Shape-based individual/group detection for sport videos categorization

Costas Panagiotakis, Emmanuel Ramasso, Georgios Tziritas, Michèle Rombaut, Denis Pellerin

Shape-based Individual/Group Detection for Sport Videos Categorization

Costas Panagiotakis [a,*], Emmanuel Ramasso [b,*], Georgios Tziritas [a], Michèle Rombaut [b] and Denis Pellerin [b]

[a] *Computer Science Department, University of Crete, P.O. Box 2208, Greece*

[b] *Laboratoire des Images et Signaux, 46 av. Félix Viallet, 38031 Grenoble, France*

Abstract

We present a shape based method for automatic people detection and counting without any assumption or knowledge of camera motion. The proposed method is applied to athletic videos in order to classify them to videos of individual and team sports. Moreover, in the case of team (multi-agent) sport, we propose a shape deformations based method for running/hurdling discrimination (activity recognition). Robust, adaptive and independent from the camera motion, the proposed features are combined within the Transferable Belief Model (TBM) framework providing a two level (frames and shot) video categorization. The TBM allows to take into account imprecision, uncertainty and conflict inherent to the features into the fusion process. We have tested the proposed scheme into a big variety of athletic videos like pole vault, high jump, triple jump, hurdling, running, etc. The experimental results of 97% individual/team sport categorization accuracy, using a dataset of 252 real videos of athletic meetings acquired by moving cameras under varying view angles, indicate the stability and the good performance of the proposed scheme.

Key words: people detection, people counting, Video analysis, Transferable Belief Model, team (multi-agent) activity recognition

1 Introduction

Video indexing is required to cope with the increasing number of videos in databases. Low level indexing (e.g. from dominant color) is not very useful nor relevant for the end-user

* Corresponding author.

Email addresses: cpanag@csd.uoc.gr (Costas Panagiotakis), Emmanuel.Ramasso@lis.inpg.fr (Emmanuel Ramasso), tziritas@csd.uoc.gr (Georgios Tziritas), Michele.Rombaut@lis.inpg.fr (Michèle Rombaut), Denis.Pellerin@lis.inpg.fr (Denis Pellerin).

(a) (b) (c) (d) (e) (f)

Fig. 1. **(a), (b)** Original image and the silhouette estimated by the method of [15] under stable camera. The silhouette quality is high since accuracy of human boundary is high and the number of wrong classified pixels is low. **(c), (e)** Original images of an individual sport (long jump) and a team sport (100 m running) and **(d), (f)** the corresponding silhouettes estimated by the method of [16] under moving camera. The silhouettes quality is low since the silhouette could be partitioned to several segments, several objects or wrong classified pixels could appear, and the estimated human boundary accuracy is low.

who prefers high level indicators such as "TV news", "cars pursuit" or "goals in soccer matches" [1].

Indexing based on human action and activity is of key of importance because can be applied in many areas such as database management [2], surveillance [3] or human-computer interface [4]. In previous work, a novel architecture utterly based on the Transferable Belief Model [5], an interpretation of Shafer's theory of evidence [6, 7], was proposed [8–11] for human action and activity recognition in athletic sports videos. As for the proposed paper, the goal is to recognize high level actions and activities based on low level shape-motion understandable features [11–13]. The database used for testing is made of real videos acquired by a moving camera under varying view angles and can concern indoor or outdoor meetings. Videos mainly comes from broadcast TV and are compressed. As for most of the methods proposed in the literature [14], the main assumption of the system is that only one main athlete is moving (little other moving objects are managed). In this paper, a solution is proposed to relieve the system and alleviate this hypothesis. Figure 1 depicts the case of one athlete and several athletes as well as the low quality of the obtained binary silhouettes.

The goal of the proposed method is to classify a video into *individual* (e.g. a high jump, a pole vault) or *team* (e.g. 100m running, 110 hurdling). As well, the system detects and counts the number of people in videos. A reliability factor is computed at each frame in order to quantify the quality of the classification and the quality is taken into account for decision concerning the number of people. A procedure is also proposed to distinguish

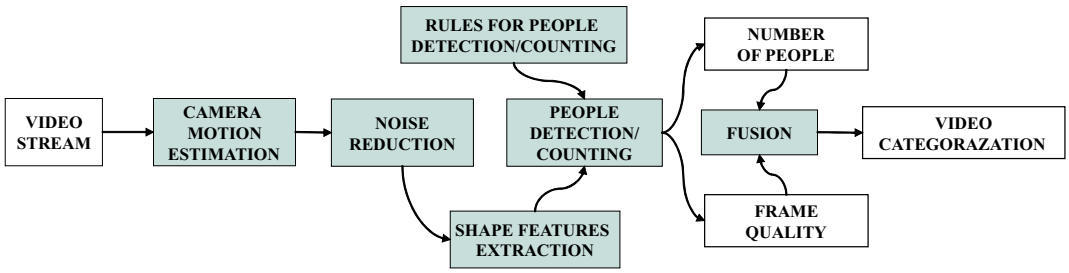


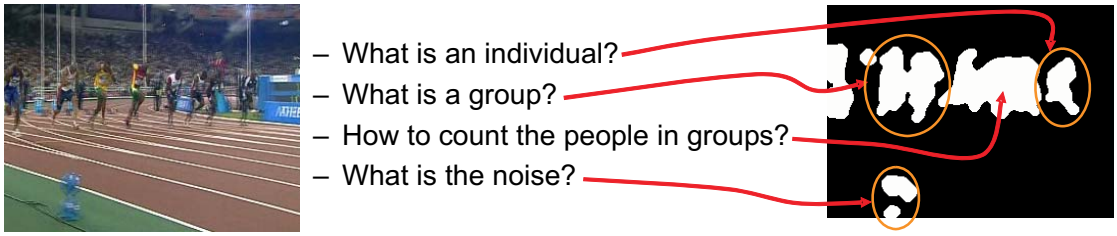
Fig. 2. Schema of the proposed system architecture.

between running and hurdling (activity recognition). In the system, no initialization step is required and no assumption or knowledge is assumed about the number of people and their motion in the scene.

The proposed system can be decomposed into several main modules illustrated in Figure 2:

(1) Silhouettes are computed using a camera motion estimation method [16], where an affine model is used to describe the camera motion. Such a model is generally sufficient for most of real video sequences. The above method that we use, was implemented by the Vista Team of IRISA and has the advantage to take into account the global variation of illumination thus it is adapted for real videos.
(2) A silhouette noise reduction procedure is executed using a short time window.
(3) People detection and counting is performed using shape features per blob (object).
(4) Finally, the video categorization procedure based on a TBM fusion process is executed taking into account the whole history of the number of the estimating people per frame and the estimated quality frame factor in order to decide between individual sport and team sport video. In the case of team sport, we propose a method for running/hurdling discrimination.

The system relies on three shape based features (concerning human): eccentricity, major axis angle and normalized area. These features are robust, adaptive and independent from the camera motion. They are estimated from binary silhouettes obtained by a robust camera motion estimation and object localization techniques (Fig. 1). Binary silhouette are analyzed by the algorithm in order to detect noise objects, groups people and individual as depicted in Figure 3. Figure 1 depicts some results. The features are combined within the *Tranferable Belief Model* (TBM) framework [5] in order to perform the video categorization. The TBM is an alternative to probability theory for knowledge modelling and the main



(a) (b)

Fig. 3. **(a)** Original image of a team sport (100 m running) and **(b)** the corresponding silhouette estimated by the method of [16] under moving camera. The red arrows show the correspondence between the questions on the middle and the estimated blobs.

advantage and power of the TBM is to explicitly model doubt and conflict. Both theories, TBM and probability, are complementary: the TBM is well suited to encode and combine the available information (like ignorance and conflict) while probability is preferred for decision [17].

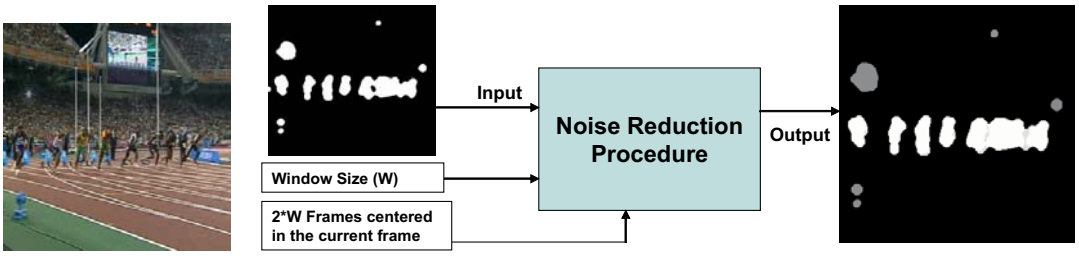
The rest of the paper is organized as follow. Section 2 present related works. Section 3 presents the silhouette estimation technique, the noise reduction procedure and the proposed features. The people detection and counting method is presented in Section 4. Section 5 describes the video categorization scheme. A running/hurdling discrimination is described in Section 6. Experimental results are given in Section 7. Finally, conclusions and discussion are provided in Section 8.

