J Math Imaging Vis (2010) 36: 159–167
DOI 10.1007/s10851-009-0178-6

A Particle Filtering and DSmT Based Approach for Conflict Resolving in case of Target Tracking with Multiple Cues

Yi Sun · Layachi Bentabet

Published online: 7 November 2009

Abstract In this paper, we propose an efficient and robust method for multiple targets tracking in cluttered scenes using multiple cues. Our approach combines the use of Monte Carlo sequential filtering for tracking and Dezert-Smarandache theory (DSmT) to integrate the information provided by the different cues. The use of DSmT provides the necessary framework to quantify and overcome the conflict that might appear between the cues due to the occlusion. Our tracking approach is tested with color and location cues on a cluttered scene where multiple targets are involved in partial or total occlusion.

Keywords Multiple targets tracking · Sequential Monte-Carlo · Dezert-Smarandache theory · Cue combination

1 Introduction

Visual tracking of moving targets has become one of the primary research issues in computer vision, due to the increasing demand for reliable activity monitoring and surveillance systems [1–3]. In order to achieve accurate tracking, several cues have been explored in the literature. These cues include color histogram, edges, motion, camera geometry (field of view), and velocity. As pointed out by [4], individual cues can potentially fail or provide paradoxical interpretations due to the occlusion in cluttered scenes, and the changes in the illumination which are unavoidable in the real world. Numerous methods suggest the use of multiple cues to increase visual robustness in cases of complex scenes [5, 6]. The integration of the extracted cues into an object representation has been performed using probabilistic methods [4–7] as well as non-probabilistic methods [8–11]. Given the prior distributions and the conditional probabilities, probabilistic methods such as the Bayesian inference offer the most complete, scalable and theoretically justifiable approach for data fusion. However, in real complex scenes such complete knowledge is difficult to obtain due to occlusion, background clutter, illumination and camera calibration problems. In [12], an extension of the probabilistic approach is proposed for data fusion; namely the Dempster-Shafer (DST) theory. The uncertainty and imprecision of a body of knowledge are represented via the notion of confidence values that are committed to a single or a union of hypotheses. The orthogonal sum rule of DST theory allows the integration of information from different sources into a single and overall representation. Unfortunately, Bayesian inference and Dempster-Shafer theories lack to provide an interesting manner of modeling conflicts and paradoxical interpretation arising between the difference information sources. The Bayesian inference assumes that all sources provide bodies of evidences using the same objective and universal interpretation of the phenomena under consideration; therefore, it cannot handle conflicts [13]. In most practical fusion applications based on DST theory, ad-hoc or heuristic techniques must always be added to the fusion process to manage or reduce the possibility of high degree conflict between the sources. Otherwise, the fusion result leads to false conclusions or cannot provide a reliable result at all. To overcome these limitations, a recent theory of reasoning with plausible and paradoxical cues has been developed in [13]. The Dezert-Smarandache Theory (DSmT)

Y. Sun · L. Bentabet (✉)
Department of Computer Science, Bishop's University,
2600 College Street, Sherbrooke, QC, J1M 1Z7, Canada
e-mail: lbentabe@ubishops.ca

Y. Sun
e-mail: ysun@ubishops.ca

can be considered as a generalization of the DST theory. In this paper, we propose a sequential particle filter approach for multiple targets tracking using multiple cues. The different cues are combined using the DSmT theory. If the targets are partially or completely occluded, the conflicts and paradoxes that arise between the cues are assessed and used in the tracking process. The proposed scheme is simple and provides an effective tracking in cluttered scenes.

2 The Dezert-Smarandache Theory

The Dezert-Smarandache Theory (DSmT) of plausible and paradoxical reasoning [13] is a generalization of the classical Dempster-Shafer theory (DST), which allows formal combining of rational, uncertain and paradoxical sources. The DSmT is able to solve complex fusion problems where the DST usually fails, especially when conflicts between sources become large. In this section, we will first review the principle of the DST before discussing the fundamental aspects of the DSmT.

2.1 Principle of Dempster-Shafer Theory

The DST makes inferences from incomplete and uncertain knowledge by combining additional sources of confidence, even in the process of partially contradictory sensors. The DST contains the Bayesian theory of partial belief as a special case. In the DST, there is a fixed set of mutually exclusive and exhaustive elements, called the frame of discernment, which is symbolized by $\Theta = \{\theta_1, \theta_2, \ldots, \theta_N\}$. The frame of discernment Θ defines the propositions for which the sources can provide confidence. Information sources can distribute mass values on subsets of the frame of discernment, $A_i \in 2^\Theta$. If an information source can not distinguish between two propositions A_i and A_j, it assigns a mass value to the set including both hypotheses $(A_i \cup A_j)$. The mass distribution for all hypotheses has to fulfill the following conditions

$$\begin{aligned} &0 \le m(A_i) \le 1 \\ &m(\phi) = 0 \\ &\sum_{A_i \in 2^\Theta} m(A_i) = 1 \end{aligned} \tag{1}$$

Mass functions $m_1, m_2, \ldots, m_d$ from d different sources are combined with Dempster's orthogonal rule. The result is new distribution, $m = m_1 \oplus m_2 \oplus \cdots \oplus m_d$, which carries the joint information provided by the d sources

$$m(A) = (1-K)^{-1} \sum_{\substack{A_1,\ldots,A_d \in 2^\Theta \\ A_1 \cap \cdots \cap A_d = A}} \left[\prod_{i=1}^{d} m_i(A_i) \right], \quad \text{for } A \neq \phi \tag{2}$$

where

$$K = m(\phi) = \sum_{\substack{A_1,\ldots,A_d \in 2^\Theta \\ A_1 \cap \cdots \cap A_d = \phi}} \left[\prod_{i=1}^{d} m_i(A_i) \right]$$

