Dezert–Smarandache theory for multiple targets tracking in natural environment

Yi Wang[1], Yingwu Fang[1], Jian Zhang[1], Pengyang Li[2]

[1]School of Information and Navigation, Air Force Engineering University, Fenghao Road 1, Xi'an 710077, People's Republic of China
[2]School of Mechanical and Precision Instrument Engineering, Xi'an University of Technology, Xi'an 710048, People's Republic of China
E-mail: fangyw72@126.com

Abstract: The aim of this article was to investigate multiple targets tracking in natural environment based on Dezert–Smarandache theory (DSmT). On the basis of establishing conflict strategy and combination model, the basic framework and algorithm of fusing multi-source information were described. The multiple targets tracking platform which embedded location and colour cues into the particle filters (PFs) was developed in the framework of DSmT. Three sets of experiments with comparisons were carried out to validate the suggested tracking approach. Results showed that the conflict strategy and DSmT combination model were available, and the introduced approach exhibited a significantly better performance for dealing with high conflict between evidences than a PF. As a result, the approach was suitable for real-time video-based targets tracking, and it had the ability to track interesting targets. Furthermore, the approach can easily be generalised to deal with larger number of targets and additional cues in a complicated environment.

1 Introduction

Video-based target tracking is one of the research hotspots in the field of computer vision, and it has many wide applications in military guidance, visual surveillance, visual navigation of robots, human–computer interaction medical diagnosis and military guidance etc. [1]. Along with the rapid growth of information techniques in the last 10 years, moving targets tracking has attracted many researchers' attention, and has become a very popular research topic. Although many effective visual target tracking methods have been proposed, there are still a lot of difficulties in designing a robust tracking algorithm because of challenging complex scenarios such as significant illumination change in natural environment, pose variations of the objects and non-linear deformations of shapes and noise and dense clutters in complex background etc. [2–4].

Particle filters (PF) have many advantages for solving those problems of non-linear and non-Gaussian system, and the filters are very suitable for moving targets tracking. In the last 15 years we have witnessed a rapid development of the theory of PF, and the corresponding algorithms of PF are widely applied in tracking fields. Godsill and Vermaak [5] presented the PF of variable sampling rate based on a specific observation, and Abdallah et al. [6] introduced the PF of box according to a non-white and biased observation. Crisan and Obanubi [7] analysed PF with random resampling times, and deduced central-limit theorem type results for the approximating particle system with random resampling times. Maroulas and Stinis [8] described an improved PF for multiple moving targets tracking based on

drift homotopy for stochastic differential equations, and the suggested algorithm improved the performance of PF. At present, lots of strategies have been developed for addressing multiple targets tracking by PF (see e.g. [9–12]). Despite the flexibility and availability of PF, there still exist many open problems facing the variation of target motion state, the appearance variation of either target or background and the serious cross and occlusion in natural environment etc. Hence, lots of algorithms have been introduced to track moving targets in different cases. Ottlik and Nagel [13] discussed the initialisation of model-based vehicle tracking in video sequences of inner-city intersections. Chen et al. [14] analysed moving targets tracking under varying illumination conditions. However, the above approaches have mainly improved local performances by optimising PF algorithms, and there still exist many key issues which need to be discussed further. Recently, the Dezert–Smarandache theory (DSmT) by Dezert and Smarandache has been viewed as a general flexible bottom–up approach for managing uncertainty and conflicts for a wide class of static or dynamic fusion problems, where the information to combine is modelled as a finite set of belief functions provided by different independent sources of evidence [15, 16]. Hence, the DSmT of plausible and paradoxical reasoning has become a very important method to deal with high conflicting, uncertain and imprecise sources of evidence in intelligent systems by information fusion. The corresponding researches showed that the conflicting focal elements were increased based on DSmT, and the computational effort was also increased in the course of reasoning. Based on this

input, improved methods were presented to reduce the computational efforts [17, 18]. At present, some researches focus on video-based target identification based on DSmT, the related studies of video-based target tracking based on DSmT are very less. The authors discuss different modified algorithms for video-based targets tracking, and more details can be found [19, 20].

In real natural environment such complete knowledge is difficult to obtain because of natural environment, which includes multiple target crosses and high occlusions, background clutter and illumination and camera calibration problems etc. Although researchers have made progress, these problems facing target tracking in natural environment are very difficult to solve effectively. Thereby, how to develop a robust and real-time video-based targets tracking approach is very necessary in natural environment. Owing to the efficiency of DSmT in combining conflict evidences, the objective of this article is to present a novel approach of video-based multiple targets tracking that will handle the targets of crosses and occlusions in order to attain excellent information fusion in the framework of DSmT.

2 Foundations of the DSmT and PF

DSmT is a generalisation of the classical Dempster–Shafer theory (DST). The DSmT framework can easily handle not only exclusivity constraints, but also non-existential constraints or mixed constraints as well which is very useful in some dynamic fusion problems where the DST usually fails; it differs from DST because it is based on free Dedekind's lattice, and the detailed examples of the types of constraints can be found [15]. In this section, the DSmT and PF were described on the basis of reviewing the DST.

Although the DST considers $\cup$ 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 different sources throughout the assignment of non-null confidence values to the intersection of distinct hypotheses. Let $\cup = \{\theta_1,\ldots,\theta_N\}$ be a set of N elements which can potentially overlap, where θ_i denotes the ith element in the finite set $\cup$. The hyper-power set $D^{\cup}$ is defined as the set of all composite propositions built from elements of $\cup$ with $\cup$ and $\cap$ ($\cup$ generates $D^{\cup}$ under operators $\cup$ and $\cap$) operators such that [16], where $D^{\cup}$ denotes Dedekind's lattice, and which is also called hyper-power set in the DSmT framework.

(1) ϕ, $\theta_1,\ldots,\theta_N \in D^{\cup}$; (2) If A, $B \in D^{\cup}$, then $A \cap B \in D^{\cup}$ and $A^{\cup}B \in D^{\cup}$; (3) No other elements belong to $D^{\cup}$, except those obtained by using rules (1) or (2).

As in the DST, the DSmT defines a map $m(\cdot):D^{\cup} \rightarrow [0, 1]$. This map defines the confidence level that each sensor associates with the element of $D^{\cup}$. This map supports paradoxical information, and the DSmT combination rules of conflict and uncertain sources are given by

$$m(\varphi) = 0, \quad \sum_{A \in D^{\cup}} m(A) = 1 \quad (1)$$

$$m(A) = \sum_{\substack{A_1,A_2,\ldots,A_N \in D^{\cup} \\ A_1 \cap A_2 \cap \ldots \cap A_N = A}} \prod_{i=1}^{D} m_i(A_i) \quad (2)$$

PF is one kind of sampling simulation based on sequential Monte Carlo. Owing to its multiple hypothesis property, PF

can be applied to any state-space model to estimate the trajectory of an object in frames effectively; it is originally developed to track objects in clutter or a variable of interest as it evolves over time, and typically with a non-Gaussian and potentially multi-modal probability density function (pdf). The basis of the PF is to construct a sample-based representation of the entire pdf. A series of actions are taken, each one modifying the state of the variable of interest according to some model. Moreover, at certain times an observation arrives that constrains the state of the variable of interest at that time. Generally, PF is an optimal Bayesian algorithm for non-linear and non-Gaussian object tracking, and it obtains the most likely posterior estimation based on sequential Monte Carlo. The detailed description of PF was given in the literature [21].

