Integrating Color and Motion to Enhance Human Detection within Aquatic Environment

Junxian Wang, How-Lung Eng, Alvin H. Kam and Wei-Yun Yau
Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613

Abstract

In this paper, an adaptive spatio-temporal filtering scheme based on a novel concept of motion frequency is proposed as the preprocessing step to enhance human detection under noisy aquatic environment. In this framework, each pixel is first classified into one of three categories quantified by its motion frequency, each of which is filtered using an appropriate filtering scheme. For regions affected by glistening reflections and glares, a color compensation filter is specially developed to improve partly hidden human detection and minimize errors due to moving background elements. Additionally, a blob-based verification procedure is introduced to remove the target's shadow. Experimental results demonstrate the effectiveness of the algorithm and its role in enhancing the robustness of an aquatic surveillance system for outdoor swimming pools at nighttime.

1 Introduction

With the advances in real-time computing, automated video surveillance is progressively gaining prominence as an important practical technology. There are generally three fundamental parts in building such a surveillance system: change detection, tracking and event inference. Effective implementation of tracking and sophisticated inference are heavily dependent on a highly reliable change detection process. Therefore, a number of methodologies have been proposed for reliable human detection, such as background subtraction [1], optical flow computation [2] and temporal differencing [3][4]. However, one of the major difficulties continuously faced by most state-of-art systems is that human targets may be partially hidden by specular reflections or sun glares under highly dynamic aquatic environment.

In order to reduce the effect of light specular reflection, much work has been done using approaches based on a polarizing filter [5] or color spaces such as HSV which is considered to be invariant to brightness. Fujikake [5] exploited polarizing filter by setting it in front of a camera to reduce light specular reflection. This method, however, needs to detect incident polarization angle of specular reflection as optical axis of the filter. In order to overcome this disadvantage of a polarizing filter, the spatio-temporal filter, which includes motion-compensation and motion-adaptive filters, was widely exploited for noisy video sequences. In contrast with motion-compensation filtering, motion-adaptive filtering utilizes a motion-detection scheme, but does not require explicit estimation of interframe motion vectors. The parameters of motion-adaptive filter can be tuned according to a motion detection signal. However, a drawback of the adaptive motion filtering is that it may easily generate some artifacts in the filtered image [6].

In order to detect partly hidden targets by light specular reflection while providing effective artifact suppression, in our paper, we develop an adaptive spatial-temporal filter and propose a color compensation scheme based on a novel concept of motion frequency. According to such motion information of pixels, color information of hidden parts of targets are compensated and further employed to human detection by block-based color difference detection algorithm. In order to evaluate the performance of the proposed scheme, extensive detection experiments have been designed and performed in swimming pools at nighttime. Experimental results demonstrate the effectiveness of the algorithm and its role in enhancing the robustness of an aquatic surveillance system for outdoor swimming pools at nighttime.

2 An adaptive color compensation based on motion frequency

The proposed filtering scheme utilizes both the spatial and temporal information to compensate value of partly hidden human by generating 'pseudo color'. For each pixel of a frame, it is first classified to be one of three identified pixel types quantified by motion frequency, then appropriate filter is invoked accordingly.

2.1 Motion frequency

The block-based color difference detection algorithm utilizes a motion-detection scheme to identify motion-compensated blocks. By analyzing the motion information of each pixel, the algorithm identifies three types of pixels: background, body pixel of foreground in reflection region and pixel of artificial lighting. Each of these pixel types is further classified into two categories: visible and invisible. Visible pixels are those that are clearly visible in the image, while invisible pixels are those that are partially hidden by specular reflections or sun glares.

In order to enhance the robustness of the algorithm, a motion frequency concept is introduced. Each pixel is classified into one of three categories based on its motion frequency: low, medium or high. The motion frequency of a pixel is calculated based on the interframe difference of the pixel and its neighbors. Pixels with low motion frequency are those that remain relatively static over time, while pixels with high motion frequency are those that undergo significant motion. The algorithm then applies an appropriate filter to each category of pixels to suppress artifacts and enhance the detection of partly hidden targets.

Experimental results demonstrate the effectiveness of the algorithm and its role in enhancing the robustness of an aquatic surveillance system for outdoor swimming pools at nighttime.
tuation, and iii) fast moving foreground, sensor and digitization noise have a high signal fluctuation. Thus, such observations suggest that the fluctuation of pixel value over time exhibits a useful motion property to be exploited for classification of pixels into various classes. In the following, we first propose a novel motion frequency measure to quantify such pixel fluctuation property for the classification pixels.

