Particle Filter-based Predictive Tracking for Robust Fish Counting

Erikson F. Morais, Mario F. M. Campos, Flávio L. C. Pádua and Rodrigo L. Carceroni
UFMG–Universidade Federal de Minas Gerais
VERLab–Computer Vision and Robotics Laboratory
Av. Antônio Carlos, 6627, Belo Horizonte, MG, Brasil
{erikson,mario,cardeal,carceron}@dcc.ufmg.br

Abstract

In this paper we study the use of computer vision techniques for for underwater visual tracking and counting of fishes in vivo. The methodology is based on the application of a Bayesian filtering technique that enables tracking of objects whose number may vary over time. Unlike existing fish-counting methods, this approach provides adequate means for the acquisition of relevant information about characteristics of different fish species such as swimming ability, time of migration and peak flow rates. The system is also able to estimate fish trajectories over time, which can be further used to study their behaviors when swimming in regions of interest. Our experiments demonstrate that the proposed method can operate reliably under severe environmental changes (e.g. variations in water turbidity) and handle problems such as occlusions or large inter-frame motions. The proposed approach was successfully validated with real-world video streams, achieving overall accuracy as high as 81%.

Figure 1. Observation room at the fish ladder located at UHE-Igarapava, MG, Brazil, used by researchers to study several aspects of the fish ladder, mainly its adequacy and efficiency for different fish species.

1. Introduction

A wide variety of human activities have altered the distribution and abundance of the native fish fauna all over the world. Extractive industries, such as mining and forestry and other intervention by man on Nature, such as the construction of dams for hydroelectric power generation are some examples of activities that have had widespread, dramatic impacts, often with deeper effects on a local scale, resulting in extensive changes in aquatic environments and often in the break-down of reproductive isolation between some species [1, 15].

Recently, in Brazil, various legal provisions and laws have passed to minimize those types of environmental impacts, mainly the ones caused by hydroelectric impoundments. In some states the construction of fish ladders is mandatory in order to assist the fish upriver journey to spawning grounds (a natural phenomenon which in Brazil is known as *Piracema*) [1]. Moreover, those fish transposition mechanisms allow the acquisition of relevant information about different fish species (see Figure 1), such as swimming ability and patterns, time of migration and peak flow rates.

Many design factors must be considered when designing fish ladders. Given that every blockage in a river represents a unique situation, it is quite hard to design a mechanism that is able to accommodate all fish species. The present work was developed as part of a wider research effort [5], and its goal is to provide efficient vision-based techniques for automatic acquisition of accurate information about the behavior of fish species that swim upstream a fish ladder. Based on the proposed methodology for fish counting, a robust system has been developed, whose overview is illustrated in Figure 2. By using the information provided by that system, engineers and fish biologists can better analyze

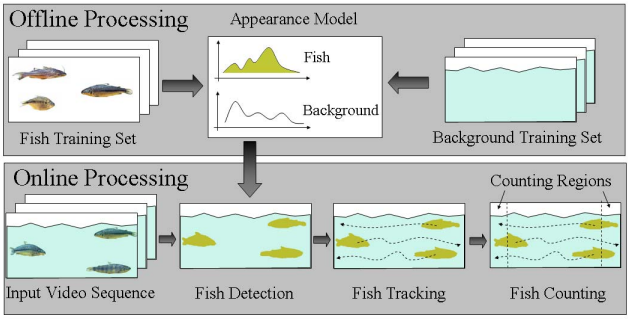

Figure 2. Overview of fish counting system developed by CTP Research Group [5].

the effectiveness of the designed transposition mechanism.

The problem we are interested in, and which lies at the background of aforementioned task of estimating the number of fishes that swim upstream a transposition mechanism, is object tracking, that is, the problem of detecting and recursively localizing objects in input video sequences. We understand that any general solution for automatic fish tracking should handle the following cases:

- **Arbitrary number of fishes**: distinct amounts of fish may enter and leave the image at any time;
- **Arbitrary fish behavior**: fishes may touch, occlude, and interact with each other;
- **Arbitrary fish size and orientation**: fish size and orientation are unknown and may be totally arbitrary;
- **3D motion**: fishes have 3D (off the ground plane) motion, making them slightly more difficult to track than people, for instance;
- **Environmental changes**: illumination variations and changes in water quality may occur;
- **Bad image quality**: the image acquisition process may be affected by noise and distortions due to the cameras optical and electronic systems;
- **Segmentation failures**: an individual fish may not be segmented reliably;

As a step towards this goal, we present a vision-based methodology that operates under all the above conditions. In particular, our approach is based on a robust Bayesian multiple-blob tracker (BraMBLe) [7], which uses a single static camera to track multiple objects as they enter, exit and move about in a scene.