Let a set of discrete particles be approximately depicted for a complicated posterior probability density, and the method level depends on the number of particles. Let us consider $X_t = (x_1, x_2,\ldots,x_t)$ as the state vector (location, size etc.) describing the target and $Z_t = (z_1, z_2,\ldots,z_t)$ as the vector of measurements (colour, texture etc.) up to time t. The tracking is based on the estimation of posterior state distribution $p(x_t|Z_t)$ at each time step. The estimation is performed using a two-step Bayesian recursion. The first step is prediction, one finally gets

$$p(x_t|Z_{t-1}) \propto \int p(x_t|x_{t-1})p(x_{t-1}|Z_{t-1})\mathrm{d}x_{t-1} \quad (3)$$

The second step is filtering, which is given by

$$p(x_t|Z_t) \propto p(z_t|x_t)p(x_t|z_{t-1}) \quad (4)$$

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 PF is very simple. Starting with a weighted set of samples, which is given by

$$S_{t-1} = \left\{ s_{t-1}^{(n)}, \ \pi_{t-1}^{(n)} \middle| \sum_{n=1}^{N} \pi_{t-1}^{(n)} = 1 \right\} \quad (5)$$

where S_{t-1} denotes the object state of time $t-1$, n denotes the nth sample.

According to $p(x_{t-1}|z_{t-1})$, 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 based on the observation, and N samples are drawn with replacement according to $\pi_t = p(z_t|x_t)$. The value will represent the best estimation of the target, given by

$$E[S_t] = \sum_{n=1}^{N} \pi_t^{(n)} s_t^{(n)} \quad (6)$$

PF has been proven to be very successful for non-linear and non-Gaussian estimation problems, and the basic tracking step includes selection of samples, propagation of samples, observation of samples and calculation and estimation of the mean state, and the detailed content can be found in the literatures [22, 23]. Although many effective target tracking methods have been proposed by PF, there are still a lot of difficulties in designing a successful tracking platform in natural environment. Hence, on the basis of the adaptive PF

[24], the following address the tracking approach based on DSmT in detail.

3 Multiple targets tracking based on DSmT

3.1 Algorithm description

Let us assume that the number of targets is τ ($\tau \gg 2$ targets), the number of cues is c and the τ and 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 target given is to combine these measurements in order to determine the best track for each candidate. It is important to note that a target candidate, refers to a particle sample, the hyper-power set $D^{\cup}$ defines the set of the hypotheses for which different cues can provide confidence values. These hypotheses can correspond to: (i) individual tracks θ_J, (ii) union of tracks $\theta_1 \cup \ldots \cup \theta_N$, which symbolises ignorance, (iii) intersection of tracks $\theta_1 \cap \ldots \cap \theta_N$, which symbolises conflict or (iv) any tracks combination obtained by $\cup$ and $\cap$ operators.

The confidence level is expressed in terms of mass function $\{m_{t,l}^{(n)}(\cdot)\}_{l=1}^{c}$ that is committed to each hypothesis and which satisfies the condition in (1) and (2). Given this framework, $m_{t,l}^{(n)}(A)$ expresses the confidence value with which cue l associates particle n to hypothesis A at time t. Based on DSmT combinational rule, a single map function $m_t^{(n)}(\cdot)$ can be derived as follows

$$m_t^{(n)}(A) = m_{t,1}^{(n)}(\cdot) \oplus m_{t,2}^{(n)}(\cdot) \oplus \cdots \oplus m_{t,c}^{(n)}(\cdot) \qquad (7)$$

where $m_t^{(n)}(A)$ denotes the overall confidence level with which all cues associate particle n to hypothesis A at time t.

Since the target candidates must be associated with individual tracks, the information contained in compound hypotheses is transferred into single hypotheses (i.e. single tracks) through the notions of the belief or plausibility functions and is given by

$$\text{Bel}_t^{(n)}(\theta_J) = \sum_{\substack{A \subseteq \theta_J \\ A \in D^{\Theta}}} m_t^{(n)}(A) \qquad (8)$$

$$\text{Pls}_t^{(n)}(\theta_J) = \sum_{\substack{A \cap \theta_J \neq \varphi \\ A \in D^{\Theta}}} m_t^{(n)}(A) \qquad (9)$$

where $\text{Bel}_t^{(n)}(\theta_J)$ (resp. $\text{Pls}_t^{(n)}(\theta_J)$) quantifies the confidence with which particle n is associated with θ_J at time t using the notion of belief (resp. plausibility).

The confidence levels are not used to determine whether a given candidate is the best estimate or not of the target, they are rather used to quantify the weight of the candidate as a sample of the state posterior distribution $p(\mathbf{X}_t|\mathbf{Z}_t)$. As a result, the PF algorithm based on DSmT is implemented, and the corresponding step is given as follows:

1. *Initialisation*: generate N samples $S_{t-1,j} = \{s_{t-1,j}^{(n)}, \pi_{t-1,j}^{(n)}\}_{n=1}^{N}$ for each target, $j = 1,\ldots,\tau$ independently, with $\pi_{t-1,j}^{(n)} = 1/N$, and set $t = 1$.
2. *Propagation*: $S_{t,j}^{(n)} = A \cdot S_{t-1,j}^{(n)} + w_{t-1,j}^{(n)}$

3. *Observation for each particle*: compute $\{m_{t-1,l}^{(n)}(A)\}_{l=1}^{c}$ and $m_{t-1,l}^{(n)}(A)$ for $A \in D^{\cup}$, calculate the particle weight $\pi_{t-1,j}^{(n)} = \text{Bel}_{t-1}^{(n)}(\theta_J)$ or $\pi_{t-1,j}^{(n)} = \text{pls}_{t-1}^{(n)}(\theta_J)$, and normalise the weight: $\tilde{\pi}_{t-1,j}^{(n)} = (\tilde{\pi}_{t-1,j}^{(n)})/(\sum_{n=1}^{N} \tilde{\pi}_{t-1,j}^{(n)})$
4. *Estimation*: target $j = 1 \ldots \tau$ is given by $E[S_{t,j}] = \sum_{n=1}^{N} \tilde{\pi}_{t,j}^{(n)} s_{t,j}^{(n)}$
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 $p(s_{t,j}^{(n)}) = \tilde{\pi}_{t,j}^{(n)}$
6. *Incrementing*: when $t = t+1$, go to (2).

3.2 Conflict strategy

Since preservation of conflicting focal elements can increase assignment of the focal element in the framework of DSmT, the convergence is very slow for assign function of main focal element in most cases, and the difficulty of tracking is increased greatly. Hence, a modified conflict strategy was presented. Namely, local conflict was not assigned globally but was assigned locally by refining global conflict into r local conflicts. Fig. 1 shows the relation of conflict and correlative factors.

Video-based multiple targets tracking may take place in many cases including target crossing, target occlusion, target scale variation and illumination change etc. Hence, the tracking process has high conflict and uncertainty in natural environment. By consulting the thoughts in the literature [25], a union operator is selected when the focal element is weak, and an extraction operator is selected when the focal element is strenuous for obtaining more information. Namely, when the conflict is strenuous, the conflict is assigned to a focal element which has a close relationship with conflict. When a conflict is weak, the conflict is mostly assigned to a focal element which has a close relationship with conflict. According to Fig. 1, r local conflicts are reassigned between conflicted focal element, and the assigned rule is basic belief assignment and belief of evidence. Finally, an inconsistent parameter R is introduced in order to efficiently describe the total level of conflicts between evidences, the conflict level based on mass ($\cdot$) function can be corrected by R and the corresponding expression of R is given by Yang *et al.* [18]. According to (10), when the parameter R tends to 1, the conflict between evidences is strenuous, and conflicts are partly transferred to a union of conflict focal elements.
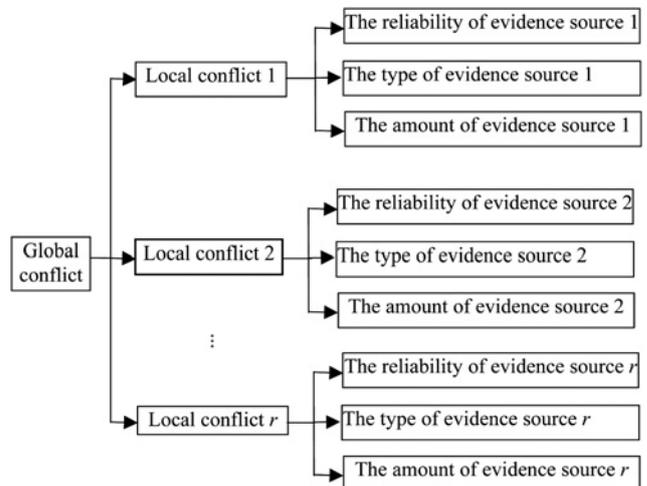

Fig. 1 *Relation of conflict and correlative factors*

When the parameter R tends to 0, the conflicts between evidences are very weak, and they are partly transferred to the conflict focal elements.

$$R = \sqrt{\sum_{A \subseteq \Theta} \left| \frac{m_1(A) - m_2(A)}{2} \right|} \qquad (10)$$

where m_1, m_2 denotes the basic probability assignment function of two evidences source, respectively.

