MVA2011 IAPR Conference on Machine Vision Applications, June 13-15, 2011, Nara, JAPAN

Detection of Swimmer Based on Joint Utilization of Motion and Intensity Information

K. L. Chan
Department of Electronic Engineering, City University of Hong Kong
83 Tat Chee Avenue, Kowloon, Hong Kong
itklchan@cityu.edu.hk

Abstract

This paper presents a method for detecting swimmers in swimming pool. A vision based system for locating individual swimmers and recognizing the activities demands accurate detection of swimmer's body parts such as head and limbs. Swimmer detection can be regarded as a background subtraction problem, which is very difficult due to a dynamic background with ripples, splashes, specular reflections, etc. Our method utilizes both local motion and intensity information estimated from the image sequence. Local motion information is obtained by computing optical flow and periodogram. We devise a scheme that can coarsely classify the image pixels into three types of motion – random/stationary, ripple and swimming. This motion map characterizes the local motion of image pixels over a short duration. We adopt the mixture of Gaussians for modeling the intensity information. Finally, by using the motion map and the Gaussian models, swimmers are detected in each video frame. Thorough tests have been performed using videos captured at daytime and nighttime, and different swimming styles (breaststroke, freestyle). Our method can detect swimmer better than using intensity information alone.

1. Introduction

To detect and track human, and subsequently to analyze and recognize human motion automatically from video sequence are challenging research topics. They are the principal concerns in various applications such as video surveillance [1], gait analysis [2], video segmentation and retrieval [3], etc. We intend to develop a vision based system for locating individual swimmers and recognizing the activities (e.g. swimming styles, drowning event). This paper presents a method for detecting swimmers in swimming pool. Swimmer detection in the aquatic environment is very difficult due to a dynamic background. The water surface exhibits random motion and ripples. Specular reflections are commonly seen in the captured video frames. That means the background modeling/subtraction, a common first step in motion analysis for detecting the foreground humans, is very difficult.

Ning *et al.* [4] use the least median of squares method for modeling individual pixels of the background scenes which are mainly static. Li *et al.* [5] propose a method for modeling background by the principle features and their statistics. The principle features for static background are color and gradient, while dynamic background is characterized by color co-occurrence. Foreground and background are classified and detected by a Bayesian framework. To detect and recognize human motion in the aquatic environment is more challenging. The background is totally non-stationary. Lu and Tan [6] model the entire background (the swimming pool) with a mixture of Gaussian distributions. The method has the advantages of minimum memory required to store the background model and is easier to update. Eng *et al.* [7] propose to model the background as a composition of homogeneous blob movements. The method can adapt to change of background regions over time. The problem of specular reflection at nighttime is also tackled by a filtering module. These methods use intensity information. Background motion such as ripple is not represented in the background models.

Figure 1 shows an overview of our swimmer detection method. Our method utilizes both local motion and intensity information estimated from the image sequence. We devise a scheme that can coarsely segment the image pixels into three types of motion – random/stationary, ripple and swimming. This motion map characterizes the local motion of image pixels over a short duration. We adopt the mixture of Gaussians for modeling the intensity information. Finally, by using the motion map and the Gaussian models, swimmers are detected in each video frame.

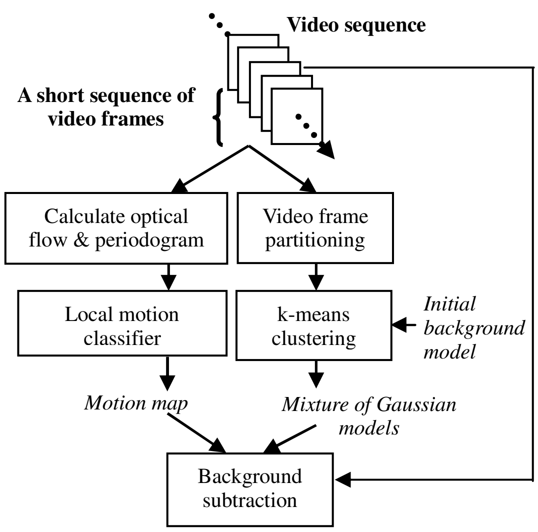

Figure 1. System overview.

