Hindawi
International Journal of Optics
Volume 2020, Article ID 4145938, 12 pages
https://doi.org/10.1155/2020/4145938

Research Article

Swimmer's Head Detection Based on a Contrario and Scaled Composite JTC Approaches

D. Benarab,[1,2] T. Napoléon ,[1] A. Alfalou,[1] A. Verney,[2] and P. Hellard[3]

[1] *ISEN Brest, Vision-AD Department, L@bISEN Lab—Yncrea Ouest, 20 Rue Cuirassé Bretagne, CS 42807, 29228 Brest Cedex, France*
[2] *ACTRIS Brest, 24 Rue Victor Grignard, 29490 Guipavas, France*
[3] *FFN (French Swimming Federation), 14 Rue Scandicci 93508 Pantin Cedex, Paris, France*

Correspondence should be addressed to T. Napoléon; thibault.napoleon@isen-ouest.yncrea.fr

Received 15 July 2019; Accepted 24 February 2020; Published 1 April 2020

Academic Editor: E. Bernabeu

In order to accompany the swimming coaches in evaluating high-level swimmers, we developed a prototype for instantaneous speed estimation. To achieve this, we proposed and validated, in a previous work, a swimmer tracking system based on data fusion. However, the initialization phase is done manually, and our aim, in this paper, is to automate this process. First, we propose a region of interest localization module that allows the detection of the first appearance of the swimmer in the lane as well as the restriction of the region of interest around him. This module is based on the method a contrario which consists of modeling the random noise corresponding to the water and detecting the structured movement relative to the swimmer motion. To do that, we calibrate the pool using DLT (Direct Linear Transform) technique, extract the concerned lane, apply the frame difference approach to detect the moving objects, and then decompose the lane into blocs and classify them into swimmer motion or noise. Second, in order to detect the swimmer's head, we propose the Scaled Composite JTC which is based on the NL-JTC correlation technique. The input plane of this latter includes a target and a reference image. The first is the region of interest detected by the method a contrario. The second consists of a Scaled Composite Reference. The tests conducted on real video sequences of French swimming championships (Limoges 2015) showed very good results in terms of region of interest localization and swimmer's head detection which allows a reliable initialization for the tracking system.

1. Introduction

Recently, a strong interest is given to kinematic and biomechanical studies in order to enhance swimmers performances. In the context of our collaboration with the French Federation of Swimming, our objective is to conceive an automatic system that estimates swimmers pace and instantaneous speed. Some constraints are imposed in order to generalize the use of the system in different cases: training and high-level competitions. In particular, this system must fulfill the following conditions: minimal user intervention, no wearable sensors, and no physical markers. We thus sought to develop an automatic swimmer tracking system using 4K video sequences.

In our previous work, we proposed an optimized simmer tracking system based on the multirelated targets approach [1]. This latter consists of tracking two targets (head and swimsuit) simultaneously using the dynamic fusion tracking approach. The main idea is to estimate the position of the occluded target taking into account the position of the visible one. For this, all the potential detections of the two targets are evaluated according to a complex criterion composed of the confidence factor of the detection of each target, the Euclidean distance between the targets, and the swimmer speed. Then, this criterion is maximized in order to choose the best couple head-swimsuit. This approach showed good results in terms of tracking.

Based on this approach, we developed a prototype of automatic swimmer tracking and instantaneous speed estimation. However, the selection of the initial reference of the swimmer's head is performed manually. Hence, in order to achieve an all-automatic robust swimmer tracking system, we proposed a novel approach called Scaled Composite JTC [2]. In this paper, we will extend this work and enhance principally the region of interest localization and the reference initialization modules.

For the first module, we propose the following process. First, we calibrate the camera using the DLT technique (Direct Linear Transform) [3, 4] which will allow us to pass from any real metric coordinates to the correspondent pixel coordinates in the image and vice versa. This will help us to extract the swimming lane containing the swimmer to be tracked. Afterwards, we apply frame difference in order to detect the swimmer's motion. For this, we decompose the difference image into blocs and analyze each bloc locally to determine its nature (swimmer motion/noise). This can be done using the method *a contrario* [5–7] which creates a model for the random noise and then detects the structured movement that appears in the lane. This allows the detection of the exact moment when the swimmer appears in the lane to launch the tracking process and the restriction of the region of interest around the swimmer all along the video sequence.

The second module consists of an automatic initialization of the tracking process with a reference image of the swimmer's head. In our previous work [1, 8, 9], this initial reference is selected manually to launch the tracking process. To detect it automatically, we propose the Scaled Composite JTC approach which is based on the NL-JTC correlation technique [10–12]. The input plane of this latter contains a reference and a target images. For the first image, we propose a Scaled Composite Reference which is generated by applying the composite filter on 3 images chosen from a pregenerated database according to the current situation (swimming direction, gender, age. . .). Then, this composite reference is scaled according to the concerned lane pixel dimensions. For the second image, we use the proposed region of interest localization module to select the target image. The application of the NL-JTC correlation technique on this input plane provides a potential detection in each frame of the video sequence. In order to initialize the swimmer tracking system, we choose the best 3 potential detections according to their PCE value (Peak to Correlation Energy).

These two modules are integrated in the prototype of swimmer tracking and instantaneous speed estimation to achieve principally two goals. First, we validated swimmer's head detection using the Scaled Composite JTC approach for the initialization of the tracking. Second, we enhanced the region of interest localization using the a contrario approach. Accordingly, these propositions allowed us to overcome some of the difficulties of the automatic swimmer tracking, namely target occlusions.

2. Environment Specification

FINA (International Swimming Federation) has established several standards for competition pools. The length of the pool is 50 meters for the long races and 25 meters for the short ones [13]. Competition pools are generally covered and heated to ensure their use throughout the year and to be adapted more easily with FINA regulations regarding temperature, lighting, and automatic arbitration equipments. Table 1 summarizes the standards imposed by FINA for 50 meters Olympic pools.

A 50-meter pool can be qualified by FINA to host big events, in case it has the following dimensions: 50 meters of length and 25 meters of width. In addition, it must be divided into eight lanes of 2.5-meter wide each. Two other additional lanes 2.5-meter wide (lanes 0 and 9) were added in each side of the pool to the traditional eight lanes at the congress FINA 2009 [13]. Moreover, the depth of the pool is not fixed but is limited to 2 meters at least. Other criteria are also imposed by FINA, for instance, the lane color and the position of the flags indicating the flipping moment for backstroke races (5 meters from each edge). The water temperature should be maintained at 25–28°C and the illumination level at more than 1500 lux. Touchpads are mounted on both edges of the pool in order to automatically measure the arrival time of the athletes to the edges of the pool. Figure 1 summarizes the dimensional norms imposed by FINA.

These norms ensure a good organization of the competitions, but they can also be used as reference landmarks in order to calibrate the pool. This can be done by calculating the geometric relationship between landmarks metric coordinates, which must be known accurately and the correspondent pixel coordinates. Once the video sequence is calibrated, we can simply pass from any metric coordinates to the correspondent pixel coordinates in the image and vice versa. Therefore, we will be able to calculate different measures, namely:

- Position of the detected swimmer in meters
- Distance covered during a particular period
- Instantaneous and average speed of the swimmer
- Passing time by the various landmarks in the pool
- Estimating the swimmers size in the different lanes in the image

In our case, we are mainly interested in the evaluation of swimmers according to their instantaneous speed. For this, we need to establish the correspondence between pixel and metric coordinates. This later can be done based on the concept of calibration.