Existing methodologies for fish counting are mainly based on the use of acoustic systems [9, 14, 11, 2], infrared sensors [4] or multiple counting tunnels where fishes are detected by sensing elements that measure the fluctuations in water conductivity caused by fish flows [12]. Normally, such solutions must be used in conjunction with physical structures that house several sensors. This leads to a limitation on the space in which fishes swim, forcing them to move through specified paths. Consequently, these devices may interfere in the fishes' swimming ability and even affect their moving decisions. Moreover, those methods are not recommended when the flow of fishes is large, as it usually happens in fish ladders in spawning periods. In that scenario, many fishes swim very close to each other and different individuals may activate the same sensor simultaneously leading to significative counting errors. Finally, any moving object may be counted by those systems, since they do not use any technique to differentiate fishes from other moving targets such as wood pieces.

The approach proposed in this paper does not need a physical infra-structure to limit the fish flow in order to perform the counting, since sensing is performed by a camera under skylight illumination, aided by two uniform linear light sources. Since all experiments were performed under this condition, the influence of artificial lighting was kept to a minimum, such that fish biologists did not perceive any significant impact on fish swimming patterns. The system also provides complete fish trajectories, which may be used to study fish swimming behaviors in regions of interest. The system can operate reliably under severe environmental changes, such as variations in water characteristics, and it can handle problems such as occlusions or large inter-frame motions.

Our system relies on a robust multi-target likelihood function [7], which assigns directly comparable likelihoods to hypotheses containing different numbers of objects, and a Bayesian filter for tracking the centroid of their corresponding blobs. The BraMBLe tracker [7] updates this function by searching for blobs such that all pixels inside the blobs "look like" a learned foreground model and all pixels outside the blobs "look like" a learned background model. Although a constrained vision system could alternatively be implemented, making the problem much simpler by using, for instance, artifacts for obtaining good and approximately constant lighting conditions or creating a bright and uniform background, we decided to not use those artifacts, since they affect the fish ladder's project and, thus, represent expensive solutions.

In spite of the fact that vision-based tracking of animals has countless applications in biology, so far only a handful of works have proposed effective solutions for that problem [3, 8]. Differently from our fish-tracking approach, previous works assume that the number of targets is constant and known *a priori*. Contrary to these works, our system handles temporal variations in the number of objects present in scene.

This paper is organized as follows. In Section 2 we de-

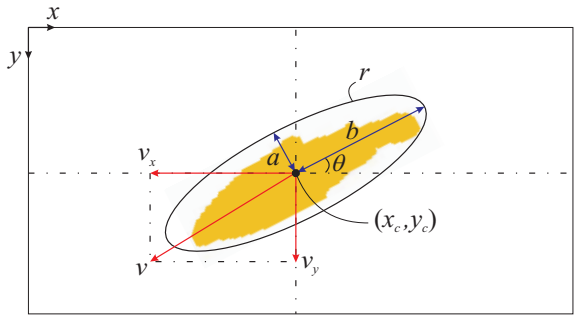

Figure 3. A fish is modelled as an ellipse and its state is described by 8 parameters: the coordinates (x_c, y_c) of the ellipse's central point, the half lengths (a, b) of the major and minor axes of the ellipse, the angle θ that measures the ellipse's rotation with respect to x-axis, a label r and the velocity components (v_x, v_y) of the fish.

scribe the observation model used by our approach for detecting fishes in the scene, which is based on the multi-blob likelihood function presented in [7]. Section 3 describes the tracking methodology used in this paper, which is based on a standard particle filter to yield *a posteriori* distribution over the number and configurations of fishes. The counting technique is presented in Section 4. Experiments with real data and their results are presented in Section 5, followed by the conclusions in Section 6.

2. Fish detection

In our fish-counting approach, the fish detection task is based on the use of a set of functions that are learned off-line, which are responsible for describing the scene appearance model, that is, the model that associates image pixels to either the scene foreground or background (see Figure 2). More specifically, by using this appearance model, we define the observation model responsible for detecting fishes, which is represented by the multi-blob likelihood function $P(I \mid S)$ described in [7] and which expresses the likelihood that a hypothesized configuration S of fishes produced an observed image I.

We represent the configuration S of a specific fish set by: $S = (n, \mathbf{s}^1, \mathbf{s}^2, \ldots, \mathbf{s}^n)$, where n is the total number of fishes present in the image plane and $\mathbf{s}^i$ is a vector encoding the state of the ith fish. By modelling a fish as an ellipse, we describe its state $\mathbf{s}^i$ with 8 parameters (see Figure 3), namely, the coordinates (x_c, y_c) of the ellipse central point, the half lengths (a, b) of the major and the minor axes of the ellipse, the angle θ that measures the ellipse's rotation with respect to image x-axis, a fish label r and the velocity components (v_x, v_y) of the centroid of the fish. Since illumination affects the image appearance and, consequently, may lead to color variation, only the chrominance information of images represented in the YUV color space is considered.