K is a measure of conflict between the sources and it is introduced as a normalization factor. The larger K is, the more the sources are conflicting and the less sense the combination has. Two functions can be evaluated to characterize the uncertainty about the hypotheses A. The belief function, *Bel*, measures the minimum uncertainty value about A, whereas, the plausibility function, *Pls*, reflects the maximum uncertainty value. Belief and plausibility functions are defined from 2^Θ to [0, 1]

$$Bel(A) = \sum_{\substack{A_i \in 2^\Theta \\ A_i \subseteq A}} m(A_i), \qquad Pls(A) = \sum_{\substack{A_i \in 2^\Theta \\ A_i \cap A \neq \phi}} m(A_i) \tag{3}$$

These measures, which have been sometimes referred to as lower and upper probability functions, have the following properties

$$Pls(A) = 1 - Bel(\neg A)$$

$$Bel(A) \le Pls(A)$$

where $\neg A$ is the complementary of A.

2.2 The Dezert-Smarandache Theory (DSmT)

While the DST considers Θ as a set of exclusive elements, the DSmT relaxes this condition and allows for overlapping and intersecting hypotheses. This allows for quantifying the conflict that might arise between the different sources throughout the assignment of non-null confidence values to the intersection of distinct hypotheses.

Let $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$ be a set of n elements which can potentially overlap. The hyper-power set D^Θ is defined as the set of all composite hypotheses obtained from Θ with $\cup$ and $\cap$ operators such that

1. $\phi, \theta_1, \theta_2, \ldots, \theta_n \in D^\Theta$.
2. If $A, B \in D^\Theta$, then $(A \cup B) \in D^\Theta$ and $(A \cap B) \in D^\Theta$.
3. No other elements belong to D^Θ, except those defined in 1 and 2.

As pointed out by F. Smarandache and Dezert [13], the cardinality of D^Θ is majored by 2^{2^n} when the cardinality of Θ equals n. The generation of the hyper-power set is closely related to the Dedekind's problem of enumerating isotone Boolean functions [14]. Since for any given finite set Θ, $|D^\Theta| \ge |2^\Theta|$, D^Θ is called the hyper-power set of Θ.

As in the DST, the DSmT defines a map $m(.) : D^\Theta \to [0, 1]$, which assigns to each hypothesis A in D^Θ a mass

function $m(A)$ that satisfies the conditions expressed in (1). The belief and plausibility functions are defined in the same way as for the DST

$$Bel(A) = \sum_{\substack{A_i \in D^\Theta \\ A_i \subseteq A}} m(A_i), \qquad Pls(A) = \sum_{\substack{A_i \in D^\Theta \\ A_i \cap A \neq \phi}} m(A_i) \quad (4)$$

The DSmT rule of combination of conflicting and on uncertain sources is given by:

$$m(A) = \sum_{\substack{A_1, A_2, \ldots, A_d \in D^\Theta \\ A_1 \cap A_2 \cap \cdots \cap A_d = A}} \prod_{i=1}^{d} m_i(A_i) \quad (5)$$

To show the difference between the DST and the DSmT combination rules, let's consider an example where $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$, and two independent sensors provide the following mass functions:

$$m_1(\theta_1) = 0.7; \qquad m_1(\theta_2) = 0.3$$

$$m_2(\theta_3) = 0.2; \qquad m_2(\theta_4) = 0.8$$

- The DST rule of combination cannot be applied because: $\forall 1 \leq j \leq 4 : m(\theta_j) = 0/0$.
- The DSmT rule of combination gives: $m(\theta_1) = m(\theta_2) = m(\theta_3) = m(\theta_4) = 0$, and $m(\theta_1 \cap \theta_3) = 0.14$, $m(\theta_1 \cap \theta_4) = 0.56$, $m(\theta_2 \cap \theta_3) = 0.06$ and $m(\theta_2 \cap \theta_4) = 0.24$. Both plausibility and belief measures are in favor of θ_4 which would be the chosen hypothesis.

The previous example shows a typical case where the classical DST theory fails to make a decision in a case where two sources are in a conflicting situation; whereas, the DSmT provides the necessary framework to deal with the paradoxical inputs and to provide the most plausible decision.

3 Sequential Monte Carlo

Sequential Monte Carlo techniques, also know as particle filtering, were introduced to track multiple objects in cluttered scenes [15]. In the following, $X_t = (x_1, x_2, \ldots, x_t)$ is a first order Markov process that describes the state vector (location, size, etc.) of the target and $Z_t = (z_1, z_2, \ldots, z_t)$ is the vector of measurements (color, texture, etc.) up to time t. The tracking is based on the estimation of the posterior state distribution $p(x_t|Z_t)$ at each time step. The estimation is performed using a two step Bayesian recursion [5, 16]. The first step is prediction,

$$p(x_t|Z_{t-1}) \propto \int p(x_t|x_{t-1}) p(x_{t-1}|Z_{t-1}) dx_{t-1} \quad (6)$$

and the second step is filtering

$$p(x_t|Z_t) \propto p(z_t|x_t) p(x_t|Z_{t-1}) \quad (7)$$

This recursion requires the specification of the state evolution $p(x_t|x_{t-1})$ and a measurement model linking the state and the current measurement $p(z_t|x_t)$. The basic idea behind the particle filter is very simple. Starting with a weighted set of samples