2 Related Work

Human motion analysis consists in [18] *detection, tracking* and *recognition*. Group detection and/or counting is generally embedded in the *detection and tracking* processes.

Object detection and tracking in complicated environments is still the key problem of the visual surveillance and it is becoming an important issue in several applications such as camera based surveillance and human machine interaction. The detection and tracking algorithms are challenged by occluding and fast/complicated moving objects, as well as illumination changes. Concerning the 2-D approaches, Wang et al. [19] propose a method to recognize and track a walker using 2D human model and both static and dynamic cues of body biometrics. Moreover, many systems use Shape-From-Silhouette methods to detect

and track the human in 2D [12] or 3D space [20]. The silhouettes are easy to extract providing valuable information about the position and shape of the person. When the camera is static, background subtraction techniques can give high accuracy measures of human silhouettes by modeling and updating the background image [21]. The temporal difference based methods [22] detect the moving objects using a temporal difference between successive frames [23], while the probabilistic based approaches [24, 25] use statistical and probability models getting high accuracy results, but they suffer from high computational cost. Otherwise, when the camera is moving, camera motion estimation methods [16, 26] can locate the independently moving objects. The system called W4 [15] is based on a statistical-background model to locate people and their parts (head, hands, feet, torso, etc.) using stable cameras and allowing multiple person groups. McKenna et al. [27] describe a computer vision system (using background substraction) for tracking multiple people from a stable camera, which is based on color information to disambiguate occlusion. Figueroa et al. [28] propose a system of tracking soccer players using multiple stable cameras. The occlusions have been treated by splitting segmented blobs based on morphological operators and a backward and forward graph representation based on human shape, motion and color features. However, in a real soccer game, there are crowd situations, where the people should be manually tracked. The M2-tracker, presented in [29], uses multiple static cameras assigning the pixels in each camera views to a particular person using color histograms. Rabaud and Belongie [30] present a method for counting moving objects without tracking them based on a highly parallelized version of the KLT tracker. It is performed in crowding situations where the tracking does not make sense to perform. Reliable counts can be produced under the constraints of stable camera and walking people based on template-based tracker [31]. Sacchi et al. [32] propose a scheme based on blod detection for people counting in a constrained environment. Rad and Jamzad [33] present a system for road traffic monitoring (vehicle classification and vehicle counting) using morphological operations for vehicle detection Kalman filtering and background differencing techniques under stable camera getting 96% accuracy. Cheng and Chen [34] propose a method for detecting and tracking multiple moving objects based on discrete wavelet transform and identifying the moving objects by their color and spatial information using a stable camera. The human detection is done using the low band of the wavelet transform of the image due to the fact that most of the fake motions in the background can be decomposed into the high



(a) (b)

Fig. 4. **(a)** Original image. **(b)** Schema of the noise reduction procedure. In the output image, the heavy gray objects and light gray pixels correspond to detected noise.

frequency wavelet sub-band.

3 Silhouette Estimation

We have used the algorithm presented in [16] in order to extract the camera motion and the binary silhouettes. This method consists of an iterative and robust multi-resolution estimation of parametric motion models between two successive frames taking into account the global variation of illumination. A gray level image is also generated by the algorithm whose pixel value gives a piece of information on their belonging to the dominant camera motion, generally the motion of the background. Using thresholding and morphological operations a binary silhouette is estimated, which is the output of the silhouette estimation procedure. Next, a noise reduction reduction method is performed in order to remove the noise blobs. Shape features are computed, which are used to detect humans, groups and to count the people in groups. At the end of this section, we describe the shape features conversion into beliefs (symbolic representation).

3.1 Noise reduction

Figure 4 illustrates the schema of the noise reduction procedure. The binary silhouettes, estimated by the camera motion estimation method, probably contain objects that do not follow the athletes motion (see Fig. 3(b)), e.g. fake objects in the background. Moreover, sometimes, fake objects appear because of instant failure of camera motion estimation method. We assume that the camera tries to track the athletes (this is a tenable assumption,

since the athlete is the object of interest), so the athletes are about in the same position in a short time window. If we suppose that the noise is appearing in random positions (white noise) over the time, then a lowpass time filter can remove the noise.

We use a short time window of size $2 \cdot W + 1$ (e.g. $W = 3$) centered at the current frame t, (time window: $\{t - W, t - W + 1, \cdots, t, \cdots, t + W\}$). The probability of a pixel (x, y), at frame t, to belong to a human ($P_r((x, y) \in H)$) is calculated by getting the number of times that (x, y) belongs to the foreground over the time window divided by the window size. Then, the probability of an object i at frame t, O_t^i, to be a human ($P_r(O_t^i \in H)$) is calculated by getting the mean value of probabilities $P_r((x, y) \in H)$, where (x, y) denotes a pixel of object O_t^i:

$$\begin{aligned} P_r((x, y) \in H) &= \frac{1}{2 \cdot W + 1} \times \sum_{k=t-W}^{k=t+W} I_k(x, y) \\ P_r(O_t^i \in H) \quad &= \frac{1}{|O_t^i|} \times \sum_{(x,y) \in O_t^i} P_r((x, y) \in H) \end{aligned} \tag{1}$$

with I_k is binary image of frame k estimated by the camera motion estimation method and $|O_t^i|$ is the number of pixels of object O_t^i. We have used a threshold of 0.5 in order to decide if an object corresponds to a human. If an object is detected as human ($P_r(O_t^i \in H) > 0.5$), then we could use the same threshold in order to erase some noisy pixels (x, y) from the human object O_t^i i.e. $P_r((x, y) \in H) \leq 0.5$ (see light gray pixels of Fig. 4(b)). Using this second thresholding, groups of people are possibly separated, facilitating the people counting procedure (see Fig. 4(b)).

3.2 Features Extraction

Shape features are computed in order to detect humans, groups and to count the people in groups. We compute for each object O_t^i, its major axis angle θ_i, its eccentricity ε_i and its normalized area s_i.

3.2.1 Major axis angle θ_i

The mass center point (x_c, y_c) of the object is first computed. This point is defined as the mass center of the object pixels, $x_c = \frac{1}{|O_t^i|} \sum_{(x,y)\in O_t^i} x$, $y_c = \frac{1}{|O_t^i|} \sum_{(x,y)\in O_t^i} y$. Next, the object major axis is computed. It is defined as the main axis of the best fitting ellipse. This axis passes from the mass center point, that already has been estimated, so we have to compute just the axis orientation. The angle of the object O_t^i major axis θ_i is defined by the three second order moments $\mu_{1,1}, \mu_{2,0}$ and $\mu_{0,2}$:

$$\theta_i = \arctan(\frac{2 \cdot \mu_{1,1}}{\mu_{2,0} - \mu_{0,2}}), \quad \theta_i \in [0, 180] \tag{2}$$

with

$$\mu_{p,q} = \sum_{(x,y)\in O_t^i} (x - x_c)^p (y - y_c)^q \tag{3}$$

This angle shows the main orientation of the object. In the whole paper, angles are measured in degrees. The robustness of θ_i estimation is determined by the object's eccentricity.

3.2.2 Eccentricity ε_i

The eccentricity ($\varepsilon_i \geq 1$) is defined by the ratio between the two principal axes of the best fitting ellipse, measuring how thin and long a region is. If ε_i is close to one, then θ_i will be unspecified. The eccentricity can be defined by the three second order moments $\mu_{1,1}, \mu_{2,0}$ and $\mu_{0,2}$:

$$\varepsilon_i = \sqrt{\frac{\mu_{2,0} + \mu_{0,2} + \sqrt{(\mu_{2,0} - \mu_{0,2})^2 + 4 \cdot \mu_{1,1}^2}}{\mu_{2,0} + \mu_{0,2} - \sqrt{(\mu_{2,0} - \mu_{0,2})^2 + 4 \cdot \mu_{1,1}^2}}} \tag{4}$$

3.2.3 Area s_i

The area feature s_i should be normalized in order to be independent from both image size and distance of the object from the camera. However, there is not any knowledge available concerning the distances and the sizes (in pixels) of the projected athletes in the image plane. Generally, it holds that the area of interest concerns the athletes, that are tracked by the camera. These athletes normally have similar distances from the camera. Therefore,

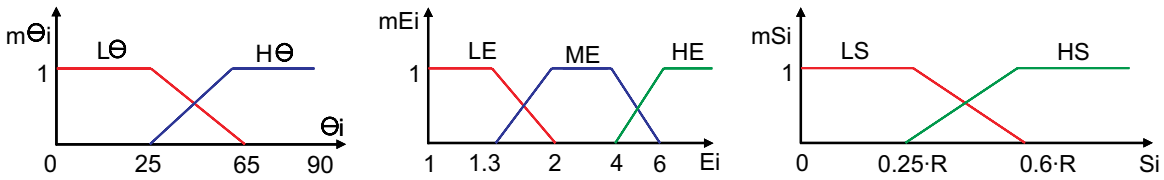


Fig. 5. From numerical features to belief. **(a)** Angle, **(b)** Eccentricity, **(c)** Area.

s_i is defined as ratio between an object area ($|O_t^i|$) and the mean object area $\frac{\sum_k |O_t^k|}{N_t}$:

$$s_i = \frac{|O_t^i| \cdot N_t}{\sum_k |O_t^k|} \tag{5}$$

where N_t (see Section 5.1) denotes the estimated number of people at frame t.