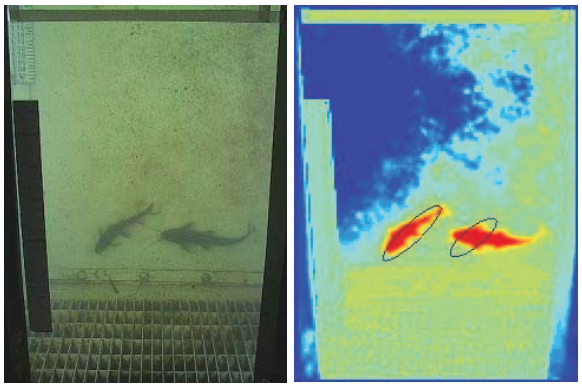

Figure 4. Log-likelihood ratios using learned foreground and background models. Redder values are more likely to be drawn from the foreground model.

The next step is similar to those in [13] and [7], where each image is overlaid with a grid of W locations spaced at 5-pixel intervals in the horizontal and vertical directions. To each image location w with coordinates (x_w, y_w) is associated an observation vector $\mathbf{z}_w$ with 4 parameters, which are obtained from the application of a Gaussian and a LoG filter to each one of the chrominance channels (U and V).

By considering the observation vectors $\mathbf{z}_w$, we define the background model by computing a Gaussian mixture for each location w. Specifically, we consider a 4-component mixture of Gaussians model [7]:

$$P(z_w \mid back) = \frac{1}{4}\sum_{k=1}^{4}\mathcal{N}\left(\mu_w^k, \Sigma_w^k\right). \qquad (1)$$

which is learned by performing a k-means clustering with k = 4 on 184 images containing only the background which were recorded at the same time as the input test sequence. The parameters μ_w^k and Σ_w^k are the mean and the diagonal approximation to the covariance of cluster k.

On the other hand, since we do not know *a-priori* the fish localizations in the image plane, we use a single foreground model for the entire image. More specifically, we use a 25-component mixture of Gaussians which were learned using k-means in the same way as the background model [7]:

$$P(z_w \mid fore) = \frac{1}{25}\sum_{k=1}^{25}\mathcal{N}\left(\mu_{fore}^k, \Sigma_{fore}^k\right). \qquad (2)$$

The training set for the foreground is generated using a training image sequence with 107 images, which contains

fishes moving against the static background of the fish ladder. Specifically, the foreground training data was acquired by applying background subtraction with images containing only the background, so that the foreground segments used in the training included only fish skin areas.

Now, before we show how to compute the multi-blob likelihood function $P(I \mid S)$ using the aforementioned foreground and background models, we would like to discuss the association of a label $l_w \in \{0, \ldots, n\}$ to each location w [7]. We associate a label $l_w = i$ ($i > 0$) to a location w if the configuration S hypothesizes that this location is centered in one of the n possible fishes present in the scene, otherwise we make $l_w = 0$, indicating that the location w is centered in the background.

By considering those labels and assuming that the observation vectors $\mathbf{z}_w$ are conditionally independent given the configuration S, we may finally define the multi-blob likelihood function $P(I \mid S)$ as follows [7]:

$$P(I \mid S) = \prod_{w=1}^{W} P(\mathbf{z}_w \mid S) = \prod_{w=1}^{W} P(\mathbf{z}_w \mid l_w). \quad (3)$$

This likelihood $P(I \mid S)$ is used as the observation model of the particle filter for fish tracking and it is responsible for assigning likelihoods to the hypothetical configurations S.

In order to optimize the computation of $P(I \mid S)$, we apply the logarithmic operation to both sides of Equation (3). This operation transforms the right side products in sums and produces a log-likelihood function L as follows:

$$L = \log(P(I \mid S)) = \sum_{w=1}^{W} h_w^l, \quad (4)$$

where $h_w^l = \log(P(\mathbf{z}_w \mid l_w))$, for $w = 1, ..., W$ and $l \in 0, 1, ..., n$. More importantly, since the particle filter described in next section requires the log-likelihood constraint only up to a multiplicative constant, the literature [7] suggests the redefinition of the log likelihood L as follows:

$$L = \sum_{w=1}^{W} (\log(P(\mathbf{z}_w \mid l)) - \log(P(\mathbf{z}_w \mid 0))), \quad (5)$$

that is, Equation (5) specifies a log-likelihood ratio comparing the hypothesis that the response was generated by an object l with the hypothesis that it was generated by the background. Observe that in this case, only foreground responses contribute to the log-likelihood, since $h_w^0 = 0$ for all w. As in [7], we consider Equation (5) in our approach. Figure 4 illustrates the log-likelihood ratios for a selected frame by considering the aforementioned foreground and background models. For additional details about the estimation process of the multi-blob likelihood function, the reader is referred to [7]. In next section, we demonstrate how to use that function together with a standard particle filter to perform robust fish tracking.

