IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 13, 2019, accepted April 15, 2019, date of publication April 23, 2019, date of current version May 1, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2912612*

# Multi-Scale Fish Segmentation Refinement and Missing Shape Recovery

**GAOANG WANG[1], JENQ-NENG HWANG[1], (Fellow, IEEE), FARRON WALLACE[2], AND CRAIG ROSE[2]**

[1]Department of Electrical and Computer Engineering, University of Washington, Seattle, WA 98195, USA
[2]Alaska Fisheries Science Center, National Oceanic and Atmospheric Administration, Seattle, WA 98115, USA

Corresponding author: Gaoang Wang (gaoang@uw.edu)

This work was supported by the Alaska Fisheries Science Center, National Oceanic and Atmospheric Administration (NOAA), Seattle, WA 98115, USA.

**ABSTRACT** Image processing and analysis techniques have drawn increasing attention since they enable a non-extractive and non-lethal approach to collecting fisheries data, such as fish size measurement, catch estimation, regulatory compliance, species recognition, and population counting. Measuring fish size accurately requires reliable image segmentation. Major challenges that can easily affect the segmentation include blurring of image areas due to water drops on the camera lens and parts of a fish body being out of the camera view. In this paper, we address each of these issues with an innovative and effective contour-based segmentation and a missing shape recovery method from an arbitrary initial segmentation. The refinement is processed from the coarse level to the fine level. At the coarse level, we align the entire fish contour of the initial segmentation with trained representative contours by using iteratively reweighted least squares (IRLS). At finer levels, we iteratively refine contour segments to represent poorly segmented or missing shape parts. This method addresses the problems listed above and generates promising results with highly robust segmentation performance and length measurement.

**INDEX TERMS** Contour, refinement, regression, segmentation.

## I. INTRODUCTION

The potential of using automatic image processing systems in fisheries has drawn attention from both industry and aquaculture science [1] –[3], [30]–[32]. Counting, measurement, and isolation of captured fish are normally carried out directly on fishing vessels. The conventional, laborious manual process is time-consuming and limits the efficiency of fisheries data collection for either commercial or research purposes. An automated chute-based fish monitoring system can systematically perform fish body segmentation and length measurement. Therefore, the development of segmentation and measurement algorithms will be beneficial and will significantly speed up this indispensable process on fishing vessels. Compared to conventional manual sorting and measurement, the automatic image processing system can be faster, less error-prone, more scalable, and more usable by those without specialized training. In one such system, the Camera Chute being developed by the Alaska Fisheries Science Center, a static camera automatically captures images of fish passed through an onboard, enclosed chute with controlled lighting, as shown in Figure 1. Images are only taken when fish trigger an infrared sensor just before exiting the chute. One important application of the Camera Chute, requiring robust segmentation, has been counting and measuring lengths of fish bycatch during release from trawlers in Alaska fisheries. Fixed limits on fish bycatch are a key constraint on Alaska trawl fisheries and accurate monitoring is difficult with conventional catch sampling. While automatic image processing systems provide many advantages in fishery or aquaculture, challenges remain for segmentation and accurate measurement [1]. Examples of challenges experienced during extensive Camera Chute testing on fish bycatch include: 1), the environment may have dynamic lighting changes with restricted visibility, 2) the cameras may be occasionally splashed by water, leaving water drops that blur parts of subsequent images, 3) parts of the fish body may be out of the camera view when the infrared sensor is triggered, and 4) the fish has some body part (e.g., tail) not completely flat on the chute surface. Some examples of

The associate editor coordinating the review of this manuscript and approving it for publication was Bora Onat.

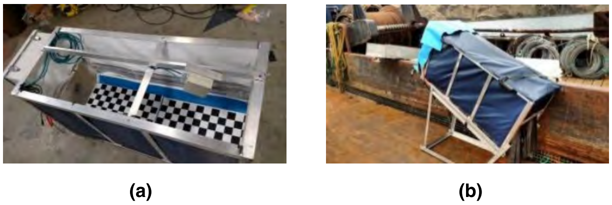


**FIGURE 1.** Chute with a static camera. (a) Interior view with checkboard for calibration. (b) Installation with cover on board.

bad segmentations and distorted fish images due to these challenges are shown in Figure 2. In order to measure the size and length of the fish, we needed to develop both a robust segmentation method and a method to estimate the fish body when part of the fish is out of the view of the camera or distorted by fish posture.

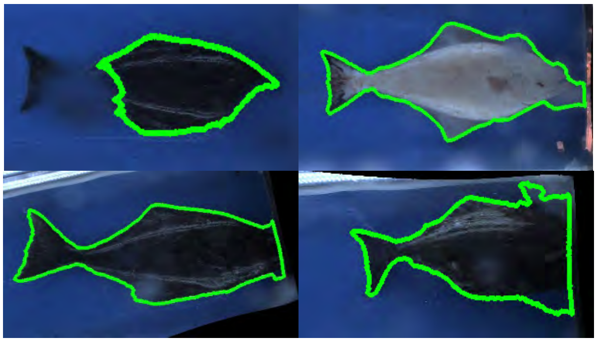


**FIGURE 2.** Examples of bad segmentation in a chute monitoring system. First row: Images with water drops. Second row: Images with part of fish out of the camera field.

In this paper, we propose a coarse-to-fine, contour-based method for segmentation refinement and missing body recovery for chute-based electronic monitoring (EM) of fisheries as shown in Figure 3. First, a segmentation method, such as [4]–[7], is applied to the input image to get the initial segmentation mask, whose contour is aligned with a pre-trained representative contour via an affine transform by iteratively solving reweighted least squares (IRLS) [33]. This constitutes the coarsest level for the entire contour alignment. At the finer level, rather than aligning the entire contour, we align two adjacent contour segments iteratively, which can recover the local structure of the fish body. At the finest level, we select the best match from all the training contours to replace each generated contour segment, which can recover the details of each contour segment. From coarse to fine, the segmentation refinement focuses more on specific parts of the fish, allowing more variations of the fish shape. We iteratively perform the minimization until convergence.

We summarize the contribution of our proposed method as follows.

1) The proposed method does not require large datasets for training. Unlike neural network-based methods, the proposed method can take advantage of the prior knowledge of the fish shape. As a result, a few samples are enough for shape modeling.

2) The segmentation refinement can be built on any initially-blurred segmentations. We can take any kind of initial segmentation as the input to our proposed refinement processes. Although the initial segmentation may be not accurate due to the blurring artifacts of water drops and lighting changes, as long as a rough contour is provided, the refinement can perform well.

3) The refinement is conducted at three levels from coarse to fine. At the coarse level, a general shape model can be easily aligned with the initial segmentation mask, while local features can be emphasized at finer levels.