3. Calibration

All along our research in the field of swimmer tracking, we have targeted various difficulties related to the aquatic environment specificities. In particular, the localization and the extraction of the swimming lanes is a difficult task that remains crucial to ensure a proper functioning of the tracking system. Indeed, such information is necessary to predict the future location of the swimmer or to estimate distances required to calculate the speed. Thus, we propose to calibrate the video in order to establish the link between metric and pixel coordinates from the pool and the image, respectively.

Table 1: FINA standards for the eligibility of 50 meters Olympic pools.

Characteristic	Conventional value
Length	50 m
Width	25 m
Depth	2 in minimum (3 m recommended)
Lanes number	10 (2,5 m of width)
Temperature	25–28°C
Lighting	1500 lux

In addition, calibration can be used to correct the distortion resulting from perspective errors and aberrations corresponding to the camera lens. Because of these distortions, the image pixels are misplaced. However, the information is not lost, and it can be partially reconstructed by measuring the extrinsic parameters of the camera (rotation and orientation). To calculate these parameters, we use the DLT calibration technique (Direct Linear Transformation) [3, 4]. This technique establishes the relation between metric and pixel coordinates that allows to correct perspectives and calculate distances.

Videos of swimming competitions can be calibrated based on the various landmarks of the swimming pool as shown in Figure 1. Given that the pool can be considered as a two-dimensional plane, it is possible to use the simplified 2D DLT technique instead of the conventional DLT [3, 4], which is calculated using the following equation:

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -u_1x_1 & -u_1y_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -v_1x_1 & -v_1y_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & 1 & 0 & 0 & 0 & -u_nx_n & -u_ny_n \\ 0 & 0 & 0 & x_n & y_n & 1 & -v_nx_n & -v_ny_n \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_7 \\ L_8 \end{bmatrix} = \begin{bmatrix} u_1 \\ v_1 \\ \vdots \\ u_n \\ v_n \end{bmatrix}. \quad (1)$$

To solve this equation, we need to calculate the calibration parameters $L_{1..8}$ which represent the unknowns of this system of equations. For this, it is sufficient to know the pixel coordinates (x_i, y_i) and metric coordinates (ui, v_i) of n points of the recorded scene, with $n \geq 4$.

In our case, the Olympic pools that organize big events usually conform to the international norms in terms of size and color of the lanes. The cameras used for recording are fixed during all the shooting sessions. Hence, to calibrate all the video sequences of each session, it is sufficient to calibrate only one frame and apply the same calibration parameters on the rest. To do this, we manually selected four points which we have accurately measured their metric coordinates, as shown in Figure 2. Note that the origin (0, 0) in our calculations corresponds to the top/right corner of the pool, as shown in Figure 2. The coordinates of the selected points replace the variables (u_i, v_i) and (x_i, x_i) in the equation (1). This generates a system of 8 equations and 8 unknowns $(L_{1..8})$. The resolution of this system allows to calculate these which represent the calibration parameters. Once they are calculated, the passage from pixel to metric coordinates is ensured using the following equation:

$$\begin{cases} u = \dfrac{L_1x + L_2y + L_3}{L_7x + L_8y + 1}, \\ \\ v = \dfrac{L_4x + L_5y + L_6}{L_7x + L_9y + 1}. \end{cases} \quad (2)$$

As shown in Figure 2, the calibration allows an accurate localization of the swimming lane.

4. Region of Interest Localization

Our aim is to localize the swimmer globally and restrict the region of interest around the swimmer during the video sequence. This module is important to prepare the video for a robust swimmer tracking which yields the ability to perform accurate measurements and evaluate his performance. The process of localizing the region of interest solves two problems. The starting time corresponds to the first movement of the swimmer in the swimming lane and the restriction of the region of interest by prelocalizing the swimmer during the race.

For this, we propose an automatic process based on motion detection [14–16] of the swimmer in the concerned lane. To do this, we take into account the specificities associated with swimming and the characteristics of the pool in order to apply the adapted preprocessing techniques. As input data, this process requires the knowledge of the camera calibration parameters, the lane number, and the dimensions of the pool. In this section, we will present the method *a contrario* used for region of interest localization, and then we will detail the different steps of this process.

4.1. Motion Detection Using the Method A Contrario: Adaptation. *A contrario* is a statistical approach based on hypothesis tests to detect significant geometric events in images. The basic idea of this approach is based on the principle of visual perception called Non – accidental, which is also known as the principle of Helmholtz [5, 6]. In their book, Desolneux et al. [6] summarize this principle as follows: "whenever a deviation of the randomness aspect appears, a structure is perceived". Here, the structure is defined by its opposite, namely, the noise. In the case of the absence of a structure, the events are independent, and they behave randomly while the structure differs in a more organized behavior. *A contrario* method was applied to various detection problems. We cite for example, edge detection in [7, 17], pattern recognition in [18], and detection of rigid points of interest for the matching between images in [19].

In this work, we propose the adaptation of the method *a contrario* in order to automatically detect the swimmer motion starting from his first diving in the water. Indeed, the noise model is considered as an independent uniform distribution. In our case, we consider the random movement of water in the empty lane as a noise model. For this reason, we will establish a dynamic and relevant threshold corresponding to the water movement in the concerned lane which will allow the detection of the structured movement of the swimmer. Subsequently, we will present in detail the different steps of the region of interest localization process based on the method *a contrario*, as shown in Figure 3.