2. Local Motion Information

In order to perform background subtraction for aquatic environment, the background motions of water must be characterized in the background model. According to airy wave theory, the elevation of waving water η can be assumed to be superposition of sinusoidal waves, each as a function of position (x, y) and time t, with different wavelengths and velocities:

$$\eta(x,y,t)=\sum_i a_i sin(\mu_i x+v_i y-\omega_i t) \quad (1)$$

where (μ_i , v_i) is the wavenumber, ω_i is the angular frequency, a_i is the amplitude. The surface normal is:

$$n=\frac{[p,q,1]^{\mathrm{T}}}{\sqrt{p^2+q^2+1}} \quad (2)$$

where p, q are surface gradients at a certain position (x, y) and can be obtained by:

$$p=-\frac{\partial\eta}{\partial x}=-\sum_i a_i\mu_i cos(\mu_i x+v_i y-\omega_i t) \quad (3)$$

$$q=-\frac{\partial\eta}{\partial y}=-\sum_i a_i v_i cos(\mu_i x+v_i y-\omega_i t) \quad (4)$$

Assume the water surface receives direct illumination which is stationary over a short duration. Scene radiance, therefore, depends on surface normal. The observed intensities of waving water over this duration can also be modeled as superimposed sinusoidal waves.

The motion field of each video frame can be estimated via optical flow computation. Given the observed intensity I = $I(x, y, t)$, the motion field v = $(v_x, v_y)^{\mathrm{T}}$, and the assumption of brightness constancy,

$$(\nabla I)^{\mathrm{T}} v+I_t=0 \quad (5)$$

Each motion vector is computed within a small patch of N points.

$$v=\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}b \quad (6)$$

where $A=\begin{bmatrix} I_x(x_1,y_1) & I_y(x_1,y_1) \\ \vdots & \vdots \\ I_x(x_N,y_N) & I_y(x_N,y_N) \end{bmatrix}$ and $b=-\left[I_t(x_1,y_1),\ldots,I_t(x_N,y_N)\right]^{\mathrm{T}}$. In our method, we set the patch size to 15 x 15 pixels and all gradient terms are computed by centered difference method.

We expect that the motion field of waving water over a short duration is characterized as superposition of sinusoidals, while other parts of the scene exhibiting no motion, random motion or swimming are not. For each scene point, periodogram is computed for each zero-meaned Gaussian tapered optical flow component over this duration. Figure 2 shows one video frame of a swimmer swimming in breaststroke style. Samples of ripple, randomly moving water surface, and swimmer's head and hand are selected as shown by the "**o**", "**x**", and "**+**" marks respectively. Figures 3-5 show the periodograms of the selected points computed from a short duration of 41 frames. Periodograms of ripples exhibit fundamental and harmonic frequencies. Periodograms of randomly moving water exhibit small and random fluctuations. The swimmer's head and hands move rapidly and the corresponding periodograms exhibit a single peak at low frequency.

Based on these observations, we devise a classification scheme for coarsely representing the local motion in a motion map. In each periodogram, the peak frequencies are identified. Those peaks with amplitude less than 10% of the range are rejected. To find out the harmonic relationship among the peak frequencies for ripple point identification is possible but too tedious. Instead it is easier to identify the random motion and the strong linear motion of swimmer's head. The remaining points can be temporary considered as ripples. Based on the structural knowledge, the points of swimmer's shoulders and hands can be identified in the second phase. In the first phase, we try to roughly identify the stationary/random motions, ripples and swimmer's head. If the mean amplitude of the periodogram is less than a certain threshold t_r, that point is regarded as stationary/random motion. If the amplitude of the first peak is larger than a certain threshold t_h and the periodograms of neighboring points are mainly of the same type, that point is regarded as the swimmer's head. Similarity of neighboring periodograms is measured by counting the number of peaks in eight neighboring periodograms which should be less than or equal to eight. Points not identified as either of two previous classes are temporary regarded as ripple. In the second phase, we further identify the swimmer's shoulders and hands. The point nearby the swimmer's head with the amplitude of the maximum peak of periodogram higher than a certain threshold t_s is regarded as swimmer's shoulder. Then, the point nearby the swimmer's shoulder with the amplitude of the maximum peak of periodogram higher than a certain threshold t_a is regarded as swimmer's hand.

