Optics Communications 356 (2015) 256–268

Contents lists available at ScienceDirect

Optics Communications

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

Optimized swimmer tracking system by a dynamic fusion of correlation and color histogram techniques

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 9 June 2015
Accepted 21 July 2015

Keywords:
Correlation
NL-JTC
Tracking systems
Swimmer tracking
Color histogram
Normalization
Dynamic fusion

ABSTRACT

To design a robust swimmer tracking system, we took into account two well-known tracking techniques: the nonlinear joint transform correlation (NL-JTC) and the color histogram. The two techniques perform comparably well, yet they both have substantial limitations. Interestingly, they also seem to show some complementarity. The correlation technique yields accurate detection but is sensitive to rotation, scale and contour deformation, whereas the color histogram technique is robust for rotation and contour deformation but shows low accuracy and is highly sensitive to luminosity and confusing background colors. These observations suggested the possibility of a dynamic fusion of the correlation plane and the color scores map. Before this fusion, two steps are required. First is the extraction of a sub-plane of correlation that describes the similarity between the reference and target images. This sub-plane has the same size as the color scores map but they have different interval values. Thus, the second step is required which is the normalization of the planes in the same interval so they can be fused. In order to determine the benefits of this fusion technique, first, we tested it on a synthetic image containing different forms with different colors. We thus were able to optimize the correlation plane and color histogram techniques before applying our fusion technique to real videos of swimmers in international competitions. Last, a comparative study of the dynamic fusion technique and the two classical techniques was carried out to demonstrate the efficacy of the proposed technique. The criteria of comparison were the tracking percentage, the peak to correlation energy (PCE), which evaluated the sharpness of the peak (accuracy), and the local standard deviation (Local-STD), which assessed the noise in the planes (robustness).

1. Introduction

Research in the field of people tracking systems has resulted in a multitude of applications [4,1,6,21,12], notably in security, health-care and sports. For example, video surveillance systems are used to detect and record suspicious activities and enhance public safety [34,28]. In the health-care field, systems are used to track patients e.g. elderly to detect falls or other disturbing behaviors and to send an alarm message to the appropriate caregiver [23]. In sports, research into tracking devices has essentially been dedicated to the evaluation and improvement of athletic performances [29,18,26]. The French Swimming Federation (FFN, Fédération Française de Natation) [14] is one of the leaders in research on swimmer evaluation and the optimization of swimming strategies in France [13,30]. A recent FFN project was to develop an automatic swimmer tracking system without physical markers or sensors. For that, we explored the possibility of using vision sensors and tracking the swimmers via video sequences. Swimmer tracking is a difficult task, in great part because of the aquatic environment: splashing, water movement and the movement of the swimmers through the water. In the literature, various object tracking systems have been proposed [4,12,23,31,17], such as those based on correlation techniques like the nonlinear joint transform correlator (NL-JTC) [3,23]. These systems take a plane containing a reference image and a target image as an input and provide a correlation plane that contains two peaks corresponding to the location of the tracked object as an output. Other tracking systems are based on color histograms [18,17], where the reference object is sought in the target scene through comparisons based on each color coded in the histogram. In the same vein, information on textures (local binary patterns: LBP) [10,31] or gradients (histogram of oriented gradient: HOG) [12,34] can be used to determine the similarity between two images.

* Corresponding author.
** Principal corresponding author.
E-mail addresses: djamel.benarab@isen-bretagne.fr (D. Benarab), ayman.al-falou@isen-bretagne.fr (A. Alfalou).

http://dx.doi.org/10.1016/j.optcom.2015.07.056
0030-4018/

Unfortunately, most of these approaches are not robust enough to be used in an uncontrolled environment such as swimmer tracking in swimming pools with a range of limitations. In particular, correlation-based techniques, as well as the LBP and HOG, are sensitive to contour deformations and sometimes generate artifacts in the correlation plane, related to the correlation between the reference and target edges. On the other hand, color-based techniques can confuse objects having the same color and they are inaccurate in terms of localization. Last, the calculation time for techniques based on histograms is often long because an exhaustive search is needed to locate the reference image in the target scene. To overcome these problems and design a robust tracking system, we combined the correlation and color histogram techniques. Similar to Alfalou et al. [5], who applied a nonlinear function to the correlation plane in order to denoise it, we aimed to improve detection by optimizing the decision. Alfalou et al. [5] proposed a non-linear function that enhanced the correlation peak and reduced the noise, which improved the value of their decision criterion based on the Peak-To-Correlation energy (PCE) [4]. However, this optimization is blind in that it does not take into account any information related to the tracked object, nor does it allow the decision to be influenced in any way.