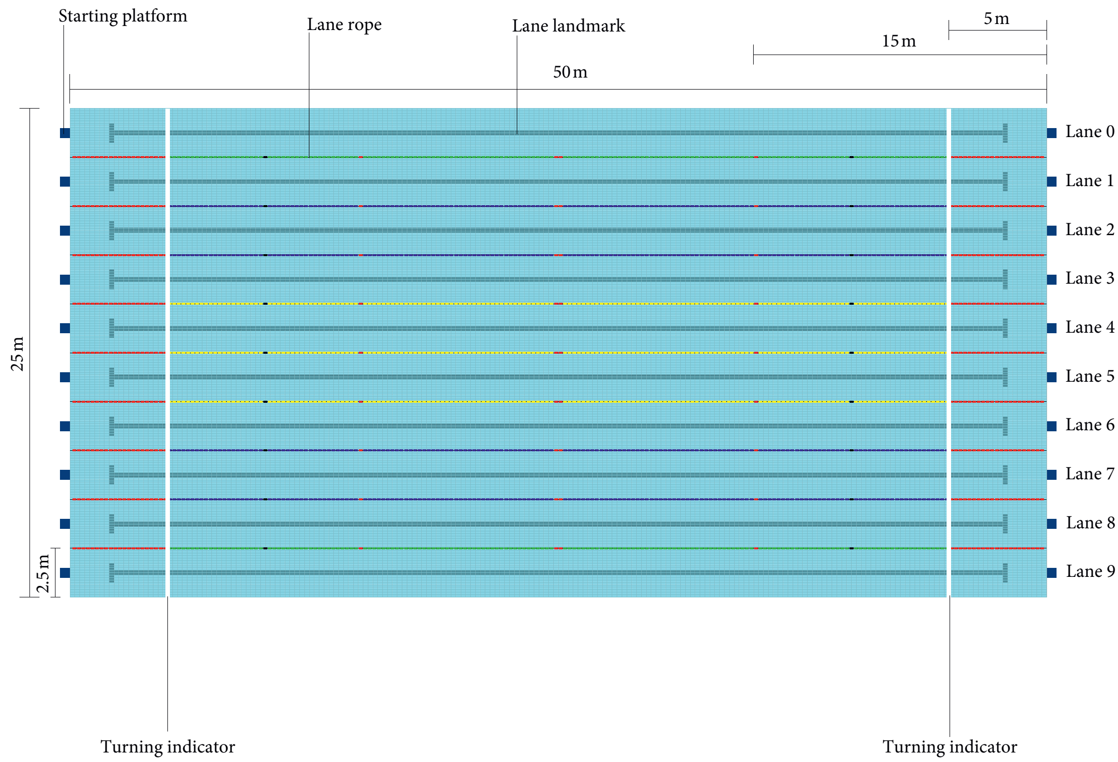


FIGURE 1: Conventional dimensional measures recognized by FINA for 50 meters Olympic pools.

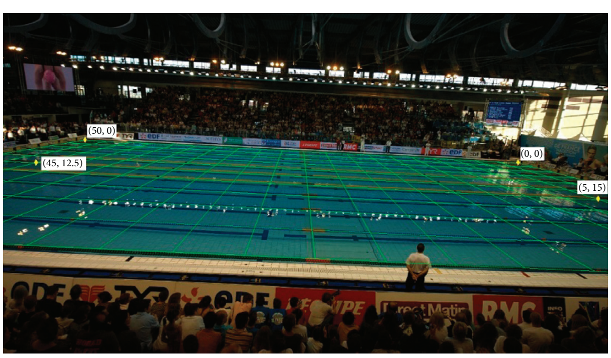


FIGURE 2: Calibration of the pool by the DLT technique using four selected points. Displaying a grid of $50 \times 25\,\text{m}^2$ where each cell represents an area of $2.5 \times 2.5\,\text{m}^2$. Swimming Championships in Limoges, France, April 2015.

4.2. Lane Extraction. Knowing that the pool and the lanes dimensions are known, we can precisely extract the lane containing the concerned swimmer. This can be done by defining the number of the lane to be extracted. For this we use the following formula:

$$\begin{cases} x_1 = 0, \\ x_2 = \text{length}, \\ y_1 = \text{NB}_{\text{lane}} * \text{width}, \\ y_2 = (\text{NB}_{\text{lane}} + 1) * \text{width}, \end{cases} \tag{3}$$

where NB_{lane} represents the lane number of the the swimmer to be tracked. length (50 m) and width (2.5 m) represent the dimensions of the lane. x_1, x_2, y_1, and y_2 represent the coordinates of the 4 corners of the lane containing the concerned swimmer. It should be noted that, in our case, the origin (0; 0) corresponds to the top/right corner of the pool. These measures are calculated in metric domain and it is, therefore, necessary to obtain the corresponding pixel coordinates. For this, we use the calibration results and more specifically the equation (2) to ensure this passage and get

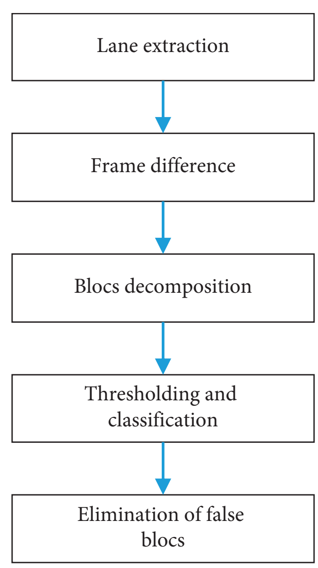


FIGURE 3: General process of the region of interest localization based on the method *a contrario*.

the coordinates of the 4 corners of the lane. Finally, we apply a mask in order to keep only the concerned lane. The result is shown in Figure 4.

4.3. Frame Difference. Frame difference [16, 20] is usually used as a preprocessing technique that is used for motion detection, especially for the videos captured by fixed cameras, which corresponds to our case. In order to detect the motion, the moving object can be segmented and extracted by performing a frame difference between the current frame i of the video sequence and the background, where the latter corresponds to an image of the scene without the object to be tracked. In our case, the first frame when the lane is empty, before the swimmer diving, can be considered as a background image. According to the literature [16, 20], this technique shows very good tracking results in case of static background. However, it remains highly sensitive to variations in lighting and movement of the various components of the scene background.

In our case, the background is not completely static, especially when the swimmer starts swimming. The latter generates a lot of splashes and waves along the entire length of the lane. This creates a significant noise after the frame difference between the frame i containing the swimmer and the first image of the empty lane, as shown in Figure 5. However, we noticed minimal variations between successive frames except in the area containing the swimmer where we noticed a significant variation. Therefore, we choose the difference of successive frames to detect the swimmer based on his movement. In order to reduce the noise caused by water movement, we propose to apply the Median filter on the two successive frames before the subtraction. Hence, the frame difference is calculated by the following equation:

$$\mathrm{Diff}_i = \left|\mathrm{Median}\left(im_i\right) - \mathrm{Median}\left(im_{i-1}\right)\right|. \tag{4}$$

The result Diff_i is called the difference image; it contains high intensities which correspond mainly to the swimmer's movement area. This is illustrated in Figure 6 where we clearly notice that the difference of successive frames is less noisy than the difference between the frame i and the empty lane frame which is shown in Figure 5. For this reason, we retain the difference of successive images for swimmer motion detection in the rest of our study.

4.4. Blocs Decomposition. The difference image contains pixels of different intensities. Indeed, the intensity value corresponds to variation level of the pixel color which allows us to detect moving objects in the scene. However, in our case, it is not only the swimmer who is in movement but also the water and the light reflections surrounding it which may falsify swimmer motion detection. To overcome this problem, we propose to decompose the difference image to blocs in order to study the pixels intensities locally and take a decision concerning the motion detection in each bloc of the lane. In our case, we decompose the difference image to blocs of size $b \times b$ as shown in the following equation:

$$b = \frac{\left|y_{11} - y_{22}\right|}{\mathrm{NB}_{\mathrm{bloc}}}, \tag{5}$$

knowing that the distance $|y_{11} - y_{22}|$ represents the lane width and $\mathrm{NB}_{\mathrm{bloc}}$ represents the desired number of blocs, in our case $\mathrm{NB}_{\mathrm{bloc}} = 20$. This allows a consistent local study focusing on the areas containing a significant movement. To take a decision concerning the nature of the bloc (swimmer motion or noise), it is necessary to establish an adapted threshold which we define subsequently.

