ARTICLE IN PRESS

Optics and Lasers in Engineering ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

journal homepage: www.elsevier.com/locate/optlaseng

Optimized swimmer tracking system based on a novel multi-related-targets approach

D. Benarab [a,b], T. Napoléon [a], A. Alfalou [a,*], A. Verney [b], P. Hellard [c]

[a] ISEN Brest, Département Vision, L@bISEN, 20 rue Cuirassé Bretagne, CS 42807, 29228 Brest Cedex 2, France
[b] ACTRIS Brest, 24 Rue Victor Grignard, Guipavas – BP 143, 29803 Brest Cedex 9, France
[c] FFN (Fédération Française de Natation), 14 Rue Scandicci, 93508 Pantin Cedex, France

ARTICLE INFO

Article history:
Received 2 February 2016
Received in revised form
9 May 2016
Accepted 17 May 2016

Keywords:
Tracking system
Swimmer tracking
Dynamic fusion
Correlation
NL-JTC
Color histogram
Multi-related-targets

ABSTRACT

Robust tracking is a crucial step in automatic swimmer evaluation from video sequences. We designed a robust swimmer tracking system using a new multi-related-targets approach. The main idea is to consider the swimmer as a bloc of connected subtargets that advance at the same speed. If one of the subtargets is partially or totally occluded, it can be localized by knowing the position of the others. In this paper, we first introduce the two-dimensional direct linear transformation technique that we used to calibrate the videos. Then, we present the classical tracking approach based on dynamic fusion. Next, we highlight the main contribution of our work, which is the multi-related-targets tracking approach. This approach, the classical head-only approach and the ground truth are then compared, through testing on a database of high-level swimmers in training, national and international competitions (French National Championships, Limoges 2015, and World Championships, Kazan 2015). Tracking percentage and the accuracy of the instantaneous speed are evaluated and the findings show that our new appraoach is significantly more accurate than the classical approach.

1. Introduction

Swimming is a very popular sport and the competition between national teams can be fierce. Unsurprisingly, swimming performance is much investigated [1–4] during training sessions and competitions, mainly through statistical and kinematic studies (study of the movements regardless of the causes that produce them). The current systems evaluate swimmers in terms of speed, breathing cycles, diving, turning and swimming techniques, etc. These data are used by coaches to assist the swimmers in analyzing their swimming strategies in order to improve their performance.

Most of these systems acquire data from sensors that the swimmers wear. These sensors provide accurate biomechanical data, but their use is prohibited in competition. Moreover, even in training sessions, wearing the sensors restricts movement. Other systems are based on the manual annotation of the swimmer's position in each video frame in order to estimate various measures. This type of systems yields accurate analysis without imposing constraints on the swimmers, but annotation is both time-consuming and labor-intensive. We thus sought to design an automatic swimmer tracking system with minimal user intervention and no physical markers.

For accurate analysis of swimmer performance, we first sought to optimize the tracking module. In the literature, the object tracking systems have been based on such detection techniques as the Non-Linear Joint Transform Correlator (NL-JTC) [5,6,19], color histograms [7,8], Local Binary Patterns (LBP) [9,10] or Histograms of Oriented Gradients (HOG) [11,12]. Starting from the principle that each of these techniques has advantages and limitations, Benarab et al. proposed a swimmer multi-tracking system [13] that combines several tracking approaches. The decision is made after several iterations and the best detected target is chosen according to a similarity criterion based on the color histogram and the history of detections.

This approach improved tracking performance compared with the mono-tracking approaches that use only one descriptor to characterize the target. However, the decision criterion is not optimal because it favors color information. In order to improve this criterion and better merge the data, Benarab et al. then proposed a dynamic fusion approach [3] motivated by the complementarity between the NL-JTC and color histograms technique. This approach significantly enhanced the tracking results in the case of complete or partial visibility of the target. Nevertheless, tracking is impossible in the case of total occlusion.

In order to solve this issue, we propose a multi-related-targets approach that tracks two targets in parallel (head and swimsuit). Among other things, this approach takes into account the distance

* Corresponding author.

http://dx.doi.org/10.1016/j.optlaseng.2016.05.013
0143-8166/© 2016 Published by Elsevier Ltd.

between the two targets as well as the average swimming speed. This approach has two main advantages: it improves tracking accuracy and makes it possible to find even an entirely occluded target based on the other target when this latter is visible. In addition to our multi-related-targets approach, we propose to calibrate the video in order to optimize the prediction of the swimming trajectory. Last, we estimate the swimmer's instantaneous speed in order to study the performance.

