Pattern Analysis and Applications (2018) 21:829–846
https://doi.org/10.1007/s10044-017-0669-9

INDUSTRIAL AND COMMERCIAL APPLICATION

Unsupervised detection of ruptures in spatial relationships in video sequences based on log-likelihood ratio

Abdalbassir Abou-Elailah[1] · Isabelle Bloch[1] · Valerie Gouet-Brunet[2]

Received: 14 April 2017 / Accepted: 6 November 2017 / Published online: 18 November 2017

Abstract

In this work, we propose a new approach to automatically detect ruptures in spatial relationships in video sequences, based on low-level primitives, in unsupervised manner. The spatial relationships between two objects of interest are modeled using angle and distance histograms as examples. The evolution of the spatial relationships during time is estimated from the distances between two successive angle or distance histograms and then considered as a temporal signal. The evolution of a spatial relationship is modeled by a linear Gaussian model. Then, two hypotheses "without change" and "with change" are considered, and a log-likelihood ratio is computed. The distribution of the log-likelihood ratio, given that H_0 is true, is estimated and used to compute the p value. The comparison of this p value to a significance level α, expressing the probability of false alarms, allows us to detect significant ruptures in spatial relationships during time. In addition, this approach is generalized to detect multiple object events such as merging, splitting, and other events that contain ruptures in their spatial relationships evolution. This work shows that the description of spatial relationships across time is a promising step toward event detection.

Keywords Spatial relationships · Distances between histograms · Detection of ruptures · Hypotheses testing · Log-likelihood ratio · Significance level

1 Introduction

1.1 Context

Nowadays, the growth of video content is exponential and methods for intelligent video systems are needed. For this reason, many intelligent video surveillance systems are developed in the literature, and each system is dedicated to a specific application, such as sport match analysis, people counting, analysis of personal movements in public shops, behavior recognition in urban environments, and drowning detection in swimming pools.[1] The VSAM project [1] was probably one of the first projects dedicated to surveillance from video sequences. The ICONS project [2] aimed to recognize the incidents in video surveillance sequences. The goal of the three projects ADVISOR [3], ETISEO [4], and CareTracker [5] was to analyze record streaming video, in order to recognize events in urban areas and to evaluate scene understanding. The AVITRACK project [6] was applied to the monitoring of airport runways, while the BEWARE project [7] aimed to use dense camera networks for monitoring transport areas (railway stations, metro). This paper aims at contributing to this domain, and in particular to the question of event detection, by exploiting structural information.

✉ Abdalbassir Abou-Elailah
abd.bassir@gmail.com

Isabelle Bloch
isabelle.bloch@telecom-paristech.fr

Valerie Gouet-Brunet
valerie.gouet@ign.fr

[1] LTCI, Télécom ParisTech, Université Paris-Saclay, 75013 Paris, France

[2] LaSTIG MATIS, IGN, ENSG, Univ. Paris-Est, 94160 Saint-Mande, France

[1] See http://www.cs.ubc.ca/~lowe/vision.html for examples of companies and projects on these topics.

1.2 Objective and motivation

Recently, increasing efforts have been made to address "event" detection problems. Event generally means something unexpected, unusual, an abrupt change in some elements of the scene, etc. Here, we address this problem from the point of view of change in relationships. This contrasts with existing approaches, summarized next. The motivation is to account for structural information, as a complement to the information taken into account in existing approaches, which has already proved useful in many computer vision problems. This new point of view in the present context will allow us to propose a method for detecting strong changes in spatial relationships.

In the two next sections, we describe related works and then provide an overview of the proposed approach.

2 Related work

In this section, we summarize related work and introduce our previous approach, on which we build the newly proposed one, while overcoming its drawback.

2.1 State of the art

In [8], a method was proposed to detect anomalous events based on learning 2-D trajectories. In this approach, a single-class support vector machine (SVM) clustering was used to identify anomalous trajectories. A probabilistic model of scene dynamics was proposed in [9] for applications such as anomaly detection and improvement in foreground detection. A system was proposed in [10] based on learning the statistical motion patterns from trajectories of tracking objects. Then, statistical approaches were used to detect deviations from the learned patterns as unusual behaviors. In [11, 12], histograms of optical flow were used as descriptors with nonlinear one-class SVM to detect abnormal events in video sequences. Unusual events were detected in [13] based on low-level motion features on multiple local monitors.

Tracking moving objects in crowded scenes is very challenging due to the large number of persons and background clutter. In the literature, there are many approaches proposed for abnormal event detection, based on spatio-temporal features. In [14], a statistical approach was proposed for extremely crowded scenes based on modeling the local spatio-temporal motion pattern behavior. In [15], an unsupervised approach was proposed based on motion contextual anomaly of crowd scenes. The authors in [16] used a social force model for abnormal crowd behavior detection. In [17], an abnormal event detection framework in crowded scenes was proposed based on spatial and temporal contexts. The same authors proposed in [18] a similar approach based on sparse representations over normal bases. Recently, Hu et al. [19] proposed a local nearest neighbor distance descriptor to detect anomaly regions in video sequences. The authors in [20] have proposed a video event detection approach based on spatio-temporal path search. It was also applied for walking and running detection

An approach [21] was proposed based on appearance and motion DeepNet framework, to discover anomalous activities in video surveillance scene. In [22], behavior-specific dictionaries through unsupervised learning were proposed for abnormal event detection. A sparse combination learning framework was proposed in [23] for speedy abnormal event detection. In [24], an approach was proposed based on unsupervised dynamic sparse coding for detecting unusual events in videos. In [25], videos are represented by a sparse set of spatio-temporal interest points, and similar spatio-temporal relationships of interest points are merged to form deformable interaction templates. Then, Gaussian process regression was used to model the geometric relations of the features and detect global anomaly. Many approaches [26–28] were proposed for detecting anomaly based on motion pattern analysis and behavior analysis. In [29], SVM with Gaussian sample uncertainty was presented for the problem of event detection. A semantic video representation based on freely available social tagged videos was presented in [30] for video event detection. In [31, 32], a robust structured subspace learning is proposed to integrate image understanding and feature learning into a joint learning framework. These approaches showed encouraging results in image tagging, clustering, classification, and image retrieval applications. Relationships remain implicit in these approaches, while our aim is to make them explicit and to build a detection system on them.

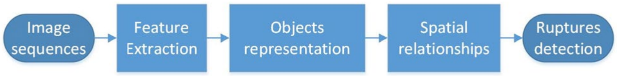


Fig. 1 Overall structure of the proposed approach

2.2 Preliminary work

This section describes our previous work in [33] to manually detect ruptures in the spatial relationships between two objects. First, a fuzzy representation of the objects is estimated exploiting only feature points. Then, spatial relationships between objects are computed, using this sparse representation of the objects. Finally, the evolution of the spatial relationships during time is described by a signal. The block diagram of this approach is depicted in Fig. 1.

Specifically, the spatial distribution of the feature points that are extracted using a detector such as Harris or SIFT is studied for a given object. Feature points can be used to

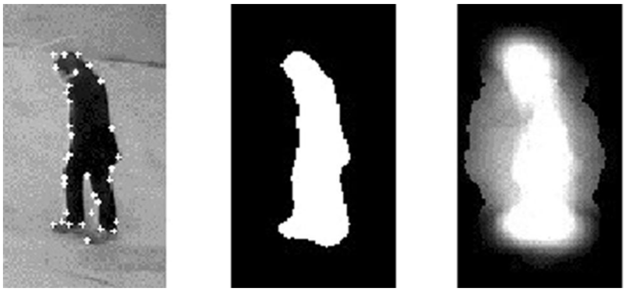


Fig. 2 Original object with the feature points, ground truth of the object, and fuzzy representation of the object [33]

isolate and track objects in video sequences [34, 35]. Thus, it is supposed that each moving object is represented by a set of interest points isolated from others with the help of such techniques. Here, two different criteria are proposed to represent the objects as regions, exploiting only the feature points. The first one is based on the **depth** of the feature points, by assigning a value to each point based on its centrality with respect to the feature points. The second one assigns a value to each point depending on the **density** of its closest feature points. Finally, the depth and density estimations are combined together to form a fuzzy representation of the object, where the combined value at each pixel represents the membership degree of this pixel to the object. This allows reasoning on the feature points or on the fuzzy regions derived from them, without needing a precise segmentation of the objects. Figure 2 shows an example of the fuzzy representation of an object by combining the depth and density estimations [33].

The computation of the spatial relationships between two objects is based on the fuzzy representation of the objects. The angle [36] or distance histogram h between two different objects is computed. The obtained histograms are normalized such that the sum of all bins is equal to 1. Then, the Quadratic-Form (QF) distance [37] is used to assess the distance between the angle or distance histograms during time. Note that other methods for comparing distributions could be used [38, 39]. Let f_i $(i = 0, 1, \ldots, N-1)$ be the frames of the video sequences and h_i be the computed angle or distance histogram between the objects A and B in frame f_i. The function y describing the evolution of the angle or distance histograms over time is defined as $y_i = d(h_i, h_{i+1})$, for each $i = 0, 1, \ldots, N-1$, from the QF distance between two successive histograms h_i and h_{i+1}. The QF distance is defined as $d(h_1, h_2) = \sqrt{ZSZ^T}$, where $Z = h_1 - h_2$ and $S = \{s_{ij}\}$ is the bin-similarity matrix. This distance is commonly used for normalized histograms (the distance histogram for example). Here, we propose an approach to adapt it to the case of angle histograms just by adjusting the elements of the similarity matrix S. We consider that the two histograms h_1 and h_2 defined on $[0, 2\pi]$ consist of k bins B_i. Usually, for a distribution on the real line, the distance between B_i and B_j is defined as follows: $x_{ij} = |B_i - B_j|$, where $1 \le i \le k$ and $1 \le j \le k$. However, in the case of angle histograms, the distance between B_i and B_j is defined as follows: $x^c_{ij} = \min(x_{ij}, 2\pi - x_{ij})$ to account for the periodicity on $[0, 2\pi]$. Thus, the elements of the matrix S are simply defined, in the case of angle histograms, using x^c_{ij} instead of x_{ij} as follows:

$$s_{ij} = 1 - \frac{x^c_{ij}}{\max_{i,j}\left(x^c_{ij}\right)} \tag{1}$$

If a strong change in the spatial relationships occurs at instant t_r $(t_r < N)$, where t_r denotes the instant of rupture; this means that the angle or distance histogram h_r significantly changes compared to previous angle or distance histograms $(h_i, i < t_r)$. Thus, the instant of rupture t_r can be effectively detected from the analysis of the function y.

In this paper, an automatic algorithm is proposed to automatically detect the ruptures in the spatial relationships based on the function (considered as a signal) y.

3 Summary of the proposed approach

According to our motivation in this paper, by rupture we mean a significant change in spatial relations between two objects. Significant is intended in a statistical meaning and is assessed using statistical hypothesis checking. In case of several objects, as soon as there is a change in spatial relations between any two objects, a rupture is identified. As shown in [40–43], incorporating spatial constraints between the objects reveals significant performance improvements in multiple object detection and tracking. Our goal is to detect in an unsupervised way strong changes in spatial relationships in video sequences. This rules out supervised learning-based algorithms which require specific training data. This is useful in all situations where an action or an event can be detected based on such changes or ruptures. Here, we propose to use low-level generic primitives, such as Harris or SIFT detectors [44, 45], which are suitable to efficiently detect and track moving objects during time in video sequences [34, 35].

In our previous work [33], a derivative filter was used to detect the ruptures with a fixed threshold. In this context, a different threshold is needed for each event. Building on this preliminary work, the work presented in this paper consists in automatically detecting ruptures in the spatial relationships using the obtained signal over time. First, a linear Gaussian model is considered to represent the evolution of the spatial relationships during time, whose parameters are estimated from samples from the actual signal. Then, two hypotheses are defined : "$\mathbf{H}_0$: there is no rupture in the considered samples" and "$\mathbf{H}_1$: there is a rupture at an unknown time t_r in the considered samples." Then, a criterion is defined that maximizes the probability of selecting a hypothesis when it is actually true. The second point is the estimation of the unknown time t_r, when $\mathbf{H}_1$ is decided. Here, the log-likelihood ratio is considered to decide which hypothesis is best and to estimate the unknown time t_r, if hypothesis $\mathbf{H}_1$ is selected.

