Motion Detection Using Color Space-Time Interest Points

Insaf Bellamine and Hamid Tairi

Abstract Detecting moving objects in sequences is an essential step for video analysis. Among all the features which can be extracted from videos, we propose to use Space-Time Interest Points (STIP). STIP are particularly interesting because they are simple and robust low-level features providing an efficient characterization of moving objects within videos. In general, Space-Time Interest Points are based on luminance, and color has been largely ignored. However, the use of color increases the distinctiveness of Space-Time Interest Points. This paper mainly contributes to the Color Space-Time Interest Points (CSTIP) extraction and detection. To increase the robustness of CSTIP features extraction, we suggest a pre-processing step which is based on a Partial Differential Equation (PDE) and can decompose the input images into a color structure and texture components. Experimental results are obtained from very different types of videos, namely sport videos and animation movies.

Keywords Color space-time interest points · Color structure-texture image decomposition · Motion detection

1 Introduction

Detecting moving objects in dynamic scenes is an essential task in a number of applications such as video surveillance, traffic monitoring, video indexing, recognition of gestures, analysis of sport-events, sign language recognition, mobile robotics and the study of the objects' behavior (people, animals, vehicles, etc….). In the literature, there are many methods to detect moving objects, which are based

I. Bellamine (✉) · H. Tairi
Department of Computer Science, Sidi Mohamed Ben Abdellah University LIIAN,
Fez, Morocco
e-mail: insaf.bellamine@usmba.ac.ma

H. Tairi
e-mail: hamid.tairi@usmba.ac.ma

A. El Oualkadi et al. (eds.), *Proceedings of the Mediterranean Conference on Information & Communication Technologies 2015*, Lecture Notes in Electrical Engineering 380, DOI 10.1007/978-3-319-30301-7_11

on: optical flow [1], difference of consecutive images [2, 3], Space-Time Interest Points [4] and modeling of the background (local, semi-local and global) [5].

For grayscale sequences, the notion of Space-Time Interest Points (STIP) is especially interesting because they focus information initially contained in thousands of pixels on a few specific points which can be related to spatiotemporal events in an image. Laptev and Lindeberg were the first who proposed STIP for action recognition [4], by introducing a space-time extension of the popular Harris detector [6]. They detect regions having high intensity variation in both space and time as spatio-temporal corners. The STIP detector usually suffers from sparse STIP detection [4]. Later, several other methods for detecting STIP have been reported. Dollar et al. [7] improved the sparse STIP detector by applying temporal Gabor filters and selecting regions of high responses. Dense and scale-invariant Space-Time Interest Points were proposed by Willems et al. [8]. An evaluation of these approaches has been proposed in [9].

For color sequences, we propose a color version of Space-Time Interest Points extension of the Color Harris detector [10] to detect what they call "Color Space-Time Interest Points detector" (CSTIP). To increase the robustness of CSTIP features extraction, we propose a color version of Structure-Texture image decomposition extension of the Structure-Texture image decomposition technique [11].

The Color Structure-Texture image decomposition technique is essential for understanding and analyzing images depending on their content. This one decomposes the color image f into a color structure component u and a color texture component v, where f = u + v. The use of the color structure-texture image decomposition method enhances the performance and the quality of the motion detection.

Our contribution is twofold: First, after giving two consecutive images, we split each one into two components (Structure, Texture). Second, we compute the Color Space Time Interest Points (CSTIP) associated to the color structure (respectively texture) components by using the proposed algorithm of the detection of Color Space- Time Interest Points. The fusion procedure is implemented to compute the final Color Space- Time Interest Points. One of the aims of the present paper is to propose a new parallel algorithm so as to generate good results of moving objects.

This paper is organized as follows: Sect. 2 presents the materials and methods, Sect. 3 presents our proposed approach, and finally, Sect. 4 shows our experimental results.