4) The segmentation can also be refined when part of the fish is out of the camera view. By taking advantage of prior knowledge of the fish shape, the missing part of the fish can be reliably recovered. This recovery of missing parts is very important when we want to obtain accurate size and length measurement of the fish.

The outline of the paper is as follows: In Section II, we review some previous related work of image segmentation. The segmentation refinement method is then introduced in Section III. Experimental results are presented in Section IV. Finally, we provide some conclusions and future work in Section V.

## II. RELATED WORK

Measuring fish size and length requires a robust segmentation approach. Depending on whether annotated training data are required, the segmentation methods can be roughly divided into two major categories, i.e., unsupervised approaches [13]–[19] and supervised approaches [6], [7], [20], [21], [24], [25], [35].

The advantages of unsupervised approaches are obvious. First, it does not need any human effort to annotate the ground truth labels for each image pixel. Second, it can easily segment a new class object that is not represented in the labeled training data. The drawbacks are also obvious. Without the supervision of the ground truth, such methods can not achieve very good performance. Most unsupervised segmentation methods adopt clustering approaches to cluster pixels into different groups. For example, Achanta *et al.* [13] proposed a simple linear iterative clustering (SLIC) approach, which adapts a k-means clustering to efficiently generate superpixels. However, this method can only be used to generate superpixels, but not to segment the entire object. Graph-cut based approaches [14], [15], [18], [19] use graph models to minimize the energy on the edges to perform the segmentation. However, graph-based methods usually require a lot of computation time, especially for images with high resolution. Moreover, the mean-shift method [16] is also utilized to cluster pixels of the image, but it is sensitive to the bandwidth of the mean-shift kernels. While Arbelaez *et al.* [17] reduce the problem of image segmentation to contour detection, the segmented objects are not class specific and it is not clear how to extract the foreground object out of the image. Although these

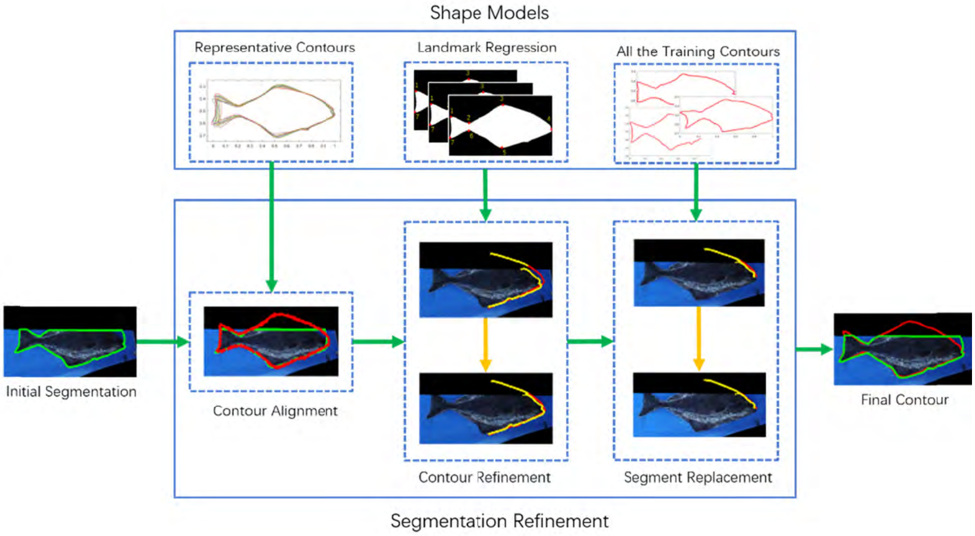


**FIGURE 3.** The flowchart of the contour refinement. From left to right is a coarse-to-fine process.

unsupervised methods are efficient even without training, they are not suitable for our task. They are still easily affected by water drops and do not address the situation when the part of fish is out of the view of the camera.

With more and more labeled datasets [22], [23] available in recent years, supervised learning-based approaches become more powerful in dealing with segmentation. With the supervision of the ground truth labels, more complex models can be trained to achieve higher segmentation accuracy, like neural network-based approaches [6], [7], [20], [21], [24]–[26], [35]. Long *et al.* [6], [7] adopt fully convolutional networks (FCNs) for image segmentation, which can take an input of arbitrary size and produce correspondingly-sized output with efficient inference and learning. However, this approach treats the segmentation as pixel classification without any other constraints in the cost function, resulting in non-sharp boundaries and blob-like shapes. In [21], Zheng *et al.* combine the strengths of FCN and Conditional Random Fields (CRFs)-based probabilistic graphical modeling and formulate mean-field approximate inference for the Conditional Random Fields with Gaussian pairwise potentials as Recurrent Neural Networks. Similarly, Chen *et al.* [24] combine deep networks with fully connected CRFs to perform the segmentation. Combining CRF can recover detailed information near the boundaries, which achieves better performance. However, FCN-based architectures still cannot perform as well as deconvolutional networks [35]. Badrinarayanan *et al.* [25] propose encoder-decoder architecture networks, in which the decoder uses pooling indices computed in the max-pooling step of the corresponding encoder to perform non-linear upsampling. This eliminates the need for learning to upsample in the deconvolution. Although neural network-based methods show promising results, they rely on large training data and cannot segment a new class object that is not in the training ground truth labels. Still, such methods are not suitable for our task, which require large datasets for training, cannot better utilize the prior knowledge of fish shapes, and cannot segment blur regions well or recover missing fish parts.

Apart from the general segmentation methods, there are also some methods that specifically deal with water drop segmentations and fish body segmentations. Some works [8]–[12] have been done for water drop detections or blur detections. For example, Alippi *et al.* [8] propose a method which detects external disturbances on camera lenses by comparing the blur measures of a series of frames which contain the same scene acquired from a static camera. Kanchev *et al.* [9] propose an algorithm for detecting blurred regions in images by using wavelet-based histograms and SVM. Blur measure is defined on the fish contour to locate the water drop region by Huang *et al.* [10]. Moreover, Liu *et al.* [11] develop several blur features to detect blur regions for general images. However, such water drop detection methods do not address how to deal with the issue in terms of segmentation. Chan *et al.* [12] propose a two-stage image segmentation method for blurry images with Poisson or multiplicative gamma noise. However, the water drop issue in our images for fish length measurement only occurs locally, which is not suitable for this setting. As for fish body segmentation, Chuang *et al.* [5] propose the double local thresholding (DLT) method which uses the color histogram to distinguish foreground and background. Similarly, Huang *et al.* [10] combine the DLT and Gaussian mixture model (GMM) [4] to model the static background to help extract the foreground fish object. However, these two methods purely rely on unsupervised approaches and do not take advantage of fish shape priors, and therefore still cannot recover the missing fish bodies.