3. Fish tracking

We perform tracking of fishes represented in a configuration S_t at time t, by propagating in time the function $P(S_t \mid I_{1...t})$, which represents the *a posteriori* probability of configuration S_t, given a sequence of observations (images) from times 1 to t. The key idea behind this approach is based on the methodology presented in [6]. In that work, the function $P(S_t \mid I_{1...t})$ is defined as follows:

$$P(S_t \mid I_{1...t}) = k_t P(I_t \mid S_t) P(S_t \mid I_{1...t-1}), \quad (6)$$

where k_t is a multiplicative constant for data normalization and $P(S_t \mid I_{1...t-1})$ is given by

$$P(S_t \mid I_{1...t-1}) = \int_{S_{t-1}} P(S_t \mid S_{t-1}) P(S_{t-1} \mid I_{1...t-1}).$$

We use a particle filter to approximate the function $P(S_t \mid I_{1...t})$ in Equation (6). Specifically, we represent that function by a weighted set of N random hypotheses (particles) [6] of the configuration S. To each particle S_t^i at time t, for $i = 1, ..., N$, we assign a weight π_t^i, which is given by the log-likelihood ratio L in Equation (5). In this sense, we construct a discrete representation of $P(S_t \mid I_t)$ at time t by using the weighted particle set $\{(S_t^i, \pi_t^i) : i = 1, ..., N\}$. The accuracy of this representation depends on the number of particles used. In fact, the larger the particle set the higher will be the accuracy of that representation.

Consider now the above-described discrete representation for $P(S_{t-1} \mid I_{t-1})$. To perform its propagation to time t, we initially perform a resampling (with replacement) of the current particle set [6], generating a new set with N samples. Next, each sample S_{t-1}^i chosen for the new particle set is subjected to the prediction model $P(S_t \mid S_{t-1})$ in a predictive step. The prediction function f preserves the unique identifier r of a fish and uses the following models for the fish parameters represented in Figure 3:

$$f(r, (x_c, y_c), (a, b), \theta, (v_x, v_y)) =$$
$$(r, (x_c^{'}, y_c^{'}), (a^{'}, b^{'}), \theta^{'}, (v_x^{'}, v_y^{'})).$$
$$x_c^{'} = x_c + v_x + \delta_x, \;\; y_c^{'} = y_c + v_y + \delta_y$$
$$v_x^{'} = x_c^{'} - x_c, \;\; v_y^{'} = y_c^{'} - y_c$$
$$a^{'} = a + \delta_a, \;\; b^{'} = b + \delta_b, \;\; \theta^{'} = \theta + \delta_t,$$

where δ_x, δ_y, δ_a, δ_b, δ_t simulate white Gaussian noises with zero mean and whose corresponding standard deviations are presented in Table 1.

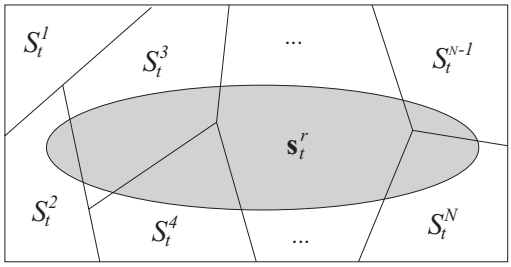

Figure 5. A specific fish r with state $\mathbf{s}_t^r$ must be considered as an element of the configuration $\hat{S}_t$, if its probability computed by using the theorem on total probability is higher than μ_s.

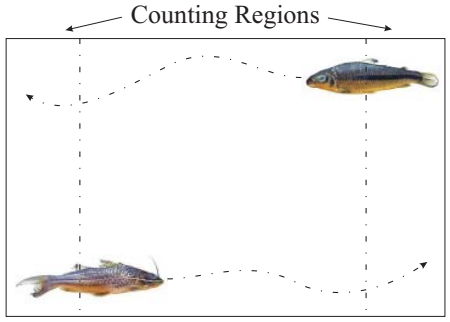

Figure 6. Illustration of the two virtual counting regions in the image plane, where the estimations of the amounts of fishes that swim up- and downstream the transposition mechanism are performed.

As in [7], our prediction model states that each fish will remain in the scene with probability μ_r at each time step, and additionally that there is a probability μ_e that a new fish will enter the scene at each time step. The parameters μ_r and μ_e work as decision parameters in the multi-object prediction algorithm proposed in [7] and are not used in any equations. The values considered for μ_r and μ_e are presented in Table 1, which are conservative values that lead to accurate tracking results.

Note that after the prediction stage, the sample set $\{S_t^i\}$ for the new time-step has been generated but, for now, without its weights. Therefore, the next step consists in assigning appropriate weights π_t^i to all the hypothetical configurations $\{S_t^i\}$. Those weights are determined by the log-likelihood ratios L that are computed according to Equation (5). Importantly, all the weights π_t^i determined by the log-likelihood function are normalized, so that they satisfy $\sum_{i=1}^{N} \pi_t^i = 1$. Therefore, the normalized weights may be considered as probability values, as in [6].