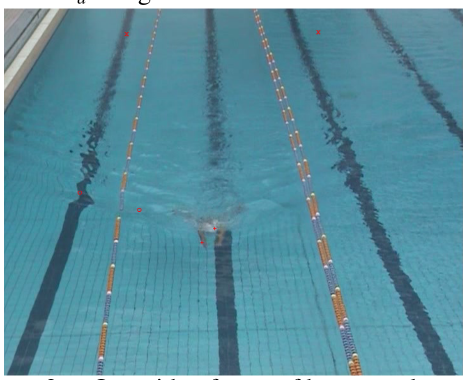

Figure 2. One video frame of breaststroke swimming with sample points.

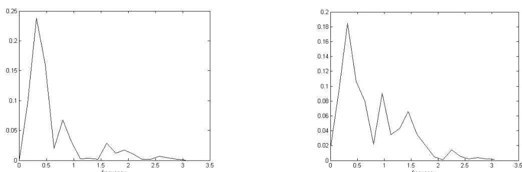

Figure 3. Periodograms of ripple.

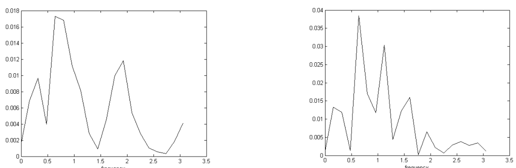

Figure 4. Periodograms of randomly moving water.

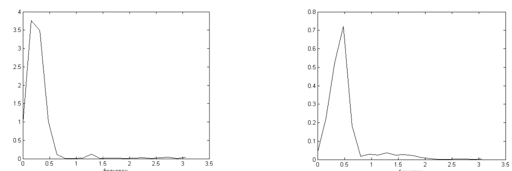

Figure 5. Periodograms of swimmer's head (left) and hand (right).

Figure 6a shows the motion map of swimmer's head of the 41-frame breaststroke swimming sequence. White color represents the trajectory of swimmer's head. The threshold t_h is 3. If the threshold is too small, there will be points wrongly identified as swimmer's head as shown in Figure 6b. If the threshold is too large, there will be much less points identified as swimmer's head as shown in Figure 6c. Figure 7a shows the motion map of swimmer's shoulders. The threshold t_s is 0.1. The search range from swimmer's head is +/- 15 pixels. Figure 7b shows the motion map of swimmer's hands. The threshold t_a is 0.2. The search range from swimmer's shoulder is +/- 30 pixels. Figure 8 shows the final motion map. The stationary/random motion points are labeled as grey. The threshold t_r is 0.04. The ripple points are labeled as black. The points of swimmer are labeled as white.

(a) (b) (c)

Figure 6. Motion map of swimmer's head identified at (a) appropriate threshold, (b) small threshold, (c) large threshold.

(a) (b)

Figure 7. Motion map of (a) swimmer's shoulders and (b) hands.

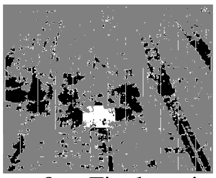

Figure 8. Final motion map.

3. Intensity Information And Detection Of Swimmer

We adopt and simplify the method of Eng *et al.* [7] for modeling local intensity information. Each video frame is partitioned into n_1 x n_2 nonoverlapping blocks $B_{a,b}$, where $1 \le a \le n_1$ and $1 \le b \le n_2$. All pixels of the same block position (a, b) over a short duration are collected and clustered via k-means into H homogeneous regions. Each homogeneous region of that block position is characterized by the mean and standard deviation of the Gaussian distribution, $\mu_{B^h_{a,b}} = \left\{\mu^R_{B^h_{a,b}} \quad \mu^G_{B^h_{a,b}} \quad \mu^B_{B^h_{a,b}}\right\}^T$ and $\sigma_{B^h_{a,b}} = \left\{\sigma^R_{B^h_{a,b}} \quad \sigma^G_{B^h_{a,b}} \quad \sigma^B_{B^h_{a,b}}\right\}^T$ respectively, where $1 \le h \le H$. We set the block size to 30 x 32 pixels.