$$S_{t-1} = \left\{ s_{t-1}^{(n)}, \pi_{t-1}^{(n)} \middle| \sum_{n=1}^{N} \pi_{t-1}^{(n)} = 1 \right\} \quad (8)$$

representing target candidates and distributed according to $p(x_{t-1}|Z_{t-1})$, where $x_{t-1}^{(n)}$ is the state vector of sample n at time $t-1$. Similarly, $z_{t-1}^{(n)}$ is the measurement vector of sample n at time $t-1$. In the prediction step new samples are obtained by propagating each sample according to the target's state model, $p(x_t|x_{t-1})$. In the filtering step, each sample is weighted given the observation and N samples are drawn with replacement according to $\pi_t = p(z_t|x_t)$. In order to achieve a smooth tracking, the target's estimate is defined as the weighted average of the particle set, i.e.

$$E[S_t] = \sum_{n=1}^{N} \pi_t^{(n)} s_t^{(n)}$$

Particle filtering has proven to be very successful for nonlinear and non-Gaussian estimation problems. The tracking iteration using particle filtering can be summarized as follows.

Step 1: *Select* N samples

$$S_{t-1} = \left\{ s_{t-1}^{(n)}, \pi_{t-1}^{(n)} \middle| \sum_{n=1}^{N} \pi_{t-1}^{(n)} = 1 \right\} \quad (9)$$

Step 2: *Propagate* each sample according to

$$s_t^{(n)} = H \cdot s_{t-1}^{(n)} + w_{t-1}^{(n)} \quad (10)$$

where H is a square matrix defining the deterministic component of the target's motion model and $w_{t-1}^{(n)}$ is the random component of the target's motion model.

Step 3: *Observe* the samples and evaluate $\pi_t^{(n)} = p(z_t^{(n)}|x_t)$.

Step 4: *Estimate* the mean state of S_t

$$E[S_t] = \sum_{n=1}^{N} \pi_t^{(n)} s_t^{(n)} \quad (11)$$

4 DSmT-Based Tracking

Let's assume that the number of targets, τ, and the number of cues, c, are known. Up to time $t-1$, each target is associated with a track $\{\theta_j\}_{j=1}^{\tau}$. At time t, an image frame is

extracted from the video sequence and a number of measurements are obtained for each target candidate. Thus, the objective is to combine these measurements in order to determine the best track for each candidate. It is important to notice that a target candidate, in this paper, refers to a particle sample. The hyper-power set D^{Θ} defines the set of the hypotheses for which the different cues can provide confidence values. These hypotheses can correspond to: (1) individual tracks θ_j, (2) union of tracks $\theta_r \cup \cdots \cup \theta_s$, which symbolizes ignorance, (3) intersection of tracks $\theta_r \cap \cdots \cap \theta_s$, which symbolizes conflict or (4) any tracks combination obtain by $\cup$ and $\cap$ operators. The confidence level is expressed in terms of mass functions $\{m_{t,l}^{(n)}(.)\}_{l=1}^{c}$ that are committed to each hypothesis and which satisfy the condition in (1). Given this framework, $m_{t,l}^{(n)}(A)$ express the confidence with which cue l associates particle n to hypothesis A at time t. According to DSmT combinational rule in (5), a single map function $m_t^{(n)}(.)$ can be derived as follows

$$m_t^{(n)}(A) = m_{t,1}^{(n)} \oplus m_{t,2}^{(n)} \oplus \cdots \oplus m_{t,c}^{(n)}(A) \tag{12}$$

where $m_t^{(n)}(A)$ is the overall confidence level with which all cues associate particle n to hypothesis A at time t. Since the target candidates are associated with individual tracks, only single hypotheses (i.e. single tracks) are considered for decision making which is done using the notions of the belief or plausibility functions

$$\begin{aligned} Bel_t^{(n)}(\theta_j) &= \sum_{\substack{A \in D^{\Theta} \\ \theta_i \subseteq A}} m_t^{(n)}(A) \\ Pls_t^{(n)}(\theta_j) &= \sum_{\substack{A \in D^{\Theta} \\ \theta_i \cap A \neq \phi}} m_t^{(n)}(A) \end{aligned} \tag{13}$$

where $Bel_t^{(n)}(\theta_j)$ (resp. $Pls_t^{(n)}(\theta_j)$) quantifies the confidence with which particle n is associated to θ_j at time t using the notion of belief (resp. plausibility). The confidence levels in (13) are not used to determine whether a given a candidate is the best estimate or not of the target, they are rather used to quantify the weight of the candidate (or particle) as a sample of the state posterior distribution $p(x_t|Z_t)$. The DSmT-based particle filtering algorithm implemented in this paper is given below.

Step 1: *Initialization*

- Generate N samples $S_{t,j} = \{s_{t,j}^{(n)}, \pi_{t,j}^{(n)}\}_{n=1}^{N}$ for each target $j = 1, \ldots, \tau$ independently, with $\pi_{t,j}^{(n)} = \frac{1}{N}$.
- Set $t = 1$.