At the same time, an efficient approach of settling conflict assignment problem is adopted based on the idea of the literature [26]. If the support degree of m_1 and m_2 for involving conflict focal element A was greater than focal element B, $m(A)$ value of assembled basic probability assignment should be greater than $m(B)$. Especially, most conflicts should be assigned to the union of involving a conflict focal element when the conflict is strenuous, or most conflicts should be assigned to the focal element of involving conflict when the conflict is weak. As a result, the following conflict strategies between evidences are defined based on the above idea.

If the basic focal elements are A and B, then the conflict strategies are defined as follows

$$m(A \cap B)^c = m_1(A)m_2(B) + m_1(B)m_2(A) \qquad (11)$$

$$m(B) = \sum_{A \cap B = B} m_1(A)m_2(B) + m(A \cap B)^c$$

$$\frac{m_0(B)}{m_0(A) + m_0(B) + m_0(A \cup B)}(1 - R) \qquad (12)$$

$$m(A \cup B) = \sum_{A \cap B = A \cup B} m_1(A)m_2(B) + m(A \cap B)^c$$

$$\frac{m_0(A \cup B)}{m_0(A) + m_0(B) + m_0(A \cup B)}(1 - R) \qquad (13)$$

$$m(A \cap B) = m(A \cap B)^c \cdot R \qquad (14)$$

where m_0 denotes the basic probability assignment function based on the D–S evidence theory.

3.3 Dynamic combination model

According to the above analysis, the following established dynamic combination model of multiple targets tracking. In order to describe conveniently, the cues of colour and location were used to track two targets in this section. For two targets, $\cap \cup$ was defined as follows

$$\cup = \left\{ \theta_1, \theta_2, \overline{\theta_1 \cup \theta_2} \right\} \qquad (15)$$

In (15), θ_1 refers to 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. Since this latter can change during tracking, we will refer to $\overline{\theta_1 \cup \theta_2}$ as the false alarm hypothesis. Besides, $\theta_1 \cap \theta_2 \neq \phi$ because of the possible occlusion, and $\theta_j \cap \overline{\theta_1 \cup \theta_2} = \varphi$ for $j = 1, 2$.

The number of picture elements of the ith colour and the total number of picture elements of an image ratio are called normalised colour histograms. Let us assume that both target models are known and given by normalised colour histograms $\{q_j(u)\}_{u=1}^m$, where u is a discrete colour index, and m is the number of histogram bins. At time t, the

normalised colour histogram of particle $s_{t,j}^{(n)}$ is given by $\{h_{t,j}^{(n)}(u)\}_{u=1}^m$. The probability that particle $s_{t,j}^{(n)}$ belongs to target $j = 1, 2$ according to the colour histogram is derived from the following Gaussian pdf.

$$p_{t,j}^{(n)} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(d_{t,j}^{(n)}\right)^2/(2\sigma^2)}, \quad j = 1, 2 \qquad (16)$$

$$d_{t,j}^{(n)} = \sqrt{1 - \sum_{u=1}^m h_{t,j}^{(n)}(u)q_j(u)} \qquad (17)$$

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

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

$$q_{FA}(u) = \max\{q_{scene}(u) - q_1(u) - q_2(u), \, 0\} \qquad (18)$$

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

$$p_{t,FA}^{(n)} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(d_{t,FA}^{(n)}\right)^2/(2\sigma^2)} \qquad (19)$$

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

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

$$m_{t,2}^{(n)}\left(\overline{\theta_1 \cup \theta_2}\right) = \frac{p_{t,FA}^{(n)}}{p_{t,1}^{(n)} + p_{t,2}^{(n)} + p_{t,FA}^{(n)}} \qquad (20)$$

$$m_{t,2}^{(n)}(\theta_j) = \frac{p_{t,j}^{(n)}}{p_{t,1}^{(n)} + p_{t,2}^{(n)} + p_{t,FA}^{(n)}}, \quad j = 1, 2 \qquad (21)$$

The targets locations at time $t - 1$ are known and given by $(x_{t-1,1}, y_{t-1,1})$ and $(x_{t-1,2}, y_{t-1,2})$. At time t, the probability that a particles $s_{t,j}^{(n)}$ located at $\left(x_{t,j}^{(n)}, y_{t,j}^{(n)}\right)$ belongs to target $j = 1, 2$ according to the location information is defined from a Gaussian pdf as follows

$$p_{t,j}^{(n)} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(x_{t,j}^{(n)} - x_{t-1,j}\right)^2 + \left(y_{t,j}^{(n)} - y_{t-1,j}\right)^2}{2\sigma^2}} \qquad (22)$$

where σ denotes a parameter of bandwidth, the probability that a given particle does not belong to θ_1 and θ_2 is inversely proportional to the distance between particles and both targets. Since Θ is exhaustive, a particle that does not belong to θ_1 and θ_2, and does belong to $\overline{\theta_1 \cup \theta_2}$. This leads us to the definition of a new pdf, $p_{t,FA}^{(n)}$, which measures the membership of a particle $n = 1,...,N$ to the false alarm hypothesis.

$$p_{t,FA}^{(n)} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(d_{max} - d_{1-2}^{(n)}\right)^2/(2\sigma^2)} \qquad (23)$$

$$d_{1-2}^n = \sqrt{\left(x_{t-1}^{(n)} - \frac{x_{t-1,1} - x_{t-1,2}}{2}\right)^2 + \left(y_{t-1}^{(n)} - \frac{y_{t-1,1} - y_{t-1,2}}{2}\right)^2} \qquad (24)$$

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

The mass function of particle n according to its location is given as follows

$$m_{t,1}^{(n)}(\theta_j) = \frac{p_{t,j}^{(n)}}{p_{t,1}^{(n)} + p_{t,2}^{(n)} + p_{t,\mathrm{FA}}^{(n)}}, \quad j = 1, 2 \qquad (25)$$

$$m_{t,1}^{(n)}(\overline{\theta_1 \cup \theta_2}) = \frac{p_{t,\mathrm{FA}}^{(n)}}{p_{t,1}^{(n)} + p_{t,2}^{(n)} + p_{t,\mathrm{FA}}^{(n)}} \qquad (26)$$

Based on the modified combination strategy introduced, the combination rule leads to the mass function $m_t^{(n)}(\cdot)$, and the corresponding combination rules of colour and location are defined in Table 1.