These features are uncorrelated, independent from camera view, and their values can be estimated robustly under low quality silhouettes. The proposed features are independent from camera view under the assumption that the angle between the image plane and ground is almost stable. This assumption is generally true in sport videos since the camera tracks the athletes (by zooms, translations, rotations) without changing its elevation.

3.3 Numeric-to-symbolic conversion

The used features (area, angle and eccentricity) are simple and well understandable. Thus, they can be converted easily into beliefs (symbolic representation). Using symbolic representation, the people detection and counting can be performed based on appropriate table rules. We have proposed the numeric-to-symbolic conversion presented in Fig. 5, where L is used for low value, M for medium values and H for high values.

Fig. 5(a) presents the angle Θ_i numeric-to-symbolic conversion, with $\Theta_i = min(\theta_i, 180 - \theta_i)$, $\Theta_i \in [0, 90]$. When Θ_i is close to 90 degrees the object major axis direction is vertical, otherwise when Θ_i is close to 0, the object major axis direction is horizontal. There are two beliefs for angle feature: low angle ($L\Theta$), which is true when $\Theta_i \leq 25$, and high angle ($H\Theta$), which is true when $\Theta_i \geq 65$. The red and blue curves correspond to the probability of $L\Theta$ and $H\Theta$ symbols respectively.

Fig. 5(b) presents the numeric-to-symbolic conversion of eccentricity. There are three beliefs for eccentricity feature: low eccentricity (LE), which is true for $\varepsilon_i \leq 1.3$ indicating to the unspecified direction, medium eccentricity (ME), which is true for $2 \leq \varepsilon_i \leq 4$, and high eccentricity (HE), which is true for $\varepsilon_i \geq 6$. An individual eccentricity is normally medium. The red, blue and green curves correspond to the probability of LE, ME and HE respectively.

Fig. 5(c) presents the numeric-to-symbolic conversion of area feature s_i. Two beliefs are concerned: low area (LS), which is true for $s_i \leq 0.25 \cdot R$ [1] indicating a little area objects (possibly noise), and high area (HS), which is true for $s_i \geq 0.6 \cdot R$, indicating an object of normal area (possibly humans). The red, blue and green curves correspond to the probability of LE, ME and HE respectively.

4 People Detection and Counting

4.1 People detection

The noise reduction procedure removes the most of noise objects of the image (white noise). However, it is possible that some objects appear in the scene although they are not humans. These objects can be detected using their shape features. The goal of the proposed procedure is to remove such objects. We have used the rules of Table 6(a) in order to detect and remove such objects, combining the symbolic beliefs. This table can be estimated by a learning stage using an EM procedure for instance. The problem is to obtain references which require to manually annotate each frame of the videos. In the case that there are more symbols per cell, then the value behind each symbol is analog to the symbol's probability (after normalization, the sum of symbol values on each cell will be one). In the case that there is just one symbol per cell, then its probability is one. Using this table, the probability of human $P_r(O_t^i \; is \; H)$ can be estimated by the cells where the

[1] R is an adaptive factor, denoting the probability of an object, which has $s_i > 0.05$, to be a human object. It is robustly estimated using previous results of the people detection procedure.

	LE	ME	HE
LS,LΘ	N	N	N
LS,HΘ	N	H , 0.15 N , 0.85	N
HS,LΘ	N , 0.2 H , 0.8	H	H
HS,HΘ	N , 0.1 H , 0.9	H	N , 0.05 H , 0.95

(a)

	LE	ME	HE
HS,LΘ	1 , 0.3+ max(1-|K-min(1,K)|,0) 2 , 1 + max(1-|K-2|,0) 3 , 0.8 +max(1-|K-max(3,|K|)|,0)	$\lfloor K+1 \rfloor$,$1-K+\lfloor K+1 \rfloor$ $\lfloor K \rfloor$,*rest*	$\lfloor K+1 \rfloor$,$1-K+\lfloor K+1 \rfloor$ $\lfloor K \rfloor$,*rest*
HS,HΘ	1 , 0.6 + max(1-|K-min(1,K)|,0) 2 , 1 + max(1-|K-2|,0) 3 , 0.8 + max(1-|K-max(3,|K|)|,0)	1 , 0.6+0.4*max(1-|K-1|,0) 2 , rest	1 , 0.6+0.4*max(1-|K-1|,0) 2 , rest

(b)

Fig. 6. The table rules for **(a)** human/noise detection, N, H denote *noise*, *humans*, respectively, and **(b)** people counting.

probability of H is positive:

$$P_r(O_t^i \; is \; H) = 0.15 \cdot ME_i \cdot LS_i \cdot H\Theta_i + 0.8 \cdot LE_i \cdot HS_i \cdot L\Theta_i + ME_i \cdot HS_i \\ + HE_i \cdot HS_i \cdot L\Theta_i + 0.95 \cdot HE_i \cdot HS_i \cdot H\Theta_i \tag{6}$$

An object O_t^i will be detected as human if $P_r(O_t^i \; is \; H) > 0.5$ (it holds that $P_r(O_t^i \; is \; N) + P_r(O_t^i \; is \; H) = 1$). Otherwise, it will be detected as noise.

4.2 People counting

The people counting procedure is executed for each human detected object. The people counting is based on the assumption that each human major axis (in the most time) is mainly vertical, which is true during running and hurdling and most of the time true in jumping and falling. Thus, an individual object probably has high angle and medium eccentricity. Otherwise, the object is probably a group of people containing two, three or

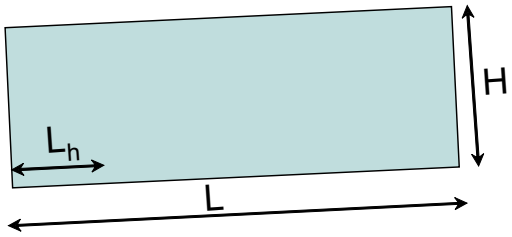
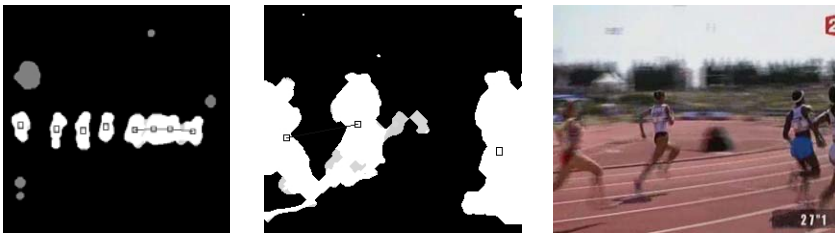


Fig. 7. Group model of length L, height H and human length L_h.



(a) (b) (c)

Fig. 8. **(a), (b)** The little black boxes corresponds to the mass centers of the detected humans. **(a)** Four individual and one group of four people are detected. The original image is shown in Fig. 4(a). **(b)** An individual and a group of two people are detected. **(c)** The original image of **(b)** is shown.

more people. Using the rules of Table 6(b), where the proposed features are combined, the number of people per object can be estimated. K denotes the number of people (real value) in groups estimated by the algorithm described hereafter.

The number of people K_i (real value) of a horizontally directed object O_t^i is estimated by using its eccentricity (ε_i) and its area ($|O_t^i|$). Fig. 7 illustrates the used group model.

- Let L, H and L_h be the mean group length (major axis), height, and human length. The object is horizontally directed, therefore L can be computed directly using $|O_t^i|$, ε_i, since $|O_t^i| = L \cdot H$, $\varepsilon_i = \frac{L}{H}$. Therefore, it holds that, $L = \sqrt{|O_t^i| \cdot \varepsilon_i}$.
- Next, we set $K_i = s_i$, getting an eccentricity per human $e_h = \frac{H}{L_h}$, where $H = \frac{|O_t^i|}{L}$ and $L_h = \frac{L}{K_i}$.
- Finally, if e_h is higher than four, which is the maximum individual eccentricity, the number of humans will be recomputed by enforcing the eccentricity per human to be four. Thus, it holds that, $L_h = \frac{H}{4}$ and $K_i = \frac{L}{L_h}$.

Fig. 8 illustrates the results of people detection and counting algorithm.