To decide which hypothesis is true, a threshold can be applied on the log-likelihood ratio, and this threshold may depend on the event underhand. Thus, instead of using a fixed threshold in the selection of hypotheses, we find the distribution of the log-likelihood ratio given that $\mathbf{H}_0$ is true. Using this distribution, the probability of false alarm (p value) for a given log-likelihood ratio can be computed and compared to a significance level α, in order to detect the ruptures in the spatial relationships. Moreover, the proposed approach is generalized to handle detection of ruptures in multiple object scenarios. This generalization may model interaction among the objects and can be applied to detect multiple object events such as merging, splitting, and crossing. These kinds of events exhibit strong ruptures in the evolution of the spatial relationships among objects. In addition, the detection of multiple ruptures in spatial behaviors is addressed. The efficiency of the proposed approach is demonstrated on synthetic and real video sequences.

The contributions of this paper are:

- Defining two hypotheses H_0 and H_1 in an efficient manner and estimating their parameters.
- Estimating the distribution of the log-likelihood ratio given that H_0 is true and computing the p value.
- Estimating the time(s) of rupture.
- Generalizing our approach to handle multiple object scenarios.
- Detecting multiple ruptures in spatial behaviors during time.

Note that the proposed approach is based on low-level primitives such as Harris or SIFT detectors, and the goal is to show the efficiency of spatial behavior in event detection frameworks. To the best of our knowledge, this work is the first one that handles ruptures in spatial relationships over time in video sequences. However, the proposed approach in this paper may be used as the first step to detect areas of interest in video sequences that contain ruptures in their spatial behaviors, and then high-level approaches can be applied on these areas for further processing. The proposed approach is hence not competing with existing works, but is rather complementary. The work presented in this paper shows the validity of the approach in the case of mono-view video sequences, which is a necessary step toward extending the approach to multi-view video sequences.

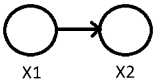


Fig. 3 Probabilistic graphical model of two variables X_1 and X_2

The rest of this paper is organized as follows. The proposed method for automatically detecting ruptures in the spatial relationships is described in Sect. 4. Experimental results are shown in Sect. 5 in order to evaluate the performance of the proposed approach. Finally, conclusions and future work are presented in Sect. 6.

4 Proposed method for detection of ruptures

In this section, the goal is to detect the ruptures in the function y in an automatic way, based on hypothesis testing. More specifically, a simple probabilistic model is introduced to describe the evolution of the spatial relationships over time in Sect. 4.1. In Sect. 4.2, two hypotheses and their log-likelihood ratio are defined. In Sect. 4.3, the distribution of the log-likelihood ratio given that H_0 is true is estimated. The log-likelihood ratio and its distribution are reformulated under equal standard deviations assumption in Sect. 4.4. Finally, the generalization of our approach and the procedure for detecting ruptures over time are described in Sect. 4.5.

4.1 Parameters estimation

The rupture occurs when the value of the function y rapidly changes. For this reason, a simple graphical probabilistic model $\mathbf{G}$ with two nodes X_1 and X_2 is constructed that represent the values of the function y at instants t and $t+1$, respectively. It is assumed that the value of the function y at instant $t+1$ is only dependent on the value at instant t (Markov assumption). This model is shown in Fig. 3. In this model, the conditional distribution of each variable X_i is assumed to be a Gaussian distribution with mean μ_i and variance σ_i^2.

Under this assumption, the variables X_1 and X_2 can be parameterized as follows:

$$\begin{cases} X_1 & \sim \mathcal{N}(\mu_1, \sigma_1^2) \\ X_2|X_1 & \sim \mathcal{N}(\mu_2 = \alpha_0 + \alpha_1 X_1, \sigma_2^2) \end{cases}$$

The set θ of parameters of the graph $\mathbf{G}$, i.e., μ_1, σ_1, α_0, α_1, and σ_2 can be estimated by maximizing the likelihood function.

Given a dataset $\mathbf{D} = d_1 = (x_{11}, x_{21}), d_2 = (x_{12}, x_{22}),$ $\mathbf{D} = \{d_1 = (x_{11}, x_{21}), d_2 = (x_{12}, x_{22}), \dots, d_n = (x_{1n}, x_{2n})\}$ of the two variables (X_1, X_2), the likelihood function is:

$$\mathcal{L}(d_1, d_2, \dots, d_n|\theta) = \prod_{i=1}^{n} P(d_i|\theta), \tag{2}$$

assuming independence between the samples conditionally to θ. It is often simpler to work with the log-likelihood function:

$$\ell(d_1, d_2, \dots, d_n|\theta) = \sum_{i=1}^{n} \log P(d_i|\theta) \tag{3}$$

From the joint distribution $P(X_1, X_2) = P(X_1)P(X_2|X_1)$ of the graph $\mathbf{G}$, the log-likelihood function is derived as:

$$\begin{aligned}\ell_G(D|\theta) &= \ell_G(d_1, d_2, \dots, d_n|\mu_1, \sigma_1, \alpha_0, \alpha_1, \sigma_2) \\ &= \sum_{i=1}^{n} \log P(d_i|\mu_1, \sigma_1) + \sum_{i=1}^{n} \log P(d_i|\alpha_0, \alpha_1, \sigma_2)\end{aligned} \tag{4}$$

The two terms of this function can be independently maximized. For a Gaussian distribution, the maximization of the first term leads to the following classical expressions:

$$\begin{cases} \mu_1 &= \mathbb{E}[X_1] = \frac{1}{n} \sum_{i=1}^{n} x_{1i} \\ \sigma_1^2 &= \mathrm{Var}[X_1] = \frac{1}{n} \sum_{i=1}^{n} (x_{1i} - \mu_1)^2 \end{cases}$$

The second term of the log-likelihood function $\ell_G(D|\theta)$ can be written as follows:

$$\ell_{X_2|X_1}\left(d_1, d_2, \dots, d_n|\alpha_0, \alpha_1, \sigma_2\right) = \sum_{i=1}^{n} -\frac{1}{2}\left(\log\left(2\pi\sigma_2^2\right) + \frac{\left(\alpha_0 + \alpha_1 x_{1i} - x_{2i}\right)^2}{\sigma_2^2}\right) \tag{5}$$

Computing the gradient of $\ell_{X_2|X_1}$ with respect to α_0 and equating the gradient to 0, we get:

$$\alpha_0 + \alpha_1 \mathbb{E}[X_1] = \mathbb{E}[X_2] \tag{6}$$

Then, computing the gradient of $\ell_{X_2|X_1}$ with respect to α_1 and equating the gradient to 0, we get:

$$\alpha_0 \mathbb{E}[X_1] + \alpha_1 \mathbb{E}\left[X_1^2\right] = \mathbb{E}\left[X_1 X_2\right] \tag{7}$$

Now, there are two linear equations with two unknowns α_0 and α_1 that can be solved to obtain the values of α_0 and α_1. Finally, the value of σ_2^2 parameter is obtained by computing the gradient of $\ell_{X_2|X_1}$ with respect to σ_2 and equating to 0:

$$\sigma_2^2 = \mathrm{Var}[X_2] - \alpha_1^2 \mathrm{Var}[X_1] \tag{8}$$

At this stage, the parameters of the graph can be estimated using the actual data. In the next section, the hypotheses and their log-likelihood ratio are described.

4.2 Hypotheses testing

In the literature, there are a lot of approaches to detect ruptures in a signal. The reader is referred to [46] for more details. Inspired by these approaches, we propose a new method based on the log-likelihood ratio between two hypotheses to detect and determine the instants of ruptures in the spatial relationships between two moving objects in video sequences. The two hypotheses are defined as follows:

$$H_0 : \left\{ y_{i+1}|y_i \sim \mathcal{N}\left(\mu_{i+1} = \alpha_0 + \alpha_1 y_i, \sigma_\alpha^2\right) \text{ for } i \in [1, N] \right.$$

$$H_1 : \begin{cases} y_{i+1}|y_i \sim \mathcal{N}\left(\mu_{i+1} = \beta_0 + \beta_1 y_i, \sigma_\beta^2\right) \text{ for } i \in [1, t_r - 1] \\ y_{i+1}|y_i \sim \mathcal{N}\left(\mu_{i+1} = \gamma_0 + \gamma_1 y_i, \sigma_\gamma^2\right) \text{ for } i \in [t_r + 1, N] \end{cases}$$

The hypothesis H_0 means that there is no rupture during the considered time interval [1, N], and the evolution of the spatial relationships in this window can be represented by a single model of three parameters α_0, α_1 and σ_α. The alternative hypothesis H_1 means that there is a rupture at instant t_r, and the evolution of the spatial relationships before t_r is represented by a model of parameters β_0, β_1 and σ_β and the evolution after t_r is represented by a different model of parameters γ_0, γ_1 and σ_γ.

In order to decide which hypothesis is best, the log-likelihood ratio is computed as:

$$\rho(t_r) = \log \frac{\mathcal{L}_{H_1}}{\mathcal{L}_{H_0}} = \ell_{H_1} - \ell_{H_0} \tag{9}$$

This log-likelihood ratio is computed for each instant t_r in the window [1, n], and if $\max_{t_r}(\rho(t_r))$ is greater than a fixed threshold η, the hypothesis H_1 is decided and a rupture is detected at instant $t_r^* = \arg\max_{t_r}(\rho(t_r))$. Otherwise, H_0 is decided and no rupture is detected in this window. This formulation is valid under the assumption that there is at most one rupture in the considered window [1, n].

Using this formulation, two important issues must be addressed. First, all the parameters in the hypotheses H_0 and H_1 are unknown, and these parameters can be estimated using the observed data. The parameters in the hypothesis H_0 (α_0, α_1 and σ_α) can be estimated using N samples. The parameters in H_1 before the instant t_r (β_0, β_1 and σ_β) can be estimated using t_r samples and after the instant t_r (γ_0, γ_1 and σ_γ) using $N - t_r$ samples. This can lead to overfitting the data in the hypothesis H_1 compared to hypothesis H_0. To illustrate this behavior, let us consider $N = 4$ points and $t_r = 2$. In the estimation of parameters in the hypothesis H_1, a perfect fit is obtained since 2 points are used to fit a linear model before and after the instant t_r. For the H_0 parameters, 4 points are used to fit the linear model. Thus, we always end up by deciding the hypothesis H_1 since a perfect fit is achieved even if there is no real rupture at instant t_r.

To avoid this issue, n (even) samples are always considered in the estimation of the parameters in hypotheses H_0 and H_1. More specifically, at the instant t_r, the samples in $[t_r - \frac{n}{2}, t_r + \frac{n}{2}[$ are considered in the estimation of the parameters in H_0. The samples in $[t_r - n, t_r[$ are used to estimate the parameters in H_1 before the instant t_r and in $]t_r, t_r + n]$ after the instant t_r. In the computation of the log-likelihood ratio, the samples in $[t_r - \frac{n}{2}, t_r + \frac{n}{2}]$ are considered.

According to the new formulation, the two hypotheses are defined as follows:

$$H_0 : \left\{ y_{i+1}|y_i \sim \mathcal{N}\left(\mu_{i+1} = \alpha_0 + \alpha_1 y_i, \sigma_\alpha^2\right) \text{ for } i \in [1, n] \right.$$

$$H_1 : \begin{cases} y_{i+1}|y_i \sim \mathcal{N}\left(\mu_{i+1} = \beta_0 + \beta_1 y_i, \sigma_\beta^2\right) \text{ for } i \in [1, \frac{n}{2} - 1] \\ y_{i+1}|y_i \sim \mathcal{N}\left(\mu_{i+1} = \gamma_0 + \gamma_1 y_i, \sigma_\gamma^2\right) \text{ for } i \in [\frac{n}{2} + 2, n] \end{cases}$$

The second issue is that a threshold η must be fixed to decide the best hypothesis. According to our experimental results, the threshold η depends on the event underhand or more generally on the type of event. To handle this issue, instead of comparing the log-likelihood ρ with a fixed threshold η to select the best hypothesis, the distribution of ρ given the null hypothesis H_0 is true is investigated. Based on this distribution, a p value can be computed, which is defined as the probability of observing the data or more extreme outcome given the null hypothesis is true. If the p value is smaller than a significance level α, we can say that the observed data provide a convincing evidence to reject the null hypothesis H_0 in favor of the alternative hypothesis H_1 and a rupture is detected. Otherwise, we can say that the observed data do not provide a convincing evidence to reject the null hypothesis H_0 and no rupture can be detected. Note that the major benefit of using a significance level α instead of setting directly a threshold η is the interpretability of α. As an example, the user can set the probability of false alarm α according to his needs, and then a specific threshold for each video sequence is automatically derived. The next section describes the distribution of ρ given that H_0 is true.