where

$$m_t^{(n)}(\theta_1) = m_{t,1}^{(n)}(\theta_1) \cdot m_{t,2}^{(n)}(\theta_1) \qquad (27)$$

$$m_t^{(n)}(\theta_2) = m_{t,1}^{(n)}(\theta_2) \cdot m_{t,2}^{(n)}(\theta_2) \qquad (28)$$

$$m_t^{(n)}(\theta_1 \cap \theta_2) = m_{t,1}^{(n)}(\theta_1) \cdot m_{t,2}^{(n)}(\theta_2) \\ + m_{t,1}^{(n)}(\theta_2) \cdot m_{t,2}^{(n)}(\theta_1) \qquad (29)$$

$$m_t^{(n)}(\overline{\theta_1 \cup \theta_2}) = m_{t,1}^{(n)}(\overline{\theta_1 \cup \theta_2}) \cdot m_{t,2}^{(n)}(\overline{\theta_1 \cup \theta_2}) \qquad (30)$$

$$m_t^{(n)}(\varphi) = m_{t,1}^{(n)}(\overline{\theta_1 \cup \theta_2})\left(m_{t,2}^{(n)}(\theta_1) + m_{t,2}^{(n)}(\theta_2)\right) \\ + m_{t,2}^{(n)}(\overline{\theta_1 \cup \theta_2})\left(m_{t,1}^{(n)}(\theta_1) + m_{t,1}^{(n)}(\theta_2)\right) \qquad (31)$$

According to Table 1, (27) is the confidence level with which both cues associate $s_{t,j}^{(n)}$ to target 1. Equation (28) is the confidence level with which both cues associate $s_{t,j}^{(n)}$ to target 2. Equation (29) is the conflict value between the cues for membership of $s_{t,j}^{(n)}$ to targets 1 or 2. Equation (30) expresses the confidence value with which both cues agree that the particle corresponds to a false alarm. Equation (31) quantifies the conflict between the targets and the false alarm hypothesis.

The weight of $s_{t,j}^{(n)}$ particle within the posterior $p(X_t|Z_t)$ distribution is calculated using belief (or the plausibility) function for target J, given by

$$\pi_{t,J}^{(n)} = \mathrm{Bel}_t^{(n)}(\theta_J) = m_t^{(n)}(\theta_J) + m_t^{(n)}(\theta_1 \cap \theta_2) \\ J = 1, 2 \qquad (32)$$

Based on the above discussion, the generalisation of the tracking scheme described in this section to τ targets can be carried out by defining a frame of discernment $\cup = \{\theta_1, \ldots, \theta_\tau, \theta_1 \cup \cdots \cup \theta_\tau\}$, where θ_J denote an individual target, and $\overline{\theta_1 \cup \cdots \cup \theta_\tau}$ denote the false alarm hypothesis. Hence, the same method can be adopted to establish a dynamic combination model of τ targets in the framework of DSmT.

4 Tracking experiments

4.1 Tracking realisation

Based on the above analysis of combination strategy and dynamic model for multiple targets tracking in natural environment, an improved algorithm of multi-source information fusion was realised in the framework of DSmT. The algorithm by merging location and colour cues was embedded in the tracking frame of PF, and the whole tracking process included read module of video sequence, setting parameter module, preprocessing module of video image, detecting module of moving targets and tracking module of moving targets. Fig. 2 shows the basic framework of the tracking process.

According to Fig. 2, the read module of the video sequence was used to read video data by video collecting equipment and provide right data format for latter tracking. For the setting parameter module, it was used to provide correlative parameters for every module. For the preprocessing module of video image, it was used to suppress noise and
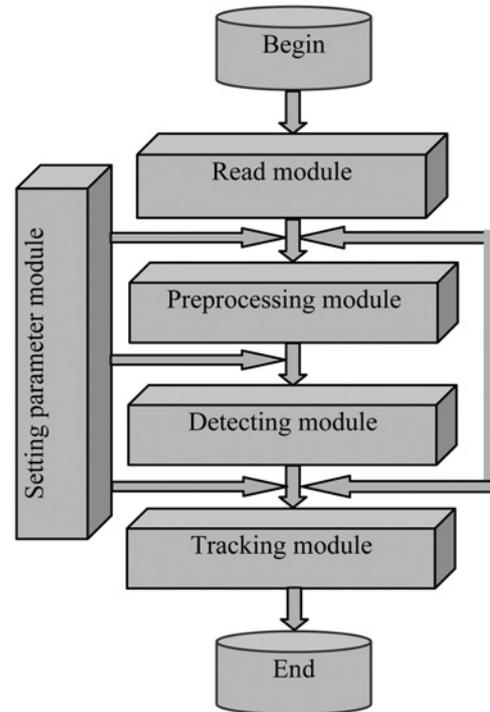

Fig. 2 *Basic framework of tracking process*

Table 1 Combination rules of colour and location

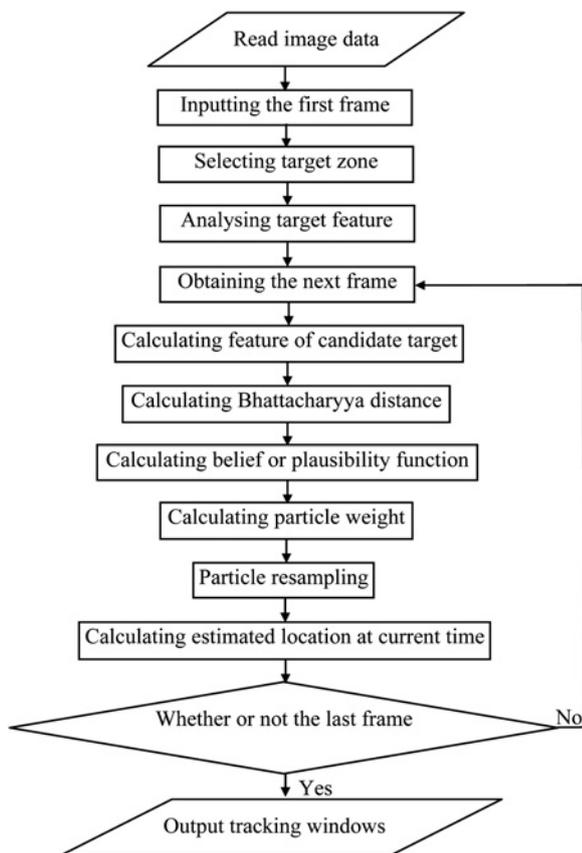| | Colour cue | | |
	$m_{t,2}^{(n)}(\theta_1)$	$m_{t,2}^{(n)}(\theta_2)$	$m_{t,2}^{(n)}(\overline{\theta_1 \cup \theta_2})$
Location cue $\quad m_{t,1}^{(n)}(\theta_1)$	$m_t^{(n)}(\theta_1)$	$m_t^{(n)}(\theta_1 \cap \theta_2)$	ϕ
$m_{t,1}^{(n)}(\theta_2)$	$m_t^{(n)}(\theta_1 \cap \theta_2)$	$m_t^{(n)}(\theta_2)$	ϕ
$m_{t,1}^{(n)}(\overline{\theta_1 \cap \theta_1})$	ϕ	ϕ	$m_t^{(n)}(\overline{\theta_1 \cup \theta_2})$

Fig. 3 *Main algorithm flowchart of tracking target*

4 GB memory, 300 G hard disk and Windows XP. The tracking procedure can be performed when a new video frame comes. Fig. 4 shows the operation panel of the tracking platform.

4.2 Tracking examples and results

Since the purpose of this article was to address tracking and not the target detection, targets were selected manually. In this section, tracking experiments with comparisons were carried out to validate the introduced approach. Especially, this section mainly illustrated how to use the location and colour cues for multiple targets tracking in natural environment, and the introduced approach how to improve the efficiency and robustness for the following factors: (i) crossing targets, (ii) occlusion targets and (iii) scale variation of targets. Finally, three sets of tracking experiments based on different scenes were carried out to validate the suggested approach.