2 Materials and Methods

2.1 Color Structure-Texture Image Decomposition

Let f be an observed image which contains texture and/or noise. Texture is characterized as repeated and meaningful structure of small patterns. Noise is

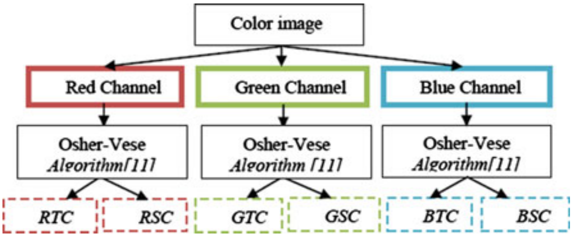

Fig. 1 The color structure-texture image decomposition algorithm

characterized as uncorrelated random patterns. The rest of an image, which is called cartoon, contains object hues and sharp edges (boundaries). Thus an image f can be decomposed as f = u + v, where u represents image cartoon and v is texture and/or noise. In recent years, several models based on total variation, which are inspired by the ROF model, were created [12]. In the literature there is also another model called Mayer [13, 14] that is more efficient than the ROF model. Many algorithms have been proposed to solve numerically this model. In the following, we represent the most popular algorithm, Osher-Vese algorithm [11] for grayscale sequences:

This algorithm is based on the decomposition model as follows:

$$F^{OV}_{\lambda,\mu,p}(u,g) = J(u) + \lambda \|f - (u + \mathrm{div}(g))\|^2_{L^2} + \mu \left\| \sqrt{g_1^2 + g_2^2} \right\|_{L^p} \tag{1}$$

The additional term $\|f - (u + \mathrm{div}(g))\|^2_{L^2}$. ensures that we have the constraint f = u + v where f is the original image that consists of two components: a component u containing all the objects, a component v: is the sum of textures and noise.

In this paper, we propose an adapted decomposition Osher-Vese algorithm for color images shown in Fig. 1. where RTC, RSC, GTC, GSC, BTC and BSC are respectively the red texture component, red structure component, green texture component, green structure component, blue texture component and blue structure component .The color texture component of the color image will be calculated by the combination between RTC, GTC and BTC. The color structure component of the color image will be calculated by the combination between RSC, GSC and BSC.

The Color Aujol Structure-Texture decomposition method has been applied on the karate's fight image of size 352 × 288. The results of the decomposition are shown in Fig. 2.

2.2 Color Space-Time Interest Points

The idea of color interest points [10] in the spatial domain can be extended into the color spatio-temporal domain by requiring the image values in space-time to have large variations in both the spatial and the temporal dimensions. For grayscale

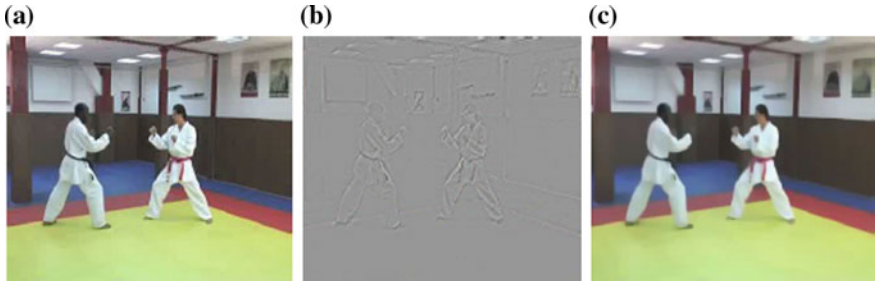

Fig. 2 The color structure-texture image decomposition algorithm: **a** original images (karate's fight image), **b** the color texture component of karate's fight image, **c** the color structure component of karate's fight image

sequences, Laptev et al. [4], proposed a spatio-temporal extension of the Harris detector to detect what they call "Space-Time Interest Points", denoted STIP in the following. Detection of space-time interest points is performed by using the Hessian-Laplace matrix H, which is defined by:

$$H(x, y, t) = g(x, y, t; \sigma_s^2, \sigma_t^2) \otimes \begin{pmatrix} \frac{\partial^2 I(x,y,t)}{\partial x^2} & \frac{\partial^2 I(x,y,t)}{\partial x \partial y} & \frac{\partial^2 I(x,y,t)}{\partial x \partial t} \\ \frac{\partial^2 I(x,y,t)}{\partial x \partial y} & \frac{\partial^2 I(x,y,t)}{\partial y^2} & \frac{\partial^2 I(x,y,t)}{\partial y \partial t} \\ \frac{\partial^2 I(x,y,t)}{\partial x \partial t} & \frac{\partial^2 I(x,y,t)}{\partial y \partial t} & \frac{\partial^2 I(x,y,t)}{\partial t^2} \end{pmatrix} \tag{2}$$

where I (x, y, t) is the intensity of the pixel s (x, y) at time t denotes the convolution. $g(x, y, t; \sigma_s^2, \sigma_t^2)$ is the Gaussian smoothing (see Eq. (4)) and the two parameters (σ_s and σ_t) control the spatial and temporal scale.