4.3 Null distribution

In this section, we investigate the probability density function of the log-likelihood ratio ρ given that H_0 is true. Given this distribution and a probability of false alarm α, a threshold η can be obtained. The probability of false alarm α is interpreted as the probability of deciding H_1 given that H_0 is true (i.e., α is the probability of $\rho > \eta$ given that H_0 is true).

The log-likelihood ratio between the two defined hypotheses is computed as:

$$\begin{aligned} \rho(H_0, H_1) &= \log \frac{\mathcal{L}_{H_1}}{\mathcal{L}_{H_0}} = \ell_{H_1} - \ell_{H_0} \\ &= \sum_{i=1}^{n/2-1} -\frac{1}{2}\left(\log\left(2\pi\sigma_\beta^2\right) + \frac{(\beta_0 + \beta_1 y_i - y_{i+1})^2}{\sigma_\beta^2} \right) \\ &\quad + \sum_{i=n/2+2}^{n} -\frac{1}{2}\left(\log\left(2\pi\sigma_\gamma^2\right) + \frac{(\gamma_0 + \gamma_1 y_i - y_{i+1})^2}{\sigma_\gamma^2} \right) \\ &\quad - \sum_{i=1}^{n} -\frac{1}{2}\left(\log\left(2\pi\sigma_\alpha^2\right) + \frac{(\alpha_0 + \alpha_1 y_i - y_{i+1})^2}{\sigma_\alpha^2} \right) \end{aligned} \tag{10}$$

The difficulty here is to estimate the probability density function of ρ given that H_0 is true. Let us reformulate the log-likelihood ratio as follows:

$$\begin{aligned} \rho(H_0, H_1) &= \frac{n}{2} \log\left(2\pi\sigma_\alpha^2\right) \\ &\quad - \frac{1}{2}\left(\frac{n}{2} - 1\right)\left(\log\left(2\pi\sigma_\beta^2\right) + \log\left(2\pi\sigma_\gamma^2\right)\right) \\ &\quad + \frac{1}{2} \sum_{i=1}^{n/2-1} \left(\frac{(\alpha_0 + \alpha_1 y_i - y_{i+1})^2}{\sigma_\alpha^2} - \frac{\left(\beta_0 + \beta_1 y_i - y_{i+1}\right)^2}{\sigma_\beta^2} \right) \\ &\quad + \frac{1}{2} \sum_{i=n/2+2}^{n} \left(\frac{(\alpha_0 + \alpha_1 y_i - y_{i+1})^2}{\sigma_\alpha^2} - \frac{\left(\gamma_0 + \gamma_1 y_i - y_{i+1}\right)^2}{\sigma_\gamma^2} \right) \\ &\quad + \frac{1}{2} \sum_{i=n/2}^{n/2+1} \frac{\left(\alpha_0 + \alpha_1 y_i - y_{i+1}\right)^2}{\sigma_\alpha^2} \end{aligned} \tag{11}$$

Thus, the log-likelihood ratio can be split into four independent terms as shown in Eq. 11. The first term K_r is constant:

$$K_r = \frac{n}{2} \log\left(2\pi\sigma_\alpha^2\right) - \frac{1}{2}\left(\frac{n}{2} - 1\right)\left(\log\left(2\pi\sigma_\beta^2\right) + \log\left(2\pi\sigma_\gamma^2\right)\right) \tag{12}$$

Let us call the three last terms $T_k (k = 2, 3, 4)$. Before finding out the distribution of each of the three terms T_k given that H_0 is true, let us determine the distribution of an intermediary variable Y_i defined as follows:

$$Y_i = \frac{(X_i - \mu_{i0})^2}{\sigma_0^2} - \frac{(X_i - \mu_{i1})^2}{\sigma_1^2}, \tag{13}$$

with $X_i \sim \mathcal{N}(\mu = \mu_{i0}, \sigma = \sigma_0)$. After a suitable reformulation, the Y_i term can be expressed as follows:

$$\begin{aligned} Y_i &= \left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right)\left(\frac{X_i - \frac{\sigma_1^2\mu_{i0} - \sigma_0^2\mu_{i1}}{\sigma_1^2 - \sigma_0^2}}{\sigma_0}\right)^2 + \left(\frac{\mu_{i0}^2}{\sigma_0^2} - \frac{\mu_{i1}^2}{\sigma_1^2} - \frac{(\sigma_1^2\mu_{i0} - \sigma_0^2\mu_{i1})^2}{\sigma_0^2\sigma_1^2(\sigma_1^2 - \sigma_0^2)}\right) \\ &= c\left(\frac{X_i - \phi_i}{\sigma_0}\right)^2 + O_i \end{aligned} \tag{14}$$

where

$$\begin{cases} c = & 1 - \frac{\sigma_0^2}{\sigma_1^2} \\ \phi_i = & \frac{\sigma_1^2\mu_{i0} - \sigma_0^2\mu_{i1}}{\sigma_1^2 - \sigma_0^2} \\ O_i = & \frac{\mu_{i0}^2}{\sigma_0^2} - \frac{\mu_{i1}^2}{\sigma_1^2} - \frac{(\sigma_1^2\mu_{i0} - \sigma_0^2\mu_{i1})^2}{\sigma_0^2\sigma_1^2(\sigma_1^2 - \sigma_0^2)} \end{cases}$$

The distribution of $\frac{X_i - \phi_i}{\sigma_0}$ term is $\mathcal{N}(\mu_i = \frac{\mu_0 - \phi_i}{\sigma_0}, \sigma = 1)$. Thus, $(\frac{X_i - \phi_i}{\sigma_0})^2$ term follows a non-central Chi-squared distribution with one degree of freedom and a non-centrality parameter $\lambda = \mu_i^2$ (this distribution is symbolized as $\chi_1^2(\lambda)$). Thus, Y_i is a displaced and scaled non-central Chi-squared distribution.

Given that H_0 is true, $y_{i+1}|y_i \sim \mathcal{N}(\mu_{i+1} = \alpha_0 + \alpha_1 y_i, \sigma_\alpha^2)$ for $i \in [1, n]$. Thus, T_2 can be expressed in terms of Y_i as follows:

$$T_2 = \frac{1}{2}\sum_{i=1}^{n/2-1}\left(\frac{(\alpha_0 + \alpha_1 y_i - y_{i+1})^2}{\sigma_\alpha^2} - \frac{(\beta_0 + \beta_1 y_i - y_{i+1})^2}{\sigma_\beta^2}\right) = \frac{1}{2}\sum_{i=1}^{n/2-1} Y_i \tag{15}$$

where

$$\begin{cases} \sigma_0 = & \sigma_\alpha \\ \sigma_1 = & \sigma_\beta \\ \mu_{i0} = & \alpha_0 + \alpha_1 y_i \\ \mu_{i1} = & (\alpha_0 + \alpha_1 y_i) - (\beta_0 + \beta_1 y_i) \end{cases}$$

Thus, T_2 is a displaced and scaled non-central Chi-squared distribution with $(n/2 - 1)$ degrees of freedom and a non-centrality parameter $\lambda = \sum_{i=1}^{n/2-1} \mu_i^2$. T_3 can be composed by the same way with the following parameters:

$$\begin{cases} \sigma_0 = & \sigma_\alpha \\ \sigma_1 = & \sigma_\gamma \\ \mu_{i0} = & \alpha_0 + \alpha_1 y_i \\ \mu_{i1} = & (\alpha_0 + \alpha_1 y_i) - (\gamma_0 + \gamma_1 y_i) \end{cases}$$

T_3 is also a displaced and scaled non-central Chi-squared distribution with $(n/2 - 1)$ degrees of freedom and a non-centrality parameter $\lambda = \sum_{i=n/2+2}^{n} \mu_i^2$. The last term T_4 is a scaled Chi-squared distribution with two degrees of freedom (i.e., $\frac{1}{2}\chi_2^2$). We can conclude that the distribution of ρ is a weighted sum of non-central Chi-squared distributions (let us call this distribution P_ρ).

Given a probability of false alarm α, using the distribution of ρ (P_ρ) given that H_0 is true, a threshold η is obtained such that $\alpha = \int_\eta^\infty P_\rho d\rho$ ($\eta = Q_\rho(1 - \alpha)$ where Q_ρ is the quantile function of the distribution P_ρ). For a given instant k, if the log-likelihood ratio ρ_k is greater than η, the null hypothesis H_0 is rejected in favor of the alternative one H_1, and a rupture is detected at instant k. Otherwise, the null hypothesis is decided and no rupture is detected. This procedure is equivalent to compute the p value for a log-likelihood ratio ρ_k at instant k as $\int_{\rho_k}^\infty P_\rho d\rho$ and to compare it with the significance level α (probability of false alarm).

In the next section, a reasonable assumption is made about the standard deviations σ_α, σ_β, and σ_γ.

4.4 Equal standard deviations

As mentioned above, the distribution of ρ given that H_0 is true is a weighted sum of non-central Chi-squared distributions. In the case of multiple objects or using many signals simultaneously in the log-likelihood ratio, the computation cost of finding the distribution of ρ becomes very high due to summation of many weighted non-central Chi-squared distributions. In this section, all the standard deviations σ_α, σ_β, and σ_γ are assumed to be equal ($\sigma = \sigma_\alpha = \sigma_\beta = \sigma_\gamma$).

Regardless the benefit of reducing the computation cost due to the assumption of equal standard deviations, we believe that this assumption is reasonable and may improve the results. Technically speaking, two errors may affect the standard deviation in a model. The first error is due to the precision of the model (model selection), and the second error is due to the noise in the data. In our case, the same model is used in two hypotheses H_0 and H_1 and these hypotheses are applied to the same signal (i.e., same level of noise). Thus, when a huge difference in σ is observed, this means that a rupture may exist.

Under this assumption, the log-likelihood ratio can be written as:

$$\begin{aligned}\rho(H_0, H_1) = {} & \log(2\pi\sigma^2) \\ & + \frac{1}{2\sigma^2}\sum_{i=1}^{n/2-1}\left((\beta_0 + \beta_1 y_i - \alpha_0 - \alpha_1 y_i)y_{i+1} + c_\beta\right) \\ & + \frac{1}{2\sigma^2}\sum_{i=n/2+2}^{n}\left((\gamma_0 + \gamma_1 y_i - \alpha_0 - \alpha_1 y_i)y_{i+1} + c_\gamma\right) \\ & + \frac{1}{2}\sum_{i=n/2}^{n/2+1}\frac{(\alpha_0 + \alpha_1 y_i - y_{i+1})^2}{\sigma^2}\end{aligned} \tag{16}$$

where

$$\begin{cases} c_\beta = (\alpha_0 + \alpha_1 y_i)^2 - (\beta_0 + \beta_1 y_i)^2 \\ c_\gamma = (\alpha_0 + \alpha_1 y_i)^2 - (\gamma_0 + \gamma_1 y_i)^2 \end{cases}$$

Here, the log-likelihood ratio can also be split into four independent terms. The first term is constant, and the distribution of the last term remains the same $\frac{1}{2}\chi_2^2$. The distributions of the second and third terms are Gaussian since their equations become linear with y_{i+1}. The benefit of this assumption is that the sum of two Gaussian distributions is a Gaussian distribution, and the sum of two independent Chi-squared distributions is a Chi-squared distribution. Thus, under this assumption, in the case when using many signals simultaneously, the distribution of ρ given that H_0 is true is the sum of a Gaussian distribution and a Chi-squared distribution, regardless the number of terms in the log-likelihood ratio. The validity of this assumption is shown in Sect. 5.1.1.

The next section describes the generalization of our approach to the case of multiple objects or when combining many signals. Moreover, the procedure for detecting ruptures over time is described.

4.5 Generalization

In this section, we show that our approach can be used in the case of multiple objects, i.e., by considering many signals simultaneously, and for detecting several ruptures in a sequence.