To better address the segmentation challenges, the prior knowledge of the fish shape needs to be exploited in the segmentation. If we have a shape contour model with pose variations, the water drop region can be estimated and predicted with the help of other parts of the fish body based on

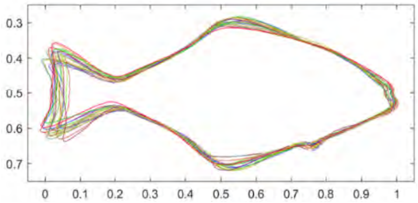


**FIGURE 4.** Twenty representative contours generated by the k-means algorithm.

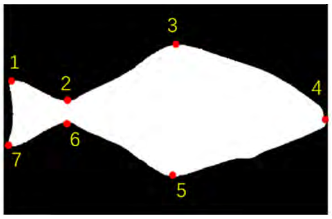


**FIGURE 5.** An example of defined seven landmarks in the shape model (red dots).

the shape contours. Similarly, when dealing with the missing shape issue, where part of the fish is out of the camera view, we can also efficiently predict missing parts with the help of the rest of the fish body based on the contour model even with very limited training data.

## III. THE PROPOSED METHOD

The flowchart of the coarse-to-fine refinement method is shown in Figure 3. Given trained shape models (see Section III.A.), first, we align the contour of the initial segmentation mask with one of the pre-trained representative contours (see Section III.B.) via the affine transform estimated by IRLS, aligning the entire contour at the coarsest level. On a finer level, the refinement proceeds by estimating the affine transform only for each set of two adjacent contour segments, using the same routine as the coarsest level (see Section III.C.). On the finest level, we match each generated contour segment with the corresponding segments from all the training data and choose the best match to replace the generated contour segment (see Section III.C.). The final contour result is shown on the right side of the figure. Note that the refinement proceeds from coarse to fine, focusing more on finer details of the fish. As a result, the segmentation can be refined even with variant postures.

### A. SHAPE MODELS TRAINING

Before processing the segmentation refinement, we illustrate how we train the shape models. As shown in the top of Figure 3, the shape models contain three components, i.e., representative contours, landmark regression, and all the training contours. For each training contour, we uniformly sample the contour points to a fixed length size, 1000. Then 20 representative contours are generated using the k-means algorithm based on all the training contours, as shown in Figure 4. Note that these representative contours are smoothed contours with diverse different poses which can roughly represent the shape of the fish, although some details are lost.

As for the landmark regression model, we first define seven landmarks with highest curvatures, on the contour of the fish segmentation. Landmarks 1 and 7 are located at the end of the tail; landmarks 2 and 6 are located at the connection between the fish tail and fish body; landmarks 3 and 5 are the turning points of the fish body while landmark 4 is located at the fish mouth. An example is shown in Figure 5.

In the case that part of the fish is missing, some landmarks may not be accessible in the initial segmentation, as shown in the second row in Figure 2. To deal with such issues, we train a linear regression to estimate the missing landmark location based on the other 6 landmarks. For example, if we want to estimate the location of the $j$-th landmark, the regression model can be expressed as

$$\hat{\boldsymbol{c}}_j(x) = arg\,\min_{c_j(x)} \left\| \boldsymbol{B}_j^T \boldsymbol{c}_j(x) - \boldsymbol{v}_j(x) \right\|_2^2, \tag{1}$$

$$\hat{\boldsymbol{c}}_j(y) = arg\,\min_{c_j(y)} \left\| \boldsymbol{B}_j^T \boldsymbol{c}_j(y) - \boldsymbol{v}_j(y) \right\|_2^2, \tag{2}$$

with

$$\boldsymbol{B}_j = \left[\boldsymbol{b}_{j,1}, \boldsymbol{b}_{j,2}, \ldots, \boldsymbol{b}_{j,i}, \ldots, \boldsymbol{b}_{j,N_{tr}}\right], \tag{3}$$

where each column of $\boldsymbol{B}_j$, i.e., $\boldsymbol{b}_{j,i}$, contains the $(x, y)$-coordinates of all the other 6 landmarks except the $j$-th landmark with 12 dimensions; $\boldsymbol{v}_j(x)$ and $\boldsymbol{v}_j(y)$ are $(x, y)$-coordinates of the $j$-th landmark of all training data, respectively; $\hat{\boldsymbol{c}}_j(x)$ and $\hat{\boldsymbol{c}}_j(y)$ are the learned coefficients of the regression model; $N_{tr}$ is the number of training data. Moreover, we also want to make the model invariant to both the scale $s$ and the shift $\left(s_x, s_y\right)^T$. Then the learned coefficients should satisfy the following conditions,

$$\left(s\boldsymbol{b}_{j,i} - \boldsymbol{d}\right)^T \boldsymbol{c}_j(x) = s\boldsymbol{v}_{j,i}(x) - s_x, \tag{4}$$

$$\left(s\boldsymbol{b}_{j,i} - \boldsymbol{d}\right)^T \boldsymbol{c}_j(y) = s\boldsymbol{v}_{j,i}(y) - s_y, \tag{5}$$

where $\boldsymbol{d} = \left(s_x, s_y, s_x, s_y, \ldots\right)^T$ is a concatenated vector for an arbitrary shift $\left(s_x, s_y\right)^T$ with 12 dimensions. Combining the above two equations with the following two solutions, i.e.,

$$\boldsymbol{b}_{j,i}^T \boldsymbol{c}_j(x) = \boldsymbol{v}_{j,i}(x), \tag{6}$$

$$\boldsymbol{b}_{j,i}^T \boldsymbol{c}_j(y) = \boldsymbol{v}_{j,i}(y), \tag{7}$$

we can derive the following constraints for the coefficients,

$$\boldsymbol{u}_1^T \boldsymbol{c}_j(x) = 1, \tag{8}$$

$$\boldsymbol{u}_2^T \boldsymbol{c}_j(x) = 0, \tag{9}$$

$$\boldsymbol{u}_1^T \boldsymbol{c}_j(y) = 0, \tag{10}$$

$$\boldsymbol{u}_2^T \boldsymbol{c}_j(y) = 1, \tag{11}$$

where $\boldsymbol{u}_1 = (1, 0, 1, 0, \ldots)^T$ with elements in odd indices equal to 1 and $\boldsymbol{u}_2 = (0, 1, 0, 1, \ldots)^T$ with elements in even indices equal to 1. Both $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ have the same dimension with $\boldsymbol{c}_j(x)$ and $\boldsymbol{c}_j(y)$. After the scale and shift invariant modification, Eq. (1) and Eq. (2) can be rewritten as,

$$\hat{\boldsymbol{c}}_j(x) = \mathop{arg\,min}_{c_j} \left\|\boldsymbol{B}_j\boldsymbol{c}_j(x) - \boldsymbol{v}_j(x)\right\|_2^2, \quad s.t.\ \boldsymbol{u}_1^T\boldsymbol{c}_j(x) = 1, \quad \boldsymbol{u}_2^T\boldsymbol{c}_j(x) = 0, \tag{12}$$

and

$$\hat{\boldsymbol{c}}_j(y) = \mathop{arg\,min}_{c_j} \left\|\boldsymbol{B}_j\boldsymbol{c}_j(y) - \boldsymbol{v}_j(y)\right\|_2^2, \quad s.t.\ \boldsymbol{u}_1^T\boldsymbol{c}_j(y) = 0, \quad \boldsymbol{u}_2^T\boldsymbol{c}_j(y) = 1. \tag{13}$$

This linear constraint least square problem can be solved by constructing a Karush–Kuhn–Tucker (KKT) matrix as illustrated in [27].