Clean background frames are generated by vector median filtering using the same video. Figure 9 shows one clean background frame. Initial background model is generated from these clean background frames, also via k-means. The background models are updated by:

$$\mu^t_{B^h_{a,b}} = (1-\rho)\mu^{t-1}_{B^h_{a,b}} + \rho\mu^t_{B^h_{a,b}} \qquad (7)$$

$$\sigma^t_{B^h_{a,b}} = (1-\rho)\sigma^{t-1}_{B^h_{a,b}} + \rho\sigma^t_{B^h_{a,b}} \qquad (8)$$

As for the breaststroke video, we select 12 frames at an interval of 10 frames for each clean background generation. We use five clean background frames for initial background model training. The update factor ρ is 0.2 and the background model is updated at an interval of 5 frames.

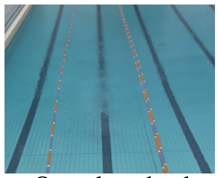

Figure 9. One clean background frame obtained from breaststroke swimming sequence.

The similarity of a pixel with respect to a homogeneous background region can be computed using the distance measure:

$$D(I_{x,y} \mid \mu_{B^h_{a,b}}, \sigma_{B^h_{a,b}}) = \sqrt{\sum_{c=R,G,B} \frac{(I^c_{x,y} - \mu^c_{B^h_{a,b}})^2}{(\sigma^c_{B^h_{a,b}})^2}} \qquad (9)$$

If the distance of the current pixel with respect to any homogeneous background region in a search space of 5 x 5 blocks is below a certain threshold, it is regarded as a background pixel. As for the breaststroke video, we set the threshold to 3.46.

Our swimmer detection method utilizes both local motion and intensity information. According to the motion map, the white label region is temporary regarded as the swimmer. Water ripples and stationary/randomly moving water surface are surely regarded as background. To detect the outline of swimmer in each video frame, each foreground point is examined with respect to the similarity of local intensity of background regions. The pixel is changed to background if the distance measure is less than the threshold. As for the breaststroke video, the best threshold is set to 2.65. The top row of Figure 10 shows some original video frames in one cycle of breaststroke swimming. The middle row shows the results obtained using local intensity information only. Many background points, especially in the water ripples and splashes nearby the swimmer, are erroneously regarded as foreground. The bottom row shows the results obtained by our method. Erroneous foreground points are drastically reduced. The outline of the swimmer is identified much better.

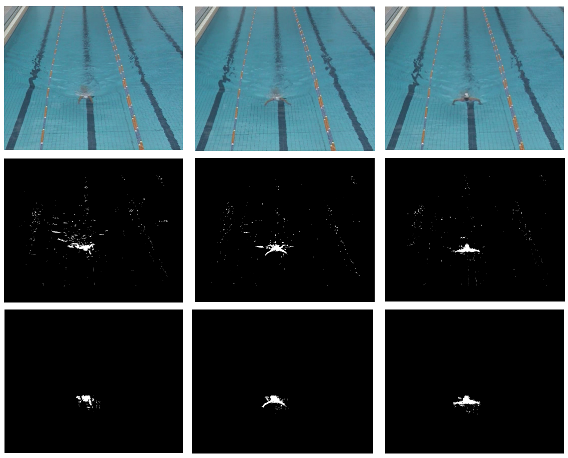

Figure 10. Results of swimmer detection.

4. Results

The second testing video is also a breaststroke swimming. It was captured at close up view and at nighttime. There are strong specular reflections in each video frame. The swimmer swims slowly and the amplitude of swimmer's movement does not differ much from the water surface. The top row of Figure 11 shows some original video frames. The middle row shows the results obtained using local intensity information only. The best threshold is set to 4.47. Many background points, especially in the water ripples and specular reflections, are erroneously regarded as foreground. The bottom row shows the results obtained by our method. The best threshold is also set to 4.47.

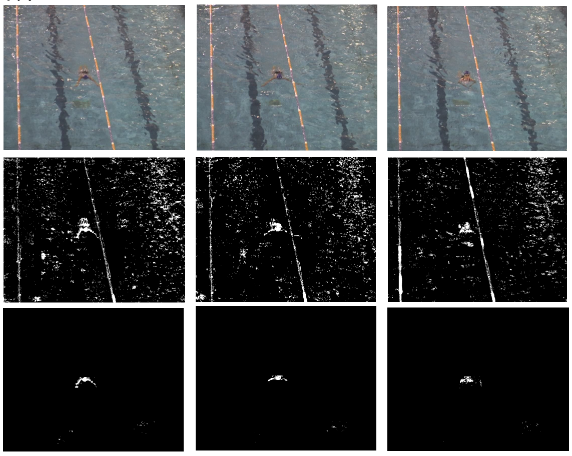

Figure 11. Results of swimmer detection in nighttime breaststroke swimming sequence.