Let us assume that there are K different signals $y^k (k = 1, \dots, K)$ over time. These signals can represent the evolution of the directional (metric) relationships among different objects or the evolution of many spatial information between two objects over time. In this generalization, the interaction among the different signals can be modeled as follows:

$$y_{i+1}^r | (y_i^1, y_i^2, \dots, y_i^K) \sim \mathcal{N}\left(\mu_{i+1}^r = a_{r0} + \sum_{k=1}^{K} a_{rk} y_i^k, \sigma_r^2\right) \tag{17}$$

where $r = 1, \dots, K$. In this equation, it is also assumed that the values of the functions y^k at instant $i + 1$ are only dependent on the values at instant i, and all the signals $y^k (k = 1, \dots, K)$ interact with the signal y^r. However, this equation can only include the signals that really interact with the signal y^r. Here, the parameters $a_{rk} (k = 0, \dots, K, r = 1, \dots, K)$ can also be estimated using the observed data by maximizing the log-likelihood function (there are $K(1 + K)$ parameters to estimate).

The two hypotheses H_0 and H_1 can be defined as in Sect. 4.4, and the log-likelihood ratio ρ can be computed. In addition, the distribution of ρ given that H_0 is true can be obtained by summing up many Gaussian and Chi-squared distributions thanks to equal standard deviations assumption. Then, the p value can be also computed in the same way, and by comparing it to a significance level α, the ruptures can be detected.

Now, let us describe our algorithm to detect many sequential ruptures in the spatial relationships during time Algorithm 1. First, we search in the interval $[t, t + L]$ for a rupture. It is assumed that there is at most one rupture in this window $W = [t, t + L]$. For each $i \in W$, the parameters in the hypotheses H_0 and H_1 are first estimated using the observed data using n samples. In order to obtain a robust estimation of the parameters in hypotheses H_0 and H_1, n must be large enough. As explained above, the samples in $[i - n/2, i + n/2[$ are used to estimate the parameters in the hypothesis H_0. For the parameters in the hypothesis H_1, the samples in the interval $[i - n, i[$ are used to estimate the parameters β_0, β_1, and σ_β, and the samples in the interval $]i, i + n]$ are used to estimate the parameters γ_0, γ_1, and σ_γ. Then, the log-likelihood ratio ρ_i is estimated for each instant $i \in W$. Then, the instant i^* which gives the maximum log-likelihood ratio is estimated as $i^* = \arg\max_{i \in W}(\rho_i)$. Afterward, the p value is computed using the estimated distribution of ρ_{i^*} given that the null hypothesis H_0 is true. Given a significance level α (a fixed probability of false alarm), the obtained p value is compared to the significance level α, and if the p value is smaller than α, the null hypothesis is rejected in favor of the alternative one H_1, and a rupture is detected at instant i^*. The window is then updated to $W = [i^* + L, i^* + 2L]$ to check for new ruptures. Otherwise, the null hypothesis is decided and no rupture is detected in this window, and the window is updated to $W = [t + L, t + 2L]$. The same procedure can be used in the case of multiple objects, where each signal y^k in Algorithm 1 represents the evolution of spatial relationships between two objects over time.

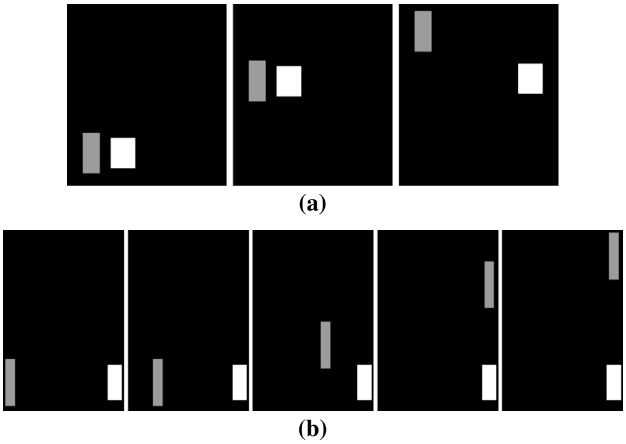


Fig. 4 Synthetic events TSE$_1$ (**a**) and TSE$_2$ (**b**). **a** Frames number 1, 30, and 50 of TSE$_1$ and **b** frames number 45, 55, 74, 95, and 105 of TSE$_2$

Algorithm 1 Ruptures detection algorithm

```
Input: signal $y$ (or signals $y^k$), significance level $\alpha$ and number of samples $n$
Output: $ruptures$
set $ruptures = \{\}$
set $L = n$, $s = n$, and $N$ to the length of $y$
while $s < N - n$ do
    set $W = [s, \min(s + L, N - n)]$
    for each $i$ in $W$ do
        compute log-likelihood ratio $\rho_i$
    end for
    find $i^* = \arg\max_{i \in W} (\rho_i)$
    compute p-value at instant $i^*$
    if p-value $< \alpha$ then
        $ruptures = ruptures \cup \{i^*\}$
        $s = s + i^*$
    else
        $s = s + L$
    end if
end while
```

5 Experiment results

In this section, some experimental results are discussed, for events such as merging, grouping, and crossing for the case of two and multiple moving objects.

To evaluate the performance of the proposed approach, some synthetic events were created, containing two or multiple objects, and also a variety of real events are used, selected from the PETS 2009 datasets [48] and PETS 2006 dataset [47] for two objects and from Friends Meet datasets [49] for multiple objects. Here, we call "event" some frames that contain a rupture in the spatial behavior. Table 1 summarizes the characteristics of the datasets used in the experiments. They have been chosen to exhibit ruptures in terms of spatial relationships and provide good illustrations of the potential and performance of the proposed approach.

First, the detection of the ruptures in the spatial relationships between two moving objects is shown, and then the obtained results for multiple objects are shown.

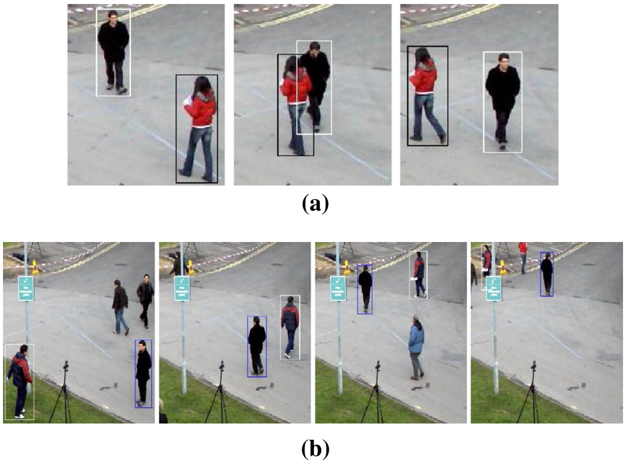


Fig. 5 Real events TRE$_1$ (**a**) and TRE$_2$ (**b**). **a** Frames number 450, 462, and 468 of TRE$_1$ selected from PETS 2009 and **b** frames number 595, 630, 670, and 700 of TRE$_2$ selected from PETS 2009

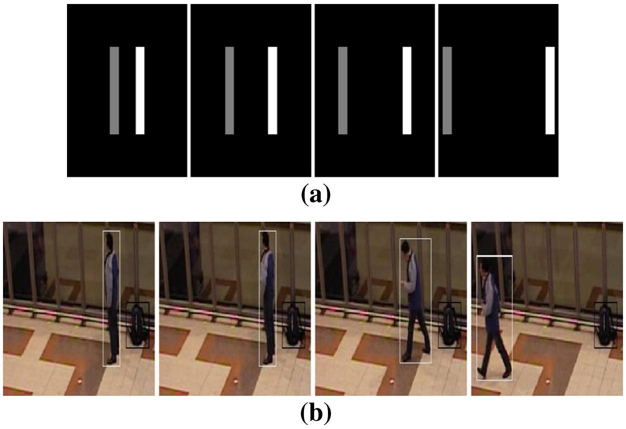


Fig. 6 Events TSE$_3$ (**a**) and TRE$_3$ (**b**). **a** Frames number 1, 5, 10, and 15 of TSE$_3$ and **b** frames number 1955, 2010, 2060, and 2100 of TRE$_3$ selected from PETS 2006 [47]

Table 1 Characteristics of the datasets used in the experiments

Sequence	# Frames	# Objects	# Ruptures
TSE$_1$	60	2	1
TSE$_2$	150	2	2
TSE$_3$	40	2	1
TRE$_1$	34	2	1
TRE$_2$	220	2	3
TRE$_3$	160	2	2
MSE$_1$	130	4	1
MRE$_1$	180	4	1

Experimental results show that the results of the proposed approach are robust to parameters change depending on a given signal. In the rest of the experiments, n is set to 21 in the case of two objects for both synthetic and real scenarios and to 41 in the case of multiple objects (there are more

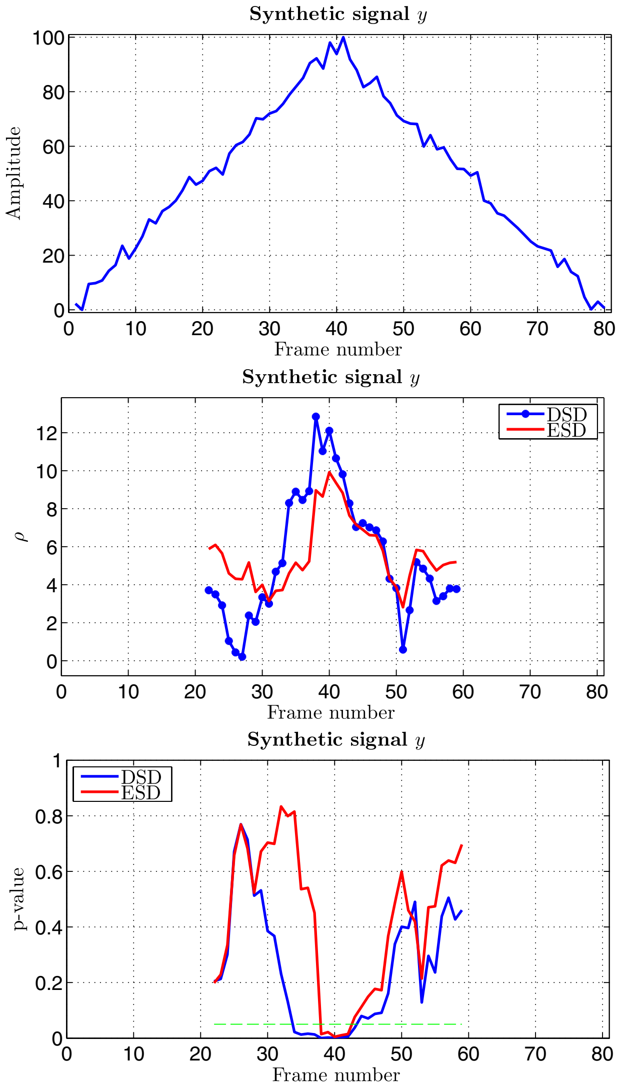


Fig. 7 Simulated signal y, the log-likelihood ratio ρ , and the obtained p value over time for the two cases: without any assumption (DSD) and under the assumption of equal standard deviations (ESD)

parameters to estimate in this case), L to n, and the significance level α to 0.05.

Note that the proposed approach is different from the state-of-the-art techniques. The proposed approach automatically detects the ruptures in the spatial relationships using low-level features. However, the state-of-the-art techniques try to address the problem of event detection using high-level features. Thus, the results presented here are not compared to any related work, since the proposed approach represents an upstream part of a complete event detection framework. In addition, we do not compare the results with [33] because a different threshold is needed for each video sequence in [33]. However, the proposed approach addresses this problem by automatically detecting all the ruptures in the spatial relationships.

5.1 Two objects

The created synthetic events that contain two objects are denoted by TSE$_i$ (illustrated in Fig. 4a, b) and by TRE$_i$ the real events selected from the PETS 2009 datasets [48] (illustrated in Fig. 5a, b). These examples are used to illustrate the ruptures in directional relations (angle histograms, as in [33]). A synthetic event and a real event selected from PETS 2006 dataset [47], displayed in Fig. 6, are used to illustrate the proposed approach with distance relations (distance histograms, as in [33]).

5.1.1 Equal standard deviations assumption

To show the validity of equal standard deviations assumption, Fig. 7 shows a simulated example of the signal y of 80 samples, the log-likelihood ratio ρ, and the obtained p value at each instant for two cases, when the distribution of ρ is obtained as described in Sect. 4.3 without any assumption (let us call it "DSD"), and under the assumption of equal standard deviations (let us call it "ESD"). In this example, $n = 20$ samples are considered when the parameters of the model are estimated. It is clear that the instant of the rupture occurs at instant $t_r = 40$, and the maximum of the log-likelihood ratio occurs at this instant in both cases. The maximum of the log-likelihood ratio in "DSD" case is bigger than the one of "ESD" case, since an equal $\sigma = \max\{\sigma_\alpha, \sigma_\beta, \sigma_\gamma\}$ is used in "ESD" case. If a smaller σ is used, this can make the maximum of the log-likelihood ratio in "ESD" bigger, but this does not change the behavior of the p value curve.