In the testing stage, we can estimate the missing location of the $j$-th landmark by

$$\boldsymbol{v}_j^*(x) = \boldsymbol{b}_j^T\hat{\boldsymbol{c}}_j(x), \tag{14}$$

$$\boldsymbol{v}_j^*(y) = \boldsymbol{b}_j^T\hat{\boldsymbol{c}}_j(y). \tag{15}$$

We will use this regression model to estimate the landmark locations when we refine the contour segments in the later section.

### B. CONTOUR ALIGNMENT

It is a challenging task to align the initial segmentation mask using shape models since water drops and missing body part can largely affect the initial segmentation. Before the alignment, we rotate the fish to the horizontal orientation based on the principal component analysis (PCA) using the initial segmentation mask as a pre-processing step. This pre-processing gives a good starting point and largely simplifies the alignment. Here, we propose an iteratively reweighted least squares (IRLS) [33] algorithm to align the segmentation mask $\boldsymbol{p}$ with representative contour models $\{\boldsymbol{p}^m\}$ via an affine transform $\boldsymbol{H}$, which is robust to outliers. We are interested in the affine transformation (6 degrees of freedom) rather than the projective transform (8 degrees of freedom) because the projective transform is more sensitive to outliers with 2 more degrees of freedom than the affine transform. Since we have 20 representative contours generated by the k-means algorithm, these 20 contours have diversity shapes, which still can address the limitations of affine transformations without skewing and shear deformations. Given the pre-trained shape models, the affine transform [29] can be estimated by

$$\hat{\boldsymbol{H}} = \mathop{arg\,min}_{\boldsymbol{H}} \sum\nolimits_i \left\|\boldsymbol{H}\boldsymbol{p}_i - \boldsymbol{p}_i^m\right\|_2^2, \tag{16}$$

with

$$\boldsymbol{H} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ 0 & 0 & 1 \end{bmatrix}, \tag{17}$$

where $\boldsymbol{p}_i = (x_i, y_i, 1)^T$ is the contour point in the initial segmentation without touching the image boundary and $\boldsymbol{p}_i^m$ is a contour point that closest to $\boldsymbol{H}\boldsymbol{p}_i$ from the $m$-th $(m = 1, 2, \ldots, 20)$ representative contour as defined in Section III.A. If we concatenate the 6 parameters in $\boldsymbol{H}$ to a vector $\boldsymbol{h}$, then the affine transform can be estimated by a least square problem as

$$\hat{\boldsymbol{h}} = \mathop{arg\,min}_{\boldsymbol{h}} \left\|\boldsymbol{A}\boldsymbol{h} - \boldsymbol{p}^m\right\|_2^2, \tag{18}$$

where $\boldsymbol{A}$ contains the information of $(x, y)$-coordinates of all contour points of the initial segmentation (from 1 to $N$), i.e.,

$$\boldsymbol{A} = \left[\boldsymbol{a}_1^T, \ldots, \boldsymbol{a}_i^T, \ldots, \boldsymbol{a}_N^T\right]^T, \tag{19}$$

$$\boldsymbol{a}_i = \begin{bmatrix} x_i & y_i & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_i & y_i & 1 \end{bmatrix} \tag{20}$$

where $N$ is the number of contour points.

If water drops occur along the segmentation contour, then the affine transform can be largely affected by these unreliable contour points. Inspired by M-estimator [28]–[34], to avoid the influence by water drops and also to avoid the outliers from the initial contour points, we reformulate the cost function and obtain the solution by using IRLS as follows,

$$f(\boldsymbol{W}, \boldsymbol{A}, \boldsymbol{h}, m) = \left\|\boldsymbol{W}\left(\boldsymbol{A}\boldsymbol{h} - \boldsymbol{p}^m\right)\right\|_2^2, \tag{21}$$

where $\boldsymbol{W}$ is a diagonal matrix in which the diagonal elements represent the weights of each observation. To minimize the cost above, we conduct the following steps iteratively.

*Step 1:* Select the best model $m^t$ that gives the lowest error based on the previous estimation, i.e.,

$$m^t = arg \mathop{min}_{m\epsilon\{1,2,\ldots,20\}} \left\|\boldsymbol{W}^{t-1}\left(\boldsymbol{A}^{t-1}\boldsymbol{h}^{t-1} - \boldsymbol{p}^m\right)\right\|_2^2. \tag{22}$$

*Step 2:* For each contour point $\boldsymbol{p}_i^{m^t}$ from the representative contour model, find the corresponding transformed point that achieves maximum weight combining both spatial distance error and gradient magnitude,

$$\boldsymbol{p}_i^t = \mathop{arg\,max}_{pi} exp\left(-\left\|\boldsymbol{H}^{t-1}\boldsymbol{p}_i - \boldsymbol{p}_i^{m^t}\right\|_2^2\right) w_g\left(\boldsymbol{p}_i\right), \tag{23}$$

where $w_g(\boldsymbol{p}_i)$ is the gradient magnitude of $\boldsymbol{p}_i$. The point selection is based on two criteria, i.e., 1) it should be close to $\boldsymbol{p}_i^{m^t}$ after the transformation, which is the spatial weighting, 2) it should have high gradient magnitude, based on the assumption that the contour point should have high gradient magnitude. Then $\boldsymbol{A}^t$ is updated based on $\boldsymbol{p}_i^t$.

*Step 3:* Obtain the weight of each selected point by

$$\boldsymbol{W}_i^t = exp\left(-\left\|\boldsymbol{H}^{t-1}\boldsymbol{p}_i^t - \boldsymbol{p}_i^{m^t}\right\|_2^2\right) w_g\left(\boldsymbol{p}_i^t\right), \tag{24}$$

where $\boldsymbol{W}_i^t$ is the $i$-th diagonal element in $\boldsymbol{W}^t$.