A measurement of frame quality (reliability) factor Q_t can be estimated using the probability of the decisions (human/noise decision and counting decision) results. If the decisions are taken with low probabilities then the Q_t should be low, otherwise the Q_t should be high. Let $P_r^{HN}(O_t^i)$ denotes the decision probability of the object O_t^i to be human or noise. Let $P_r^{NP}(O_t^i)$ be the decision probability concerning the number of people in the object O_t^i. In both of the cases, they are computed by getting the maximum of the corresponding probabilities. $Q_t \in [0, 1]$ is estimated by the product of the expected values (E_i) of $P_r^{HN}(O_t^i)$, $P_r^{NP}(O_t^i)$ over the objects:

$$Q_t = E_i(P_r^{HN}(O_t^i)) \cdot E_i(\frac{P_r^{NP}(O_t^i)}{\sqrt{\max(K_i, 1)}}) \tag{7}$$

$P_r^{NP}(O_t^i)$ is divided with the square root of the number of detected people $\sqrt{\max(K_i, 1)}$ in object O_t^i, because the accuracy of people counting procedure decreases, as the number of people increases (occlusions are appeared). The use of $\sqrt{\max(K_i, 1)}$ improves slightly the categorization accuracy (see Section 7). Q_t will be used on video categorization scheme. Fig. 9(b) presents Q_t numeric-to-symbolic conversion. There are three beliefs for quality factor: bad quality (Bad), unknown quality ($Bad \cup Good$) and high quality ($Good$).

5 Video Categorization Scheme

The results of people counting procedure and the frame quality factor are fused using TBM framework in order to discriminate the video of individual sport (I) and team sport (T). We have used the TBM framework, since it is more general than probabilities and explicitly defines the conflict and doubt. In this work, we exploit conflict to improve the modelling of the trapezes of the numeric to symbolic conversion, so that the fusion minimizes the conflict. Doubt is used for "hesitation modelling": when we are not sure about the answer, we hesitate, waiting for a "clear signal". Finally, the videos are classified by fusing the whole information during the analysis stage.

5.1 Background on belief mass

This section only provides required tools for TBM fusion. For more details, the reader can refer to [5, 35–38].

The classification concerns two classes: video of individual sport (I) and team sport (T). Therefore, $\Omega = \{I, T\}$ is the frame of discernment of the classification. A basic belief assignment (BBA) m_t^Ω at frame t is defined on the set of propositions $2^\Omega = \{\emptyset,\ I,\ T, I \cup T\}$, where $\emptyset$ and $I \cup T$ correspond to the conflict and doubt respectively. $m_t^\Omega : 2^\Omega \rightarrow [0, 1]$, $X \rightarrow m_t^\Omega(X)$ and by construction it holds that $m_t^\Omega(\emptyset) = 0$, and $\sum_{X \subseteq \Omega} m_t^\Omega(X) = 1$. A value $m_t^\Omega(X)$ is a basic belief mass which expresses a confidence proposition $X \subseteq \Omega$ according to a given feature but does not imply any additional claims regarding subsets of X [39]. It is the fundamental difference with probability theory. The previously described fuzzy-set inspired method is used to convert each numerical feature into sources of belief.

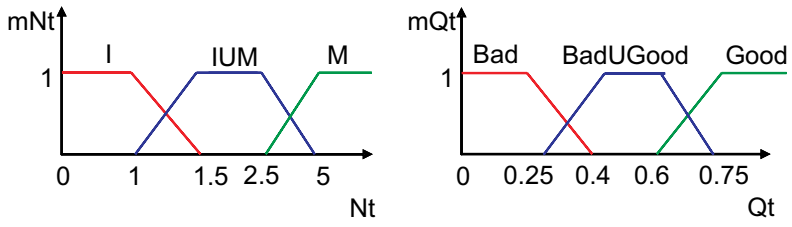
5.2 Number of people estimation

The number of people N_t (real value) at frame t is robustly estimated using quality factor,

$$N_t = (1 - Q_t) \cdot N_{t-1} + Q_t \cdot TP_t \tag{8}$$

where TP_t (integer value) denotes the number of the detected people at frame t. The estimation of number of people by N_t is more robust than by TP_t since it takes into account the quality factor.

Fig. 9(a) presents N_t numeric-to-symbolic conversion. There are three beliefs for N_t: low number of people (I), which is true when $N_t \leq 1$, middle number of people ($I \cup M$), which is true when $1.5 \leq N_t \leq 2.5$ and high number of people (M), which is true for $N_t \geq 5$. We have used the rules of Table 9(c) in order to compute the BBA in each frame.



(a) (b)

	I	**IUM**	**M**
Bad	IUT	IUT	IUT
BadUGood	I, 0.5 IUT, 0.5	IUT	T, 0.5 IUT, 0.5
Good	I	IUT	T

(c)

Fig. 9. From numerical features to belief. **(a)** Number of people, **(b)** Quality factor. **(c)** The table rules for individuals/groups detection.

5.3 Short Time Decision

The short time decision concerns a local decision that can be taken at frame t. The conjunctive rule of combination [40] is applied to obtain the belief taking all features into account. The fusion process is performed frame by frame for each proposition X yielding a new local mass $\hat{m}_t^{\Omega}(X)$:

$$\hat{m}_t^{\Omega}(X) = \hat{m}_{t-1}^{\Omega} \textcircled{\cap} m_t^{\Omega}(X) = \sum_{C \cap D = X} \hat{m}_{t-1}^{\Omega}(C) \cdot m_t^{\Omega}(D) \tag{9}$$

Using the aforementioned fusion process, the mass of the empty set (conflict) is going to increase to one while the masses of the other propositions are going to decrease to zero. This effect is due to the fact that the empty set is aborptive by the $\textcircled{\cap}$-rule. When the conflict is high, the trapezes used in the numeric to symbolic conversion are modified manually in order to decrease the conflict (by adding doubt for instance). When the conflict is not too high, we have used the Dubois & Prade's conflict redistribution rule [41] (adaptive rule) in order to manage the conflict yielding to $\hat{m}_t^{\Omega}(\emptyset) = 0$:

$$\hat{m}_t^{\Omega}(C \cup D) = \sum_{C \cap D = \emptyset} \hat{m}_{t-1}^{\Omega}(C) \cdot m_t^{\Omega}(D) \tag{10}$$

In the general case of this kind of combination, we have a proposition A on which the partial conflicting masses are assigned.

5.4 Final Decision

The final decision concerning the whole video sequence is taken by "equivalent" fusion of the beliefs at each frame. Therefore, at frame t, the mean mass $\bar{m}_t^{\Omega}(X)$ of the proposition X is computed by getting the mean of the local decision mass $\hat{m}_k^{\Omega}(X)$ over the frames $\{1, 2, \cdots, t\}$:

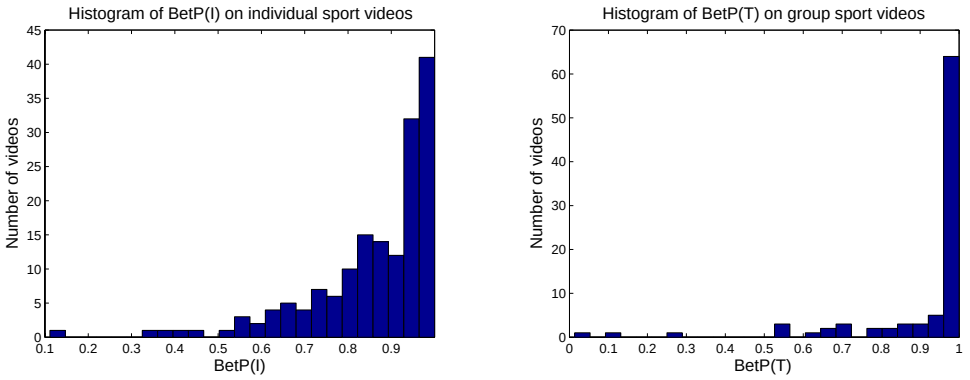
$$\bar{m}_t^{\Omega}(X) = \frac{1}{t} \cdot \sum_{k=1}^{t} \hat{m}_k^{\Omega}(X) \tag{11}$$

Finally, the decision is taken using the pignistic probability (BetP) proposed by Ph. Smets [42]. BetP is is a probability measure used for decision. $BetP(I)$, $BetP(T)$ are given as follows:

$$\begin{aligned} BetP(I) &= \frac{1}{1 - \bar{m}_t^{\Omega}(\emptyset)} (\bar{m}_t^{\Omega}(I) + \frac{\bar{m}_t^{\Omega}(I \cup T)}{2}) \\ BetP(T) &= \frac{1}{1 - \bar{m}_t^{\Omega}(\emptyset)} (\bar{m}_t^{\Omega}(T) + \frac{\bar{m}_t^{\Omega}(I \cup T)}{2}) \end{aligned} \tag{12}$$

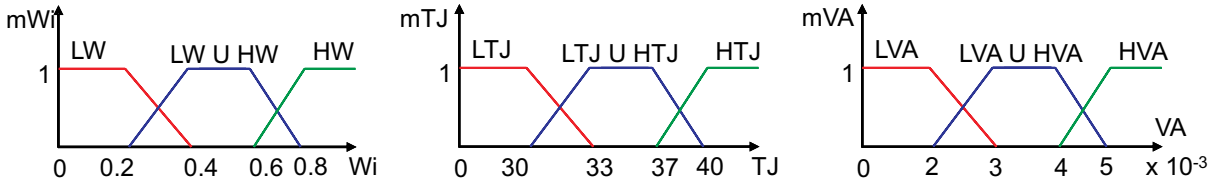
The above decision rule is equivalent with the selection of the proposition X with the highest mean mass $\bar{m}_t^{\Omega}(X)$. Concerning the conflict, it holds that $\bar{m}_t^{\Omega}(\emptyset) = 0$, since $\hat{m}_k^{\Omega}(\emptyset) = 0$. If $\bar{m}_t^{\Omega}(I \cup T)$ is the highest, this means that we can not decide (doubt is the answer). However, if we want to decide in any case, then I will be selected, if $\bar{m}_t^{\Omega}(I) > \bar{m}_t^{\Omega}(T)$, otherwise T will be selected. Using the aforementioned scheme, the value of the selected mean mass provides a final decision confidence value. This value corresponds to the probability of the final decision.