For color sequences, we propose a color version of space-time interest points extension of the Color Harris detector [10] to detect what they call "Color Space-Time Interest Points", denoted CSTIP in the following. The information given by the three RGB color channels: Red, Green and Blue.

Color plays a very important role in the stages of feature detection. Color provides extra information which allows the distinctiveness between various reasons of color variations, such as change due to shadows, light source reflections and object reflectance variations.

The H matrix can be computed by a transformation in the RGB space [15]. The first step is to determine the gradients of each component of the RGB color space. The gradients are then transformed into desired color space. By multiplying and summing of the transformed gradients, all components of the matrix are computed as follows:

$$\begin{aligned}
\frac{\partial^2 I(x,y,t)}{\partial x \partial t} &= \frac{\partial^2 R(x,y,t)}{\partial x \partial t} + \frac{\partial^2 G(x,y,t)}{\partial x \partial t} + \frac{\partial^2 B(x,y,t)}{\partial x \partial t} \\
\frac{\partial^2 I(x,y,t)}{\partial x^2} &= \frac{\partial^2 R(x,y,t)}{\partial x^2} + \frac{\partial^2 G(x,y,t)}{\partial x^2} + \frac{\partial^2 B(x,y,t)}{\partial x^2} \\
\frac{\partial^2 I(x,y,t)}{\partial y^2} &= \frac{\partial^2 R(x,y,t)}{\partial y^2} + \frac{\partial^2 G(x,y,t)}{\partial y^2} + \frac{\partial^2 B(x,y,t)}{\partial y^2} \\
\frac{\partial^2 I(x,y,t)}{\partial t^2} &= \frac{\partial^2 R(x,y,t)}{\partial t^2} + \frac{\partial^2 G(x,y,t)}{\partial t^2} + \frac{\partial^2 B(x,y,t)}{\partial t^2} \\
\frac{\partial^2 I(x,y,t)}{\partial y \partial t} &= \frac{\partial^2 R(x,y,t)}{\partial y \partial t} + \frac{\partial^2 G(x,y,t)}{\partial y \partial t} + \frac{\partial^2 B(x,y,t)}{\partial y \partial t} \\
\frac{\partial^2 I(x,y,t)}{\partial x \partial y} &= \frac{\partial^2 R(x,y,t)}{\partial x \partial y} + \frac{\partial^2 G(x,y,t)}{\partial x \partial y} + \frac{\partial^2 B(x,y,t)}{\partial x \partial y}
\end{aligned} \tag{3}$$

where R(x,y,t), G(x,y,t) and B(x,y,t) are respectively the red, green and blue components of the color pixel s (x, y) at time t.

As with the Color Harris detector, a Gaussian smoothing is applied both in spatial domain (2D filter) and temporal domain (1D filter).

$$g(x,y,t;\sigma_s^2,\sigma_t^2) = \frac{\exp(-\frac{x^2+y^2}{2\sigma_s^2} - \frac{t^2}{2\sigma_t^2})}{\sqrt{(2\pi)^3 \sigma_s^4 \sigma_t^2}} \tag{4}$$

The two parameters (σs and σt) control the spatial and temporal scale. As in [4], the color spatio-temporal extension of the Color Harris corner function, entitled "salience function", is defined by:

$$M(x,y,t) = \det(H(x,y,t)) - k \times \operatorname{trace}(H(x,y,t))^3 \tag{5}$$

where k is a parameter empirically adjusted at 0.04, det is the determinant of the matrix H and trace is the trace of the same matrix.