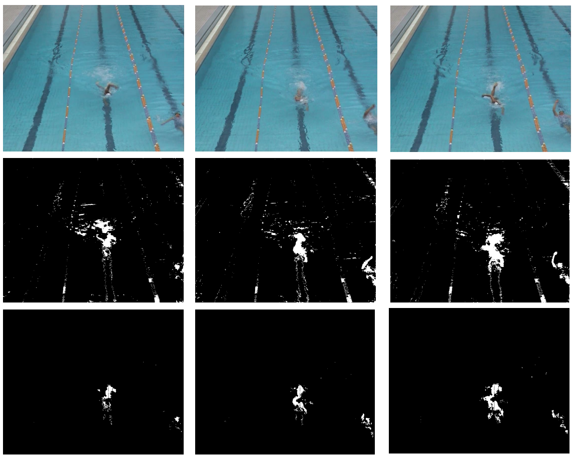

Figure 12. Results of swimmer detection in freestyle swimming sequence.

The third testing video captures two swimmers. The swimmer at the centre swims quickly in freestyle. Another swimmer at the right swims slowly in backstroke. The top row of Figure 12 shows some original video frames. The middle row shows the results obtained using local intensity information only. The best threshold is set to 3.46. Many background points, especially in the water ripples and splashes, are erroneously regarded as foreground. The bottom row shows the results obtained by our method. The best threshold is also set to 3.46.

5. Discussion And Conclusion

The motion map is generated from a short sequence of about 1 cycle of the swimming movement. If the sequence is too short, there will be insufficient optical flow data. If the sequence is too long, each periodogram will contain many different types of motion that render the classification complicated. At present, our method cannot detect swimmers in different style and speed simultaneously. In the future, we need to extent the motion map generation step for tackling multiple targets. Our method is computationally more demanding than traditional methods that only use intensity information. However, the utilization of motion information gives rise to much better results, especially in tackling various kinds of background defects such as water ripples, splashes and specular reflections.

To conclude, we develop a method for detecting swimmers in swimming pool which utilizes both local motion and intensity information estimated from the image sequence. Local motion information is obtained via optical flow and periodogram computation. We adopt a heuristic approach to generate a motion map characterizing the local motion of image pixels over a short duration. Intensity information is modeled by a mixture of Gaussians. By using the motion map and the Gaussian models, our method can detect swimmers better than using intensity information alone.

Acknowledgment: The work described in this paper was fully supported by a grant from City University of Hong Kong (Project No. 7002476).

References

[1] J. –W. Hsieh, Y. –T. Hsu, H. –Y. M. Liao, and C. –C. Chen: "Video-based human movement analysis and its application to surveillance systems," IEEE Transactions on Multimedia, Vol. 10, No. 3, pp. 372-384, 2008.

[2] D. Cunado, M. S. Nixon, and J. N. Carter: "Automatic extraction and description of human gait models for recognition purposes," Computer Vision and Image Understanding, Vol. 90, pp. 1-41, 2003.

[3] C. M. Lu and N. J. Ferrier: "Repetitive motion analysis: segmentation and event classification," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26, No. 2, pp. 258-263, 2004.

[4] H. Ning, T. Tan, L. Wang, and W. Hu: "Kinematics-based tracking of human walking in monocular video sequences," Image and Vision Computing, Vol. 22, pp. 429-441, 2004.

[5] L. Li, W. Huang, I. Y. –H. Gu, and Q. Tian: "Statistical modeling of complex backgrounds for foreground object detection," IEEE Transactions on Image Processing, Vol. 13, No. 11, pp. 1459-1472, 2004.

[6] W. Lu and Y. P. Tan: "A vision-based approach to early detection of drowning incidents in swimming pools," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, No. 2, pp. 159-178, 2004.

[7] H. -L. Eng, J. Wang, A. H. K. Siew Wah, and W. -Y. Yau: "Robust human detection within a highly dynamic aquatic environment in real time," IEEE Transactions on Image Processing, Vol. 15, No. 6, pp. 1583-1600, 2006.