\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k}) = \boldsymbol{L}\} \tag{3}$$

Here, different from the $\boldsymbol{x_i}$ in equation (1) denoting the color, the point $\boldsymbol{p_i}$ is a two element vector denotes the position of the pixel.

Since it is possible that some other objects with almost the same color as lobster or the noisy pixel happened to be with similar color, further process such as connected component could be applied to tackle with this problem. Because the segmented point set belongs to lobster should be connected together and should be with considerably large cardinality.



(a) input image (b) K-means segmentation (c) proposed GMM segmentation

Figure 1. Comparison of three lobsster segmentaion methods.

Figure 1 illustrates the effectiveness of the proposed GMM segmentation method. The input image is a challenging case in which the lobster is walking near the brim of the water tank and the shadow casted on the lobster body caused many algorithm fail to segment out the lobster. Figure 1(b) is obtained by K-means clustering algorithm. In this case, the shadow is segmented as part of the lobster which would cause the posture estimation unreliable . As can be seen in Figure 1(c), the proposed method is robust to pick the right color region to segment out the lobster with shadow.

2.2 Morphological Processing

Because the GMM segmentation is only considered in the RGB space without the pixel position information in the image, it is possible that there are small holes in the binary image of the lobster body caused by the noise. The holes are small and could be filled by checking the majority of the pixels around and set the corresponding pixels to one if the majority condition holds.

$$\boldsymbol{I(x)} = \mathbf{T}[\sum_{\boldsymbol{y}\in \boldsymbol{N(x)}} (\mathbf{T}[\boldsymbol{I(y)} = \mathbf{1}] - \mathbf{T}[\boldsymbol{I(y)} = \mathbf{0}]) > \mathbf{0}] \tag{4}$$

where $\boldsymbol{I}$ is the binary image and $\boldsymbol{I(x)}$ is the value at image position $\boldsymbol{x} \in \boldsymbol{R^2}$. $\boldsymbol{N(x)}$ denotes the neighbor points set of $\boldsymbol{x}$. $\boldsymbol{T}[\cdot]$ equals one if the expression inside the parenthesis is true and zero otherwise.

After the holes are filled, the distance transform [8] and the skeleton extraction are performed in parallel. The skeleton extraction is based on the algorithm described in [9]. The skeleton is obtained by applying the thinning operation till its convergence. We denote the skeleton points set as $\boldsymbol{S = \{s_i\}}$

2.3 Posture Estimation

The posture estimation is based on the skeleton and the distance transform from the last step and the aim is to estimate five line segments and three joints to present the claws, arms and the body of the lobster.

The estimation of each part is in the order of body, arm and claw. The joint point is regarded as a constraint for the line segment searching in the following step. First of all, we need to find the central point of the lobster body. The idea is that we first assume the centroid point of the distance transform is the central point as our first guess and then we could refine this guess after we find a line segment which fits the main body of the lobster.

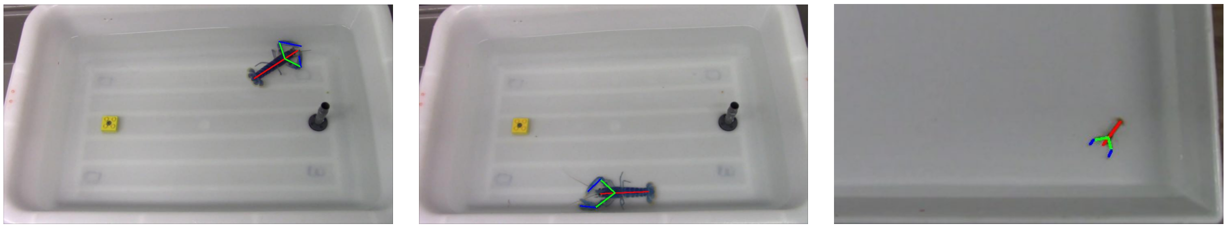
The initial centroid point $\mathbf{C}$ is given by

$$\mathbf{C} = \frac{\sum_i \boldsymbol{p_i \cdot Dist(p_i)}}{\sum_i \boldsymbol{Dist(p_i)}} \tag{5}$$

where $\boldsymbol{Dist(p_i)}$ is the distance value at location $\boldsymbol{p_i}$ obtained by distance transform of the segmented lobster image.