Finally, by considering the weighted particle set $\{(S_t^i, \pi_t^i) : i = 1, ..., N\}$ determined above, we must estimate a configuration $\hat{S}_t$ of that set, which leads to the best representation of an observed image I_t at time t. To determine $\hat{S}_t$ we use the theorem on total probability. More specifically, assume that the configurations S_t^1,..., S_t^N are mutually exclusive events represented by specific areas, whose corresponding sizes are given by the weights π_t^i of those configurations, as illustrated in Figure 5. Moreover, let a specific fish r with state $\mathbf{s}_t^r$ be represented by the gray ellipse in Figure 5. The size of the area of that ellipse represents the probability that the fish state $\mathbf{s}_t^r$ occurs in an observed image I_t at time t. Therefore, from the theorem on total probability, we know that this probability is given by:

$$\begin{aligned} P(\mathbf{s}_t^r) &= P(\mathbf{s}_t^r \cap S_t^1) + ... + P(\mathbf{s}_t^r \cap S_t^N) \\ &= P(\mathbf{s}_t^r \mid S_t^1)P(S_t^1) + ... + P(\mathbf{s}_t^r \mid S_t^N)P(S_t^N) \\ &= P(\mathbf{s}_t^r \mid S_t^1)\pi_t^1 + ... + P(\mathbf{s}_t^r \mid S_t^N)\pi_t^N. \end{aligned} \quad (7)$$

A specific fish r with state $\mathbf{s}_t^r$ is considered as an element of the configuration $\hat{S}_t$ if its corresponding probability $P(\mathbf{s}_t^r)$ is larger than a pre-determined threshold μ_s. The parameter μ_s determines whether or not an ellipse is displayed, and has no effect on the particle filter itself [7]. After several tests we considered $\mu_s = 0.4$, which lead to the most accurate experimental results.

Given the estimated configuration $\hat{S}_t$ for each time t, we compute the fish trajectories. It is important to notice that, as the quantity of fishes may vary over time (e.g., a fish enters or leaves the image plane at any time), each hypothesized fish configuration S_t^i must consider that possible variation. As far as the hypotheses that indicate fishes entering the scene are concerned, we assign the following values for the parameters of a new fish state $\mathbf{s}_t^r$ included in a configuration S_t at time t: $a = 20$, $b = 5$, $\theta = \pi$, $v_x = 0$, $v_y = 0$. The coordinate (x_c, y_c) of the centroid of the fish's ellipse is determined by the coordinates of the localizations w in the counting regions (see Figure 6) which have a log-likelihood ratio L larger than zero, indicating that those locations are probably centered in the foreground.

4. Counting approach

Counting fishes may now be performed as an almost effortless task. This is due to the reliable tracking phase described in the previous sections. Therefore, one only needs to define which events will lead to incrementing a fish counter. As far as the count is concerned, the overall system is able to provide as output the number of individuals for each of the selected species. However, the problem of classifying a fish as a member of a given species is outside the scope of this work, but it is thoroughly dealt with in [10]. Therefore, we focus here only on providing a total number of fishes regardless of their type. So, in the fish counting task, we divide the image plane into three regions

of interest as depicted in Figure 6. We consider the right and left regions as our fish *detection* and *counting* zones, while the central region is the *tracking zone*. Once the central point of the ellipse of a fish crosses one of the counting regions and then passes through the tracking zone and finally crosses the opposite counting region, a fish counter is incremented. In the many hundreds of hours of image taken at the inspection window on the fish ladder at UH-Igarapava, it has been observed that the dominant fish flow occurred from the right counting region to the left counting region, which is expected since this moving direction corresponds to the fish migration upstream (against the water flow) in the transposition mechanism.

5. Experimental Results

To demonstrate the effectiveness of our fish counting approach, we tested it on challenging real-world datasets, such as the one illustrated in Figure 9, where two fishes occlude each other. Image dimensions in all datasets were 320×240 pixels. The values for the parameters used in our experiments are summarized in Table 1. Particularly, the parameters σ_1, σ_2 and σ_3 were empirically defined by using our knowledge about the problem. The probability values for μ_r and μ_e were set to values used in [7]. As pointed out in [7], those two parameters are consistent with a Poisson distribution on object arrivals and an exponential distribution on their survival times. Regarding the number of particles used by the Bayesian filtering algorithm, we chose a particle set with $N = 2000$ samples. This number of particles lead to a computational cost only slightly larger than the one observed when a particle set with only $N = 200$ particles was used. However, the results returned by particle filtering were significantly different. Therefore, we decided to perform our experiments with $N = 2000$ particles, for which the results of the particle filtering algorithm was stable. This number of particles was also used in [3] to track multiple identical mice from video of the side of their cage.