4.5. Thresholding and Classification. This step consists of applying a threshold on the blocs to classify them into swimmer motion or noise. For each bloc of the difference image, we calculate the local mean of its intensity, which is compared to a defined threshold (Thresh). This threshold is calculated before the diving phase in order to measure the water movement and the variation associated with light reflection in an initial state. In order to calculate the threshold, we proceed in the same way but on two successive frames of the empty lane. We calculate the difference between the two frames filtered by a Median filter. Then, we decompose the difference image into blocs and calculate the mean of each block. Finally, the threshold (Thresh) corresponds to the maximum value of the noise associated with the random motion of the water. In other words, Thresh is the maximum value of the blocs means measured for an empty lane. In this context, Figure 7 shows an example of classification according to the noise model, where the noise threshold corresponds to the bloc located in the middle. The blocs on the left are considered as noise given that their mean intensity is lower than the threshold Thresh. However, the blocs on the right represent areas containing a significant movement that can match swimmer motion because their mean intensity is greater than Thresh. These blocs are localized and labeled for further processing to localize the swimmer.

Throughout the period in which the difference image Diff_i contains random noise, we consider it a state of rest. Once a structured movement appears between the two lines delimiting the lane, the mean intensity of the concerned blocs increases and exceeds the threshold Thresh. This allows us to automatically localize the swimmer and determine his direction.

4.6. Elimination of False Blocs. Thanks to the previous steps, we can detect elements of the scene that are in motion between two successive images. However, we noticed after several tests that these detected areas correspond to the swimmer motion, the movement of the lines delimiting the lane and the light reflections. To eliminate these last two cases, we rely on two main criteria: the position and surface. The position of each detected bloc helps to determine its nature and whether it may be a swimmer or not. For example, in Figure 8, we can distinguish the blocs that match the movement of the lines delimiting the lane according to their position. To refine the detection, we eliminate two lines of blocs around each line delimiting the lane. On the other hand, to solve the case of the reflections, we use the surface

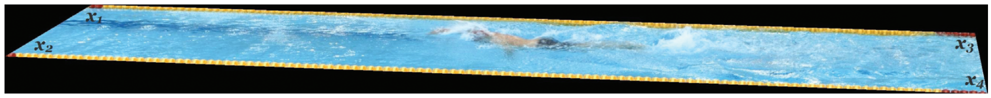


FIGURE 4: Lane extraction and mask application taking into account the four points: x_1, x_2, x_3, and x_4. The metric coordinates of these points are known and pixel coordinates are calculated using the calibration results. Swimming Championships in Limoges, France, April 2015.

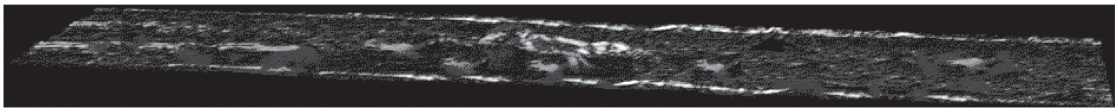


FIGURE 5: Frame difference between the frame and (i) the first frame of the video sequence containing an empty lane. Swimming Championships in Limoges, France, April 2015.

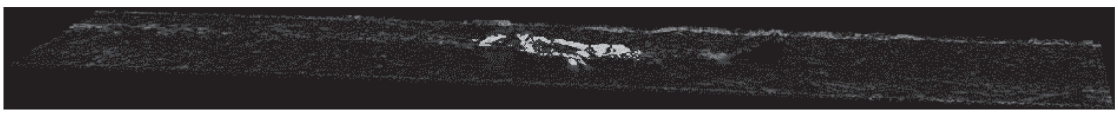


FIGURE 6: Successive frames difference between the frames $i-1$ and i. The high intensities correspond to the movement of the swimmer or the water. Swimming Championships in Limoges, France, April 2015.

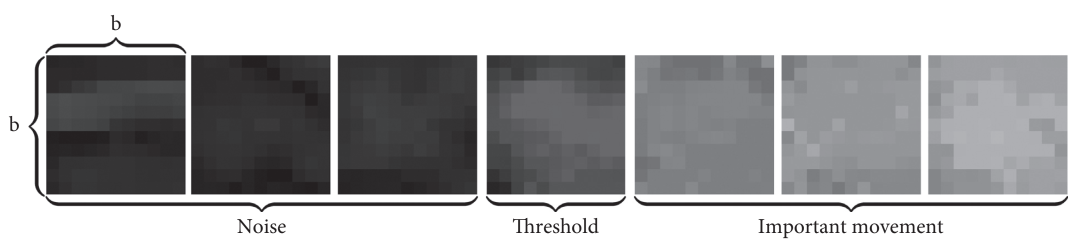


FIGURE 7: Example of bloc classification according to a noise model. The central bloc is the threshold. The blocs in the left are classified as noise. The blocs in the right are classified as an important movement that will be analyzed to determine its nature (swimmer or not). Swimming Championships in Limoges, France, April 2015.

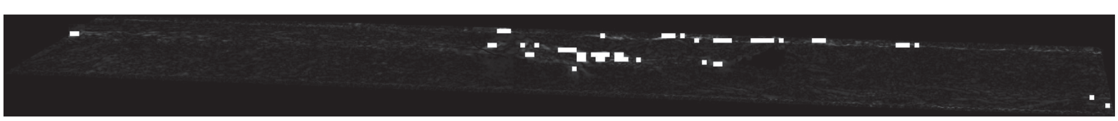


FIGURE 8: Blocs decomposition, classification, and thresholding. The white blocs represent areas with significant movement that may correspond to a swimmer. Swimming Championships in Limoges, France, April 2015.

criterion knowing that the blocs representing the reflections are usually detected as isolated blocs. All remaining blocs correspond to the swimmer motion detected in the frame i, as shown in Figure 9.

Figure 10 shows the different stages of the process of localizing the region of interest using the method *a contrario* for swimmer motion detection, starting with the noise model calculated on the empty lane to the restriction of the region of interest around the swimmer.

4.7. Discussion. The region of interest localization approach introduced in this section can detect the appearance of the swimmer in the lane as well as restrict the region of interest, as shown in Figure 10. On the other hand, our main goal is to develop an accurate automatic swimmer tracking and evaluation system. For this, we use the method *a contrario* to detect the swimmer, determine the exact time of his appearance in the lane, and localize him globally throughout the video. However, it remains to treat the aspect of tracking accuracy. To do this, it is necessary to consider the various difficulties that correspond mainly to contour deformation and the occlusion of the swimmer. Indeed, the swimmer's head is the part that allows obtaining the best compromise between visibility and rigidity (less deformation). For this reason, we propose in the following section initializing the tracking based on the detection of the of the swimmer's head using the Scaled Composite JTC approach.

5. Tracking Initialization Using the Scaled Composite JTC Approach

We proposed in our previous work [1, 8, 9] an optimized swimmers tracking system based mainly on the head as the body part to be tracked. This system requires, as an input, a

Figure 9: Refining detections by the elimination of false blocs corresponding to the reflections and the movement of the lines delimiting the lanes. Swimming Championships in Limoges, France, April 2015.