*Step 4:* Update the affine transform by

$$\boldsymbol{h}^t = \left(\boldsymbol{A}^{t^T}\boldsymbol{W}^{t^T}\boldsymbol{W}^t\boldsymbol{A}^t\right)^{-1}\boldsymbol{A}^{t^T}\boldsymbol{W}^{t^T}\boldsymbol{W}^t\boldsymbol{p}^{m^t}. \tag{25}$$

We use IRLS to measure the affine transform, which is similar to M-estimator. However, there are two major differences between the proposed approach and the standard M-estimator. First, the weight function is not only based on the spatial error, but also combines the gradient magnitude of the contour point. Second, the matched contour points are allowed to change for each iteration, which can largely remove the outliers, while the samples from the standard M-estimator is always fixed. We conduct the above optimization steps iteratively until convergence. An example of contour alignment is shown in Figure 6.

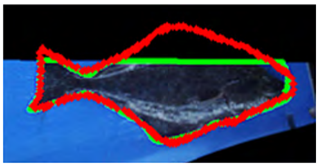


**FIGURE 6.** **Contour alignment. Green points are the contour points from the initial segmentation. Red contour is the best match from the representative contour models.**

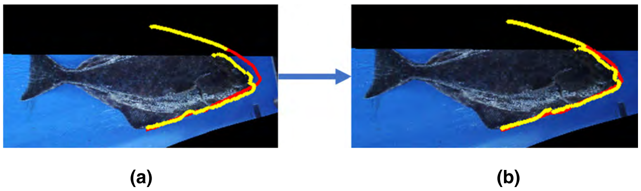
### C. CONTOUR REFINEMENT

For the coarsest level of contour alignment, the initial segmentation contour is represented by a general fish shape defined by the selected representative contour model, which cannot precisely match the segmented fish. In this subsection, the segmentation contour is refined with more local details recovered at two finer levels.

For the finer level refinement, the affine transform is estimated only for two adjacent contour segments as shown in Figure 7. To be specific, the affine transform is estimated from

$$f\left(\boldsymbol{W}_{j-1,j+1}, \boldsymbol{A}_{j-1,j+1}, \boldsymbol{h}_j, m\right) = \left\| \boldsymbol{W}_{j-1,j+1}\left(\boldsymbol{A}_{j-1,j+1}\boldsymbol{h}_j - \boldsymbol{p}^m_{j-1,j+1}\right)\right\|_2^2 \quad (26)$$

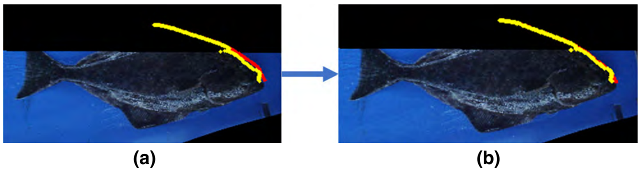
where $\boldsymbol{p}^m_{j-1,j+1}$ are the representative contour points between the $(j-1)$-th landmark and $(j+1)$-th landmark, which contains two adjacent contour segments; $\boldsymbol{A}_{j-1,\,j+1}$ contains the information of $(x, y)$-coordinates of matched contour points between the $(j-1)$-th landmark and $(j+1)$-th landmark and $\boldsymbol{W}_{j-1,\,j+1}$ is the diagonal matrix with the diagonal elements equal to the weights of the two contour segments. To constrain the contour in a regular shape, the same affine transform $\boldsymbol{h}_j$ between two adjacent contour segments is shared in the estimation.



(a)

(b)

**FIGURE 7.** **Contour refinement. An affine transform is estimated for every two adjacent contour segments. (a) Before (b) After.**



(a)

(b)

**FIGURE 8.** **Contour segment replacement. Use the best contour segment chosen from one segment of all the training data to match the generated contour segment. (a) Before (b) After.**

Additionally, if the $j$-th landmark is missing, the above estimation is not reliable. To deal with such issue, a regularization term is added to the cost function to restrict the missing landmark location to follow the general fish shape from shape models, i.e.,

$$f\left(\boldsymbol{W}_{j-1,j}, \boldsymbol{A}_{j-1,j}, \boldsymbol{h}_j, m\right) = \left\| \boldsymbol{W}_{j-1,j}\left(\boldsymbol{A}_{j-1,j}\boldsymbol{h}_j - \boldsymbol{p}^m_{j-1,j}\right)\right\|_2^2 + \lambda \left\| \left(\boldsymbol{D}_j\boldsymbol{h}_j - \boldsymbol{v}^*_j\right)\right\|_2^2, \quad (27)$$

and

$$\boldsymbol{D}_j = \begin{bmatrix} v(x_j) & v(y_j) & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & v(x_j) & v(y_j) & 1 \end{bmatrix} \quad (28)$$

where $\boldsymbol{v}^*_j$ is the estimated location of the $j$-th landmark which can be computed by Eq. (14) and Eq. (15) from the representative contour models and $\boldsymbol{D}_j$ contains the information of the $j$-th landmark coordinates. Similarly, we adopt IRLS to estimate the affine transform as presented in III.B. In the step 4, the affine transform is estimated by

$$\boldsymbol{h}^t_j = \left(\boldsymbol{A}^{t^T}_{j-1,j}\boldsymbol{W}^{t^T}_{j-1,j}\boldsymbol{W}^t_{j-1,j}\boldsymbol{A}^t_{j-1,j} + \lambda\boldsymbol{D}^T_j\boldsymbol{D}_j\right)^{-1} \times \left(\boldsymbol{A}^{t^T}_{j-1,j}\boldsymbol{W}^{t^T}_{j-1,j}\boldsymbol{W}^t_{j-1,j}\boldsymbol{p}^{m^t}_{j-1,j} + \boldsymbol{D}^T_j\boldsymbol{v}^*_j\right). \quad (29)$$

At the finest level, we search all the segments from all the training data to find the best contour segment that matches the generated contour segment. Then we replace the generated contour segment with the matched segment from the training data. In this way, the detail information can be recovered. An example is shown in Figure 8.

Note that for each iteration, the affine transform $\hat{h}_j$ is estimated alternatively with $j = 1, 2, \ldots, 7$. Then we alternatively update generated contour points between finer and finest levels iteratively until convergence.

## IV. EXPERIMENTS AND RESULTS

### A. IMPLEMENTATION DETAILS

We manually label 583 segmentation masks in the experiments. 489 segmentation masks are used to train the shape models while the remaining 94 images are used for testing. We store seven landmark locations and the segments between every two adjacent landmarks for all the training data. Moreover, we use linear interpolation for all the segments to make them have equal points across the training data. For hypeparameters, we set $\lambda = 0.1$.