Our experimental results demonstrate that the tracking algorithm accurately tracks fishes, even when their motions are particularly erratic and the distribution of their current positions given their past trajectories presents a high variance. For instance, consider Figure 7, where we show some frames of an input dataset (all datasets and results can be found at http://www.dcc.ufmg.br/∼erikson/sibgrapi-2005/) containing a single fish, as well as their corresponding log-likelihood ratios computed according to the BraMBLe algorithm. Redder values are more likely to be drawn from the foreground models. Note that the ellipses in those frames show the regions in which the fish motion and appearance were estimated, giving us a visual perception of the good tracker's accuracy. Figure 8 quantifies this visual perception by comparing the horizontal and vertical positions of the ellipse's central point estimated by the tracker with their corresponding manually defined "ground-truth" positions. Since all ground-truth positions were determined by a manual identification of the fish blobs' centroids and the central points of the computed fish ellipses do not coincide exactly with those centroids, we obtained small offsets between the estimated and manual measures, as we may see in Figure 8(a) and 8(b).

Symbol	Meaning	Value
μ_r	Fish survival probability	0.99
μ_e	New fish arrival probability	0.02
μ_s	Fish ellipse display threshold	0.4
σ_1	Standard deviation (in pixels) of the gaussian noises δ_x and δ_y added to the fish coordinates (x_c, y_c)	1.2
σ_2	Standard deviation (in pixels) of the gaussian noises δ_a and δ_b added to the fish ellipse's axes (a, b)	0.9
σ_3	Standard deviation (in degrees) of the gaussian noise δ_t added to the fish ellipse's rotation angle θ	11.5
N	Number of particles	2000

Table 1. Parameter values for experiments.

Another important evidence of the robustness of the tracking methodology is illustrated in Figure 9. In this case, two fishes occlude each other and despite the occurrence of such an occlusion, the tracker successfully follows both fishes. This result is particularly important for demonstrating the actual applicability of our approach in fish counting applications where fishes occlude each other severely and often, especially when the fish flow is large, what usually happens in fish ladders during spawning periods. Finally, Figure 10 illustrates the robustness of the tracker to complex fish motions, characterized by quick turns and accelerations that are not fit by our simple dynamics model.

By using several input video sequences with distinct fish amounts, our approach was successfully validated and achieved an overall accuracy as high as 81%. The observed percentage of counting errors (19%) was determined by the occurrence of false negatives. Considering that the main goal of our fish counting system consists in providing engineers and fish biologists with information about the adequacy and efficiency of the designed transposition mechanism for different fish species, we note that the observed accuracy of 81% may be considered as a satisfactory estimation, since it provides a reasonable approximation of the actual fish flow.

6. Concluding Remarks

Our experimental results suggest that the Bayesian tracking algorithm used, whose development is based on the mul-

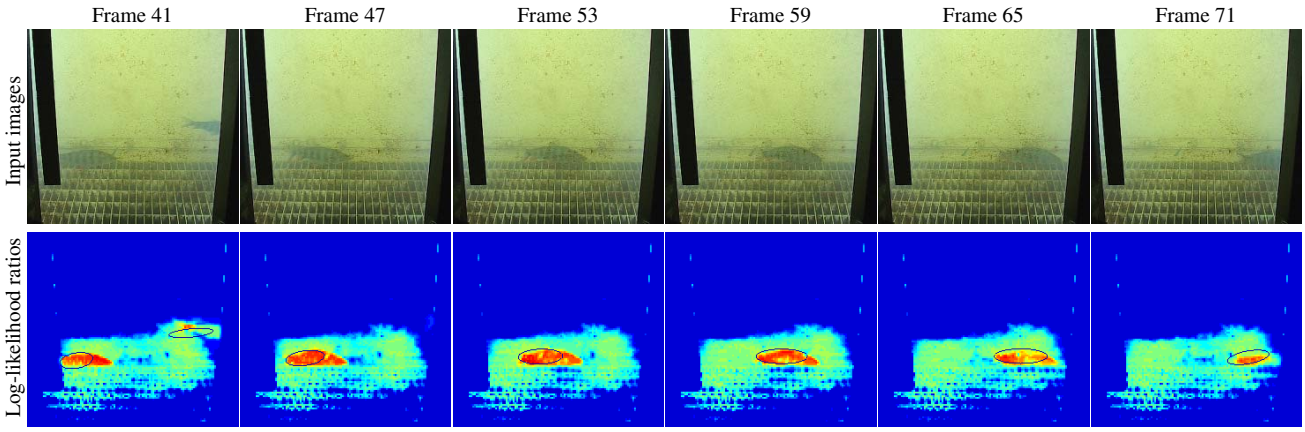

Figure 7. Frames 41, 47, 53, 59, 65 and 71 of an input dataset containing a single fish, as well as their corresponding log-likelihood ratios computed according to the BraMBLe algorithm. Note that the ellipses in those frames show the regions in which the fish motion and appearance were estimated, giving us a visual perception of the good tracker's accuracy.