Then, observing that the longest line segment of the posture is from the central point to the tail, we search the tail point $\boldsymbol{p_t}$ from the points set of lobster skeleton $\boldsymbol{\{s_i\}}$.

$$\boldsymbol{p_t} = \underset{\boldsymbol{s}}{\mathbf{argmax}} \sum_{\boldsymbol{x \in line(s,C)}} \boldsymbol{Dist(x)} \tag{6}$$



(a) normal case (b) with shadow casting on the body (c) result for lobster lavar

Figure 2. Experiments for lobster posture estimation in different cases.

where $\boldsymbol{line(s, C)}$ denotes the point set along the line segment with end points $\boldsymbol{s}$ and central point $\boldsymbol{C}$. The head point $\boldsymbol{p_h}$ is obtained by extend the line segment we find to the points on the edge of segmented lobster body.

After obtaining the initial estimation of body line segment, the subsequent refine process is to find a more accurate line segment based on the skeleton points $\{\boldsymbol{s_i}\}$. First, collect the points in $\{\boldsymbol{s_i}\}$ that are within a certain distance to line segment between points $\boldsymbol{p_t}$ and $\boldsymbol{p_h}$. This point set is written as $\{\boldsymbol{s_{in}}\}$. Then calculate the best line fitting for $\{\boldsymbol{s_{in}}\}$ in terms of perpendicular offsets. Finally, find the neighbor points again. Run the steps until convergence. Finally, the previous $\boldsymbol{C}$, $\boldsymbol{p_t}$ and $\boldsymbol{p_h}$ is projected onto the new body line to obtained the refined central point, tail point and head point. The detailed algorithm is described in Algorithm. 1.

Algorithm 1: refine the body line segment

Input: $\boldsymbol{C}$,$\boldsymbol{p_t}$,$\boldsymbol{p_h}$,$\boldsymbol{Dist}$,$\{\boldsymbol{s_i}\}$,distance threshold $\boldsymbol{d}$
Output: refined $\boldsymbol{C}$,$\boldsymbol{p_t}$,$\boldsymbol{p_h}$,$\{\boldsymbol{s_{in}}\}$
repeat
  compute line segment $\boldsymbol{y = a + bx}$ between $\boldsymbol{p_t}$,$\boldsymbol{p_h}$;
  $\boldsymbol{w} = \begin{bmatrix} -\boldsymbol{b} & \boldsymbol{1} \end{bmatrix}$;
  $\{\boldsymbol{s_{in}}\} = \{\boldsymbol{s_i} | \frac{|\boldsymbol{w \cdot s_i - a}|}{||\boldsymbol{w}||} < \boldsymbol{d}\}$;
  $\widetilde{\boldsymbol{a}}, \widetilde{\boldsymbol{b}} = \mathbf{argmin}_{\boldsymbol{a,b}} \sum_{\boldsymbol{s} \in \{\boldsymbol{s_{in}}\}} \frac{|\boldsymbol{w \cdot s - a}| \boldsymbol{Dist(s)}}{||\boldsymbol{w}||}$;
  Project $\boldsymbol{p_t}$,$\boldsymbol{p_h}$ onto line $\boldsymbol{y} = \widetilde{\boldsymbol{a}} + \widetilde{\boldsymbol{b}}\boldsymbol{x}$ to get new $\boldsymbol{p_t}$,$\boldsymbol{p_h}$;
until *reach maximun iteration* ***OR*** *Convergence*;
Project $\boldsymbol{C}$ onto line $\boldsymbol{y} = \widetilde{\boldsymbol{a}} + \widetilde{\boldsymbol{b}}\boldsymbol{x}$ get new $\boldsymbol{C}$;
return $\boldsymbol{C}$,$\boldsymbol{p_t}$,$\boldsymbol{p_h}$,$\{\boldsymbol{s_{in}}\}$

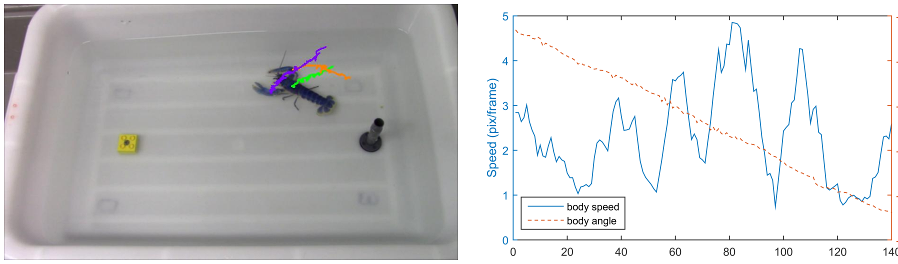
Before it proceeds to the searching of arms and claws, the skeleton that explains the lobster body is deducted. Therefore,

$$\{\boldsymbol{s_i}\} \longleftarrow \{\boldsymbol{s_i}\} - \{\boldsymbol{s_{in}}\} \tag{7}$$