Step 2: *Propagation*

- $s_{t^*,j}^{(n)} = H \cdot s_{t-1,j}^{(n)} + w_{t-1,j}^{(n)}$

Step 3: *Observation* (for each particle)

- Compute $\{m_{t^*,l}^{(n)}(A)\}_{l=1}^{c}$ for $A \in D^{\Theta}$
- Compute $m_{t^*}^{(n)}(A)$ for $A \in D^{\Theta}$ according to (5)
- Calculate the particle weight $\pi_{t^*,j}^{(n)} = Bel_{t^*}^{(n)}(\theta_j)$ (or $\pi_{t^*,j}^{(n)} = Pls_{t^*}^{(n)}(\theta_j)$)
- Normalize the weight: $\tilde{\pi}_{t^*,j}^{(n)} = \frac{\pi_{t^*,j}^{(n)}}{\sum_{n=1}^{N} \pi_{t^*,j}^{(n)}}$

Step 4: *Estimation*

- Target $j = 1, \ldots, \tau$ is given by $E[S_{t^*,j}] = \sum_{n=1}^{N} \tilde{\pi}_{t^*,j}^{(n)} s_{t^*,j}^{(n)}$.

Step 5: *Resampling* (for each target)

- Generate $S_{t,j} = \{s_{t,j}^{(n)}, \pi_{t,j}^{(n)}\}_{n=1}^{N}$ by resampling N times from $S_{t^*,j}$ where $p(s_{t,j}^{(n)} = s_{t^*,j}^{(m)}) = \tilde{\pi}_{t^*,j}^{(m)}$.

Step 6: *Incrementing*

- $t = t + 1$, go to step 2.

5 Tracking Multiple Target Using Location and Color

In the following, the particle filtering approach described in the previous section is applied to perform multiple target tracking using color and location. These cues are combined using the DSmT rule to resolve the conflict that might arise due to occlusion. Color cue is continuously measured from the scene using an RGB video camera. The location cue is not directly measured from the scene. Indeed, we consider the target's estimated location at time $t-1$ as the best available indication of where the target should be at time t. This implies that the targets are not moving fast.

The framework described above does not provide the optimal condition where the information to be fused originate from multiple sensors. However, it leads to a conflict between the location and the color cues at occlusion time. This is actually the type of tracking scenarios which will allow us to show how our approach resolves the conflict in order to achieve an accurate tracking in cluttered scenes.

For the sake of simplicity, let's assume that the scene contains two targets. we can define D^{Θ} as follows

$$D^{\Theta} = \{\theta_1, \theta_2, \overline{\theta_1 \cup \theta_2}\} \tag{14}$$

In (14), θ_1 refers the first target, θ_2 refers to the second target and $\overline{\theta_1 \cup \theta_2}$ refers to the rest of the scene. Actually, hypothesis $\overline{\theta_1 \cup \theta_2}$ can refer to the background information. However, since this latter can change during the tracking, we will refer to $\overline{\theta_1 \cup \theta_2}$ as the false alarm hypothesis. Beside, $\theta_1 \cap \theta_2 \neq \phi$ due to the possible occlusion, and $\theta_j \cap \overline{\theta_1 \cup \theta_2} = \phi$ for $j = 1, 2$.

5.1 The Location Cue

The estimate of the targets' locations at time $t-1$ is given by $(x_{t-1,1}, y_{t-1,1})$ and $(x_{t-1,2}, y_{t-1,2})$. The likelihood that a particle $s_{t,j}^{(n)}$, located at $(x_{t,j}^{(n)}, y_{t,j}^{(n)})$ at time t, belongs to target $j=1,2$ is defined according to the location cue as a Gaussian p.d.f.

$$p_{t,j}^{(n)} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_{t,j}^{(n)} - x_{t-1,j})^2 + (y_{t,j}^{(n)} - y_{t-1,j})^2}{2\sigma^2}} \tag{15}$$

where σ is a bandwidth parameter. Similarly, the likelihood that a given particle does not belong to θ_1 and θ_2 is inversely proportional to the distance between the particle and both targets. Since D^{Ω} is exhaustive, a particle that does not belong to θ_1 and θ_2 do belong to $\overline{\theta_1 \cup \theta_2}$. This leads us to the definition of a new p.d.f., $p_{t,FA}^{(n)}$, which measures the likelihood that a particle $n=1,\ldots,N$ is a false alarm hypothesis

$$p_{t,FA}^{(n)} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(d_{\max} - d_{1-2}^{(n)})^2}{2\sigma^2}} \tag{16}$$

where σ is the bandwidth parameter, $d_{\max}$ is the radius of a circle centered on the mid-point of targets 1 and 2, and which contains all the particles used for tracking at the time $t-1$, $d_{1-2}^{(n)}$is the distance separating particle n and the mid-point

$$d_{1-2}^{n} = \sqrt{\left(x_{t,1}^{(n)} - \frac{x_{t-1,1} - x_{t-1,2}}{2}\right)^2 + \left(y_{t,1}^{(n)} - \frac{y_{t-1,1} - y_{t-1,2}}{2}\right)^2} \tag{17}$$

The mass functions of particle n according to its location are given by

$$m_{t,1}^{(n)}(\theta_j) = \frac{p_{t,j}^{(n)}}{p_{t,1}^{(n)} + p_{t,2}^{(n)} + p_{t,FA}^{(n)}}, \quad j=1,2 \tag{18}$$

$$m_{t,1}^{(n)}(\overline{\theta_1 \cup \theta_2}) = \frac{p_{t,FA}^{(n)}}{p_{t,1}^{(n)} + p_{t,2}^{(n)} + p_{t,FA}^{(n)}} \tag{19}$$

$$m_{t,1}^{(n)}(\phi) = 0 \tag{20}$$

5.2 The Color Cue

Let's assume that both target models are known and given by normalized color histograms $\{q_j(u)\}_{u=1}^{m}$, where u is a discrete color index and m is the number of histogram bins. The normalized color histogram $\{h_{t,j}^{(n)}(u)\}_{u=1}^{m}$ of particle $s_{t,j}^{(n)}$ is calculated from frame t of the image sequence. The likelihood that particle $s_{t,j}^{(n)}$ belongs to target $j=1,2$ according to the color histogram is derived from the following Gaussian p.d.f.