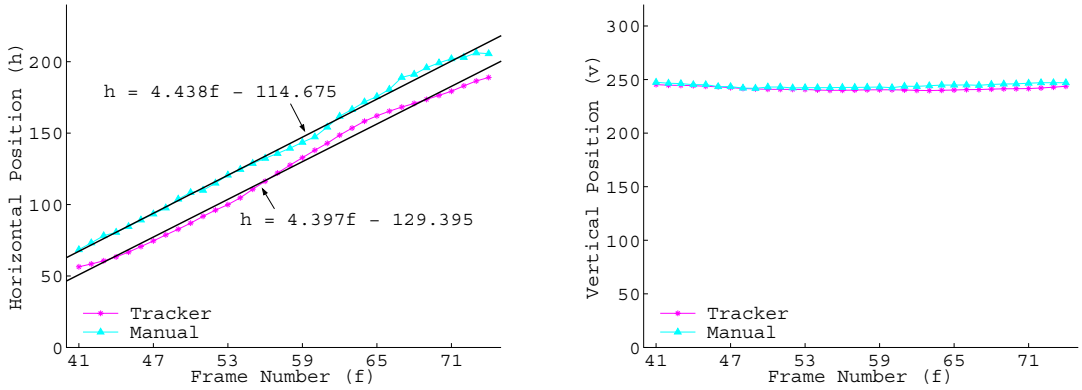

Figure 8. Quantification of the quality of the fish tracking executed in Figure 7. One may compare the horizontal and vertical positions of the ellipse's central point estimated by the tracker with their corresponding manually defined "ground-truth" positions.

titarget likelihood function proposed in [7], can be successfully applied for performing accurate fish counting in transposition mechanisms. Unlike previous fish-counting works, our approach provides adequate means for the acquisition of relevant information about characteristics of different fish species such as swimming ability, time of migration and peak flow rates. Our system is also able to estimate fish trajectories over time, which can be further used to study their behaviors when swimming in regions of interest. The experimental results demonstrate that the proposed method can operate reliably under severe environmental changes and handle problems such as occlusions. Our approach was successfully validated with real-world video streams, achieving overall accuracy as high as 81%. We are currently adapting the implementation code of our fish counting system, so that it can be executed in real-time together with the fish classification module developed in [10].

Acknowledgments The authors thank Kristin Branson and Serge Belongie for providing their implementation of BraMBLe, which was used as the basis for the development of our fish-counting system. They also thank Michael Isard for his help in finding this implementation of BraMBLe. Mario Campos, Flávio Pádua and Rodrigo Carceroni thank the support of CNPq-Brazil and FAPEMIG. Erikson Morais gratefully acknowledges the support of DOCTUM Institute in Caratinga, MG, Brazil.

References

[1] H. G. and. A. Godinho. Fish Communities in Southeastern Brazilian River Basins Submitted to Hydroelectric Impoundments. *Acta Lim. Brasiliensia*, 5:187–197, 1994.

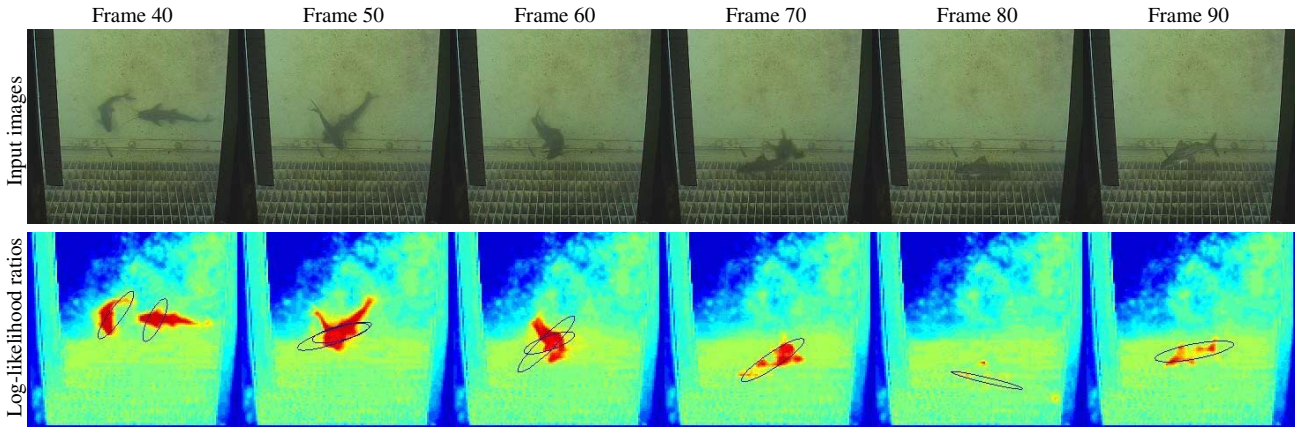

Figure 9. Robust tracking despite occlusion. Tracking results for frames 40, 50, 60, 70, 80 and 90 are shown. In this case, two fishes occlude each other and despite the occurrence of such an occlusion, the tracker successfully follows both fishes.