The next step is to find the line segments to fit the two arms. One end of the arm line segment is fixed at central point $\boldsymbol{C}$, the other end $\boldsymbol{p_a}$ is searched within the set of lobster skeleton $\{\boldsymbol{s_i}\}$. The criterion to find the best point is the same as equation (6) and refinement is done in a similar way as we refine the body parts. The skeleton points which explain the arm are deducted also. Following this procedure, the claw is searched by fixing one end of line segment at $\boldsymbol{p_a}$ and the other end among $\{\boldsymbol{s_i}\}$ with further refinement.

3. EXPERIMENTS

We set up the experiment at a typical laboratory in a marine research institute. In order to test the robustness of our algorithm, we deliberately set the camera in the position that the image plane do not parallel to the bottom of the water tank. Therefore, the lobster is under perspective projection. In this situation, the size of the lobster appears on the image is variable. In addition, the light is also positioned to have shadow on the brim of the water tank in order to increase the difficulty for testing. The background color of the water tank is also not



(a) typical pose trajectory of a lobster (b) pose measurement

Figure 3. Trajectory of lobster posture and the behavior measurement.

strictly controlled to be uniform. The reason in doing so is that we intend to develop a robust algorithm that is handy and easily setup for biologists. The less requirement for the lighting and the position of the camera will bring more convenience to the algorithm users.

In a normal situation, as shown in Figure.2(a), the posture is successfully estimated as we intended to extract five line segment to presents their pose. Although, there are pereiopods and antenna as disturbances appearing differently between frames.

A more challenging situation is shown in Figure.2(b). The lobster in this frame is another individual other than the previous one with slightly lighter blue color. In this case, the lobster is partly in the shadow casted by the brim of the water tank. Meanwhile, two claws appear very differently for one claw holds in flat position and the other holds in side position. Even in this situation, our algorithm is still successfully estimated the posture.

Also, we test the same algorithm to the lobster larva as show in Figure.2(c) The larva looks similar to the lobster but with much smaller size and different color. This case also shows the robustness of our algorithm with broad application range for the lobster.

The trajectory of a lobster posture in a video clip consisting of 140 consecutive frames is shown in Figure.3. Only three points including the central point of body and two joint points between arm and claw are depicted in Figure.3(a) for clarity of vision. In this example, the lobster is turning around in the video clip with nearly constant angle speed as shown in Figure.3(b). The periodic body speed which is caused by the periodic movement of the pereiopods is also clearly shown in the pose measurement.

REFERENCES

[1] Gherardi, F., Cenni, F., Parisi, G., and Aquiloni, L., "Visual recognition of conspecifics in the american lobster, homarus americanus," *Animal Behaviour* **80**(4), 713 – 719 (2010).

[2] Dalal, N. and Triggs, B., "Histograms of oriented gradients for human detection," *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* **1**, 886–893 (2005).

[3] Benenson, R., Mathias, M., Timofte, R., and Van Gool, L., "Pedestrian detection at 100 frames per second," *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* , 2903–2910, IEEE (2012).

[4] Freund, Y. and Schapire, R. E., "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences* **55**(1), 119–139 (1997).

[5] Lepetit, V. and Fua, P., "Keypoint recognition using randomized trees," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **28**(9), 1465–1479 (2006).

[6] Ruzon, M. and Tomasi, C., "Alpha estimation in natural images," *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on* **1**, 18–25 vol.1 (2000).

[7] Dempster, A. P., Laird, N. M., and Rubin, D. B., "Maximum likelihood from incomplete data via the em algorithm," *Journal of the royal statistical society. Series B (methodological)* , 1–38 (1977).

[8] Maurer, C.R., J., Qi, R., and Raghavan, V., "A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **25**(2), 265–270 (2003).

[9] Lam, L., Lee, S.-W., and Suen, C. Y., "Thinning methodologies-a comprehensive survey," *IEEE Transactions on pattern analysis and machine intelligence* **14**(9), 869–885 (1992).