$$p_{t,j}^{(n)} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(d_{t,j}^{(n)})^2}{2\sigma^2}}, \quad j=1,2 \tag{21}$$

where σ is a color bandwidth parameter, $d_{t,j}^{(n)}$is the Bhattacharyya distance between $h_{t,j}^{(n)}(u)$ and $q_j(u)$ at time t

$$d_{t,j}^{(n)} = \sqrt{1 - \sum_{u=1}^{m} h_{t,j}^{(n)}(u) q_j(u)} \tag{22}$$

Let's define $\{q_{FA}(u)\}_{u=1}^{m}$ as the histogram of the scene from which we subtract the histogram of targets 1 and 2

$$q_{FA}(u) = \max\{q_{scene}(u) - q_1(u) - q_2(u), 0\} \tag{23}$$

The likelihood that $s_{t,j}^{(n)}$ belongs to the false alarm hypothesis will be given by

$$p_{t,FA}^{(n)} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(d_{t,FA}^{(n)})^2}{2\sigma^2}} \tag{24}$$

where

$$d_{t,FA}^{(n)} = \sqrt{1 - \sum_{u=1}^{m} h_{t,j}^{(n)}(u) q_{FA}(u)} \tag{25}$$

The mass functions of particle n according to color can be evaluated as follows

$$m_{t,2}^{(n)}(\theta_j) = \frac{p_{t,j}^{(n)}}{p_{t,1}^{(n)} + p_{t,2}^{(n)} + p_{t,FA}^{(n)}}, \quad j=1,2 \tag{26}$$

$$m_{t,2}^{(n)}(\overline{\theta_1 \cup \theta_2}) = \frac{p_{t,FA}^{(n)}}{p_{t,1}^{(n)} + p_{t,2}^{(n)} + p_{t,FA}^{(n)}} \tag{27}$$

$$m_{t,2}^{(n)}(\phi) = 0 \tag{28}$$

5.3 The Cue Combination

The combination rule leads to the mass functions $m_t^{(n)}(.)$ given in Table 1 where

$$m_t^{(n)}(\theta_1) = m_{t,1}^{(n)}(\theta_1) \cdot m_{t,2}^{(n)}(\theta_1) \tag{29}$$

$$m_t^{(n)}(\theta_2) = m_{t,1}^{(n)}(\theta_2) \cdot m_{t,2}^{(n)}(\theta_2) \tag{30}$$

$$m_t^{(n)}(\theta_1 \cap \theta_2) = m_{t,1}^{(n)}(\theta_1) \cdot m_{t,2}^{(n)}(\theta_2) + m_{t,2}^{(n)}(\theta_1) \cdot m_{t,1}^{(n)}(\theta_2) \tag{31}$$

$$m_t^{(n)}(\overline{\theta_1 \cup \theta_2}) = m_{t,1}^{(n)}(\overline{\theta_1 \cup \theta_2}) \cdot m_{t,2}^{(n)}(\overline{\theta_1 \cup \theta_2}) \tag{32}$$

Table 1 Combination rule for color and location

Location cue	Color cue		
	$m_{t,2}^{(n)}(\theta_1)$	$m_{t,2}^{(n)}(\theta_2)$	$m_{t,2}^{(n)}(\overline{\theta_1 \cup \theta_2})$
$m_{t,1}^{(n)}(\theta_1)$	$m_t^{(n)}(\theta_1)$	$m_t^{(n)}(\theta_1 \cap \theta_2)$	$m_t^{(n)}(\phi)$
$m_{t,1}^{(n)}(\theta_2)$	$m_t^{(n)}(\theta_1 \cap \theta_2)$	$m_t^{(n)}(\theta_2)$	$m_t^{(n)}(\phi)$
$m_{t,1}^{(n)}(\overline{\theta_1 \cup \theta_2})$	$m_t^{(n)}(\phi)$	$m_t^{(n)}(\phi)$	$m_t^{(n)}(\overline{\theta_1 \cup \theta_2})$

$$m_t^{(n)}(\phi) = m_{t,1}^{(n)}(\overline{\theta_1 \cup \theta_2})(m_{t,2}^{(n)}(\theta_1) + m_{t,2}^{(n)}(\theta_2)) + m_{t,2}^{(n)}(\overline{\theta_1 \cup \theta_2})(m_{t,1}^{(n)}(\theta_1) + m_{t,1}^{(n)}(\theta_2)) \quad (33)$$

Equation (29) (respectively (30)) is the confidence level with which both cues associate $s_{t,j}^{(n)}$ to target 1 (respectively target 2). Equation (31) is the conflict value between the cues for the membership of $s_{t,j}^{(n)}$ to target 1 or target 2. Equation (32) expresses the confidence value with which both cues agree that the particle corresponds to a false alarm. Equation (33) quantifies the conflict between the targets and the false alarm hypothesis.