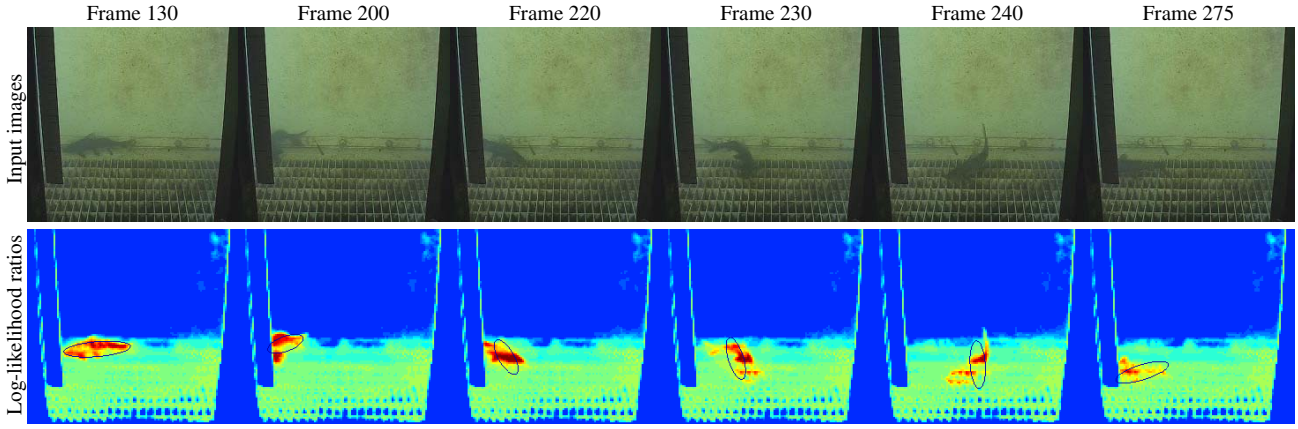

Figure 10. Illustration of the robustness of the tracker to complex fish motions, characterized by quick turns and accelerations that are not fit by our simple dynamics model.

[2] H. Braithwaite. Sonar fish counting. In *European Inland Fisheries Advisory Commission Symp.*, 2–4 May 1974.

[3] K. Branson and S. Belongie. Tracking multiple mouse contours (without too many samples). In *Proceedings of IEEE Computer Vision and Pattern Recognition Conference*, San Diego, June 2005.

[4] S. Cadieux, F. Lalonde, and F. Michaud. Intelligent System for Automated Fish Sorting and Counting. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2000.

[5] Centro de Transposição de Peixes. Endereço eletrônico: http://www.ctpeixes.ufmg.br/. Instituições de fomento: UFMG, CEMIG e FUNBIO, 2005.

[6] M. Isard and A. Blake. Condensation – conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.

[7] M. Isard and J. MacCormick. Bramble: A bayesian multiple-blob tracker. In *Proc. IEEE Int. Conf. on Computer Vision*, pages 34–41, 2001.

[8] Z. Khan, T. Balch, and F. Dellaert. An mcmc-based particle filter for tracking multiple interacting targets. In *Proceedings of IEEE European Conference on Computer Vision*, volume 4, pages 279–290, Prague, May 2004.

[9] A. Menin and R. D. Paulus. Fish counting by acoustic means. In *IEEE Oceans*, volume 6, pages 166–168, 1974.

[10] M. S. Nery. Classificação de espécies de peixes da bacia do rio grande baseada em visão computacional. Master's thesis, UFMG, Departamento de Ciência da Computação, Belo Horizonte, MG, Brasil, May 2004.

[11] D. G. Pincock and N. W. Easton. The feasibility of doppler sonar fish counting. *Proc. IEEE Journal of Oceanic Engineering*, OE–3(2):37 – 49, April 1978.

[12] I. Smith-Root. Fish counting systems. `http://www.smith-root.com/`, 2005.

[13] J. Sullivan, A. Blake, M. Isard, and J. MacCormick. Object localization by bayesian correlation. In *Proc. IEEE Conf. on Computer Vision*, volume 2, pages 1068–1075, 1999.

[14] B. White. A transportable acoustic fish census system for lake and river studies. In *Proc. IEEE Oceans*, volume 8, 1976.

[15] E. Woynarovich. The Hydroelectric Power Plants and the Fish Fauna. *Verh. Int. Verein. Limnol.*, 24:2531–2536, 1991.