From the p value curves, we can see that the p value is almost 0 at instant $t_r = 40$ in both "DSD" and "ESD" cases. This means that the data provide a very strong evidence to reject the null hypothesis H_0 in favor of the alternative one H_1, and a rupture is detected at this instant. It is also observed that the p value becomes very small when the instant of the rupture is included in the hypothesis H_0, even if it is also included in hypothesis H_1 (see p value at [39, 43] except 40). This behavior is not bad since a rupture exists in the considered window, and the p value becomes even smaller when the instant of the rupture is only included in the hypothesis H_0 (at instant 40). As shown, this behavior is more restricted when assuming equal standard deviations. Another observation is that the p value in the "ESD" case is often larger than the one of "DSD" case in the area when there is no real rupture in the considered window (a noise is added to the signal y). This behavior shows that the "ESD" case is more robust to noise.

Figure 8 shows the same curves for a smoothed signal y obtained from a real event TRE$_2$ of 220 frames. The QF distance is used between two successive angle histograms to generate the signal y. There are three ruptures in the directional spatial relationships as shown by the signal y at 50,

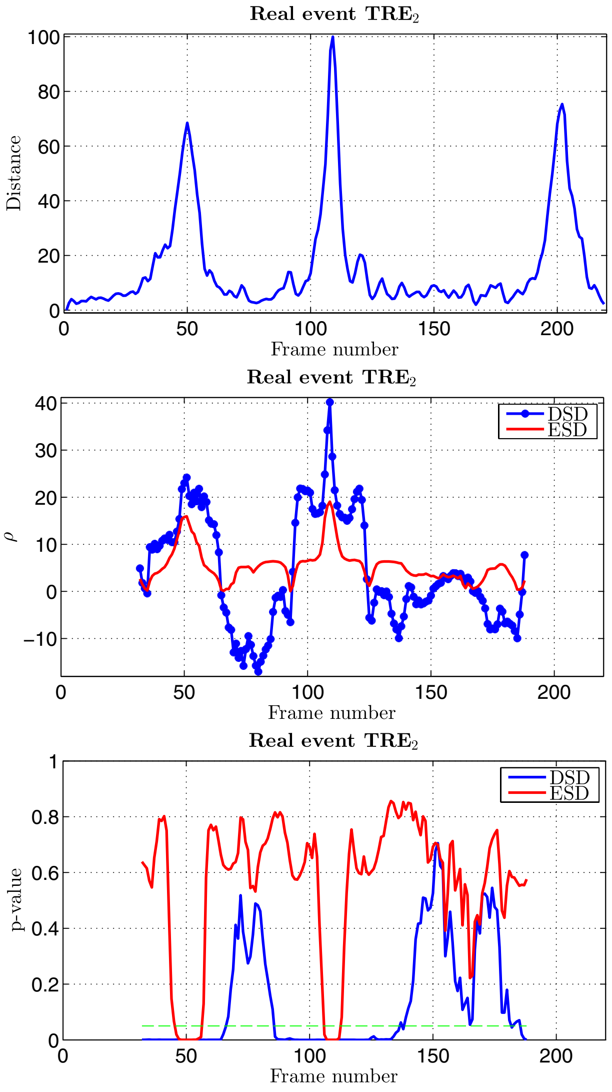
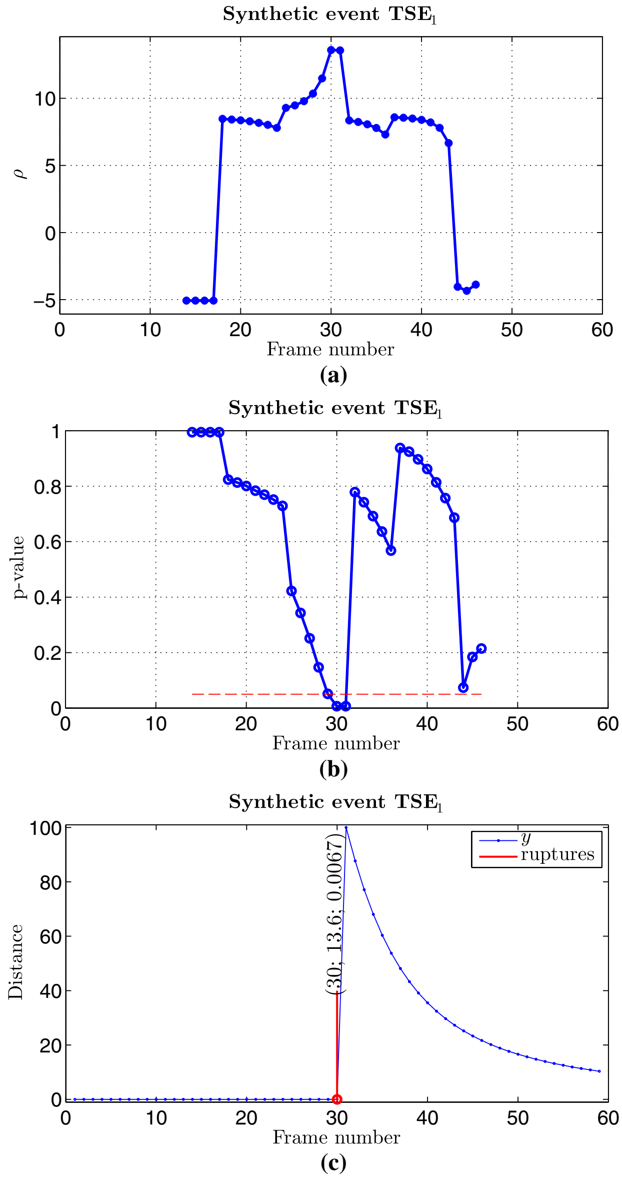


Fig. 8 Smoothed real signal y, the log-likelihood ratio ρ and the obtained p value over time for the two cases: without any assumption (DSD) and under the assumption of equal standard deviations (ESD)

110, and 200 (the two moving objects converge and diverge three times). In this figure, $n = 30$ samples are considered when the parameters of the model are estimated. In this real scenario, the mentioned behaviors above (for the case of simulated signal in Fig. 7) becomes even stronger, since the level of the noise is higher and n is bigger. As shown, the p value suddenly becomes almost 0 at the instant of ruptures in the case of equal standard deviations. Note that the last rupture is not detected since the instant 200 is not evaluated (the last evaluated instant is 187). Hereafter, equal standard deviations assumption is always considered due to robustness and low computational cost.



(a)

(b)

(c)

Fig. 9 Results for synthetic event TSE$_1$. **a** Log-likelihood ratio ρ at each possible instant for TSE$_1$, **b** p value at each possible instant for TSE$_1$ and **c** ruptures detected in TSE$_1$ for a significance level of 0.05

5.1.2 Directional relationships

Three snapshots of the first synthetic event (TSE$_1$) of 60 frames are shown in Fig. 4a (two objects moving together and then separately). In this case, there is a rupture in the directional spatial relationships, when the two objects diverge. Figure 4b shows five snapshots of the second TSE$_1$. In this event, the object B moves toward the object A (fixed) from the left to the right. Then, the object B changes its direction (frame 74), and when the object B becomes above the object A, it goes toward the top.

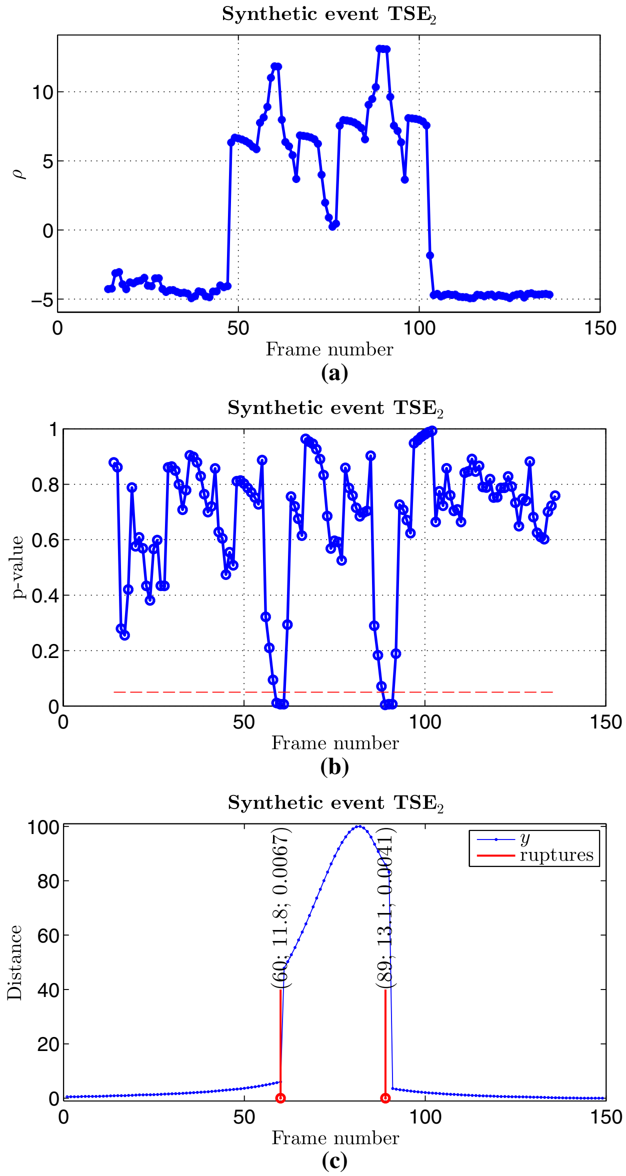


Fig. 10 Results for synthetic event TSE$_2$. **a** Log-likelihood ratio ρ at each possible instant for TSE$_2$, **b** p value at each possible instant for TSE$_2$ and **c** ruptures detected in TSE$_2$ for a significance level of 0.05

Figures 9 and 10 show the log-likelihood ratio ρ at each possible instant for TSE$_1$ and TSE$_2$,respectively (a), the corresponding p value for each instant (b), and the signal y with the ruptures detected by our algorithm (c). Note that it is not necessary to compute the p value at each instant when applying our algorithm (the p value is only computed at the instant which gives the largest log-likelihood ratio as shown in Alg. 1). For event TSE$_1$, the function y shows a strong variation at frame number 31. At this instant, there is a rupture in the spatial relationships (the two objects begin to separate). Our algorithm efficiently detects the instant of rupture, shown in red (a log-likelihood ratio of 13.6 is obtained at the instant of rupture with a p value of 0.0067). For the second event TSE$_2$ (150 frames), two strong variations can be seen in the function y; the first strong variation (frame 61) occurs when B changes its direction with respect to A, the second strong variation (frame 91) occurs when B becomes above A and changes its direction toward the top. Our algorithm clearly shows the two strong variations (log-likelihood ratios of 11.8 and 13.1 are obtained at these instants of ruptures, respectively, with p values of 0.0067 and 0.0041). Thus, the proposed method can efficiently detect the instants of ruptures in the spatial relationships. Several other synthetic events were created and tested using the proposed approach, and similar results were obtained.