For the particular case of tracking with two targets using the frame of discernment in (14), the plausibility and belief measures are identical. In what follows, they are used to calculate the weight of particle $s_{t,j}^{(n)}$:

$$\pi_{t,j}^{(n)} = Pls_t^{(n)}(\theta_j) = Bel_t^{(n)}(\theta_j) = m_t^{(n)}(\theta_j) + m_t^{(n)}(\theta_1 \cap \theta_2), \quad j = 1, 2 \quad (34)$$

The generalization of the tracking scheme described in this section to τ targets can be carried out by defining a frame of discernment

$$\Theta = \{\theta_1, \ldots, \theta_\tau, \overline{\theta_1 \cup \cdots \cup \theta_\tau}\}, \quad (35)$$

where θ_j are individual targets and $\overline{\theta_1 \cup \cdots \cup \theta_\tau}$ is the false alarm hypothesis.

6 Experimental Results

The tracking approach is tested on a cluttered scene that contains five walking persons as shown in Fig. 1a. In this figure each person is surrounded by a rectangle that represents the initial estimation of the target's location and size. Since the objective of this paper is tracking and not target detection, all the targets (location and size) are initialized manually.

As shown in Fig. 1b, a set of $N = 20$ particles are randomly generated around each target. A particle n is a rectangle that is represented at time t by its state vector $x_t^{(n)} = (location, size)$. An increased number of particles will result in a smoother tracking, while increasing the processing time. The frame of discernment is generated according to (35) with $\tau = 5$. For the location cue, the distance measure

(a) (b)

Fig. 1 (**a**) A cluttered scene with five targets. (**b**) Initial particles

in (15) is calculated between the centers of particle n at time t and the estimate of the target location at time $t-1$. For the color cue, the normalized histogram of particle n in (21) is calculated as the frequency of each color bin within the rectangle of the particle. The target's estimate is provided by the algorithm described in Sect. 4. The tracking result, shown in Fig. 2, demonstrates that our approach is able to deal with a cluttered scene, where a relatively large number of targets are involved in partial or total occlusion.

In order to analyze the behavior of our algorithm during the phase of occlusion, we performed another test with two targets only as indicated in Fig. 3.

For the sake of clarity, the person on the right (with the yellow rectangle) is denoted as target 1 and the person on the left (with the white rectangle) is denoted as target 2. The tracking sequence is divided into three phases. Phase 1 is the pre-occlusion sequence, phase 2 corresponds to the occlusion sequence, and phase 3 is the post-occlusion sequence. Tracking in phase 2 is challenging due to the closeness of the targets, which perturbs the measured cues and might lead to a false identification. The location cue loses gradually its ability to separate targets 1 and 2 as they converge to the intersection point. However, the location cue remains a valid measurement because it is independent from the relative location of targets with respect to the camera (occluding or occluded). The color cue is extremely sensitive to the occlusion. During phase 2, target 1 is partially or totally occluded by target 2. As a result, the color measurement for particles associated with target 1 is corrupted by the presence of target 2. When the occlusion is total, target 1 disappears from the scene and the color measurement becomes invalid. The occlusion affects also the behavior of particles associated with target 2 since the presence of target 1 in its neighborhood will be interpreted by the algorithm as

Fig. 2 Tracking of five targets in a cluttered scene

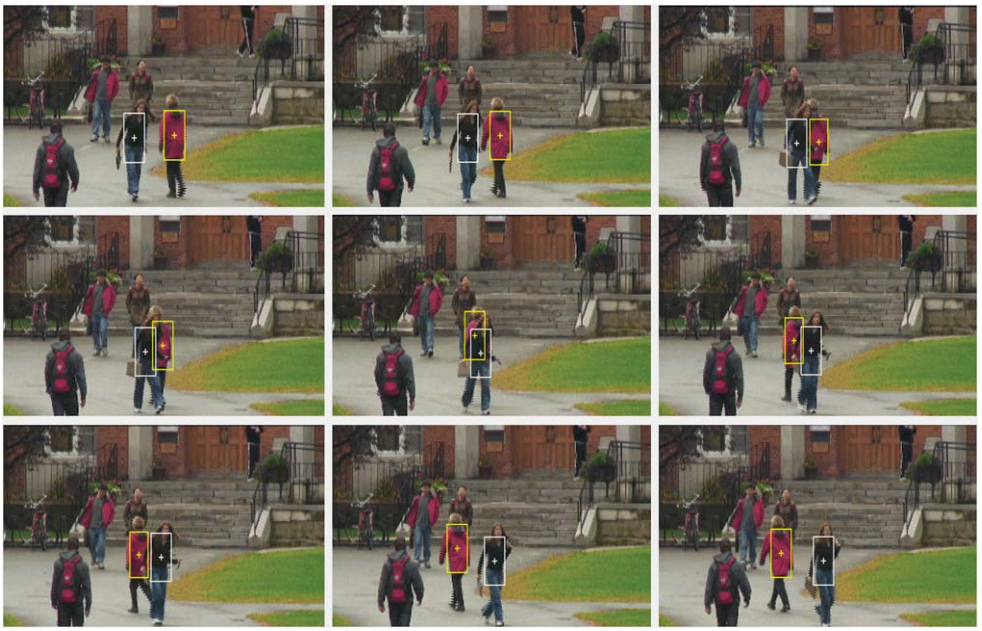

Fig. 3 *First row*: Tracking during the pre-occlusion phase. *Second row:* Tracking during the occlusion phase. *Third row*: Tracking during the post-occlusion phase

a rapid change in the background information. The tracking performances in phase 3 depend on the outcome of tracking during phase 2. A successful tracking would result in a correct identification; whereas, a failure would result in a bad identification of the occluded target (target 1). As shown in Fig. 3, the approach proposed in this paper accurately identifies the targets during the three phases of the tracking. This is due to the effective handling of the conflicting information provided by the location and color cues during the second phase of tracking by the DSmT model. Figure 4 shows the variation of the average value of the confidence levels calculated on the particles associated with each target during the tracking.