Let \( f(x, y, i - N), f(x, y, i - N + 1), \ldots, f(x, y, i) \) be intensity values of pixel at spatial location \((x, y)\) for \(N + 1\) consecutive frames, motion frequency, \(\varpi(x, y, i)\), is given by:

\[
\varpi(x, y, i) = \frac{1}{M} \sum_{j=0}^{N} P(x, y, j).
\]

where \(M\) is the sample rate of video sequence, \(D(x, y, i, j)\) is interframe difference of a pixel in spatial location \((x, y)\),

\[
D(x, y, i, j) = |f(x, y, i) - f(x, y, j)|.
\]

\(f(x, y, j)\) is the intensity of a pixel in the reference image. \(T\) is an experimentally determined threshold.

### 2.2 Filtering scheme

The motion frequency of a pixel provides motion information about moving targets, background and lighting reflection. Based on different motion information of pixels, different filtering schemes are then invoked accordingly. \(p(x, y, i)\), the output intensity after adaptive spatiotemporal filtering with respect to the current frame’s intensity \(f(x, y, i)\), is thus given by:

\[
p(x, y, i) = \begin{cases} 
  s(x, y, i); & \text{mean filter} \\
  g(x, y); & \text{color compensation filter} \\
  f(x, y, i); & \text{no filter} 
\end{cases} \quad \varpi > T_a, \quad T_b < \varpi \leq T_a,
\]

\[
p(x, y, i) = \begin{cases} 
  s(x, y, i) + \varpi; & \text{mean filter} \\
  g(x, y); & \text{color compensation filter} \\
  f(x, y, i); & \text{no filter} 
\end{cases} \quad \varpi \leq T_b.
\]

where \(\varpi\) is the motion frequency of pixels at spatial position \((x, y)\). \(T_a\) and \(T_b\) are experimental thresholds, respectively.

Figures inside a \((2n+1) \times (2n+1)\) window centered at position \((x, y)\) are smoothed by the mean filter, \(s(x, y, i)\).

The color compensation filter thus provides color estimation based on the dominant color of a pixel across consecutive frames. The filter is specially effective in removing the effects of light reflections glinting from the pool’s surface. It however occasionally produces erroneous colors for moving areas of a dynamic background (which are categorized as compensated regions based on the motion frequency criteria). These errors inevitably lead to falsely classified foreground areas by the subsequent target detection algorithm; nevertheless these erroneous regions are always small in size and can be removed using techniques based on connected-component analysis.
3 Target detection and tracking

In this section, a brief overview will be given about our target detection and tracking algorithm for hostile outdoor pool environments [7]. The overall architecture of the methodology is shown in Figure 1. It is basically comprised of three main tasks: 1) preprocessing the video frame using adaptive spatio-temporal filtering, as described in the previous section and generating a block-based background model; 2) extracting swimmers using block-based color difference algorithm and thresholding-with-hysteresis methodology; 3) swimmer tracking while maintaining background model through rapid updating.

In the preprocessing phase, spatio-temporal filtering is applied on each pixel in the current frame in the manner described previously. The detection phase actually consists of a ‘learning’ and a ‘normal’ mode. In the learning mode, temporal median filtering and a skin color model are used to isolate swimmer pixels to build a relatively clean background reference. A block-based methodology is then used to build the final background model after it captures the large-scale spatial dependencies and non-stationarity features of the background pixels. In the normal mode, a thresholding-with-hysteresis approach [8] is utilized to select thresholds for the framework of background subtraction. This approach fills in the gaps between high-confidence foreground swimmer regions in a robust and meaningful way compared to using morphological operators. Detected swimmers are tracked based on minimum spatial Mahalanobis distances between consecutive frames. We found this scheme to be fairly robust and effective in handling fragmented swimmer parts due to detection errors.

In the detection phase, object formation is completed by the blob-based verification procedure. The verification process is performed by considering the whole color distribution of pixels in the blob, rather than pixel-based color property. Generally, a person under pool environment has multi-color regions (e.g., dark color of the head, skin color of the body and various colors of the swimwear). The color distribution of pixels in the shadow region is more uniform than the real targets, thus, the range of color variation of pixels in the blob is exploited to distinguish the target and the shadow. Figure 2 demonstrates that the swimmer’s shadow is removed based on the standard deviation of blob’s color distribution. The 1st row of Figure 2 is a sample frame of a typical nighttime sequence of an outdoor pool and detected blobs; 2nd row: blobs’ color distributions and detected swimmers while removing the swimmer’s shadow.