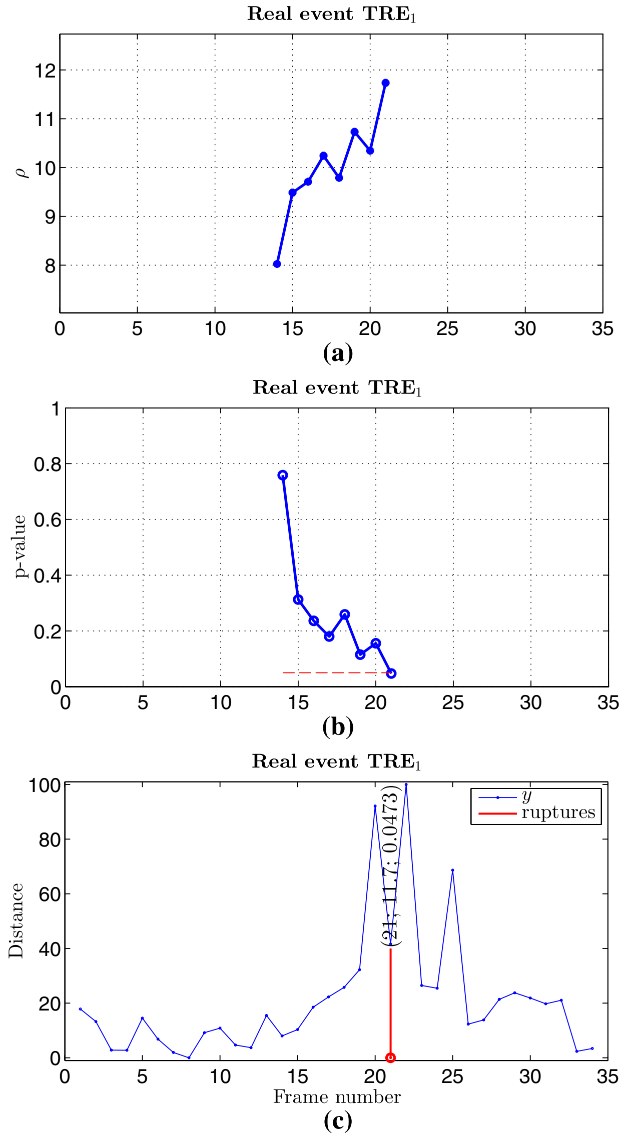


Fig. 11 Results for real event TRE$_1$. **a** Log-likelihood ratio ρ at each possible instant for TRE$_1$, **b** p value at each possible instant for TRE$_1$ and **c** ruptures detected in TRE$_1$ for a significance level of 0.05

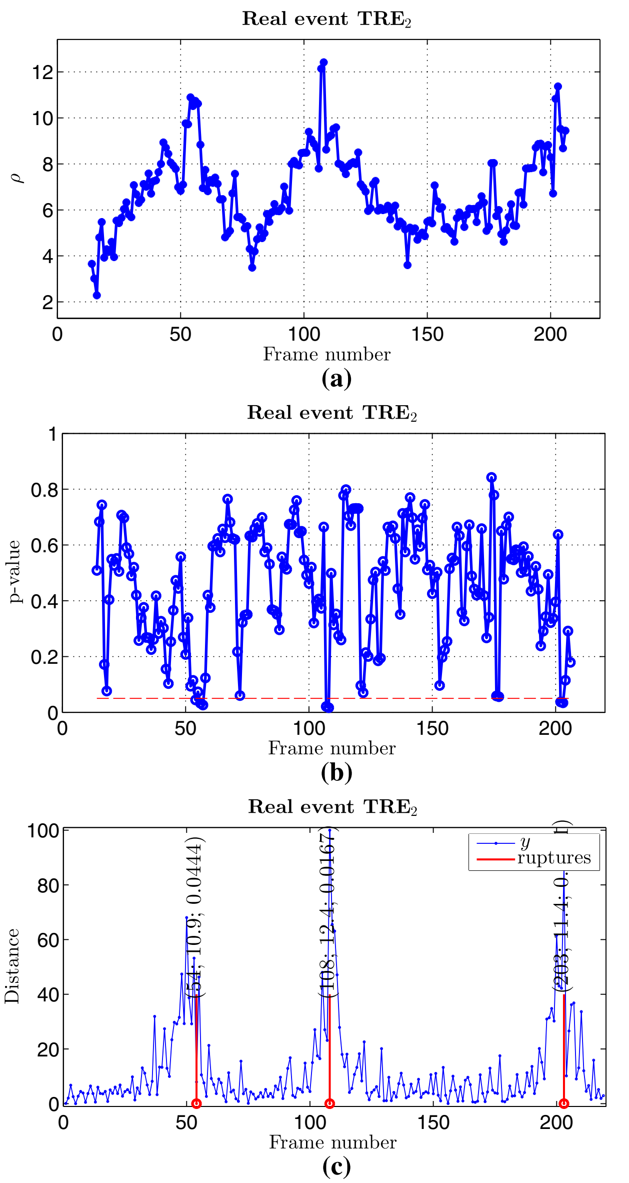


Fig. 12 Results for real event TRE$_2$. **a** Log-likelihood ratio ρ at each possible instant for TRE$_2$, **b** p value at each possible instant for TRE$_2$ and **c** Ruptures detected in TRE$_2$ for a significance level of 0.05

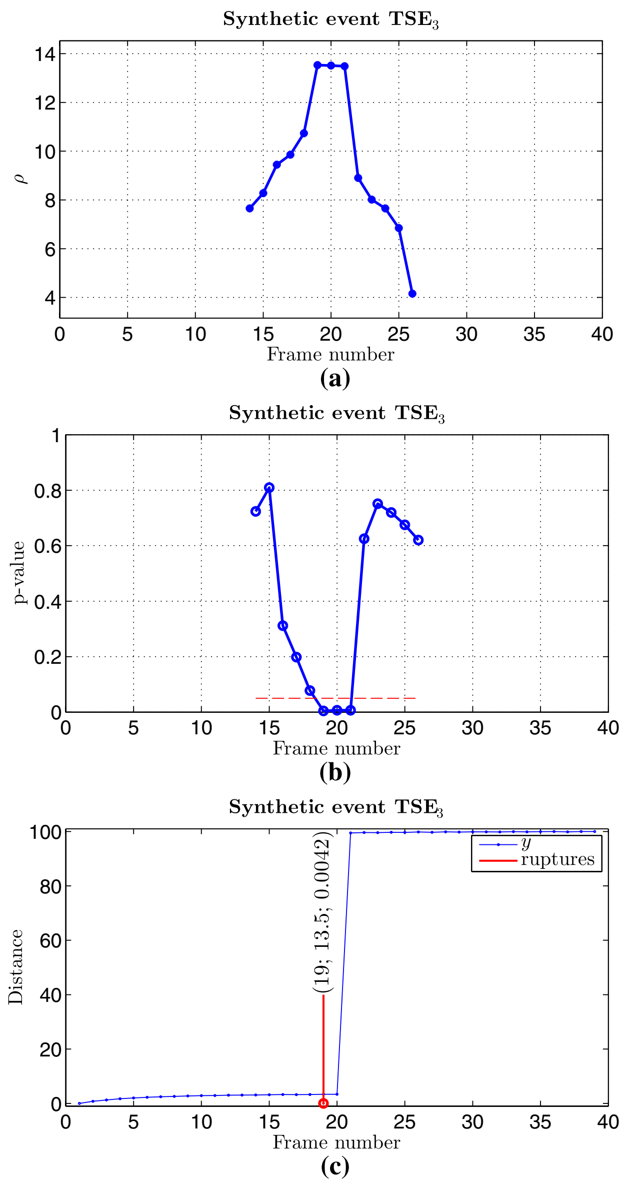


Fig. 13 Results for synthetic event TSE$_3$ **(a)** Log-likelihood ratio ρ at each possible instant for TSE$_3$. **(b)** p value at each possible instant for TSE$_3$. **(c)** Ruptures detected in TSE$_3$ for a significance level of 0.05

Let us now evaluate the proposed detection of ruptures in the spatial relationships in the presence of noise (deformation of objects, etc.) in real events. For the real event TRE$_1$ of 34 frames (Fig. 5a), the two persons converge and then diverge spatially. In the event TRE$_2$ (Fig. 5b) of 220 frames, the two persons (surrounded by white and blue bounding boxes) converge and diverge several times. Figures 11 and 12 show the obtained results for events TRE$_1$ and TRE$_2$, respectively. In the event TRE$_1$, a rupture in the directional spatial relationships occurs when the two persons meet and separate. Our algorithm can efficiently detect the instant of rupture (a log-likelihood ratio of 11.7 is obtained with p value of 0.047) in the directional spatial relationships (Fig. 11c). Figure 12c shows the function y over time and the instants of ruptures in red that are obtained by our algorithm, for the event TRE$_2$. Our algorithm detects several instants of ruptures shown in red, for log-likelihood ratios of 10.9, 12.4, and 11.4, respectively. All the ruptures in the directional spatial relationships can be efficiently detected by our algorithm. It is important to note that our algorithm is applied on the events TRE$_1$ and TRE$_2$ without applying any smoothing on the functions y, and $n = 12$ samples are always used. This shows the efficiency of the algorithm in the presence of strong noise in real scenarios. The results can

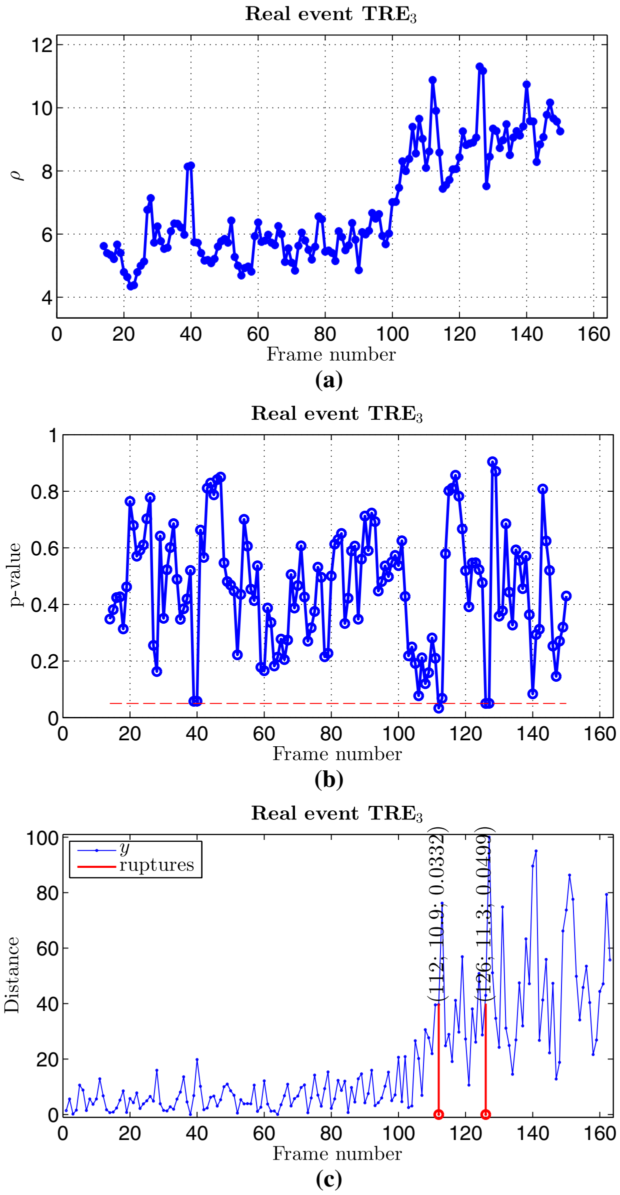


Fig. 14 Results for real event TRE$_3$. **a** Log-likelihood ratio ρ at each possible instant for TRE$_3$, **b** p value at each possible instant for TRE$_3$ and **c** Ruptures detected in TRE$_3$ for a significance level of 0.05

be improved by applying a smoothing on the functions y and using more samples in the estimation of the parameters, in the case of real events (Fig. 8).

5.1.3 Distance relationships

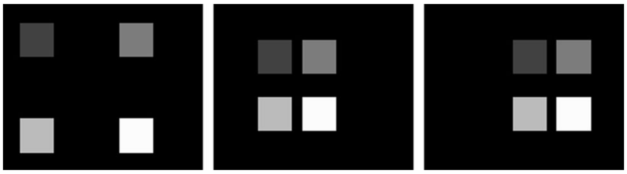


Fig. 15 Frames number 60, 65, and 85 of synthetic event MSE$_1$

Four snapshots of the third synthetic event TSE$_3$ (40 frames) are shown in Fig. 6a. At the beginning of this event, the two objects diverge at a speed of 4 pixels/frame, and at a given instant (precisely at frame 20), the speed of the two objects becomes 8 pixels/frame. Thus, the velocity of the objects is suddenly increased. Figure 6b shows four snapshots of the third real event TRE$_3$ selected from PETS 2006. In this event, the luggage is attended to by the owner for a moment, and then the person leaves the place and goes away.

In Fig. 13, the obtained results for event TSE$_3$ are shown. As shown in this figure, the function y (Fig. 13c) shows a strong variation at frame number 20, when the velocity of the objects changes. At this instant, a rupture in the metric spatial relationships is detected by our algorithm.