In this paper, we start by a general overview of tracking techniques in Section 2. We then detail the steps of the swimmer tracking system based on the dynamic fusion technique in Section 3. In Section 4, we introduce the concept of the multi-related-targets approach and its complex decision criterion. Finally, in Section 5, we present the experimental protocol and the results in the form of a comparative study.

2. Methodology of tracking

2.1. JTC-based techniques

The joint transform correlator (JTC) is a correlation-based technique that consists of comparing two images in order to detect, localize or identify the reference in a target image. This method has proven its efficiency by the numerous publications in the literature [5,6,14]. Most often, this originally all-optical method is implemented numerically for reasons of simplicity and portability.

Different JTC variants have been proposed, the most common being the NL-JTC technique, which is based on a non-linear thresholding function applied to the joint spectrum. This allows for greater control of the accuracy and robustness of the technique by varying the sharpness of the correlation peaks. However, this technique is sensitive to contour deformation, blur and noise.

2.2. Histogram-based techniques

Histogram-based methods are designed to numerically describe the color, texture or contour characteristics of the target to be tracked. To do so, they encode the occurrence of particular information in the image in each component of a histogram. It should be noted that a single histogram can be calculated on the whole image, in which case we speak of global description, but it is also possible to calculate several histograms, one for each region of the image, and in this case we speak of local description. The comparison between the histograms corresponding to the similarity between the reference image and the target image is obtained through a calculation of distance (Bhattacharya [10], Chi squared [15], etc.).

The well-known techniques as reported in the literature are color histograms, HOG and local LBP. The essential notion behind the color histogram technique is that the target appearance can be characterized by the distribution of its color. This technique is robust with respect to the orientation and deformation of the object, but it is sensitive to changes in lighting conditions and the confusion between same-color objects. As for the HOG descriptor, an image can be described by the distribution of intensity gradients or edge directions. The LBP technique describes the target texture by calculating the relative gray level of the neighboring pixels. These latter two techniques are robust with respect to lighting change. However, in the case of swimmer tracking, the deformation of the target due to splashing will strongly affect the contours and make these methods less discriminative.

2.3. Dynamic fusion-based technique

The above-mentioned methods for automatic swimmer tracking are often insufficient due to the particularities of the swimming environment. In order to develop a robust system, Benarab et al. [3] proposed to fuse the NL-JTC and color histogram techniques, as they showed a high degree of complementarity. Indeed, the NL-JTC is accurate for localization but it is sensitive to contour deformation, while the color histogram approach is imprecise for localization but robust regarding the deformations of contour.

Therefore, the dynamic fusion technique was used in the design of a tracking system with accurate localization and robustness with respect to contour deformation. The fusion provides a richer description of the target by combining contour and color characteristics. This technique showed improved performance compared with the techniques mentioned above. However, it remains sensitive to strong occlusions and is insufficient in these situations.

3. Dynamic fusion-based swimmer tracking system

In order to propose a robust swimmer tracking system, we set up a system based on the following successive steps: swimming pool calibration, reference selection, trajectory prediction and swimmer recognition and tracking. It should be noted that the last three steps are repeatedly applied to all of the analyzed video.

The calibration module ensures the passage from pixel coordinates to metric coordinates and vice versa. The tracking process is then initialized in the step of selecting a reference, which is the part of the swimmer to be tracked: head, swimsuit, etc. Then, the trajectory prediction module is applied based on the calibration to restrict the region of interest. Next, the images are analyzed to detect and recognize the target using the dynamic fusion technique. Last, the reference is replaced by the detected target to relaunch the next iteration of the tracking process.

3.1. Two-dimensional direct linear transformation (2D DLT) calibration

Swimmer performance analysis from a video sequence consists of measuring and studying the swimming movements. To do this, we need to calculate the spatial coordinates of the target from the pixel coordinates in the image. This passage from pixel coordinates to metric coordinates is called calibration. It relies on the prior knowledge of the camera projection parameters, which vary depending on the camera settings and positioning. For this, the DLT approach was proposed by Abdel-Aziz and Karara [16]. In the case of calibrating a swimming pool, which is a 2D space, it is sufficient to apply a simplified 2D DLT equation:

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

where u and v are the pixel coordinates of a point in the image, x and y are the real metric coordinates of the same point. $L_{1..8}$ represent the calibration parameters which are the set of the unknowns of this system of equations. In order to calculate these eight parameters and solve this system of equations, it is required to select four points whose metric and pixel coordinates are known, as shown in Fig. 1. Once we calculate $L_{1..8}$, the following equation allows the passage from pixel to metric coordinates and vice versa.