To check the efficiency of the proposed method, we use two initial segmentation masks as a comparison, namely double local thresholding (DLT) [5] and fully convolutional networks (FCN) [7]. For training FCN, we first resize each

**TABLE 1. Average IoU before and after refinement.**

| Method | Average IoU (%) |
|---|---|
| DLT [5] | 89.65 |
| FCN [7] | 92.35 |
| DLT +Refinement | 95.59 |
| FCN+ Refinement | 95.47 |

training image into 224 × 224 and adopt the VGG network with 8-stride (FCN-8s). In the testing stage, after obtaining the predicted segmentation mask, we resize it back to the original image size. For the data augmentation, we use random rotation (±10°), random flip (horizontal and vertical) and random crop (with at most 1/4 variations of width and height of the image size) in the training. In total, we augment the training data 10 times. We set the batch size to 2, the learning rate to 1e-4 with Adam Optimizer. We plot the loss curve for training and validation in Figure 9. After around 1000 iterations, the validation loss does not decrease, hence, we adopt the model trained at the 1000-th iteration as the final FCN model.

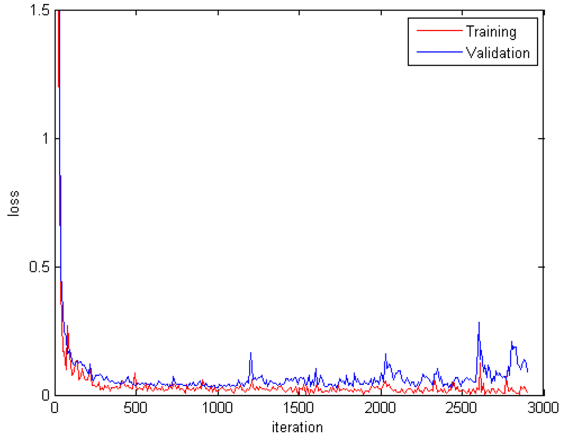


**FIGURE 9. Training and validation loss for FCN.**

### B. HANDLING WATERDROPS

We use the average intersection over union (IoU) to test the performance of the proposed refined segmentation method. The average IoU is defined as,

$$Avg\ IoU = \frac{1}{N}\sum_{i=1}^{N}\frac{Area\left(GT_i \cap PredSeg_i\right)}{Area\left(GT_i \cup PredSeg_i\right)}, \tag{30}$$

where $GT_i$ is the ground truth segmentation of the $i$-th image, $PredSeg_i$ is the predicted segmentation, ∩ and ∪ represent intersection and union, respectively. Note that our optimization is based on a weakly supervised approach, which means we do not use pixel labels as the supervision in the training. Without the pixel label supervision, the IoU metric cannot be directly used in the cost function. Instead, the cost function we are using is to align the generated fish shape with representative contours as much as possible. In general cases, a good alignment with the object contours usually ensures a good measure of IoU, although these two metrics are not exactly the same. In the experiments, we will show that our proposed method can also achieve promising results with IoU metric.

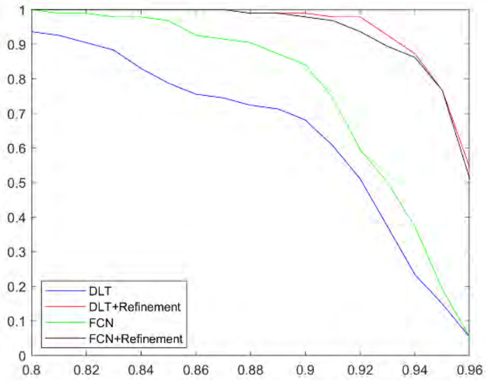


**FIGURE 10. Acceptance rate along with the IoU threshold.**

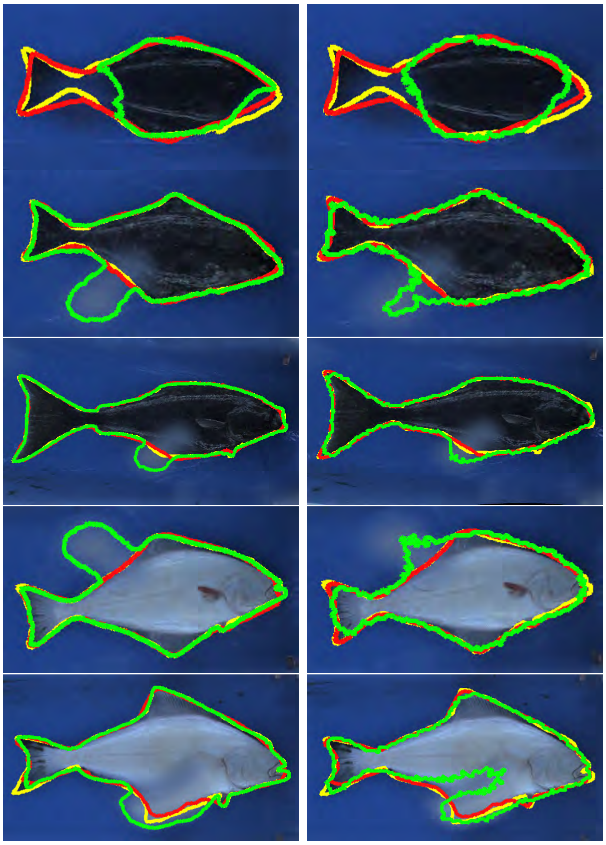


**FIGURE 11. Examples of segmentation refinement related to water drops. Left column: Green contours are initial input segmentation contours with DLT; red contours are refined segmentation contours; yellow contours are ground truth. Right column: Green contours are initial input segmentation contours with FCN; red contours are refined segmentation contours; yellow contours are ground truth.**

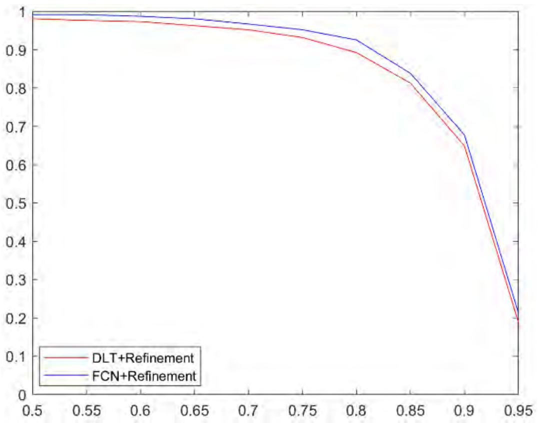


**FIGURE 12.** Acceptance rate along with the IoU threshold.

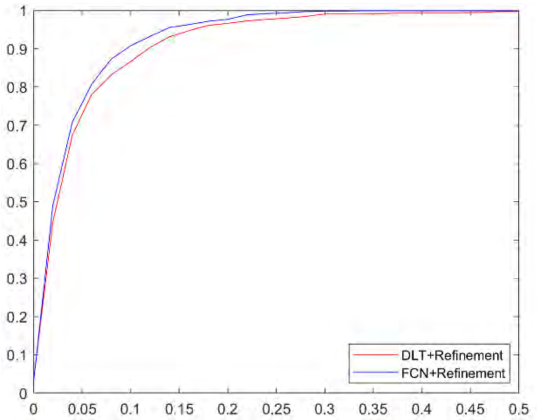


**FIGURE 13.** Acceptance rate along with the length error threshold.

The results of refinement are shown in Table 1. We can see the average IoU achieves over 95% after refinement with both of the two initial segmentations. Although FCN seems more reliable than DLT, the results after refinement are roughly similar, which means the proposed refinement is robust to the initial segmentation.