In the presence of noise, Fig. 14 shows the results for event TRE$_3$ (160 frames). When the person leaves the place and goes away, a strong change can be seen in the function y. By applying our algorithm, two instants of ruptures are detected in the metric spatial relationships for a significance level α of 0.05. The first instant is when the person leaves the place and goes away, but the second instant is a false detection due to the high level of noise at this instant. We can see that the obtained p value at this instant is 0.0499 (very close to the significance level $\alpha = 0.05$). Thus, this false detection can be avoided by using a harder significance level α.

These results can be used to indicate events occurring in the video sequences, such as escaping in Fig. 6a and left luggage in Fig. 6b.

5.2 Multiple objects

Here, the created synthetic events are denoted by MSE$_i$ and contain multiple objects. Figure 15 shows the first synthetic event MSE$_1$ (130 frames) with four objects which merge and then walk together (i.e., merging event). Twelve histograms (angle and distance) are computed between each objects pair (1–2, 1–3, 1–4, 2–3, 2–4, and 3–4). The functions $y^k (k = 1, 2, \ldots, 12)$ are obtained by computing the QF distance between two successive histograms for each pair. In this context, the generalization of our approach (Sect. 4.5) is applied to detect the ruptures in the spatial relationships.

Figure 16 shows the functions $y^k (k = 1, 2, \ldots, 12)$, the log-likelihood ratio ρ at each possible instant and the correspondent p value for each possible instant. As shown, the instant when the objects meet can be efficiently detected by our algorithm. Here, the directional (angle) and the metric (distance) relationships are used together in the generalization approach to detect ruptures. In this event, both metric

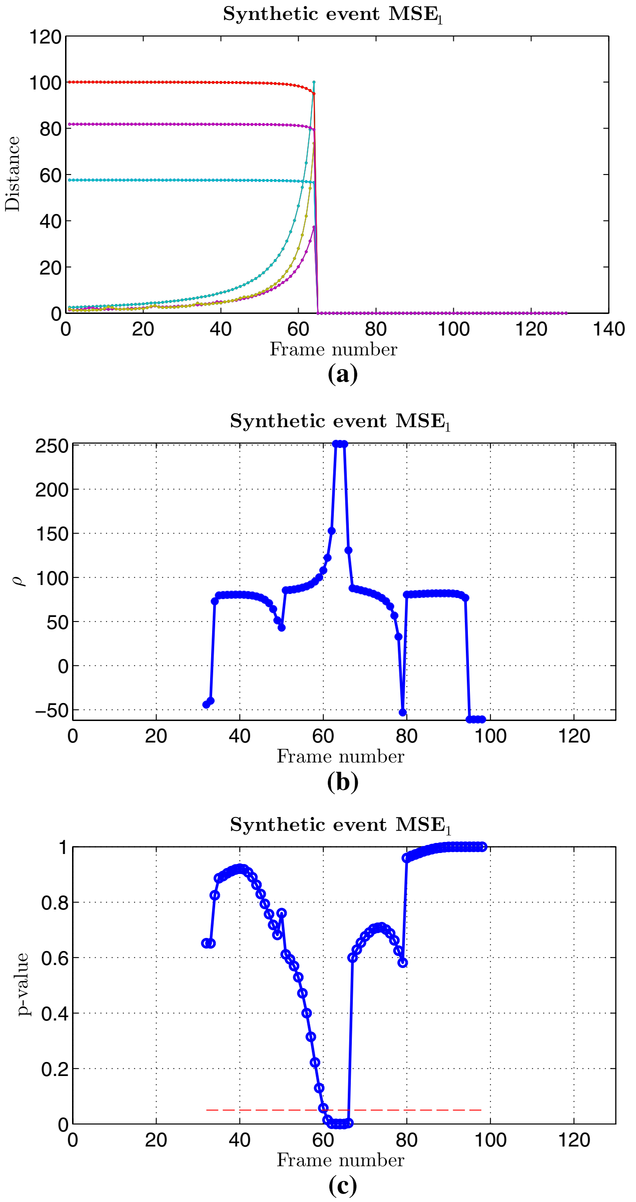


Fig. 16 Results for event MSE_1. **a** Functions $y^k (k = 1, \ldots, 12)$ obtained from angle and distance histograms, **b** log-likelihood ratio ρ at each possible instant for MSE_1 and **c** p value at each possible instant for MSE_1

and directional relationships show a rupture when the persons meet. Both relationships are separately tested by our approach, and the instant of rupture is efficiently detected in both spatial relationships. Similar results are obtained for splitting and crossing synthetic events.

The generalization of our approach is also tested on real scenarios. The real events with multiple objects are selected from Friends Meet datasets [49] and denoted by MRE_i. Figure 17 shows three snapshots of the first real event with four objects (MRE_1) ((a) two moving persons cross two other moving persons) and three snapshots of the second

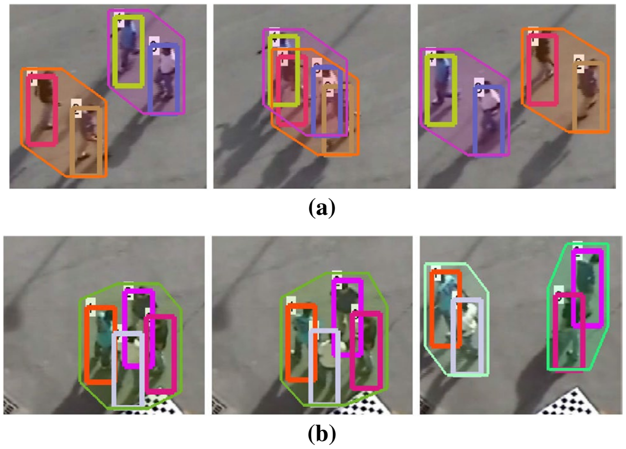


Fig. 17 Real events MRE_1 (**a**) and MRE_2 (**b**) selected from Friends Meet datasets [49]. **a** Frames number 150, 170, and 190 of MRE_1 and **b** frames number 200, 260, and 310 of MRE_2

real event with four objects (MRE_2) ((b) four persons talk together for a moment and then diverge). Figure 18 shows the log-likelihood ratio ρ at each possible instant (a), the corresponding p values (b), and the signals y^k with the ruptures detected by our algorithm (c), for MRE_1 (180 frames). In this event, the angle histograms are used to obtain the signals $y^k (k = 1, 2, \ldots, 6)$. As shown, the instant of rupture in the spatial relationships is efficiently detected by our algorithm, when the two persons 1 and 2 cross the persons 3 and 4.

Figure 19 shows the log-likelihood ratio ρ at each possible instant (a), the corresponding p values (b), and the signals y^k with the ruptures detected by our algorithm (c), for MRE_2 (422 frames). The metric relationships (distance histogram) is used here to obtain the signals $y^k (k = 1, 2, \ldots, 6)$. As we can see, two instants of ruptures are detected by our algorithm. The first instant is when the four persons begin to diverge, and the second instant when they are still diverging. Figure 20 shows some snapshots containing the frames at the instants of ruptures detecting (with a red border) by our algorithm for MRE_1 and MRE_2.

Note that the proposed approach for a single signal y can be always used to study the evolution in the spatial relationships, for a given pair of objects in multiple objects scenarios. As shown by our experimental results, the proposed approach detects efficiently the instants of ruptures when an average level of noise is present. More samples n can be used to account for high level of noise. In our previous work, a different threshold was needed for each sequence to detect the ruptures in the spatial relationships. Figure 21 shows the functions y and g (the derivative of y) over time for real event TRE_3. The function g is used to detect the ruptures in our previous work [33] using a threshold. As we can see, a small change in these thresholds could lead to large false positives, and fixing a threshold to detect the significant ruptures is a

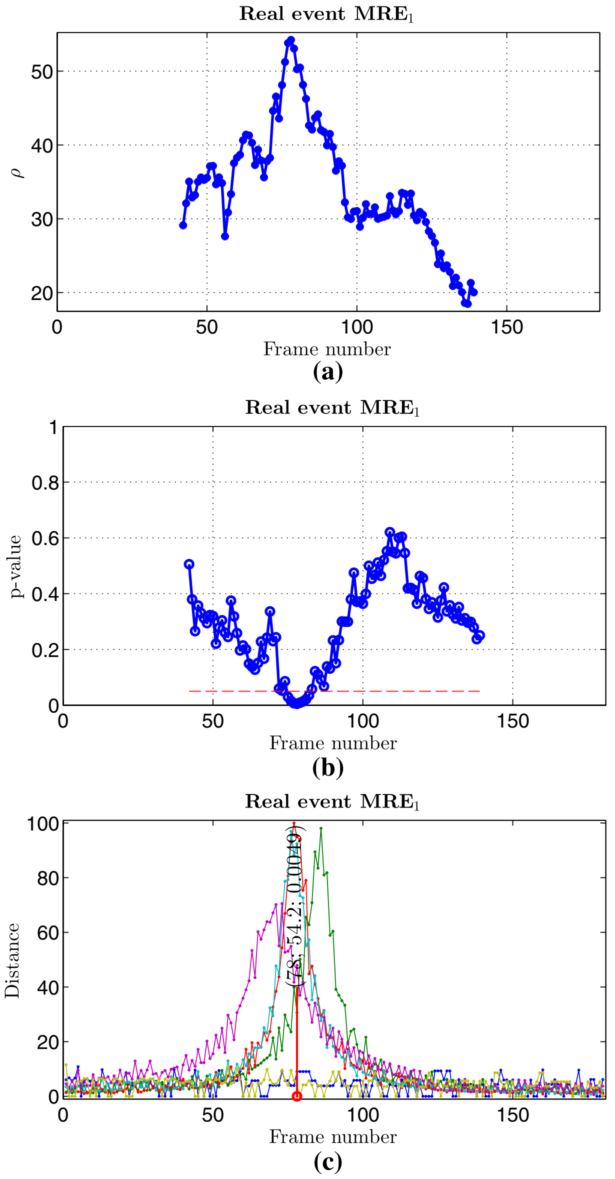


Fig. 18 Results for real event MRE$_1$. **a** Log-likelihood ratio ρ at each possible instant for MRE$_1$, **b** p value at each possible instant for MRE$_1$ and **c** ruptures detected in MRE$_1$ for a significance level of 0.05

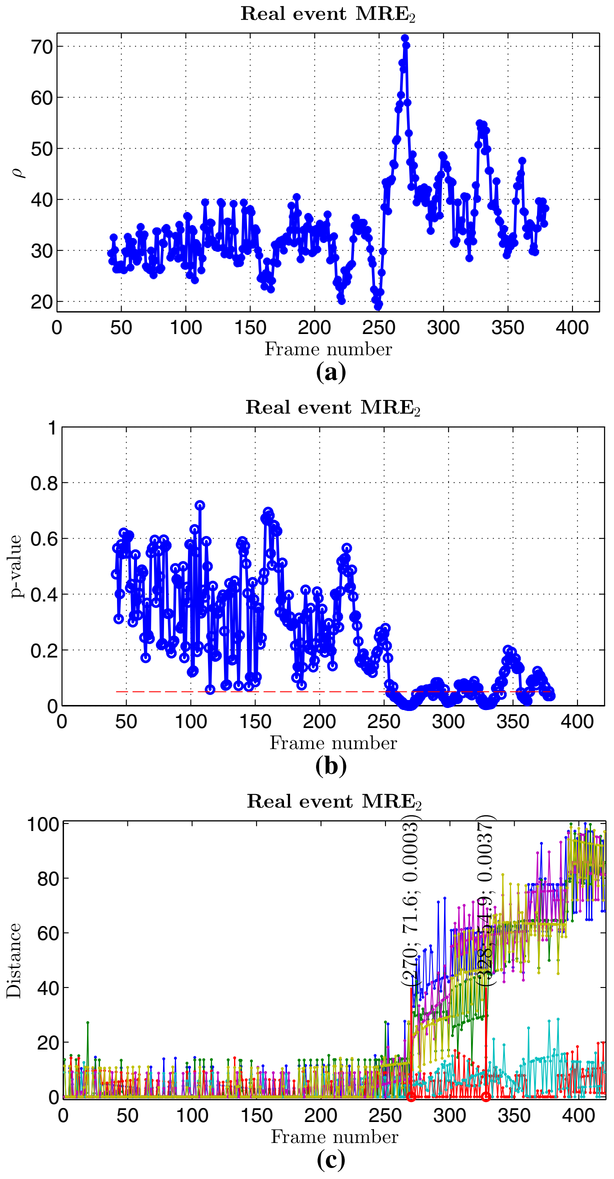


Fig. 19 Results for real event MRE$_2$. **a** Log-likelihood ratio ρ at each possible instant for MRE$_2$, **b** p value at each possible instant for MRE$_2$ and **c** ruptures detected in MRE$_2$ for a significance level of 0.05

tedious task, without any concrete clue on the probability of false alarm. In contrast, here a significant level (a probability of false alarm) can be fixed by the user according to his needs to detect the ruptures thanks to the proposed approach, and the corresponding threshold is then derived automatically and adaptively for each sequence. In our experimental results, a probability of false alarm of 5% was used for all the tested sequences. The experimental results show that all the strong ruptures are efficiently detected using a probability of false alarm of 5%.