Let $X = argmax(BetP(I), BetP(T))$, then $BetP(X)$ corresponds to a confidence value of the final the decision. Fig. 10 illustrates the histograms of $BetP(I)$ on individual sport videos and $BetP(T)$ on team sport videos. The mean confidence value for team sport videos is higher than for individual sport videos. This is an expected result, since team videos are more easily detected because of multiple people, while on individual sport videos other people can be appeared confusing the whole process.



(a) (b)

Fig. 10. Histograms of **(a)** $BetP(I)$ on individual sport videos and **(b)** $BetP(T)$ on team sport videos.

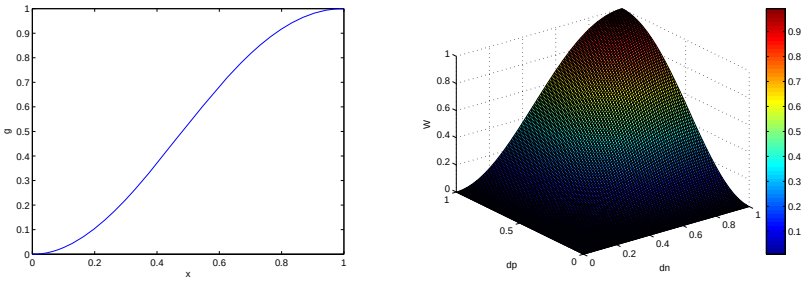


(a) (b) (c)

Fig. 11. From numerical features to belief. **(a)** Extremum reliability factor $W(\tau_k)$, **(b)** Jumping period (TJ). **(c)** Variance of humans' area (VA).

6 Running/Hurdling Discrimination

In the case of a team sport detection, running and hurdling classification can be performed. Running and hurdling can be discriminated using shape deformations and mass center trajectory (vertical variation of mass center) during the jumping stage of the hurdling. The mass center trajectory is not a robust feature because it depends on the camera motion. Moreover, it varies slightly during the jumping stage, so it is very sensitive on noise under low quality silhouettes. Concerning eccentricity, it is a robust feature that can be used in discrimination. It decreases during the jumping stage independently from camera view because the human silhouette is deformed to a circular-like shape. This deformation is easier to be recognized at the start of the hurdling sequence, because the athletes are synchronized. We have used the TBM framework on this classification similarly with the individual/team sport categorization scheme.



(a) (b)

Fig. 12. **(a)** Plot of function $g(x) = x^2 \cdot e^{1-x^2}$. **(b)** Plot of function $w(dp, dn)$.

6.1 Feature extraction

We have used a global eccentricity E_t for a frame t, since we do not perform human tracking in order to measure and track eccentricities per human object. E_t is estimated by the mean of human eccentricities. Median and low pass filtering are performed on E_t reducing the noise level. As experiments show, in a hurdling sequence, E_t is decreasing at least 30% during the jumping stage, while it remains almost unchange during running stage (see Fig. 15). In order to measure this property, we detect the times of jumping and the jumping period during the sequences based on eccentricity variations. We have used a modified version of an extremum analysis method, proposed in [13], where was applied to measure the gait period.

First, we estimate the times $(\tau_k, k \in \{1, \cdots, K\})$ of the local maximum and minimum of E_t. These times are estimated in a short time window (red points on Fig. 15 (xvii)). Their minimum values correspond to jumping times on hurdling videos. If τ_k is a valid extremum (not noise), then the quantities defined as:

$$dp_k = 5 \cdot |\frac{E_{\tau_k} - E_{\tau_{k-1}}}{E_{\tau_k} + E_{\tau_{k-1}}}|, \quad dn_k = 5 \cdot |\frac{E_{\tau_k} - E_{\tau_{k+1}}}{E_{\tau_k} + E_{\tau_{k+1}}}|$$

should be at least one. Thus, we introduce the reliability factor of the extremum, τ_k:

$$W(\tau_k) = w(dp_k, dn_k) = dp_k^2 \cdot dn_k^2 \cdot e^{2-dn_k^2-dp_k^2}$$

getting values between one and zero. It is close to one when (dp_k, dn_k) are close to one (see Fig. 12). In the case of $dp_k > 1$ or $dn_k > 1$ we use in the formula one instead of dp_k or

dn_k. If $W(\tau_k)$ is close to one, then the measurement τ_k is probably valid, so there is a true detection (in hurdling case).

The jumping period can be estimated as follows. Let $T_k = 2 \cdot (\tau_k - \tau_{k-1})$ be a measurement of the jumping period using the τ_k, τ_{k-1}. The jumping period (TJ) is determined by the weighted mean of T_k, $TJ = \frac{1}{\sum_{k=1}^{K-1} W(\tau_k)} \sum_{k=1}^{N-1} W(\tau_k) \cdot T_k$. It holds that the jumping period on hurdling sequences is normally less than 30 frames.

Because of low quality silhouettes, eccentricity is possible to change in running videos. However, in many cases this phenomenon occurs since the humans' area evolves at the same time. It holds that the humans area remains almost unchanged during jumping stage in hurdling videos and eccentricity varies because of shape deformation. Therefore, we measure the variance (VA) of detected human objects during the recognized jumping period increasing the accuracy of discrimination.

6.2 Numeric-to-symbolic conversion

Next, numeric-to-symbolic conversion is performed. Fig. 11(a) presents $W(\tau_k)$ numeric-to-symbolic conversion. There are three beliefs for $W(\tau_k)$: LW (low) for an invalid extremum, $LW \cup HW$ for an unspecified extremum and HW (high) for a valid extremum. Fig. 11(b) presents TJ numeric-to-symbolic conversion. There are three beliefs for TJ: LTJ (low) for an normal jumping period, $LTJ \cup HTJ$ for an unspecified jumping period and HTJ (high) for an invalid jumping period. Fig. 11(c) presents VA numeric-to-symbolic conversion. There are three beliefs for VA: LVA (low) for an normal hurdling sequence, $LVA \cup HVA$ for an unspecified hurdling sequence and HVA (high) for an invalid hurdling sequence. The values for the trapezoids have been estimated by statistical analysis of the dataset.

6.3 Extremum mass

Based on $W(\tau_k)$, an extremum τ_k can be classified into valid and invalid categories using TBM framework. A valid and invalid extremum correspond to hurdling (H) and running

(R) propositions respectively. Therefore, the mass $\hat{m}_k^{\Omega_B}(Y)$, $\Omega_B = \{R, H\}$, $Y \subseteq \Omega_B$ is estimated using $W(\tau_k)$, where the propositions R, $R \cup H$ and H correspond to LW, $LW \cup HW$ and HW respectively. Fig. 15 (xvii) illustrates an example of $\hat{m}_k^{\Omega_B}(Y)$ estimation in a hurdling sequence.

6.4 Final decision

The final decision, concerning the whole video sequence, is taken by "equivalent" fusion of beliefs at each frame similarly with the Section 5.4. Therefore, at extremum k, the mean mass $\bar{m}_k^{\Omega_B}(Y)$ of the proposition Y is computed. Finally, the decision is taken using the pignistic probability (BetP). Results using just eccentricity feature are illustrated on fourth line of Table 2.

Because of low quality silhouettes, eccentricity is possible to change on running videos. Better results can be achieved by fusion the whole estimated features E_t, TJ and VA (see Table 2). Based on properties of TJ described in Section 6.2, the mass of TJ, $\bar{m}_{TJ}^{\Omega_B}(Y)$, $Y \in \{R, R \cup H, H\}$ is estimated using the beliefs of TJ. H, $R \cup H$ and R correspond to LTJ, $LTJ \cup HTJ$ and HTJ, respectively. Similarly, the mass of VA, $\bar{m}_{VA}^{\Omega_B}(Y), Y \in \{R, R \cup H, H\}$ is estimated using the beliefs of VA. H, $R \cup H$ and R correspond to LVA, $LVA \cup HVA$ and HVA, respectively. We suppose that the above variables are independent. Therefore the pignistic probability of H $(BetP(H))$ is given by:

$$BetP(H) = (\bar{m}_k^{\Omega_B}(H) + \frac{\bar{m}_k^{\Omega_B}(R \cup H)}{2}) \cdot (\bar{m}_{TJ}^{\Omega_B}(H) + \frac{\bar{m}_{TJ}^{\Omega_B}(R \cup H)}{2}) \cdot (\bar{m}_{VA}^{\Omega_B}(H) + \frac{\bar{m}_{VA}^{\Omega_B}(R \cup H)}{2})$$

The above decision rule is equivalent with the selection of proposition H, if $BetP(H) > 0.5$, otherwise R is selected, since $BetP(R) + BetP(H) = 1$. The value of the selected BetP provides a confidence value concerning the final decision.