4 Experimental results and analyses

A real-time surveillance system has been set up on trial for continual monitoring at an outdoor Olympic size public swimming pool from early morning to nighttime. Note that light specular reflections gradually appeared at nighttime because overhead lamps in the swimming pool were turned on for lighting the monitored region.

Extensive testing on the proposed algorithm has been carried out in the swimming pool during nighttime as shown in Figure 3. Promising segmentation results have been obtained from 6pm to 8pm while water reflections were spread from sparse region to concentrated region. Figure 4 pro-

Figure 2: Removing a shadow of a swimmer based on the standard deviation of blob’s color distribution. 1st row: sample frame of a typical nighttime sequence of an outdoor pool and detected blobs; 2nd row: blobs’ color distributions and detected swimmers while removing the swimmer’s shadow.

Figure 3: Detected swimmers under aquatic environment at nighttime from 6pm to 8pm. Water reflections were spread from sparse region to concentrated region.
vides a good comparison of the detection performance of the system without and with the proposed filtering scheme. As it shows in the 2\textsuperscript{nd} row of Figure 4, without the pre-
processing scheme, it is very difficult to correctly extract parts of swimmers that are occluded by light specular re-
flections. Moreover moving elements of the dynamic back-
ground such as the apparent movements of the lane dividers
due to water ripples are mistakenly classified as foreground
regions. If the proposed adaptive spatio-temporal filtering
is implemented, detection results are improved as shown in
the 3\textsuperscript{rd} row.

![Figure 4: A comparison of detected swimmers without and with filtering.](image)

To evaluate the performance quantitatively, we ex-
plot the false positive rate (FPR) [9]:
\[
\text{FPR} = \frac{\text{Number of segmented foreground}}{\text{Number of benchmark foreground}}.
\]
The proposed algorithm

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
Video & Only color without & Color and motion & Proposed method \\
& filtering & without compensation & \\
\hline
\hline
Video 1 & 38.1\% & 61.3\% & 85.8\% \\
Video 2 & 47.7\% & 67.3\% & 85.4\% \\
\hline
\end{tabular}
\caption{Detection performance comparison using color information without filtering, color and motion information without color compensation and proposed spatio-temporal filtering scheme.}
\end{table}

was tested on two video sequences of length 2000 frames
each (4 frames per second). Table 1 shows the detection
performance comparison based on color without filtering, color and motion without color compensation and proposed method for detection and tracking swimmers within hos-
tile aquatic environments at night in terms of FPR. To be
fair, we obtained their different false positive rate for a
given situation in which their false negative rate is minimum
based on color without filtering, color and motion without
out color compensation and the proposed spatio-temporal
method. Note that the number of segmented foreground
noise and the number of benchmark foreground were man-
ually counted. We can see that the FPR is poor for detec-
tion and tracking swimmers with only color without filter-
ing. Our proposed color compensation based on motion fre-
cquency achieves 85\% of FPR.

5 Conclusions

This paper provides spatio-temporal filtering scheme to re-
duce the noise effect caused by light specular reflection
within aquatic environment. Especially, a novel color com-
penation filtering is designed to generate ‘pseudo color’ for
partly hidden human detection. Experimental results show
the effectiveness of the proposed filtering scheme when in-
corporated within an actual real-time outdoor swimming
pool surveillance that operates under various challenging
conditions at nighttime.

Acknowledgments

This work is supported by the Enterprise Challenge Unit,
Singapore’s Prime Minister’s Office and the Singapore
Sports Council.

References

Moving Object and Background Separation”, Signal Process-
Unattended Outdoor Environment”, IEEE Trans. On circuits
and system for video technology, 1998, no. 8, vol. 1, pp. 697–
704.
ent Motion: the Statistics of Temporal Continuity”, IEEE
Kobayashi. Video camera system using liquid-crystal po-
larizing filter to reduce reflected light. IEEE Trans. Pattern
419–426.
[7] H. L. Eng, Alvin H. Kam, Junxian Wang and Wei-Yun Yau
“Human detection and tracking within hostile aquatic envi-
ronment”, Proc. the 12\textsuperscript{th} Conf. on Image Analysis and Pro-
679–698.
tion with Moments Under Time-varying illumination”, IEEE