Assume the segmentation mask is acceptable if the IoU is above a certain threshold. Then we define the acceptance rate as the ratio between the number of acceptable segmentation masks and the total number of the testing data given the threshold. We plot the acceptance rate with the increase of the threshold as shown in Figure 10. We can see that the refined segmentation result is much better than the initial segmentation.

We also show some qualitative results of the refined segmentation dealing with water drops in Figure 11, where the refinement can effectively recover the contour in the blur regions.

### C. MISSING PART RECOVERY

To test the performance of the proposed method when a part of the fish body is out of the camera view, we randomly cut off about 1/3 of the fish body 10 times for each testing data. For 940 (94 × 10) testing data in total, we evaluate the segmentation performance of the proposed method. We measure the mean absolute error of the length and IoU of the segmentation, whose results are shown in Table 2.

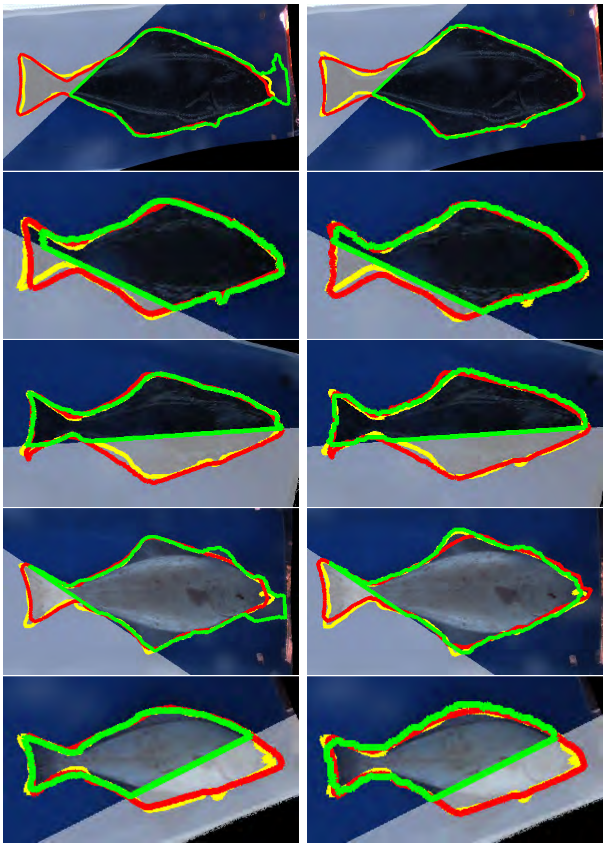


**FIGURE 14.** Examples of segmentation refinement related to part of fish out of camera view. Transparent part of the image is cut off in the simulation. Left column: Green contours are initial input segmentation contours with DLT; red contours are refined segmentation contours; yellow contours are ground truth. Right column: Green contours are initial input segmentation contours with FCN; red contours are refined segmentation contours; yellow contours are ground truth.

**TABLE 2.** Measurement of IoU and length error.

| Method | DLT+Refinement | FCN+Refinement |
|---|---|---|
| Average IoU (%) | 88.45 | 89.67 |
| Mean length error (%) | 4.68 | 3.90 |

Similarly, we define the acceptance rate for different thresholds of IoU and length error. We plot the results in Figure 12 and Figure 13, where we can see no large difference with two initial segmentation approaches, which further proves that the proposed refinement is also robust to the initial segmentation when recovering the missing part. We also show some qualitative examples in Figure 14.

## V. CONCLUSIONS AND FUTURE WORK

In this paper, a coarse-to-fine contour-based segmentation refinement is proposed to deal with segmentation challenges in fish measurement. There are three key components for the proposed refinement: 1) we refine the segmentation from the coarse level to the fine level to deal with variant fish shapes; 2) the refinement is processed iteratively via affine transforms; 3) shape models provide rich prior knowledge when estimating the segmentation contours. In our experiments, we compare two initial segmentation approaches in the refinement, which shows the robustness of the refinement to different initial segmentations. The refinement also shows the effectiveness in the segmentation dealing with water drops and missing part recovery for flat fish.

There are also some limitations of our proposed method. In the current settings, all the testing images are from flat fish, like halibut and flounder. When we have more categories that have large shape difference, we may need to use more clusters in the k-means to generate more representative contours, which will increase the running time in the testing stage to find the best alignment. In future work, we are going to combine the fish classifier to address this issue and create the shape models for each fish category.

## REFERENCES

[1] B. Zion, "The use of computer vision technologies in aquaculture—A review," *Comput. Electron. Agriculture*, vol. 88, pp. 125–132, Oct. 2012.

[2] J. R. Mathiassen, E. Misimi, M. Bondø, E. Veliyulin, and S. O. Østvik, "Trends in application of imaging technologies to inspection of fish and fish products," *Trends Food Sci. Technol.*, vol. 22, no. 6, pp. 257–275, Jun. 2011.

[3] M.-C. Chuang, J.-N. Hwang, K. Williams, and R. Towler, "Tracking live fish from low-contrast and low-frame-rate stereo videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 1, pp. 167–179, Jan. 2015.

[4] Z. Zivkovic and F. Van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognit. Lett.*, vol. 27, no. 7, pp. 773–780, May 2006.

[5] M.-C. Chuang, J.-N. Hwang, K. Williams, and R. Towler, "Automatic fish segmentation via double local thresholding for trawl-based underwater camera systems," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 3145–3148.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.

[7] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[8] C. Alippi, G. Boracchi, R. Camplani, and M. Roveri, "Detecting external disturbances on the camera lens in wireless multimedia sensor networks," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 11, pp. 2982–2990, Nov. 2010.

[9] V. Kanchev, K. Tonchev, and O. Boumbarov, "Blurred image regions detection using wavelet-based histograms and SVM," in *Proc. IEEE 6th Int. Conf. Intell. Data Acquisition Adv. Comput. Syst.*, Prague, Czech, Sep. 2011, pp. 457–461.

[10] T.-W. Huang, J.-N. Hwang, and C. S. Rose, "Chute based automated fish length measurement and water drop detection," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, Mar. 2016, pp. 1906–1910.