7 Experimental Results

The method has been implemented using C and Matlab. For experiments, we have been using a Pentium 4 CPU at 2.8 GHz. A typical processing time for the execution of the

Methods	**Ind. Sports**	**Team Sports**
Proposed Scheme	96.9	96.7
Without Quality factor	93.8	97.8
Without $\sqrt{\max(K_i, 1)}$ on (Eq. 8)	95.6	97.8
Without TBM (Thr = 2)	88.8	97.8
Without TBM (Thr = 2.35)	96.9	94.5
Without TBM (Thr = 2.5)	98.1	92.3

Table 1
Accuracy results for team sports and individual sports discrimination under several methods.

proposed scheme is about 8 frames per second.

We have tested the proposed algorithm on a data set containing 252 athletic videos: 161 video sequences from individual sports like pole vault, high jump, triple jump, long jump, shot, javelin and 91 video sequences from team sports like running and hurdling. The database is characterized by its heterogeneity with a panel of view angles as well as unconstrained indoor or outdoor environments (other moving people can be appeared), and athletes (male, female with different skills, skin colors). The most of the videos are in low quality (having resolution 352 x 288) captured from broadcast TV. The number of frames per shots are varied mainly between 100 and 600.

Using the proposed scheme, the accuracy of the team sports detection was 96.9% (156/161) and the accuracy of the individual sports was 96.7% (88/91). In Fig. 10, the bad classified results have $BetP(I) < 0.5$ on individual sport videos or $BetP(T) < 0.5$ on team sport videos. We have performed several tests in order to make comparisons between the proposed scheme versus several variations of this scheme (see Table 1). First, we tested the proposed scheme without using quality criterion (setting quality factor equal to one), getting 93.79% for individual sports and 97.8% for team sports. Next, we tested the proposed scheme without using the $\sqrt{\max(K_i, 1)}$ on Eq. 8, getting similar results. Next, we tested the proposed scheme without using TBM framework, deciding using a threshold (Th_r) on the mean of N_t over the frames, getting about 3% less performance. Conclusively, the aforementioned comparisons show the importance of using quality Q_t, TBM framework and the robustness of the proposed features under several decisions rules.

Figs. 13, 14 show frames from the original sequences and the corresponding results of the proposed scheme. The small black boxes correspond to the mass center detected humans. The people in group are connected with straight lines. In Fig. 13, two athletes are initially appearing in the scene making the method confusing to decide (at first frames). Finally, the camera tracks one of them and the system responds that it was an individual sport video. The belief mass $\hat{m}_t^{\Omega}(X)$ gives an instant decision for a current frame. According to the $\hat{m}_t^{\Omega}(X)$ until the frame 35 multiple people are appearing in the scene, which is very close to the ground truth. The global decision for a period can be taken using $\bar{m}_t^{\Omega}(X)$. According to this mass, after the 70th frame, the video is classified as an individual sport, since it contains more frames of single athlete rather than multiple athletes. The value of quality feature is very close to what a human expert will decide for a quality function, since it is maximized when one athlete is appearing without noise (middle frames of the video). During these frames, the system is able to decide. While it is minimized at the end of the video, where a lot of noise objects are appearing and the human silhouette is segmented into more objects.

Fig. 14 illustrates a 100 m running video. First, the athletes are separated providing high accuracy results to the people counting procedure and high values on Q_t. After the middle of the sequence, a lot of occlusions and bad quality silhouettes are appearing. The occluded athletes correspond to one or two groups of people, and at the same time Q_t has low values. This example shows the accuracy of people counting procedure under several conditions and the usefulness of Q_t in order to be able to give a confidence value about the people counting at each frame. The video is correctly classified into team sports, since at least two people are detected at each frame of the video.

7.1 Running/Hurdling discrimination

The proposed method has been tested on a data set containing 88 athletic videos: 71 videos of running and 17 videos of hurdling. Using the proposed scheme, the accuracy of running detection was 90.2% and the accuracy of hurdling was 82.4%, getting a mean accuracy of 86.3%. The method fails on hurdling detection when the athletes are very far from the camera. On these cases, they appear as small objects in the image plane.

Methods	**Running**	**Hurdling**
Proposed Scheme	90.2	82.4
Using Eccentricity, Area	77.5	88.2
Using Eccentricity, Period	73.3	88.2
Using Eccentricity	59.2	88.2

Table 2
Accuracy results for running and hurdling discrimination under several features.

Thus, the eccentricity variations are very sensitive to silhouette noise and they can not be detected. We have performed several tests in order to make comparisons between the proposed scheme versus several variations of this scheme (see Table 2) and to see the usefulness of the proposed features. According to this table, the feature order concerning their contribution on discrimination, starting from the most important one, is eccentricity, VA and TJ.

Fig. 15 illustrates a hurdling video. At the start of this video, the athletes are synchronized, so the jumping times are easily detected by the variations of E_t. At the end of this video, the athletes are running, without jumping, or jumping without synchronization. Therefore, a short time decision (at these frames) can not be taken using $\hat{m}_k^{\Omega_B}(Y)$. Finally, on this example, H is selected by the method, since $BetP(H)$ is 96.1%.

8 Conclusion and future work

We have proposed a shape based method for unsupervised-automatic people detection and counting applied to athletic videos in order to classify them to videos of individual sports and team sports. Robust, adaptive, independent from the camera motion and well understandable by humans features, are estimated using silhouettes. Finally, the features are combined within the Transferable Belief Model (TBM) framework for video categorization yielding at the same time confidence values about the final decision.

The main contribution of this work concerns the definition of appropriate robust features and the TBM based fusion of them, using a quality function, yielding high performance results without any given feature or initialization under low quality - real conditions videos.

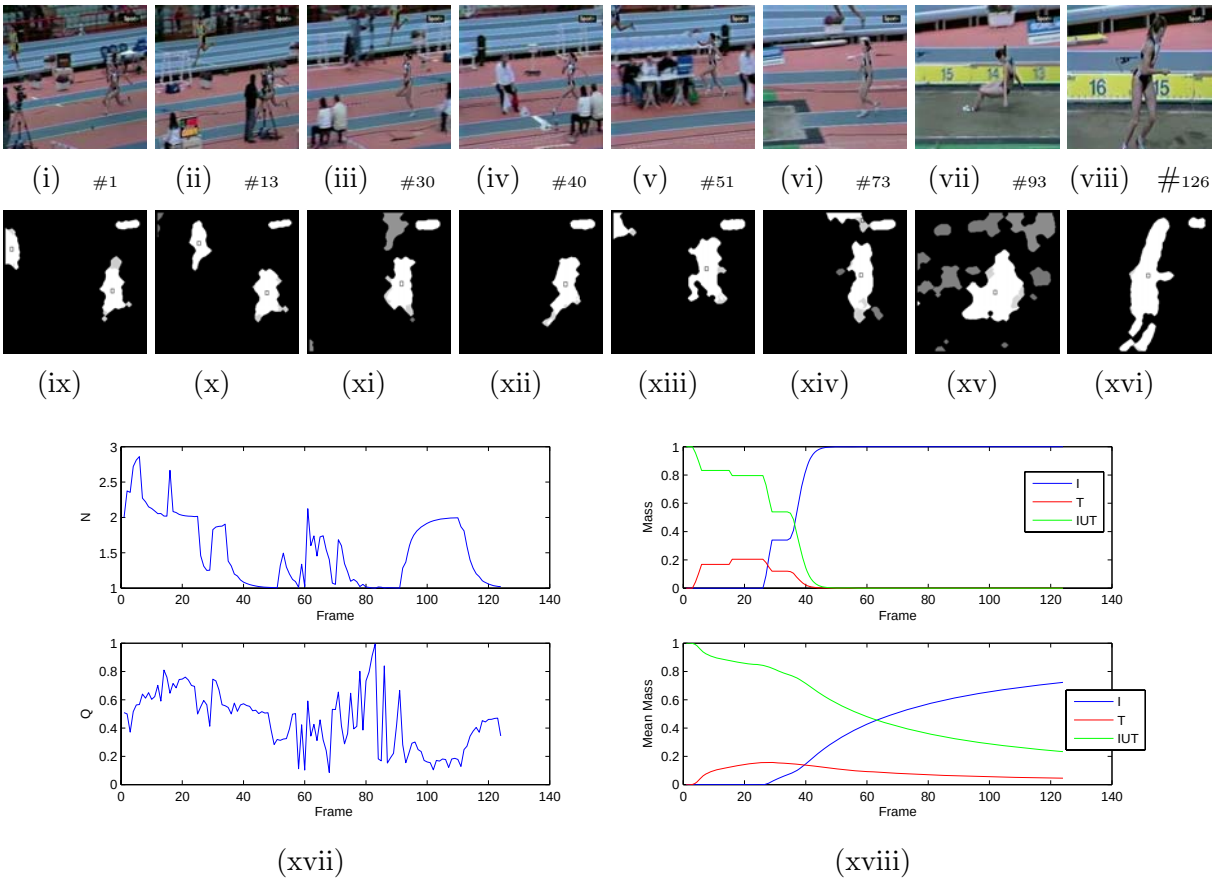


Fig. 13. **(i), ⋯, (viii)** The triple jump original sequence which contains 126 frames. **(ix), ⋯, (xvi)** The results of the people detection and counting procedure. The small black boxes corresponds to the mass center detected humans. **(xvii)** N_t, Q_t. **(xviii)** The belief masses $\hat{m}_t^{\Omega}(X)$, $\bar{m}_t^{\Omega}(X)$.