3 Proposed Approach

The most famous algorithm to detect Space-Time Interest Points is that of Laptev [4]; however we can reveal four major problems when a local method is used:

- Texture, Background, lighting changes and Objects that may influence the results;
- Noisy datasets such as the KTH dataset [16], which is featured with low resolution, strong shadows, and camera movement that renders clean silhouette extraction impossible;

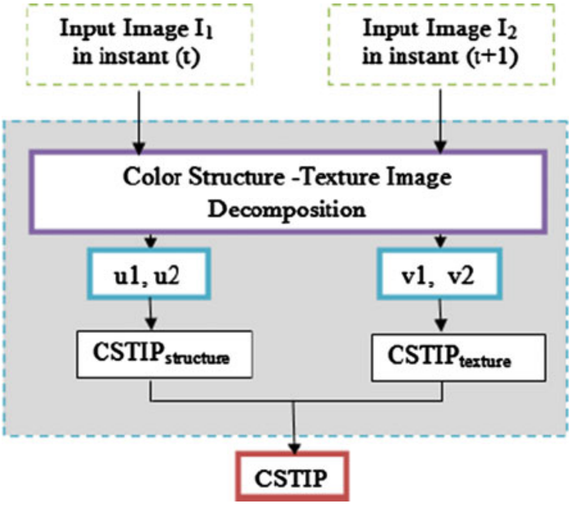
Input Image I1 in instant (t)
Input Image I2 in instant (t+1)
Color Structure -Texture Image Decomposition
u1, u2
v1, v2
CSTIP structure
CSTIP texture
CSTIP

Fig. 3 The adapted Laptev algorithm

- Features extracted are unable to capture smooth and fast motions, also they are sparse. This also explains why they generate poor results;
- RGB video must be converted into grayscale video.

However, to overcome the four problems, we propose a new technique based on the Color Space-Time Interest Points and Color Structure-Texture Image Decomposition algorithm (Fig. 1).

Let CSTIP denote the final extracted color space-time interest points; $\mathrm{CSTIP}_{\mathrm{texture}}$ and $\mathrm{CSTIP}_{\mathrm{structure}}$ denote respectively the extracted Space-Time Interest Points fields on texture components and structure components. The Color Structure-Texture Image Decomposition decomposes each input image into two components (u1, v1) for the first image I1 and (u2, v2) for the second image I2, where u1 and u2 denote the color structure components, v1 and v2 are the color texture components.

The proposed approach detects the color space-time interest points by given a pair of consecutive images in several stages; it will help to have a good detection of moving objects and even reduce the execution time by proposing a parallel algorithm (see Fig. 3). Our new Space-Time Interest Points will be calculated as the following:

$$\mathrm{CSTIP} = \mathrm{CSTIP}_{\mathrm{structure}} \cup \mathrm{CSTIP}_{\mathrm{texture}} \tag{6}$$

4 Results and Discussion

This section presents several experimental results on different kinds of sequences.

In Fig. 4, we represent some examples of clouds of Color Space-Time Interest Points detected in the sequence (sport video: “karate’s fight” lasts for 2 min and

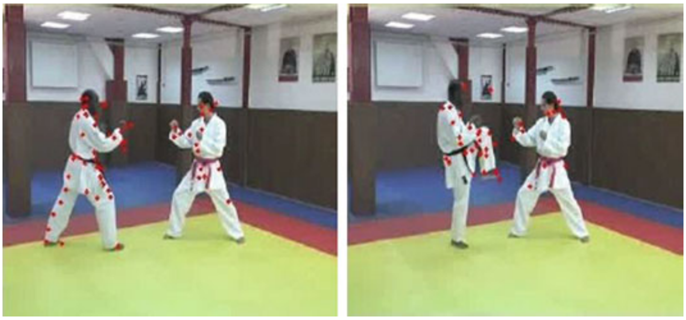

Fig. 4 Examples of clouds of color space-time interest points, in each frame (Image (t = 44)/ Image (t = 45)), we use the parameters (σt = 1.5, k = 0.04 and σs = 1.5)

49 s with 200 images and the size of each image frame is 400 by 300 pixels.). The results illustrated in Fig. 4, show that the objects moving (the two players) are detected with our approach (Fig. 3). The red points represent the extracted Color Space-Time Interest Points with our proposed approach (Fig. 3).

In order to correctly gauge performance of our algorithm (Fig. 3), we will proceed with a comparative study to use the mask of the moving object and the precision.

For each moving object, we have a number of the color space-time interest points detected in the moving object (NTP) and a number of the color space-time interest points extracted off the moving object (NFP).

The test is performed many examples of sequences [17] and gives in the following results:

The proposed approach is much less sensitive to noise, and the reconstruction of the image, it also extracts the densest features (see Fig. 5).