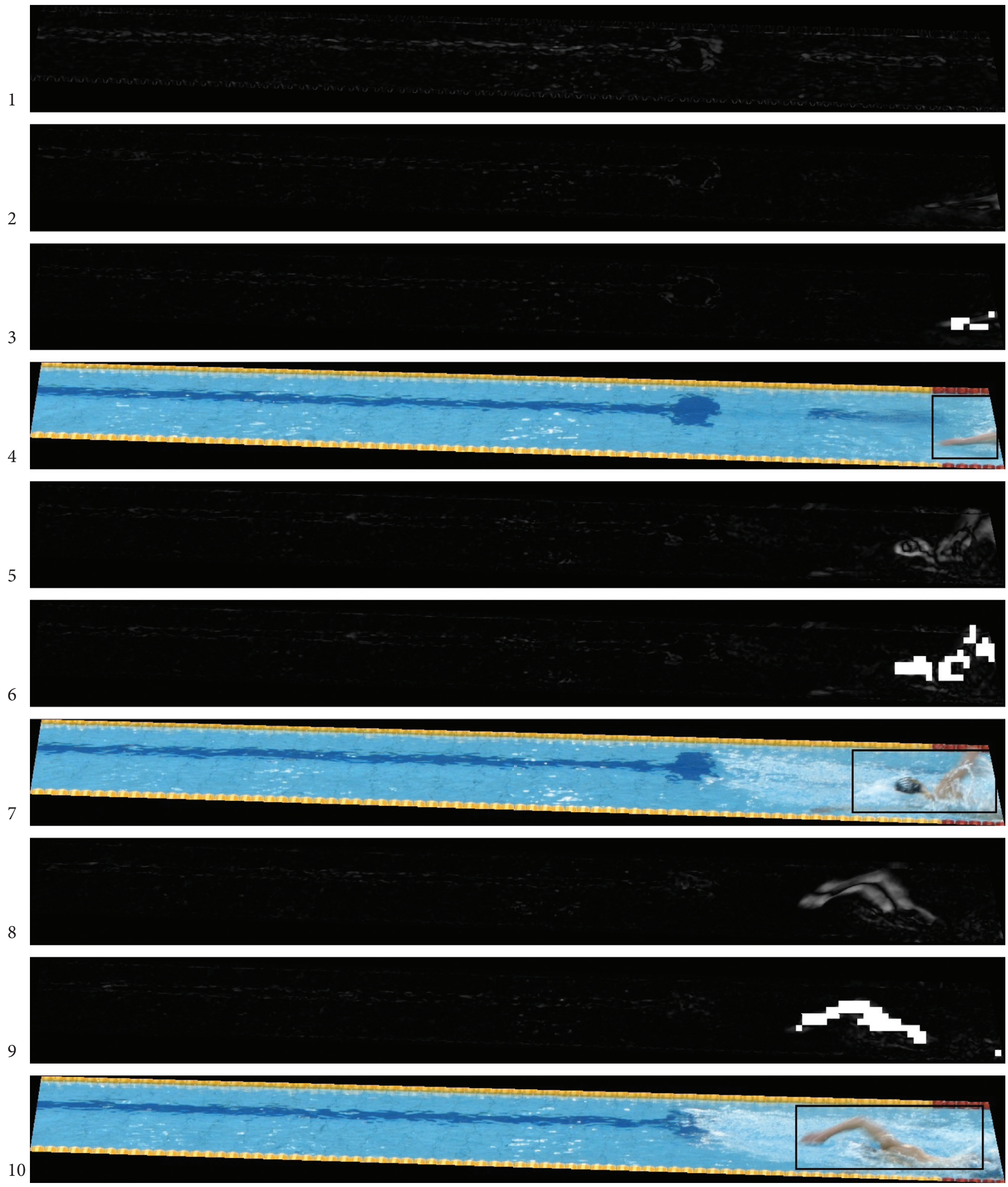


Figure 10: Localization of the region of interest based on the method *a contrario*. Image 1: noise model obtained from the difference between two images of the empty lane. Images 2, 5, and 8: frame difference between two successive frames. Images 3, 6, and 9: blocs decomposition and thresholding compared to a threshold relative to the noise model. Images 4, 7, and 10: swimmer localization. Swimming Championships in Limoges, France, April 2015.

region of interest around the swimmer to be tracked and a reference image of his head. The first input data can be found by applying the region of interest localization approach based on the method *a contrario*. While the second will be the subject of this section where we propose an automatic technique for the detection of the swimmer's head based on the correlation technique NL-JTC [10, 12] applied on a standard scaled composite reference constructed from a pregenerated database. Finally, the detected swimmer's head will be used as a reference for optimized tracking system proposed in the previous work.

5.1. Database Generation. To detect the swimmer's head, it is necessary to have a prior description of this part of the body to be tracked. To do this, we generated a learning database based on video sequences that we recorded during national and international competitions (Swimming Championships Limoges, France, 2015, World Swimming Championships, Barcelona 2013 and Kazan 2015). These videos were recorded using two 4K cameras to have more details on the images and extract the swimmer's head efficiently. Our database contains swimmers' heads extracted in different real situations occurring during swimming.

In particular, the criteria that we have taken into account for the generation of the database to cover most scenarios are

Age: senior and junior

Gender: men and women

Type of the swim: crawl, backstroke, butterfly, and breaststroke

Direction of swimming: going and coming

Figure 11 presents sample images of heads/caps of different swimmers in various situations.

5.2. Application of the NL-JTC Technique. Knowing that the color of the caps worn by swimmers may vary, we opt for their form to ensure a relevant and standard description of the swimmer's head. For this reason, we choose the NL-JTC correlation technique which is known in the literature of contour-based detection [10, 21, 22]. The input plane of the NL-JTC technique includes a reference and a target image.

The first image consists of a standard reference of the swimmer's head that we generate from a specific database (this process will be detailed subsequently). The second image represents the region of interest around the swimmer's head which is extracted using the method *a contrario* detailed in the previous section. Once the input plane is generated, we apply the NL-JTC technique, as shown in Figure 12 and we get a correlation plane. The analysis of this latter allows us to take the decision concerning the existence of a target having a shape similar to the head of the swimmer. This decision is made on a short sequence in the beginning of the race based on the PCE criterion (Peak to Correlation Energy) to select the best targets. Next, we will detail the reference image generation process, the target image, and the final decision.

5.3. Scaled Composite Reference. Using the pregenerated database, we choose n reference images of swimmers' heads relative to our case. In practice, we set $n = 3$ to have three different forms of swimmers' heads corresponding to the same situation, namely, the swimming type, direction, age.... Then, the selected reference images are converted into grayscale. This is essential for the application of the NL-JTC technique and has no influence on the detection results because the cap color information is discarded.

Then, we apply the composite filter [2, 23, 24] on n images in order to generate a single image representing a rich contour description and containing different swimmers heads. The basic idea of the composite filter (see Figure 13) is to calculate a weighted sum of n images as shown in the following equation:

$$\mathrm{REF}(x, y) = \sum \alpha_i * \mathrm{ref}_i(x, y), \quad (6)$$

with α_i the weighting factor that can be used to favor the reference ref_i.