The first video scene was a campus region, and the tracking experiment of only two targets in a natural environment was carried out to validate the introduced approach. The video was captured from internet platforms, and the target objects included two pedestrians. According to the introduced approach, the tracking process was executed by merging the location and colour cues of different targets. In the experiment, image pre-processing was employed, and the initial positions of objects were manually designated. Initialisation of tracking was executed at the beginning of every image subsequence, which included calibration of the location and space area of targets tracked, estimation of the moving direction and speed of every object and calculation of the scaling according to the trend of relative motion between targets tracked and imaging lens. Let the two pedestrians in video scene keep uniform motion along the moving direction of cross and occlusion, and the illumination change be also omitted.

In the course of tracking experiment, the colour distribution of objects had obvious difference by comparing with the surrounding environment, and the surface feature relative to the distributed location of structural distortion is very small. At the same, there existed high conflict problems including scale variations, cross and occlusion of objects. From 60 to 81 frames, the two objects underwent the following processes including cross, part and full occlusion. The number of particle was variable with conflict levels between evidences, and the maximum number of particles was 40 when the two targets were basically covered. Namely, 40 particles were only used to handle the high conflict between evidences. Fig. 5 shows the variation of particle number in different tracking stage.

pre-segment of video image. For the detecting module of moving targets, it was used to detect interesting moving target and eliminate the influence of illumination and shadow. For the tracking module of moving targets, it was mainly used to track moving target detected and gain the corresponding data including speed and location etc. The tracking algorithm was designed on the basis of analysing the tracking process. Fig. 3 shows the main algorithm flowchart of tracking target.

Based on the basic framework and algorithm of multiple tracking targets, Visual studio 2005 C++.net environment and OPENCV1.0 (open source computer vision library) were used to develop the corresponding targets tracking platform. As a result, the platform of multiple targets tracking was operated by Pentium(R)D CPU 3.00 GHz,
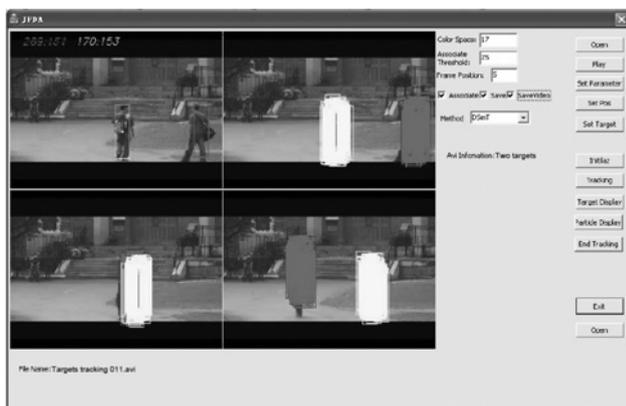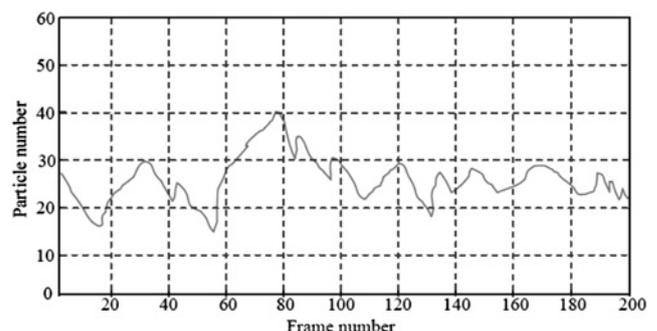

Fig. 4 *Operation panel of tracking platform*

Fig. 5 *Variation of particle number in different tracking stage*

Finally, the whole video-based tracking was accomplished, and the video demonstrating tracking results were available by the introduced approach. Fig. 6 shows the main frames with tracking particles during tracking experiment, and Fig. 7 shows the tracking process of main frames and tracking result.

In order to validate the stability for handling high conflict between evidences by the introduced approach, a mean shift approach was also applied to track the above image sequences. At last, the tracking process and result of two kinds of approaches were obtained. Fig. 8 shows the deviation of target centre (Δx, Δy) during tracking by comparing with the two approaches. It was seen from Fig. 8 that the variation of deviation of target centre was much smaller than that of the mean shift approach during almost the whole tracking. However, the mean shift approach's accuracy deteriorated rather quickly when the two targets had high cross and occlusion. Thereby, the results showed that the suggested approach could track moving targets effectively, and the approach had better adaptation to the variation of target and background.

The second tracking experiment was carried out to test the availability of the introduced approach. The corresponding video image was also captured from internet platforms, a
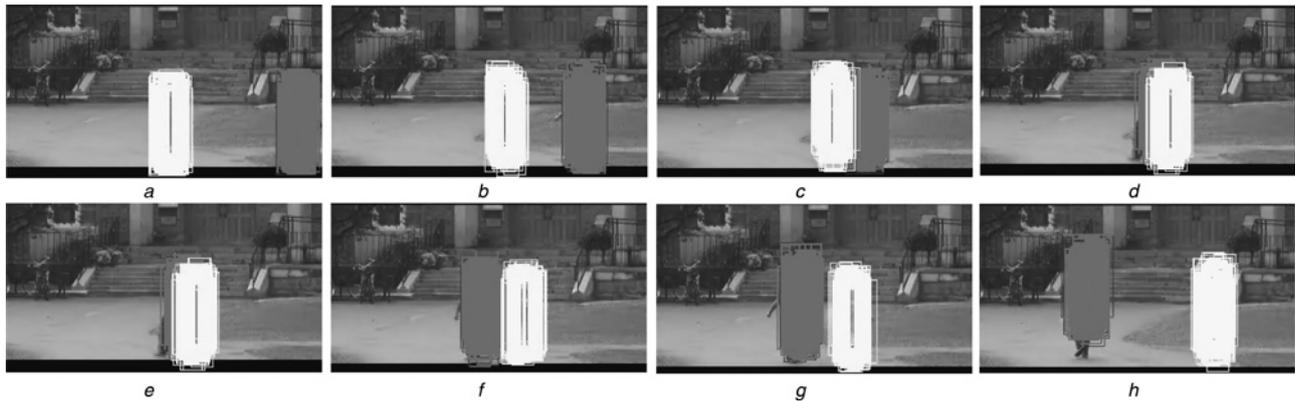

Fig. 6 *Main frames with tracking particles during tracking experiment*
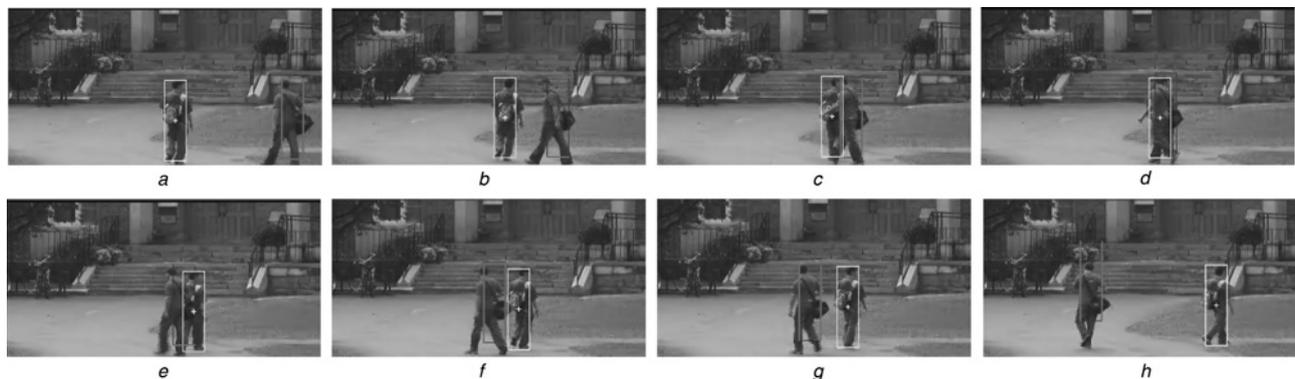