An instant decision for a current frame can be provided by the mass $\hat{m}_t^{\Omega}(X)$, while the mean mass $\bar{m}_t^{\Omega}(X)$ can be used for the global decision for the whole sequence. Apart from the human detection, this work focuses on activity recognition, analyzing big databases of videos from real and dynamic environments with unconstrained moving camera and one or multiple people under fast and complicated athlete motions.

The proposed scheme has been tested into a big variety of athletic videos like pole vault, high jump, triple jump, long jump, shot, javelin, hurdling and running. The accuracy of individual/team sports categorization was 97% under low quality videos, real unconstrained conditions/environments and the fast/complicated athlete motions. Thus, the used dataset of over than 250 real videos of athletic meetings, under a lot of variations in camera views, sports and motions indicate the stability and the good performance of the proposed scheme.

In the context of video indexing, the proposed method can be improved in order to apply

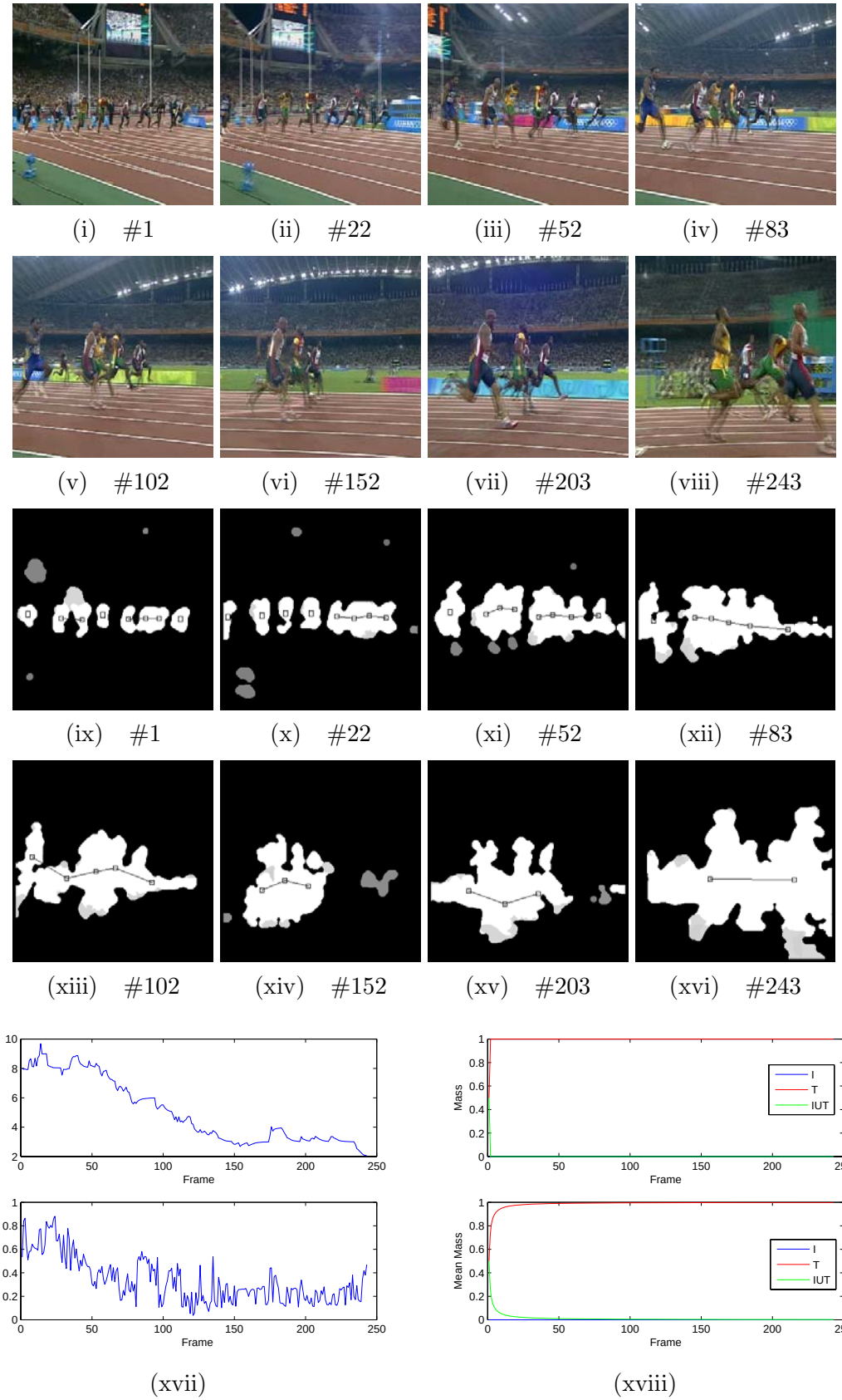
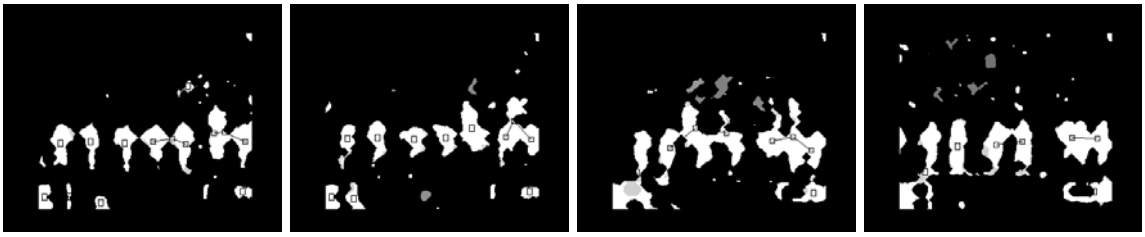


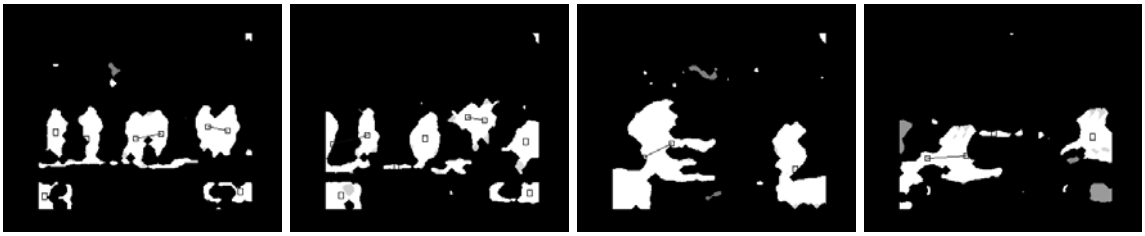
Fig. 14. **(i), ⋯, (viii)** The original running sequence which contains 243 frames. **(ix), ⋯, (xvi)** The results of the people detection and counting procedure. The small black boxes corresponds to the mass center detected humans. A group of people is detected, when the boxes are connected with a line. **(xvii)** N_t, Q_t. **(xviii)** The belief masses $\hat{m}_t^{\Omega}(X)$, $\bar{m}_t^{\Omega}(X)$.