[11] R. Liu, Z. Li, and J. Jia, "Image partial blur detection and classification," in *Proc. Conf. Comput. Vision Pattern Recognit.*, Jun. 2008, pp. 1–8.

[12] R. Chan, H. Yang, and T. Zeng, "A two-stage image segmentation method for blurry images with poisson or multiplicative gamma noise," *SIAM J. Imag. Sci.*, vol. 7, no. 1, pp. 98–127, Jan. 2014.

[13] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[14] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[15] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary region segmentation of objects in N-D images," in *Proc. IEEE 8th Int. Conf. Comput. Vision*, vol. 1, Jul. 2001, pp. 105–112.

[16] W. Tao, H. Jin, and Y. Zhang, "Color image segmentation based on mean shift and normalized cuts," *IEEE Trans. Syst., Man, Cybern., B, Cybern.*, vol. 37, no. 5, pp. 1382–1389, Oct. 2007.

[17] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.

[18] S. Vicente, V. Kolmogorov, and C. Rother, "Graph cut based image segmentation with connectivity priors," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Jun. 2008, pp. 1–8.

[19] J. Carreira and C. Sminchisescu, "Constrained parametric min-cuts for automatic object segmentation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, Jun. 2010, pp. 3241–3248.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015.

[21] S. Zheng *et al.*, "Conditional random fields as recurrent neural networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1529–1537.

[22] T.-Y. Lin *et al.*, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vision*, Sep. 2014, pp. 740–755.

[23] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Scene parsing through ADE20K dataset," in *Proc. CVPR.*, 2017, pp. 633–641.

[24] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. (2016)."DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs." [Online]. Available: https://arxiv.org/abs/1606.00915

[25] V. Badrinarayanan, A. Kendall, and R. Cipolla. (2015). "Segnet: A deep convolutional encoder-decoder architecture for image segmentation." [Online]. Available: https://arxiv.org/abs/1511.00561

[26] Z. Liu, X. Li, P. Luo, C.-C. Loy, and X. Tang, "Semantic image segmentation via deep parsing network," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1377–1385.

[27] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[28] M. M. Fouad, R. M. Dansereau, and A. D. Whitehead, "Image registration under illumination variations using region-based confidence weighted $M$-estimators," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1046–1060, Mar. 2012.

[29] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[30] G. Wang, J. N. Hwang, K. Williams, and G. Cutter, "Closed-loop tracking-by-detection for ROV-based multiple fish tracking," in *Proc. 2nd Workshop Comput. Vision Anal. Underwater Imagery (CVAUI)*, Dec. 2016, pp. 7–12.

[31] G. Wang, J.-N. Hwang, K. Williams, F. Wallace, and C. S. Rose, "Shrinking encoding with two-level codebook learning for fine-grained fish recognition," in *Proc. 2nd Workshop Comput. Vision Anal. Underwater Imagery (CVAUI)*, Dec. 2016, pp. 31–36.

[32] G. Wang, J.-N. Hwang, C. Rose, and F. Wallace, "Uncertainty sampling based active learning with diversity constraint by sparse selection," in *Proc. 19th Int. Workshop Multimedia Signal Process. (MMSP)*, Oct. 2017, pp. 1–6.

[33] A. Bjorck, *Numerical Methods for Least Squares Problems*. Philadelphia, PA, USA: SIAM, 1996.

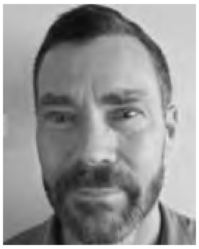
[34] S. A. Van de Geer and S. Van de Geer, *Empirical Processes in M-Estimation*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[35] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vision*, Dec. 2015, pp. 1520–1528.

**GAOANG WANG** received the B.S. degree from the Department of Electrical Engineering, Fudan University, in 2013, and the M.S. degree from the Department of Electrical and Computer Engineering, University of Wisconsin–Madison, in 2015. He is currently pursuing the Ph.D. degree with the Information Processing Lab, Department of Electrical Engineering, University of Washington. His current research interests include computer vision, machine learning, and video/image processing.

**JENQ-NENG HWANG** (F'01) received the B.S. and M.S. degrees in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1981 and 1983, respectively, and the Ph.D. degree from the University of Southern California.

In the summer of 1989, he joined the Department of Electrical Engineering, University of Washington, Seattle, where he has been promoted to a Full Professor, since 1999. He served as the Associate Chair for Research, from 2003 to 2005, and from 2011 to 2015. He is currently the Associate Chair for Global Affairs and International Development with the EE Department. He has written more than 330 journal, conference papers, and book chapters in the areas of machine learning, multimedia signal processing, computer vision, and multimedia system integration and networking, including an authored textbook on *Multimedia Networking: from Theory to Practice*, (Cambridge University Press). He has close working relationship with the industry on multimedia signal processing and multimedia networking. He received the 1995 IEEE Signal Processing Society's Best Journal Paper Award. He is a Founding Member of the Multimedia Signal Processing Technical Committee of the IEEE Signal Processing Society and was the Society's representative to IEEE Neural Network Council, from 1996 to 2000. He is currently a member of the Multimedia Technical Committee (MMTC) of the IEEE Communication Society and also a member of the Multimedia Signal Processing Technical Committee (MMSP TC) of the IEEE Signal Processing Society. He served as the Program Co-Chair of IEEE ICME 2016 and was the Program Co-Chair of ICASSP 1998 and ISCAS 2009. He served as an Associate Editor for IEEE T-SP, T-NN and T-CSVT, T-IP, and Signal Processing Magazine (SPM). He is currently on the Editorial Board of ZTE Communications, ETRI, IJDMB, and JSPS journals.



**FARRON WALLACE** is currently with NOAA Fisheries, Alaska Fisheries Science Center (AFSC), Seattle, WA, USA. He is also a Senior Fisheries Research Stock Assessment Scientist serving on the North Pacific Fisheries Management Council's Science and Statistical Committee (SSC) as the Chair, Vice Chair, and a Member. In his current role, he is leading the Electronic Monitoring Innovation project to develop new remote fishery monitoring tools. These tools incorporate the newest machine vision learning to automated image analysis to determine size and species identification.

**CRAIG ROSE** received the B.S. degree in fisheries biology from Humboldt State and the M.S. and Ph.D. degrees from the University of Washington. He led Conservation Engineering Research with the Alaska Fisheries Science Center, focusing on using *in-situ* cameras to better understand fish behavior in fishing gears and improvements to fishing methods to reduce habitat effects and improve catch selectivity. Several resulting innovations are routinely used in Alaska commercial fisheries. Leaving federal service after 37 years, in 2014, he established FishNext Research, continuing this focus on conservation improvements to commercial fishing.