6 Conclusion

In this paper, a new method was proposed to automatically detect abrupt changes in spatial relationships in video sequences. Specifically, the fuzzy representations of the objects are estimated and used to compute the angle and distance histograms. Then, the distance between the angle or distance histograms is computed during time. The evolution of relationships during time is modeled by a linear model. Afterward, two hypotheses are defined, and the log-likelihood ratio is computed. Based on the distribution of log-likelihood ratio given that the null hypothesis is true,

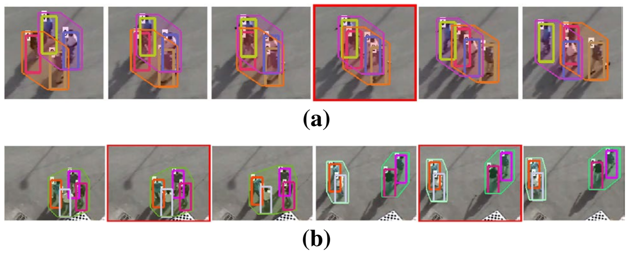

(a)
(b)

Fig. 20 Some snapshots containing the frames at the instants of ruptures (with a red border) for MRE_1 (**a**, crossing) and MRE_2 (**b**, meeting and diverging). **a** Frames number 162, 166, 168, 170, 174, and 178 of MRE_1 and **b** frames number 260, 270, 280, 322, 328, and 334 of MRE_2

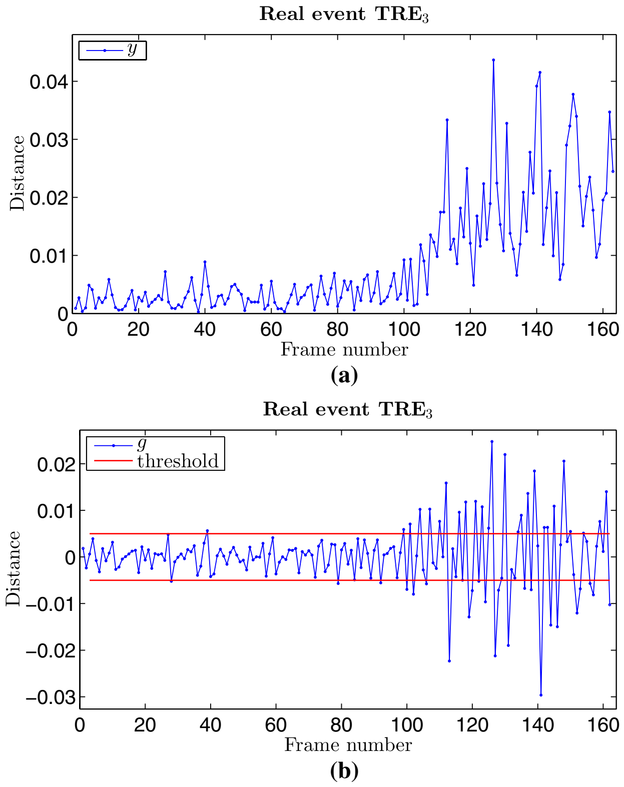


Fig. 21 Results for real event TRE_3. **a** Function y over time for TRE_3 and **b** function g (the derivative of y) [33] over time for TRE_3

the p value is computed and compared to a significance level α to detect significant changes in the spatial relationships, possibly at several instants.

In addition, the proposed approach is generalized to the case of multiple objects to model the interaction and can be used to detect merging, grouping, crossing, and other events. It shows good performances in automatically detecting ruptures in the spatial relationships for both synthetic and real video sequences.

It is important to note that our approach was tested on mono-view video sequences, but it can be extended to handle multi-view sequences. In multi-view sequences, the spatial relationships among objects can be enhanced due to availability of many video recordings of the same scene from multiple angles. As an example, occlusion problems can be efficiently addressed using multiple views of the same scene.

Future work will focus on investigating multi-view video sequences and multi-time scale analysis, in order to better detect events that take more time to happen. In addition, proposing a complete event detection framework based on spatial relationships as discriminative features seems to be promising. Note that the proposed approach can be placed upstream of the complete event detection framework.

References

1. Visam Project (1997) http://www.cs.cmu.edu/~vsam/
2. Icons Project (2000) http://www.dcs.qmul.ac.uk/research/vision/projects/ICONS/
3. Advisor Project (2000) http://www-sop.inria.fr/orion/ADVISOR/
4. Etiseo Project (2004) http://www-sop.inria.fr/orion/ETISEO/
5. Caretaker Project (2006) http://www-sop.inria.fr/members/Francois.Bremond/topicsText/caretakerProject.html
6. Avitrackr Project (2004) http://www-sop.inria.fr/members/Francois.Bremond/topicsText/avitrackProject.html
7. Beware Project (2007) http://www.eecs.qmul.ac.uk/~sgg/BEWARE/
8. Piciarelli C, Micheloni C, Foresti G (2008) Trajectory-based anomalous event detection. IEEE Trans Circ Syst Video Technol 18:1544–1554
9. Saleemi I, Shafique K, Shah M (2009) Probabilistic modeling of scene dynamics for applications in visual surveillance. IEEE Trans Pattern Anal Mach Intell 31(8):1472–1485
10. Hu W, Xiao X, Fu Z, Xie D, Tan T, Maybank S (2006) A system for learning statistical motion patterns. IEEE Trans Pattern Anal Mach Intell 28:1450–1464
11. Wang T, Snoussi H (2014) Detection of abnormal visual events via global optical flow orientation histogram. IEEE Trans Inf Forensics Secur 9(6):988–998
12. Li A, Miao Z, Cen Y, Wang T, Voronin V (2015) Histogram of maximal optical flow projection for abnormal events detection in crowded scenes. Int J Distrib Sens Netw 11:1–11
13. Adam A, Rivlin E, Shimshoni I, Reinitz D (2008) Robust real-time unusual event detection using multiple fixed-location monitors. IEEE Trans Pattern Anal Mach Intell 30:555–560
14. Kratz L, Nishino K (2009) Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In: IEEE conference on computer vision and pattern recognition, pp 1446–1453
15. Jiang F, Wu Y, Katsaggelos AK (2009) Detecting contextual anomalies of crowd motion in surveillance video. In: 16th IEEE international conference on image processing, pp 1117–1120
16. Mehran R, Oyama A, Shah M (2009) Abnormal crowd behavior detection using social force model. In: IEEE conference on computer vision and pattern recognition, pp 935–942
17. Cong Y, Yuan J, Tang Y (2013) Video anomaly search in crowded scenes via spatio-temporal motion context. IEEE Trans Inf Forensics Secur 8:1590–1599

18. Cong Y, Yuan J, Liu J (2013) Abnormal event detection in crowded scenes using sparse representation. Pattern Recognit 46:1851–1864
19. Hu X, Hu S, Zhang X, Zhang H, Luo L (2014) Anomaly detection based on local nearest neighbor distance descriptor in crowded scenes. Sci World J 1–12
20. Tran D, Yuan J, Forsyth D (2014) Video event detection: from subvolume localization to spatio-temporal path search. IEEE Trans Pattern Anal Mach Intell 36(12):404–416
21. Dan DX, Ricci E, Yan Y, Song J, Sebe N (2015) Learning deep representations of appearance and motion for anomalous event detection. British Machine Vision Conference
22. Ren H, Liu W, Olsen SI, Escalera S, Moeslund TB (2015) Unsupervised behavior-specific dictionary learning for abnormal event detection. British Machine Vision Conference, pp 1–28
23. Lu C, Shi J, Jia J (2013) Abnormal event detection at 150 fps in matlab. In: IEEE international conference on computer vision, pp 2720–2727
24. Zhao B, Fei-Fei L, Xing E (2001) Online detection of unusual events in videos via dynamic sparse coding. In: IEEE conference on computer vision and pattern recognition, pp 3313–3320
25. Cheng K, Chen Y, Fang W (2015) Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression. In: IEEE conference on computer vision and pattern recognition, pp 2909–2917
26. Basharat A, Gritai A, Shah M (2008) Learning object motion patterns for anomaly detection and improved object detection. In: IEEE conference on computer vision and pattern recognition, pp 1–8
27. Solmaz B, Moore B, Shah M (2012) Identifying behaviors in crowd scenes using stability analysis for dynamical systems. IEEE Trans Pattern Anal Mach Intell 34:2064–2070
28. Saleemi I, Hartung L, Shah M (2010) Scene understanding by statistical modeling of motion patterns. In: IEEE conference on computer vision and pattern recognition, pp 2069–2076
29. Tzelepis C, Mezaris V, Patras I (2016) Video event detection using kernel support vector machine with isotropic gaussian sample uncertainty KSVM-iGSU. In: International conference on multimedia modeling, pp 3–15
30. Mazloom M, Li X, Snoek CG (2016) TagBook: a semantic video representation without supervision for event detection. IEEE Trans Multimed 18:1378–1388
31. Li Z, Liu J, Tang J, Lu H (2015) Robust structured subspace learning for data representation. IEEE Trans Pattern Anal Mach Intell 37:2085–2098
32. Li Z, Tang J (2015) Weakly supervised deep metric learning for community-contributed image retrieval. IEEE Trans Multimed 17:1989–1999
33. Abou-Elailah A, Gouet-Brunet V, Bloch I (2015) Detection of ruptures in spatial relationships in video sequences. In: International conference on pattern recognition applications and methods, pp 110–120
34. Tissainayagam P, Suter D (2005) Object tracking in image sequences using point features. Pattern Recognit 38:105–113
35. Zhou H, Yuan Y, Shi C (2009) Object tracking using SIFT features and mean shift. Comput Vis Image Underst 113(3):345–352
36. Miyajima K, Ralescu A (1994) Spatial organization in 2D images. In: Third IEEE conference on fuzzy systems, pp 100–105
37. Hafner J, Sawhney H, Equitz W, Flickner M, Niblack W (1995) Efficient color histogram indexing for quadratic form distance functions. IEEE Trans Pattern Anal Mach Intell 17:729–736
38. Bloch I, Atif J (2015) Hausdorff distances between distributions using optimal transport and mathematical morphology. In: Mathematical morphology and its applications to signal and image processing, pp 522–534
39. Bloch I, Atif J (2016) Defining and computing Hausdorff distances between distributions on the real line and on the circle: link between optimal transport and morphological dilations. Math Morphol Theory Appl 1:79–99
40. Zhang L, van der Maaten L (2013) Structure preserving object tracking. In: IEEE conference on computer vision and pattern recognition, pp 1838–1845
41. Widynski N, Dubuisson S, Bloch I (2012) Fuzzy spatial constraints and ranked partitioned sampling approach for multiple object tracking. Comput Vis Image Underst 116:1076–1094
42. Morimitsu H, Roberto M, Bloch I (2014) A spatio-temporal approach for multiple object detection in videos using graphs and probability maps. In: International conference on image analysis and recognition, pp 421–428
43. Morimitsu H, Bloch I, Cesar RM (2016) Exploring structure for long-term tracking of multiple objects in sports videos. Comput Vis Image Underst 159:89–104
44. Harris C, Stephens M (1988) A combined corner and edge detector. In: Fourth Alvey vision conference, pp 147–151
45. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60:91–110
46. Basseville M, Nikiforov IV (1993) Detection of abrupt changes: theory and application. Prentice Hall, Englewood Cliffs, p 104
47. PETS (2006) http://www.cvg.rdg.ac.uk/PETS2006/data.html
48. PETS (2009) http://www.cvg.rdg.ac.uk/PETS2009/a.html
49. Bazzani L, Cristani M, Murino V (2012) Decentralized particle filter for joint individual-group tracking. In: IEEE conference on computer vision and pattern recognition, pp 1886–1893