(i) #11 (ii) #24 (iii) #50 (iv) #100

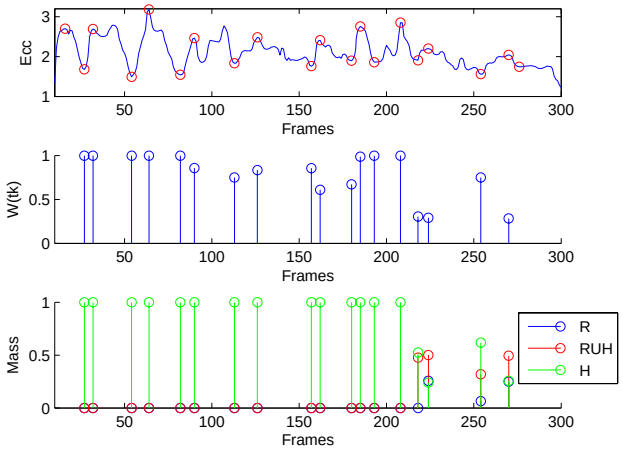
(v) #150 (vi) #200 (vii) #250 (viii) #300



(ix) #11 (x) #24 (xi) #50 (xii) #100



(xiii) #150 (xiv) #200 (xv) #250 (xvi) #300



(xvii)

Fig. 15. **(i), ⋯, (viii)** The original running sequence which contains 300 frames. **(ix), ⋯, (xvi)** The results of the people detection and counting procedure. The small black boxes corresponds to the mass center detected humans. A group of people is detected, when the boxes are connected with a line. **(xvii)** E_t (the times τ_k are shown with red cicles), $W(t_k)$ and the belief mass $\hat{m}_k^{\Omega_B}(Y)$.

it on other types of videos. In particular, the trapezes as well as tables of rules should be estimated and even adapted according to the type of videos.

Acknowledgments

This work is partially supported by SIMILAR European Network of Excellence and by the Greek PENED 2003 project.

References

[1] M. Lew, N. Sebe, J. Eakins, Challenges in image and video retrieval, Lecture notes in Computer Science, ICIVR 2383 (2002) 1–6.

[2] J. Alon, S. Sclaroff, G. Kollios, V. Pavlovic, Discovering clusters in motion time-series data, in: Proc. IEEE Computer Vision and Pattern Recognition Conference, 2003.

[3] W. Hu, T. Tan, L. Wang, S. Maybank, A survey on visual surveillance of object motion and behaviors, IEEE Trans. on Systems, man and cybernetics C 34 (3).

[4] A. Jaimes, N. Sebe, Multimodal human computer interaction: A survey, in: IEEE Int. Workshop on Human Computer Interaction in conjunction with ICCV, Vol. 3766, Beijing, China, 2005, pp. 1–15.

[5] P. Smets, R. Kennes, The Transferable Belief Model, Artificial Intelligence 66 (2) (1994) 191–234.

[6] G. Shafer, A mathematical theory of evidence, Princeton University Press, Princeton, NJ, 1976.

[7] P. Smets, Advances in the Dempster-Shafer Theory of Evidence - What is Dempster-Shafer's model ?, r.r. yager and m. fedrizzi and j. kacprzyk Edition, Wiley, 1994, pp. 5–34.

[8] E. Ramasso, D. Pellerin, C. Panagiotakis, M. Rombaut, G. Tziritas, W. Lim, Spatio-temporal information fusion for human action recognition in videos, in: 13th European Signal Processing Conf., 2005.

[9] E. Ramasso, M. Rombaut, D. Pellerin, A Temporal Belief Filter improving human action recognition in videos, in: IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP), Vol. 2, 2006, pp. 141–144.

[10] E. Ramasso, D. Pellerin, M. Rombaut, Belief Scheduling for the recognition of human action sequence, in: Proc. of the 9th Int. Conf. on Information Fusion (ICIF), Florence, Italia, 2006.

[11] C. Panagiotakis, E. Ramasso, G. Tziritas, M. Rombaut, D. Pellerin, Shape-motion based athlete tracking for multilevel action recognition, in: Proc. of AMDO 2006, 2006, pp. 385–394.

[12] C. Panagiotakis, G. Tziritas, Recognition and tracking of the members of a moving human body, in: Proc. of AMDO 2004, 2004, pp. 86–98.

[13] C. Panagiotakis, I. Grinias, G. Tziritas, Automatic human motion analysis and action recognition in athletics videos, in: European Signal Processing Conference, 2006.

[14] T. Moeslund, E. Granum, A survey of computer vision-based human motion capture, Computer Vision and Image Undertanding 81 (2001) 231–268.

[15] I. Haritaoglu, D. Harwood, L. Davis, W4: Real-time surveillance of people and their activities, IEEE Trans. on Pattern Analysis and Machine Intelligence 22 (8) (2000) 809–830.

[16] J. Odobez, P. Bouthemy, Robust multiresolution estimation of parametric motion models, J. of Vis. Comm. and Image R. 6 (4) (1995) 348–365.

[17] P. Smets, Decision making in the TBM: the necessity of the pignistic transformation, Int. Jour. of Approximate Reasoning 38 (2005) 133–147.

[18] L. Wang, W. Hu, T. Tan, Recent developments in human motion analysis, Pattern Recognition 36 (3) (2003) 585–601.

[19] L. Wang, H. Ning, T. Tan, Fusion of static and dynamic body biometrics for gait recognition, IEEE Trans. Circuits Syst. Video Techn. 14 (2) (2004) 149–158.

[20] K. Cheung, S. Baker, T. Kanade, Shape-from-silhouette across time: Part ii: Applications to human modeling and markerless motion tracking, Int. Journal of Computer Vision 63 (3) (2005) 225–245.

[21] B. Shoushtarian, H. E. Bez, A practical adaptive approach for dynamic background subtraction using an invariant colour model and object tracking, Pattern Recognition Letters 26 (1) (2005) 5–26.

[22] P. KaewTrakulPonga, R. Bowden, A real time adaptive visual surveillance system for tracking low-resolution colour targets in dynamically changing scenes, Image and Vision Computing 21 (2003) 913–929.

[23] A. Lipton, H. Fujiyoshi, R. Patil, Moving target classification and tracking from real-time video, in: Image Understanding Workshop (IUW98), 1998.

[24] N. Paragios, R. Deriche, Geodesic active contours and level sets for the detection and tracking of moving object, IEEE Trans. on Pattern Analysis and Machine Intelligence 22 (3) (2000) 266–280.

[25] H. Wang, D. Suter, A consensus-based method for tracking: Modelling background scenario and foreground appearance, Pattern Recognition, to appear.

[26] I. Grinias, G. Tziritas, Robust pan, tilt and zoom estimation, in: Int. Conf. on Digital Signal Processing, 2002, pp. 679–682.

[27] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, H. Wechsler, Tracking groups of people, Computer Vision and Image Understanding: CVIU 80 (1) (2000) 42–56.

[28] P. J. Figueroa, N. J. Leite, R. M. L. Barros, Tracking soccer players aiming their kinematical motion analysis, Computer Vision and Image Understanding: CVIU 101 (2) (2006) 122–135.

[29] A. Mittal, L. Davis, M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene using region-based stereo., in: In Proc. 7th European Conf. Computer Vision, 2002.

[30] V. Rabaud, S. Belongie, Counting crowded moving objects, in: 18th IEEE International Conference on Computer Vision and Pattern Recognition, 2006.

[31] X. Liu, P. H. Tu, J. Rittscher, A. Perera, N. Krahnstoever, Detecting and counting people in surveillance applications., IEEE Conference on Advanced Video and Signal Based Surveillance (2005) 306 – 311.

[32] C. Sacchi, G. Gera, L. Marcenaro, C. S. Regazzoni, Advanced image-processing tools for counting people in tourist site-monitoring applications, Signal Process. 81 (5) (2001) 1017–1040.

[33] R. Rad, M. Jamzad, Real time classification and tracking of multiple vehicles in highways, Pattern Recognition Letters 26 (10) (2005) 1597–1607.

[34] F.-H. Cheng, Y.-L. Chen, Real time multiple objects tracking and identification based on discrete wavelet transform, Pattern Recognition 39 (6) (2006) 1126–1139.

[35] P. Smets, The transferable belief model and other interpretations of dempster-shafer's model, in: P. Bonissone, M. Henrion, L. Kanal, J. Lemmer (Eds.), Uncertainty in Artificial Intelligence, Vol. 6, Elsevier Science, 1991, pp. 375–383.

[36] P. Smets, Imperfect information : Imprecision - uncertainty, in: A. Motro, P. Smets (Eds.), Uncertainty management in information systems, from Needs to solutions, Kluwer Academic, 1997, pp. 225–254.

[37] P. Smets, Data fusion in the Transferable Belief Model, in: Proc. Third Int. Conf. on Information Fusion, 2000, pp. 21–33.

[38] T. Denoeux, P. Smets, Classification using belief functions: the relationship between the case-based and model-based approaches, To appear in IEEE Trans. on Systems, Man and Cybernetics.

[39] G. Klir, M. Wierman, Uncertainty-based information. Elements of generalized information theory, 2nd edition, Studies in fuzzyness and soft computing, Physica-Verlag, 1999.

[40] B. Ristic, P. Smets, Target identification using belief functions and implication rules., IEEE Trans. Aerospace and Electronic Systems 41 (3) (2005) 1097–1102.

[41] D. Dubois, H. Prade, Representation and combination of uncertainty with belief functions and possibility measures, Computational Intelligence 4 (1988) 244–264.

[42] P. Smets, Decision making in the tbm: the necessity of the pignistic transformation., Int. J. Approx. Reasoning 38 (2) (2005) 133–147.