Finally, this composite reference must be scaled according to the size of the head of the concerned swimmer. This latter is unknown, in the aim of estimating it, we calculate the ratio between the standard width of the head and the width of the lane. Then, based on this ratio, the lane width in pixels and the calibration function, we estimate the dimensions of the concerned swimmer's head in pixels as shown in the following equation:

$$\mathrm{HeadSize}_{\mathrm{pixel}} = \mathrm{LaneSize}_{\mathrm{pixel}} * \frac{\mathrm{HeadSize}_{\mathrm{meter}}}{\mathrm{LaneSize}_{\mathrm{meter}}}. \quad (7)$$

Using the result of this equation, we rescale our composite reference in order to adapt it to our case and depending on the width of the concerned lane. The generation process of the scaled composite reference is summarized in Figure 14.

5.4. Localization of the Region of Interest. The localization of the region of interest is an important step for developing an accurate automatic approach for swimmers tracking. For this, we rely on the method *a contrario* presented in the previous section. This allows us to determine the moment when the swimmer appears in the lane as well as to restrict the region of interest. To do this, it is important to consider the following information: the swimming direction and the blocs corresponding to the swimmer motion. Indeed, to localize our region of interest, we select a rectangle of 2 m length and 1.5 m width that is limited by the last bloc corresponding to the swimmer motion taking into account the swimming direction. These measures are then transformed into the pixel domain and the region of interest is extracted as shown in Figure 15.

5.5. PCE-Based Decision. Our objective is to accurately detect the swimmer's head and initialize the swimmer tracking system developed in the previous work [1, 8, 9]. For this, we apply the proposed Scaled Composites JTC approach to detect the swimmer in the first images of the video sequence. This period is supposed to contain the following events: empty lane, diving, and swimming recovery phase. During the first event, we apply only the region of interest localizing process. Once the swimmer motion is detected in

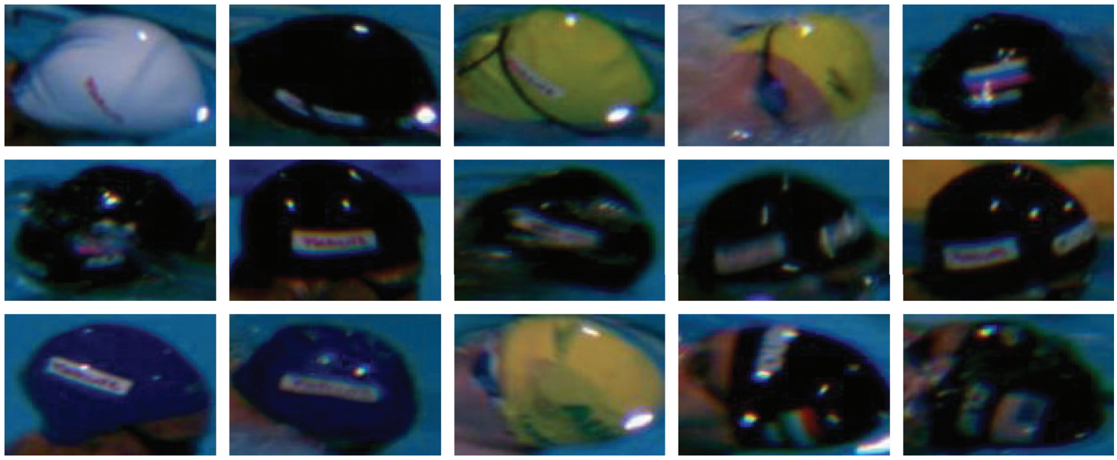
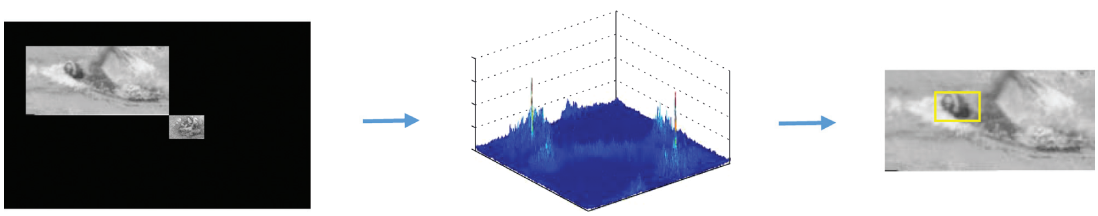


Figure 11: Examples of the training database. World Championships, Barcelona 2013 and Kazan 2015.



(a) (b) (c)

Figure 12: Application of NL-JTC technique for the detection of the swimmer's head. (a) Input plane containing a Scaled Composite Reference and the region of interest provided by the *a contrario* localization process. (b) Correlation plane. (c) Detection and localization of the swimmer's head. Swimming Championships in Limoges, France, April 2015.

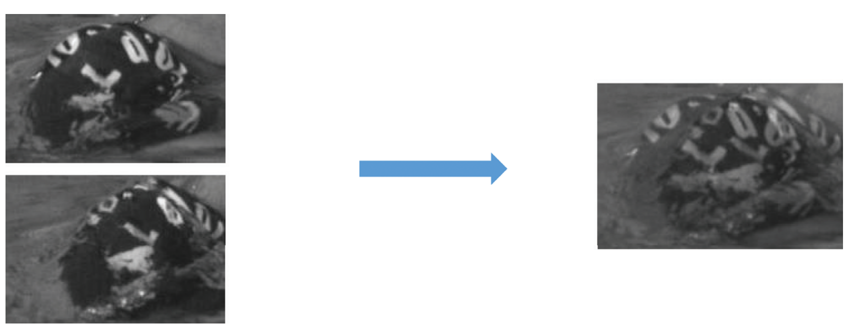


Figure 13: Principle of the composite filter. Left: reference images. Right: composite image which consists of a weighted summation of the two reference images.

the lane, we apply the Scaled Composite JTC approach to detect and localize the swimmer's head in each frame of this period. The length of this latter is set between 2 and 3 seconds (50 to 75 images). Then, for each potential swimmer's head detection, we calculate the PCE value that will be used as a confidence factor. Finally, potential detections are classified according to their PCE associated values and the final decision corresponds to those with the highest PCE. In our case, we validate three targets detected for initializing the optimized swimmer tracking system that is proposed in a previous work.

5.6. Experiments and Results. In order to prepare the video and to facilitate the swimmer tracking, we have developed two modules based on image preprocessing: the region of interest localization module and the automatic detection of the swimmer's head for initializing the swimmer tracking system. Next, we will present the evaluation of these modules.

5.6.1. Evaluation the Region of Interest Localization. The purpose of this module is to detect the swimmer motion in order to restrict the region of interest. For this, we have proposed in this paper an adapted approach based on frame difference and *a contrario* approaches.