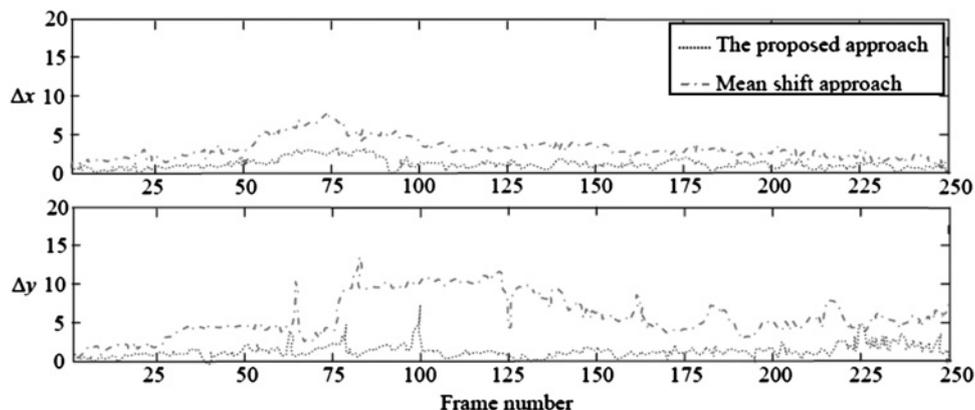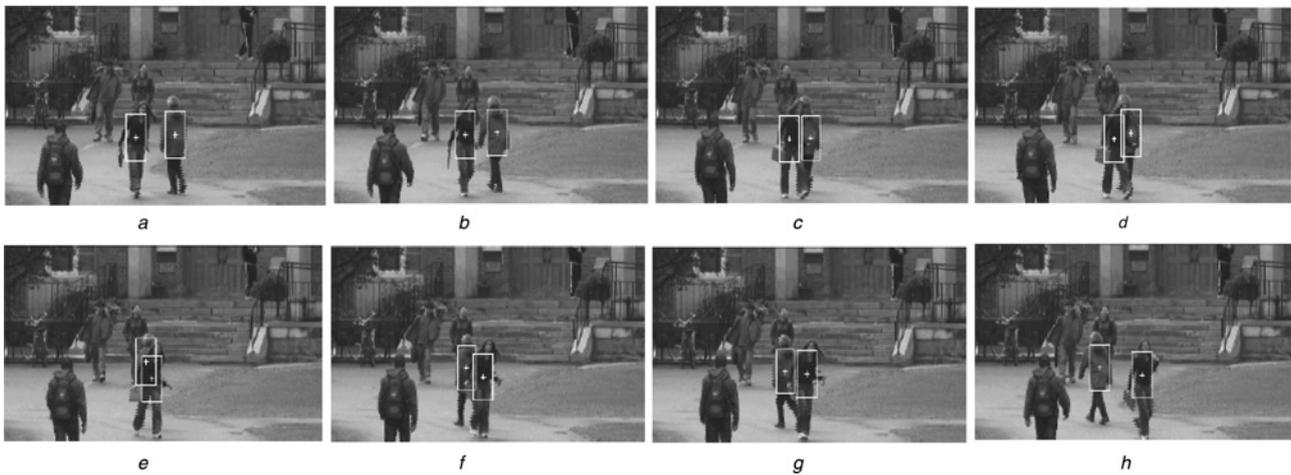**Fig. 7** *Tracking process of main frames and tracking result*

Fig. 8 *Deviation of target centre of the two approaches*

Fig. 9 *Tracking process of main frames of two moving targets and the tracking result*

group of pedestrians were walking into a campus region along different orientation and the colour character of clothing was very similar. During tracking experiment, two objects of the video scene were designated to execute tracking in a cluttered scene. In the experiment, the initial positions of objects were designated manually after pre-processing and initialisation of the image sequence. Let the objects keep uniform motion, and the illumination change be omitted, too. With the increased number of particles, the processing time was also increased, and smoother tracking was obtained. During the tracking experiment, 50 particles or so were used to handle the high conflict between evidences. For the sake of clarity, the right person was denoted target 1, and the left person was denoted target 2. At last, the tracking experiment was executed by the introduced approach. Fig. 9 shows the tracking process of main frames and the tracking result.

According to Fig. 9, the tracking sequence was divided into three stages. Namely, Figs. 9a and b were the stages of tracking pre-occlusion, Figs. 9c–f were the stages of tracking occlusion and Figs. 9g and h were the stages of tracking post-occlusion. Hence, stage 1 was the pre-occlusion sequence, stage 2 corresponded to the occlusion sequence and stage 3 was the post-occlusion sequence. Especially, tracking in stage 2 was challenging because of the closeness of the targets, and the measured cues might lead to a false identification. The location cue lost gradually its ability to separate targets 1 and 2 as they converged to the intersection point. However, the location cue remained a valid measurement because it was independent from the relative location of targets with respect to the camera (occluding or occluded). The colour cue was extremely sensitive to the occlusion. During stage 2, target 1 was partially or totally occluded by target 2. As a result, the colour measurement for particles associated with target 1 was corrupted by the presence of target 2. When the occlusion was total, target 1 disappeared from the scene, and the colour measurement became invalid. The occlusion also affected the behaviour of particles associated with target 2, since the presence of target 1 in its neighbourhood which would be interpreted by the introduced approach as a rapid change in the background information. The tracking performances in stage 3 depended on the outcome of tracking during stage 2. Finally, the tracking process imputed to a correct identification.

In order to analyse the variation of average values of the confidence levels for all particles in the course of handling occluding and occluded targets, the following equations were given by

$$m_{\mathrm{avg}}(\theta_J) = \frac{1}{N} \sum_{n=1}^{N} m_t^{(n)}(\theta_J) \qquad (33)$$

$$m_{\mathrm{avg}}(\theta_1 \cap \theta_2) = \frac{1}{N} \sum_{n=1}^{N} m_t^{(n)}(\theta_1 \cap \theta_2) \qquad (34)$$

$$\mathrm{Bel}_{\mathrm{avg}}(\theta_J) = \frac{1}{N} \sum_{n=1}^{N} m_t^{(n)}(\theta_J) \qquad (35)$$

According to (33)–(35), the average values of the confidence levels for all particles were calculated. Fig. 10 shows the variation of average values of the confidence levels for all particles during tracking.

In Fig. 10, the Figs. 10a–c denoted the variation of $m_{\mathrm{avg}}(\theta_1)$, $m_{\mathrm{avg}}(\theta_1 \cap \theta_2)$ and $\mathrm{Bel}_{\mathrm{avg}}(\theta_1)$ for the occluded targets, respectively. Figs. 10d–f denoted the variation of $m_{\mathrm{avg}}(\theta_1)$, $m_{\mathrm{avg}}(\theta_1 \cap \theta_2)$ and $\mathrm{Bel}_{\mathrm{avg}}(\theta_1)$ for the occluding targets, respectively. It was seen from Fig. 10 that the confidence level for the occluded target ($m_{\mathrm{avg}}(\theta_1)$) was high during stages 1 and 3, but the confidence level was slowly decreased in stage 2, as shown in Fig. 10a. Indeed, the colour and location cues both agreed the identity of the target in stages 1 and 3. However, the target was occluded in stage 2, and this reduced the confidence value provided by the colour cue.