The confidence level for the occluded target, $m_{avg}(\theta_1)$, is high during phases 1 and 3, but it decreases in phase 2 (see Fig. 4a). Indeed, in phases 1 and 3 the color and location cues both agree on the identity of the target. However, in phase 2 the target is occluded and this reduces the confidence value provided by the color cue. During the same phase, the location confidence remains high, which explains the increase in the conflict, $m_{avg}(\theta_1 \cap \theta_2)$, as shown in Fig. 4b. The belief function, $Bel_{avg}(\theta_1)$, is given in Fig. 4c. This curve shows the high confidence with which the target is located despite the occlusion. This is mainly due to the introduction of the conflict information through the DSmT model. Figures 4d, 4e and 4f show that the effect of the occlusion on the occluding target is small in comparison with its effect on the occluded target. The existence of such an effect can be justified by the presence of target 1 in the immediate neighborhood of target 2, which rapidly modifies the color measurement for some particles.

Finally, in order to show the importance of modeling and integrating the targets intersections, we have carried out a tracking test (see Fig. 5) where $\theta_1 \cap \theta_2$ is discarded from the combination rule, i.e.

$$m_t^{(n)}(\theta_1 \cap \theta_2) = 0$$

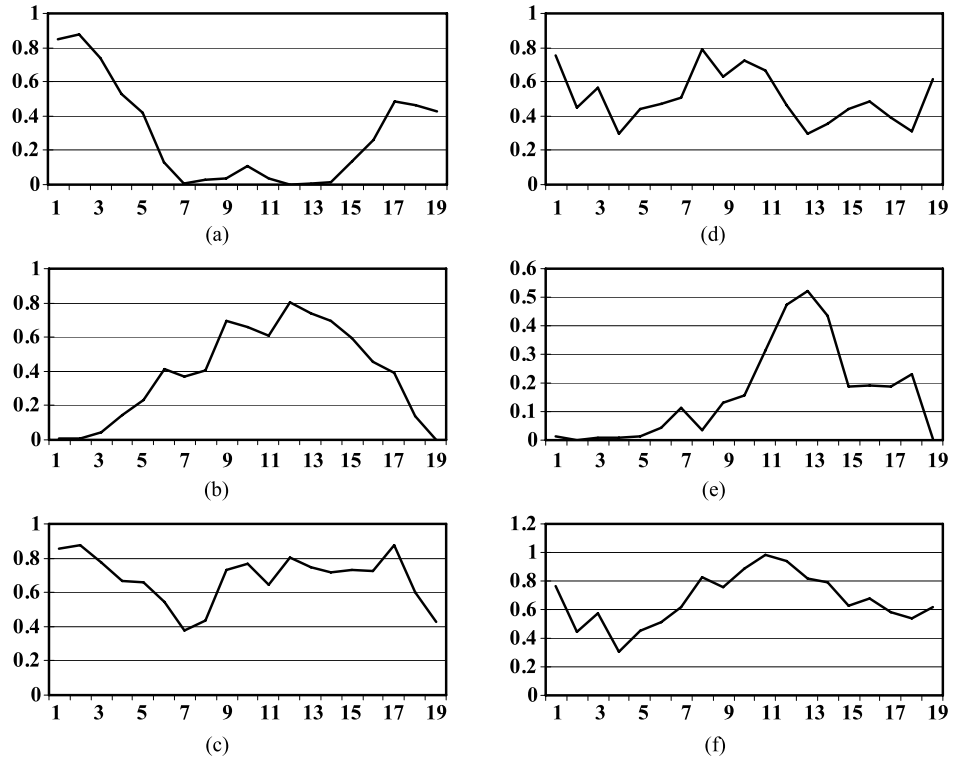

Fig. 4 The variation of: (**a**) $m_{avg}(\theta_1)$, (**b**) $m_{avg}(\theta_1 \cap \theta_2)$, (**c**) $Bel_{avg}(\theta_1)$ for the occluded target. The variation of: (**d**) $m_{avg}(\theta_1)$, (**e**) $m_{avg}(\theta_1 \cap \theta_2)$, (**f**) $Bel_{avg}(\theta_1)$ for the occluding target where $m_{avg}(\theta_j) = \frac{1}{N}\sum_{n=1}^{N} m_t^{(n)}(\theta_j)$, $m_{avg}(\theta_1 \cap \theta_2) = \frac{1}{N}\sum_{n=1}^{N} m_t^{(n)}(\theta_1 \cap \theta_2)$, and $Bel_{avg}(\theta_j) = \frac{1}{N}\sum_{n=1}^{N} Bel_t^{(n)}(\theta_j)$