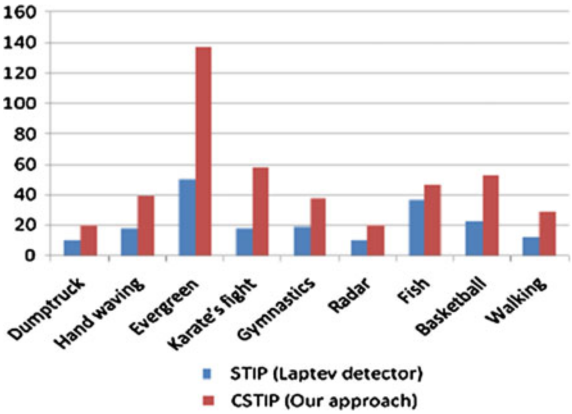

Fig. 5 Number of color space-time interest points extracted in each frame

Table 1 Compared results

Videos	Laptev detector [4]	Our approach (Fig. 3)
	Precision (%)	Precision (%)
Radar	80	84
Fish	94	96
Basketball	91	94
Walking	90	92
Dumptruck	78	89
Hand waving	92	93
Evergreen	96	98
Karate’s fight	88	96
Gymnastics	89	92

The results, illustrated in Table 1, show that our approach allows a good detection of moving objects.

5 Conclusion

This paper mainly contributes to the Color Space-Temporal Interest Points extraction and detection. To increase the robustness of CSTIP features extraction, our approach suggests a pre-processing step which can decompose the input images into a color structure component and a color texture component. The experiment results show that the proposed method outperforms the original STIP feature detector with promising results.

References

1. Simac, A.: Optical-flow based on an edge-avoidance procedure. Comput. Vis. Image Underst. **113**(2009), 511–531 (2008)
2. Galmar, E., Huet, B.: Analysis of vector space model and spatiotemporal segmentation for video indexing and retrieval. In: CIVR van Leeuwen, J. (ed.) Computer Science Today. Recent Trends and Developments. Lecture Notes in Computer Science, vol. 1000. Springer, Berlin Heidelberg New York (1995)
3. Zhou, B.: A phase discrepancy analysis of object motion, ACCV 2010
4. Laptev, I.: On space-time interest points. Int. J. Comput. Vis. **64**(2/3):107–123 (2005)
5. Nicolas, V.: Suivi d’objets en mouvement dans une séquence vidéo. Doctoral thesis, Paris Descartes university (2007)
6. Harris, C., Stephens, M.J.: A combined corner and edge detector. In: Alvey Vision Conférence (1988)
7. Dollar, P., Rabaud, V., Cottrell, G., Belongie, S.: Behavior recognition via sparse spatio-temporal features. In: VS-PETS (2005)
8. Willems, G., Tuytelaars, T., Van Gool, L.: An efficient dense and scale-invariant spatio-temporal interest point detector. Eur. Conf. Comput. Vis. **5303**(2), 650–663 (2008)

9. Wang, H.: Evaluation of Local Spatio-Temporal Features for Action Recognition. BMVC '09 London (2009)
10. Stöttinger, J., Hanbury, A., Sebe, N.: Sparse color interest points for image retrieval and object categorization IEEE Trans. Image Process. **21**(5), (2012)
11. Vese, L., Osher, S.: Modeling textures with total variation minimization and oscillating patterns in image processing. J. Sci. Comput. **19**(1–3), 553–572 (2002)
12. Gilles, J.: Décomposition et détection de structures géométriques en imagerie. Doctoral thesis, Ecole Normale Supérieure de Cachan (2006)
13. Meyer, Y.: Oscillating Patterns in Image Processing and in Some Nonlinear Evolution Equations. The Fifteenth Dean Jacquelines B. Lewis Memorial Lectures, American Mathematical Society (2001)
14. Chambolle, A.: An algorithm for total variation minimization and application. J. Math. Imaging vis. **20**(1–2), 89–97 (2004)
15. van de Weijer, J., Gevers, T.: Edge and corner detection by photometric quasi-invariants. IEEE Trans. Pattern Anal. Mach. Intell. **27**(4), 625–630 (2005)
16. Laptev, I., Caputo, B.: Recognizing human actions: a local SVM approach. ICPR **3**, 32–36 (2004)
17. Baker, et al.: A database and evaluation methodology for optical flow. Int. J. Comput. Vision **92**(1), 1–31 (2011)