During the same stage, the location confidence remained high, which explained the increase in the conflict ($m_{\mathrm{avg}}(\theta_1 \cap \theta_2)$), as shown in Fig. 10b. The variation of average values of the confidence levels was given by belief function ($\mathrm{Bel}_{\mathrm{avg}}(\theta_1)$), as shown in Fig. 10c. The curve of Fig. 10c showed the high confidence with which the target was located despite the occlusion. This was mainly because of the introduction of the conflict information through the DSmT combination model. Figs. 10d–f showed that the effect of occlusion on the occluding targets was small by comparing with its effect on the occluded targets. The existence of such an effect could be justified by the presence of target 1 in the immediate neighbourhood of
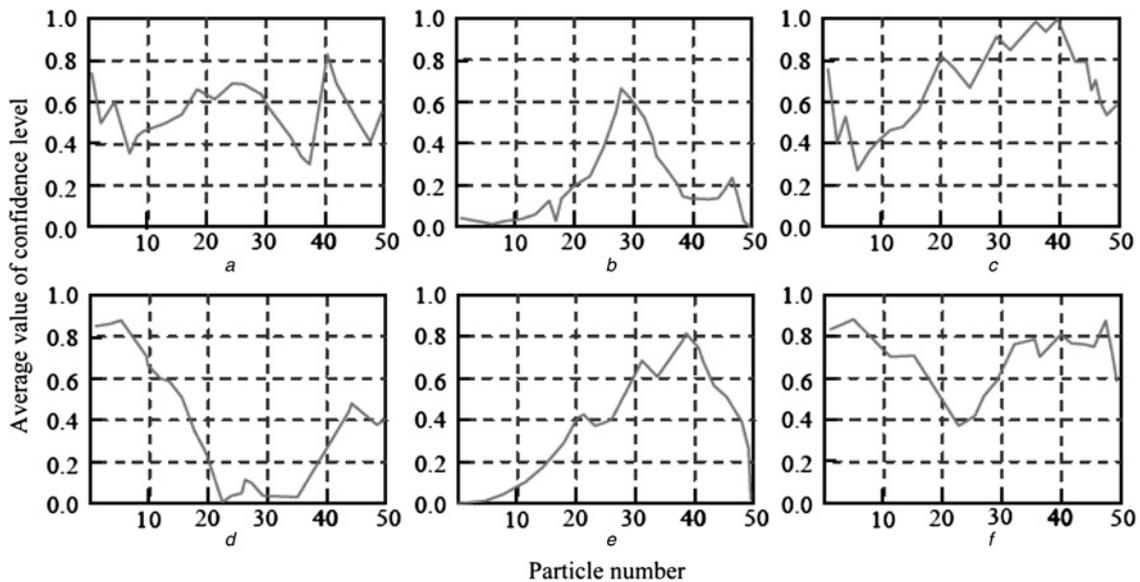
Fig. 10 *Variation of average values of the confidence levels for all particles during tracking*

a Variation of $m_{avg}(\theta_1)$ for the occluded targets
b Variation of $m_{avg}(\theta_1 \cap \theta_2)$ for the occluded targets
c Variation of $Bel_{avg}(\theta_1)$ for the occluded targets
d Variation of $m_{avg}(\theta_1)$ for the occluding targets
e Variation of $m_{avg}(\theta_1 \cap \theta_2)$ for the occluding targets
f Variation of $Bel_{avg}(\theta_1)$ for the occluding targets

target 2, which rapidly modified the colour measurement for some particles. So, it can be seen from the whole tracking experiment that the introduced approach accurately identifies the targets during the three tracking stages. This is due to the effective handling of the conflicting information provided by the location and colour cues during the second stage of tracking based on the efficient conflict strategy and excellent DSmT combination model.

Furthermore, the third video image came from a park square was executed to test the robustness of the introduced
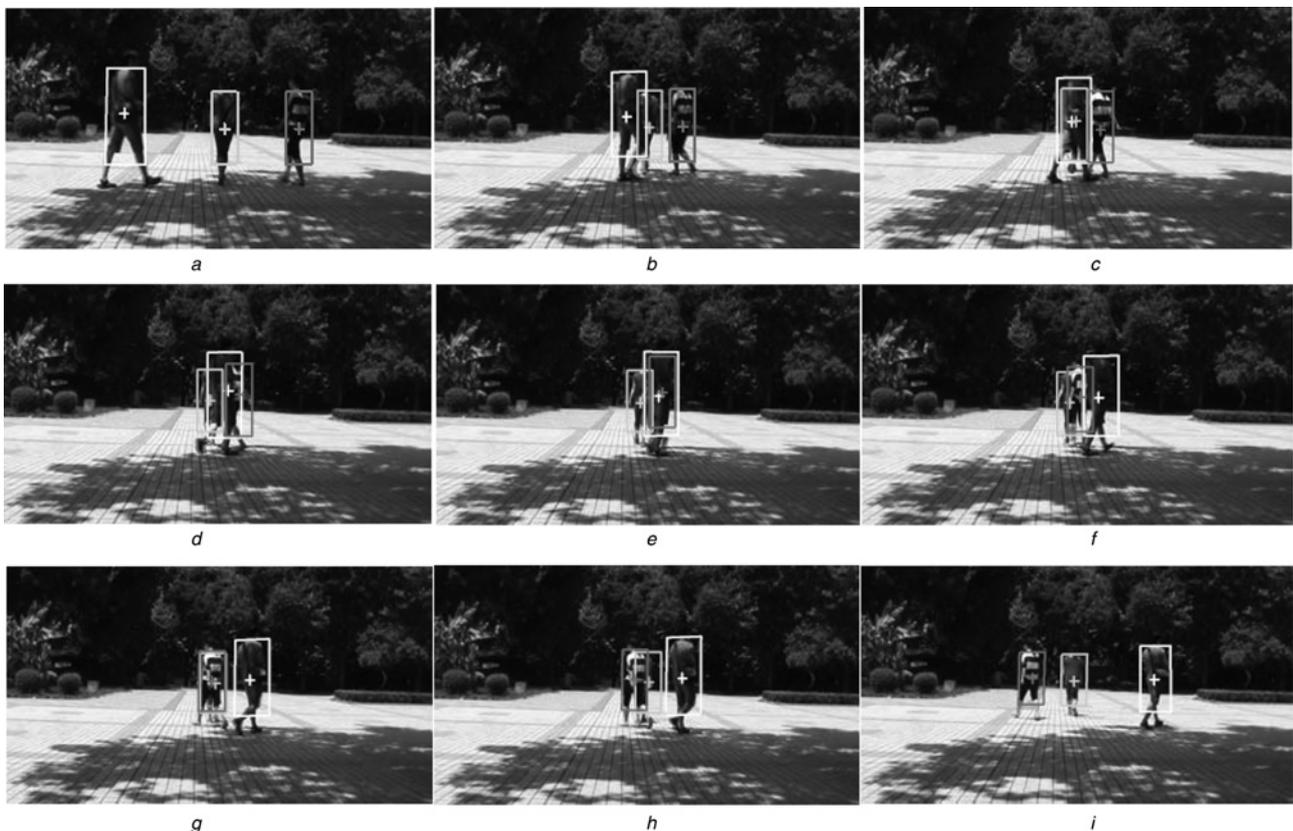
Fig. 11 *Tracking process of main frames of three moving targets and tracking result*
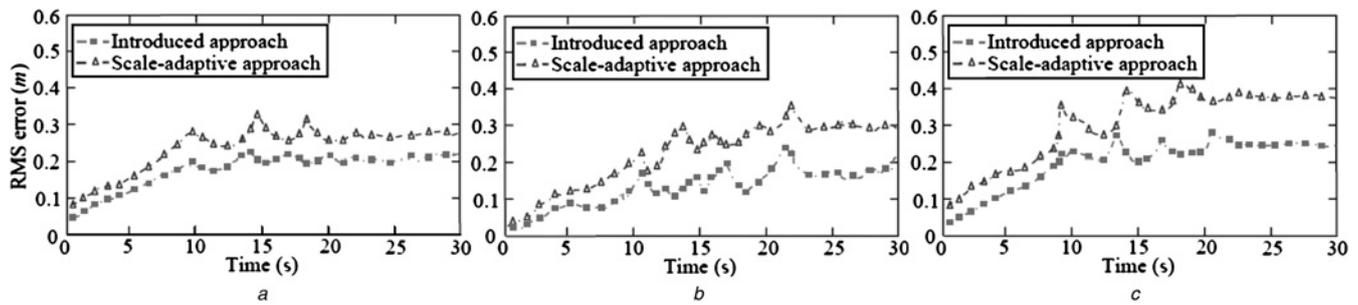
Fig. 12 *RMS error curves of estimated position*
a Target 1
b Target 2
c Target 3