We noticed that our method was unable to detect the swimmer motion in the case of minimal movement. To

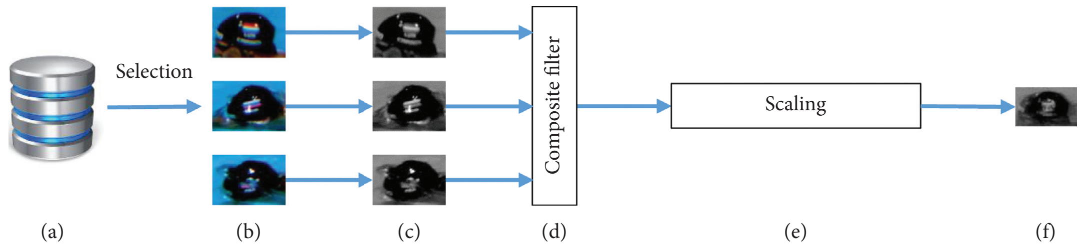


FIGURE 14: Generation of the Scaled Composite Reference. (a) Database containing head references. (b) Selection of images according to the current situation. (c) RGB color space conversion to grayscale. (d) Application of the composite filter. (e) Rescaling the composite image depending on the size of the concerned lane. (f) Result: Scaled Composite Reference. Swimming Championships in Limoges, France, April 2015.

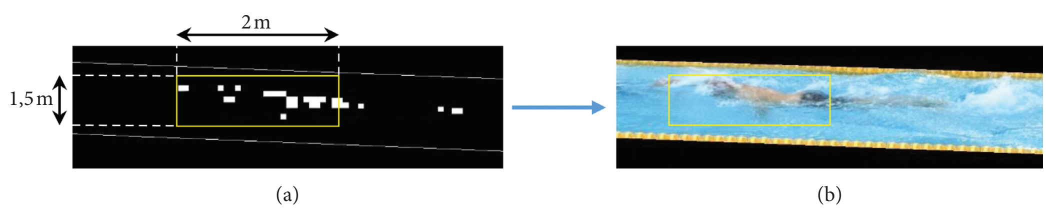


FIGURE 15: Localization of the region of interest. (a) Difference image decomposed into blocs and thresholded. The rectangle represents the region of interest centered according to the lane width, starting with the first bloc corresponding to the swimmer motion. (b) Projection of the rectangle on the original image. Swimming Championships in Limoges, France, April 2015.

overcome this problem, we proposed an adapted prediction technique based on a set of relevant criteria: the swimming direction, referential validated position, and mean speed. This prediction principle will be used in the following tests to improve the localization process.

To evaluate this region of interest localization module, we tested it on 5 crawl video sequences during the national championships in Limoges 2015. Each sequence, containing 400 images with a frame rate of 25 frames/s, begins with several frames of an empty lane before swimmer diving. These frames allow us to establish a noise model for the application of the method *a contrario*. As shown in Table 2, this module will be evaluated using the percentage of successful localizations (swimmer in the region of interest).

Table 2 presents the results of the region of interest localization module based on frame difference and *a contrario* approaches in both modes: with and without prediction. We notice that the application of the method *a contrario* provides high localization percentages ranging from 94.5% to 97%. In this context, the localized regions of interest will be used, thereafter, to accurately detect the head of the swimmer. Therefore, the region of interest localization is a crucial step that needs to ensure the existence of the head in the region of interest. In order to optimize the results, we coupled the method *a contrario* with an adapted prediction technique which allowed us to achieve very high localization percentages close to 100%.

TABLE 2: Region of interest localization percentage using method *a contrario* with and without prediction. Crawl-Swimming Championships in Limoges, France, April 2015.

Sequences	Swimmers	Localization percentage	
		A contrario without prediction	*A contrario* with prediction
Crawl 1	Swimmer 1	95.25	100
	Swimmer 2	97	100
	Swimmer 3	95.75	99.5
Crawl 2	Swimmer 1	94.5	99.75
	Swimmer 2	96	100

5.6.2. Evaluation of the Reference Initialization. The automatic swimmers tracking system proposed in our previous work [1, 8, 9] needs to have a relevant reference image of the swimmer's head to be tracked. However, since the color of swimming caps may change, the only available information is the shape of the head. Based on this, we proposed the Scaled Composite JTC approach to detect and select automatically the initial reference image of the concerned swimmer.

In order to detect the best reference image for the swimmer's head to be tracked, we apply the proposed method on short video sequences chosen at the beginning of the race, because the swimmer's head reappears quickly after the diving phase. Among the targets detected, we choose the three best targets to be used in order to initialize swimmers tracking system.

To evaluate this reference initialization module, we tested it on short sequences of 100 frames extracted from the beginning of the previous 5 crawl video sequences. As shown in Table 3, two criteria are used to evaluate this module: the detection percentage on 100 frames of each video sequence and the percentage of successful initialization of the 3 selected reference images.

Table 3 shows high detection percentages between 78.31% and 86.44% for the 5 tested sequences although we

Table 3: Evaluation of the Scaled Composite JTC for the initialization of the reference. Crawl-Swimming Championships in Limoges, France, April 2015.

Sequences	Swimmers	Detection percentage
Crawl 1	Swimmer 1	83.33
	Swimmer 2	86.44
	Swimmer 3	85.71
Crawl 2	Swimmer 1	78.31
	Swimmer 2	82.14

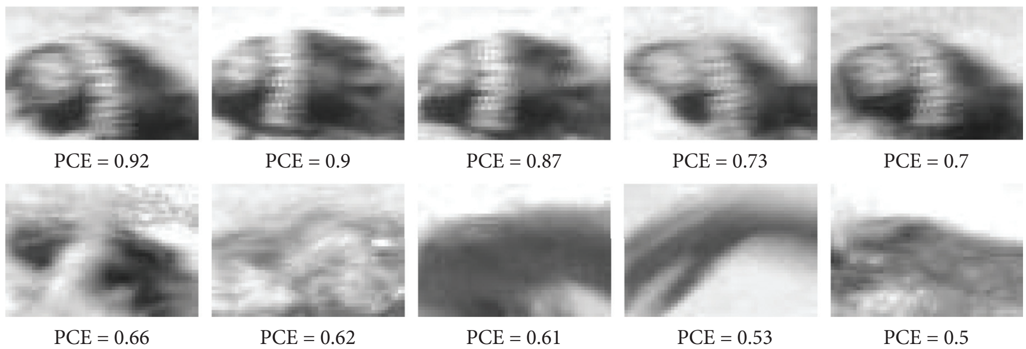


Figure 16: Validation of the initial references according to a PCE-based decision criterion. Detected targets are classified by decreasing order of PCE values. The three validated targets are those with the highest PCE. Swimming Championships in Limoges, France, April 2015.

have only applied the NL-JTC method for detection. This is explained by the visibility of the swimmer's head in this period of swimming recovery after the diving phase. The other cases of false detections correspond mainly to partial and total occlusions of the swimmer's head. Moreover, for each of the five tested sequences, we choose the best 3 detections according to their PCE values, as shown in Figure 16. In this figure, we presented the top 10 detections for the sequence crawl 1 of Table 3. Among these detections, we choose the best 3 having, respectively, the following PCE values: 0.92, 0.9, and 0.87. This enabled us to achieve an accuracy of 100% (on the true detections) among the images chosen for the initialization of the swimmer tracking system of initialization detections. Note that this accuracy is obtained thanks to the multitarget initialization of the tracking process of [1] that allows to select a set of images rather than a single one.