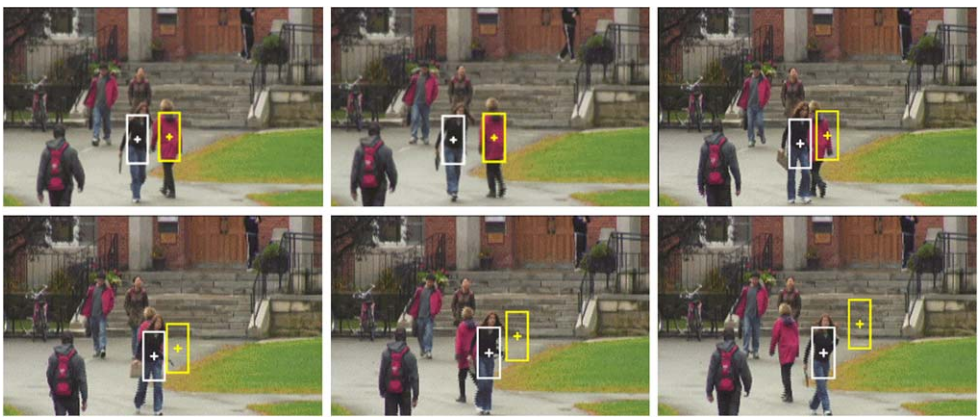

Fig. 5 Tracking using Bayesian inference

$$\pi_{t,j}^{(n)} = Pls_t^{(n)}(\theta_j) = Bel_t^{(n)}(\theta_j) = m_t^{(n)}(\theta_j)$$
$$= m_{t,1}^{(n)}(\theta_j) \cdot m_{t,2}^{(n)}(\theta_j)$$

This framework is equivalent to the Bayesian inference where confidence levels are associated only with single hypotheses. Figure 5 shows that the occluding target was accurately tracked; however, the occluded target was lost due to the occlusion.

7 Conclusion

In this paper, we addressed the problem of tracking multiple targets in a cluttered scene using multiple cues. Within this framework, we developed a model that combines the classical particle filtering approach and the novel DSmT theory. A set of particles is used to track each target. The DSmT model assigns confidence level values for the membership of each particle. This membership takes into consideration the conflict between the cues during the occlusion phase, allowing thus a better tracking. The experimental results demonstrated the effectiveness of the model in case of multiple targets tracking using location and color and its interest in cluttered scenes.

References

1. Stringa, E., Regazzoni, C.S.: Real-time video shot detection for scene surveillance applications. IEEE Trans. Image Process. **9**, 69–79 (2000)

2. Czyz, J., Ristic, B., Macq, B.: A particle filter for joint detection and tracking of color objects. Image Vis. Comput. **25**(8), 1271–1281 (2007)
3. Yang, Y., Teoh, E.K., Shen, D.: Lane detection and tracking using B-Snake. Image Vis. Comput. **22**(4), 269–280 (2004)
4. Chen, C., Lin, X., Shi, Y.: Moving object tracking under varying illumination conditions. Patern Recogn. Lett. **27**(14), 1632–1643 (2006)
5. Frank, O., Nieto, J., Guivant, J., Scheding, S.: Multiple target tracking using sequential Monte Carlo methods and statistical data association. In: Proc. of the IEEE Int. C. on Intell. Rob. and Syst., vol. 3, pp. 2718–2723 (2003)
6. Schultz, D., Burgard, W., Fox, D., Cremers, A.B.: Tracking multiple moving targets with a mobile robot using particle filters and statistical data association. In: Proc. of IEEE Int. C. on Rob. and Auto., vol. 1, pp. 1665–1670 (2001)
7. Ozyilidiz, E., Krahnstover, N., Sharma, R.: Adaptive texture and color segmentation for tracking moving objects. Patern Recogn. **35**(10), 2013–2029 (2002)
8. McCane, B., Galvin, B., Novins, K.: Algorithmic fusion for more robust feature tracking. Int. J. Comput. Vis. **49**(1), 79–89 (2002)
9. Dewasurendra, D.A., Bauer, P.H., Premaratne, K.: Evidence filtering. IEEE Trans. Signal Process. **55**(12), 5796–5805 (2007)
10. Ramasso, E., Panagiotakis, C., Pellerin, D., Rombaut, M.: Human action recognition in videos based on the transferable belief model: application to athletics jumps. Pattern Anal. Appl. **11**(1), 1–19 (2008)
11. Ramasso, E., Rombaut, M., Pellerin, D.: Forward-Backword Viterbi procedures in the transferable belief model for state sequence analysis using belief functions. Lect. Notes Artif. Intell. **4724**, 405–417 (2007)
12. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
13. Smarandache, F., Dezert, J.: Applications and Advances of DSmT for Information Fusion. Am. Res. Press, Rehoboth (2004)
14. Comtet, L.: Sperner Systems. Reidel, Dordrecht (1974), pp. 271–273
15. Arulampalam, M.S., Maskell, S., Gordon, N., Clapp, T.: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans. Signal Process. **50**(2), 174–188 (2002)
16. Hue, C., Cadre, J.-P.L., Perez, P.: Tracking multiple objects with particle filtering. IEEE Trans. Aerosp. Electron. Syst. **38**(3), 791–812 (2002)

Yi Sun received his bachelor degree from the Northern Polytechnical University in Xian, China and his M.Sc. in computer science from Bishop's University, Sherbrooke Canada in 2008. He is currently employed by Microsoft in Beijing, China.

Layachi Bentabet received the B.Eng. degree in electrical engineering from the National Polytechnic School, Algiers, Algeria in 1998, and the Ph.D. degree in computer science from Université de Sherbrooke, Sherbrooke, QC, Canada. He is currently the head of the computer science department at Bishop's university, Sherbrooke, Canada.
His research interests include image processing, computer vision and pattern recognition.