To overcome these shortcomings, we propose a new approach that directly combines the score planes of two methods: correlation and color histogram. This can be seen as a dynamic filtering of the correlation plane by a specific function – i.e. the color scores plane – that depends on the tracked object and adapts to its variations and evolution. This dynamic correlation plane containing the information on the similarity in color between the reference and the target enhances the contour information from the correlation and thereby improves the final decision. In addition to our main contribution based on the fusion of an optical method (correlation) and a numerical method (color histogram), we carefully analyzed the correlation planes and color scores and propose the following optimizations:

- Extraction of a potential correlation sub-plane that is the same size as the color scores plane to enable the dynamic fusion of the two planes.
- Optimization of the input plane by reducing its size in order to improve the calculation time for the correlation.
- Elimination of artifacts related to the correlation approach.
- Pretreatment of the input images with a Sobel filter to minimize the artifacts resulting from the blurred movement in the video sequences.
- Use of the integral images proposed by Viola and Jones [32] to improve the calculation time for the color histogram approach. In this paper, we begin by presenting the correlation and color histogram techniques in Section 2. In Section 3, we explain our technical choices for tracking and then present our dynamic fusion approach and show its advantages through experiments on a synthetic database. Several correlation plane optimizations are proposed and detailed in Section 4. The application of our fusion approach to swimmer tracking is evaluated on videos of real competitions in Section 5. We conclude in Section 6 by highlighting our contributions and the perspectives offered by our work.

2. Tracking techniques

Numerous techniques have been proposed in the literature on object tracking [4,11,18,12,9,10]. Among the best known are correlation [4,24,23], color histogram [18,8,17], histogram of oriented gradient (HOG) [12,33] and local binary patterns (LBP) [10,2]. In this section, we focus on the first two techniques, which use different but important information: contour and color for the correlation and the color histogram technique respectively. We then consider the strengths and weaknesses of each method in order to optimally combine them in the process presented in Section 3.

2.1. Correlation approaches

Correlation is a classical technique, well known in the field of optical pattern detection and recognition. Two architectures have been proposed [4]: the Vanderlugt correlator (VLC), mainly used for identification, and the joint transform correlator (JTC), which is more robust for detecting and localizing a given object in a scene. The JTC architecture is preferred as part of a tracking system because of its ability to detect and localize at the same time, thereby improving the calculation time. However, this originally optical approach is physically difficult to set up, especially in our swimmer tracking application. For this reason, we chose to implement it numerically through an image processing algorithm.

The fundamental principle of the JTC architecture is the application of a Fourier transform to an input plane made up of a reference image and a target image – i.e., a target scene – to obtain a joint spectrum of these images. Then, an inverse Fourier transform is applied to the intensity of this spectrum in order to obtain the correlation plane containing the information on the reference/target similarity. The architecture of the JTC is presented in Fig. 1 and can be formulated as follows:

$$f(x, y) = c(x, y) + r(x - d, y - d) \tag{1}$$

where $f(x, y)$ in the spatial domain represents the input plane consisting of the target image c of the coordinates (x,y) and the reference image r of the coordinates $(x - d, y - d)$ placed at a distance d from the target. The classical joint spectrum $F(u, v)$ in the frequency domain is thus obtained by the Fourier transform applied to the input plane (Eq. (1)):

$$\begin{aligned} F(u, v) &= FT(c(x, y)) + FT(r(x - d, y - d)) \\ F(u, v) &= |C(u, v)| \exp[\phi_c(u, v)] \\ &\quad + |R(u, v)| \exp[\phi_r(u, v)] \exp[-j(ud + vd)] \end{aligned} \tag{2}$$

Then, the intensity of the classical joint spectrum is recovered using the following formula:

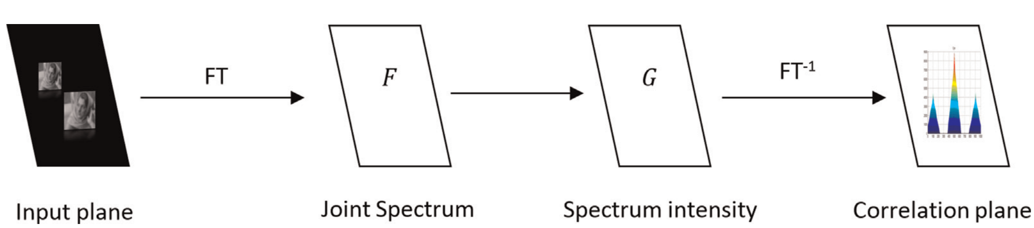

Fig. 1. C-JTC correlator architecture. The input plane contains both images reference and target. Then, the Fourier transform is applied to obtain the joint spectrum. Finally, the correlation plane is obtained by applying the inverse Fourier transform on the intensity of the joint spectrum. The obtained correlation plane contains two peaks of cross-correlation and an auto-correlation peak.