6. Conclusion

In this paper, we proposed two independent modules to optimize the swimmer tracking systems proposed [1, 8, 9]: the region of interest localization module and the automatic detection of the swimmer's head for initializing the swimmer tracking system. First we analyzed the different characteristics of the Olympic pools dedicated to high-level competitions. Among these characteristics, we were mainly interested by the pool normed dimensions which allowed us to segment the pool and extract the different objects in the scene: water, lanes, and swimmers. For this, we have introduced the DLT calibration method to model the relationship between metric and pixel coordinates. This helped us to localize the pool boundaries in the image, the lines delimiting the lanes, and to calculate accurate measurements within the image.

Then, we proposed a region of interest localization approach based on frame difference and *a contrario* approaches. To do this, we extracted the concerned lane. After that, we calculated the difference image between successive frames and decomposed it to blocs to analyze them locally. Afterwards, we detect the swimmer motion, comparing the blocs with a noise model. This approach is mainly used to restrict the region of interest and to detect the exact moment of the appearance of the swimmer in the lane. The conducted tests showed good results in terms of localization percentage. However, we noticed some errors relative to the minor movement of the swimmer in some cases. This can be solved by coupling the proposed approach with an adapted prediction technique which enhanced significantly the results.

In order to track the swimmer, we need to select automatically an initial reference of his head. For this, we have proposed the Scaled Composite JTC approach which is based on the NL-JTC technique. The basic idea consists of creating a database of swimmers' heads classified according to different situations. Then, three images are selected, depending on the current situation, to generate a standard reference using the composite filter. This composite reference is scaled according to the size of the concerned lane. Then, the NL-JTC technique is applied on an input plane containing the Composite Scaled Reference and the localized region of interest. Finally, we choose, depending on the values of PCE, three reference images that will be used to initialize the tracking system proposed in our previous work which is based on a fusion of contour and color description of the target to be tracked [1, 8, 9]. The performed tests showed the efficiency of the proposed initialization approach, where we reached a percentage of 82.2% of detections on 100 frames of 5 crawl video sequence and 100% of succeeded initialization on the true detections.

Data Availability

The dataset used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] D. Benarab, T. Napoléon, A. Alfalou, A. Verney, and P. Hellard, "Optimized swimmer tracking system based on a novel multi-related-targets approach," *Optics and Lasers in Engineering*, vol. 89, pp. 195–202, 2016.

[2] D. Benarab, T. Napoléon, A. Alfalou, A. Verney, and P. Hellard, "All-automatic swimmer tracking system based on an optimized scaled composite jtc technique," *Optical Pattern Recognition XXVII*, vol. 9845, Article ID 98450E, 2016.

[3] Y. I. AbdelAziz and H. M. Karara, "Direct linear transformation into object space coordinates in close-range photogrammetry," in *Proceedings of the Symposium on close-range Photogrammetric Systems*, pp. 1–18, Urbana, IL, USA, January 1971.

[4] M. Elipot, *"Application of new techniques of camera calibration and the cinematique analysis of swimming,"* PhD thesis, Universite Paris Descartes, Paris, France, 2010.

[5] D. Lowe, "Visual recognition as probabilistic inference from spatial relations," in *AI and the Eye*, A. Blake and T. Troscianko, Eds., John Wiley and Sons Ltd., New York, NY, USA, 1990.

[6] A. Desolneux and L. Moisan, J.-M. Morel, *From Gestalt Theory to Image Analysis: A Probabilistic Approach*, Springer Publishing Company, Incorporated, New York, NY, USA, 1st edition, 2007.

[7] I. Abraham, R. Abraham, A. Desolneux, and S. Li-Thiao-Te, "Significant edges in the case of non-stationary Gaussian noise," *Pattern Recognition*, vol. 40, no. 11, pp. 3277–3291, 2007.

[8] D. Benarab, T. Napoléon, A. Alfalou, A. Verney, and P. Hellard, "Optimized swimmer tracking system by a dynamic fusion of correlation and color histogram techniques," *Optics Communications*, vol. 356, pp. 256–268, 2015.

[9] D. Benarab, T. Napoléon, A. Alfalou, A. Verney, and P. Hellard, "A novel multitracking system for the evaluation of high-level swimmers performances," *Optical Pattern Recognition XXV*, vol. 9094, pp. 90940A–90940A–7, 2014.

[10] B. Javidi, "Nonlinear joint power spectrum based optical correlation," *Applied Optics*, vol. 28, no. 12, pp. 2358–2367, 1989.

[11] A. Alfalou, N. Ben Haj Yahia, and M. S. Alam, "Face recognition using a non-zero-order correlation plane and a nonlinear joint transform correlator," in *Proceedings of SPIE*, vol. 8498, p. 84981A, Bellingham, WA, USA, May 2012.

[12] A. Alfalou and C. Brosseau, *Face Recognition Book, Chapter Understanding Correlation Techniques for Face Recognition: From Basics to Applications*, pp. 354–380, Intech, London, UK, 2010.

[13] FINA, *Fédération Internationale de Natation*, FINA, Lausanne, Switzerland, 2016.

[14] S. Oprisescu, M. Ciuc, and V. Buzuloiu, "Histogram and motion based intrusion detection and tracking algorithms for tof cameras," in *Proceedings of the International Symposium on Signals, Circuits and Systems*, pp. 1–4, Iasi, Romania, July 2009.

[15] Z. Jiang, Du Q. Huynh, W. Moran, S. Challa, and N. Spadaccini, "Multiple pedestrian tracking using colour and motion models," in *Proceedings of the International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pp. 328–334, Sydney, Australia, December 2010.

[16] Y. Benezeth, P. M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Review and evaluation of commonly-implemented background subtraction algorithms," in *Proceedings of the 19th International Conference on Pattern Recognition*, pp. 1–4, Tampa, FL, USA, December 2008.

[17] A. Desolneux, L. Moisan, and J.-M. Morel, "Edge detection by Helmholtz principle," *Journal of Mathematical Imaging and Vision*, vol. 14, no. 3, pp. 271–284, 2001.

[18] P. Musé, F. Sur, F. Cao, Y. Gousseau, and J.-M. Morel, "An a contrario decision method for shape element recognition," *International Journal of Computer Vision*, vol. 69, no. 3, pp. 295–315, 2006.

[19] L. Moisan and B. Stival, "A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix," *International Journal of Computer Vision*, vol. 57, no. 3, pp. 201–218, 2004.

[20] M. Piccardi, "Background subtraction techniques: a review," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pp. 3099–3104, The Hague, Netherlands, October 2004.

[21] M. Sabet, R. A. Zoroofi, K. S. Niiat, and M. Sabbaghian, "Automated face tracking with self correction capability," in *2011 International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, pp. 280–284, Dalian, China, October 2011.

[22] P. Katz, M. Aron, and A. Alfalou, "Joint transform correlation for face tracking: elderly fall detection application," in *Proceedings of the SPIE*, vol. 8748, p. 87480I, San Francisco, CA, USA, March 2013.

[23] J. L. de Bougrenet de la Tocnaye, E. Quémener, and Y. Pétillot, "Composite versus multichannel binary phase-only filtering," *Applied Optics*, vol. 36, no. 26, pp. 6646–6653, 1997.

[24] B. V. K. Vijaya Kumar, "Tutorial survey of composite filter designs for optical correlators," *Applied Optics*, vol. 31, no. 23, pp. 4773–4801, 1992.