approach. In the video scene, three pedestrians were walking into the square along different orientations from dark shadow area to bright area. Owing to the tree shadow, the brightness of three pedestrians had obvious change. Especially, the trees in the scene were shaking ceaselessly, the pedestrians were occluded each other time after time and the objects scale had great change. For this complicated scene, three pedestrians in the video scene were designated as three targets to finish the tracking experiment, and 85 particles were used to handle the high conflict between evidences or so. For the sake of clarity, the left person was denoted target 1, the middle person was denoted target 2 and the right person was denoted target 3 according to Fig. 11*a*. Finally, three targets in the scene were tracked accurately by the introduced approach. Fig. 11 shows the tracking process of main frames and tracking result. According to Fig. 11, although the illumination in the scene had obvious change, the colour of target 3 clothes was similar to the scene, three targets were crossed and occluded time after time, the introduced approach overcame these interference factors and the approach avoided the effect of illumination change and disturbance between targets. As a result, the tracking performance was greatly improved for handling high conflict evidences in a cluttered scene.

In order to estimate the robustness of the introduced approach for handling high conflict between evidences, a scale-adaptive tracking approach based on literature [24] was also applied to track the above image sequence. At the same time, the root-mean-square (RMS) error of estimated position was introduced to evaluate the performance of different tracking methods, and the corresponding expression of RMS error was given by

$$\text{RMS}_t = \sqrt{\frac{1}{m}\sum_{j=1}^{m}\left[\left(x_j(t)-\hat{x}_j(t)\right)^2 + \left(y_j(t)-\hat{y}_j(t)\right)^2\right]} \quad (36)$$

where m denotes the number of tracking simulation, $(x_j(t), y_j(t))$ denotes the real position at time t in jth experiment, $\left(\hat{x}_j(t), \hat{y}_j(t)\right)$ denotes the estimated position at time t in jth experiment.

As a result, the RMS errors of estimated position for every target were analysed by the 100 tracking experiments, and Fig. 12 shows the RMS error curves of the two approaches during tracking experiments. Owing to the illumination change and confusion between targets, the processing time of the two approaches was increased. Finally, the tracking performance of the two approaches both had some decline with the mutation of video scene according to Fig. 12.

However, the tracking ability of the introduced approach was better than the approach [24]. The approach [24] mainly partitioned the reference object into several sub-regions by clustering in colour space, then the colour distribution of each sub-region was modelled as the Gaussian, and its location constituted the spatial constrain on the layout of the object. Finally, the reference model was integrated into the PF to search for the object location and detect the scale change of the object. Experimental results showed that the approach had good reliability for the scale change of the objects in a cluttered scene. However, the measurement errors of three targets were increased with the scene mutation, and the tracking errors of three targets were also increased markedly by comparing the two approaches. Particularly, the RMS errors were very bigger when three objects were completely occluded many times in 30 s or so and the brightness in the scene was changed obviously. Hence, it was very difficult for general scale-adaptive approaches to effectively settle these high conflict issues between evidences, and the tracking capability was declined markedly. However, the introduced approach had excellent robustness to the illumination and scale changes and complete occlusions between objects, tracking accuracy and stability were better than the approach [24], and the tracking capability was also greatly improved.

5 Conclusions

This article investigated the multiple targets tracking in natural environment based on DSmT by a series of experiments, and the suggested conflict strategy and DSmT combination model have been tested and evaluated. In view of the results obtained and low computational complexity, the suggested approach is suitable for real-time video-based targets tracking. The following conclusions can be drawn:

1. DSmT is a useful theory for dealing with uncertainty problems. Study showed that the DSmT could be used to handle multiple targets tracking problem in cluttered scenes, and the dynamic combination model combined a PF was developed to actualise the video-based tracking. Experimental results have been demonstrated that the introduced approach ameliorated the interference immunity when tracking multiple targets. Especially, the tracking accuracy and robustness can be improved, but the real-time characteristics of video image have not been affected. Therefore, the approach is useful for dealing with high conflict between evidences and improving the performance

of PF, and the approach can be successfully applied in dynamic targets tracking in natural environment.

2. On the basis of establishing conflict strategy and combination model, the basic framework and algorithm of fusing multi-source information were described. The multiple targets tracking platform, which embedded location and colour cues into the PF, was developed in the framework of DSmT, and this strategy helped the platform to track the objects smoothly. As a result, three sets of experiments including many difficult tracking scenes with comparisons were carried out to validate the approach. The results showed that the approach exhibited a more significant performance for tracking robustness than some conventional PF and scale-adaptive approaches, and it had the ability to track an interesting target. Hence, the approach can easily be generalised to deal with additional cues and targets in cluttered scenes.

3. The tracking results with comparisons to other representative methods had better adaptation to the target of cross and occlusion, background variation and illumination change in challenging situations. Especially, by comparison with the variation of particle number in different tracking stage, the results demonstrated that the number of particle was variable with the conflict level between evidences, and the maximum number of particles was 50 when the two targets were covered wholly. At the meantime, the variation of average values of the confidence levels for all particles was also discussed, and the maximum average value of confidence level was about 1 and 0.9 during occluded and occluding stage, respectively. Furthermore, by analysing the RMS errors of estimated position for every target, the tracking ability and accuracy of the introduced approach was very excellent.

4. Although an enhanced video-based tracking platform was established, a known issue in the suggested approach was that object detection and mutation phenomena in complicated scene had not been discussed in detail. Therefore a comprehensive study in more realistic cases for larger number of targets should be discussed. Further research can not only promote the development of multiple targets tracking technique, but also have very important theoretical meaning and practical value for pushing the applied research of video-based target tracking, and this work will be reported in a future publication.

6 Acknowledgment

7 References

1 Han, Z., Jiao, J., Zhang, B., Ye, Q., Liu, J.: 'Visual object tracking via sample-based adaptive sparse representation', *Pattern Recognit.*, 2011, **44**, pp. 2170–2183

2 Kazuhiro, H.: 'Adaptive weighting of local classifiers by particle filters for robust tracking', *Pattern Recognit.*, 2009, **42**, pp. 619–628

3 Leonid, R., Michael, R., Ehud, R.: 'Dimensionality reduction using a Gaussian process annealed particle filter for tracking and classification of articulated body motions', *Comput. Vis. Image Underst.*, 2011, **115**, pp. 503–519

4 Pan, J., Hu, B., Zhang, J.Q.: 'Robust and accurate object tracking under various types of occlusions', *IEEE Trans. Circuits Syst. Video Technol.*, 2008, **18**, pp. 223–236

5 Godsill, S.J., Vermaak, J.: 'Models and algorithms for tracking of maneuvering objects using variable rate particle filters', *Proc. IEEE*, 2007, **95**, pp. 925–930

6 Abdallah, F., Gning, A., Bonnifait, P.: 'Box particle filtering for nonlinear state estimation using interval analysis', *Automatica*, 2008, **44**, pp. 807–808

7 Crisan, D., Obanubi, O.: 'Particle filters with random resampling times', *Stoch. Process. Appl.*, 2012, **122**, pp. 1332–1368

8 Maroulas, V., Stinis, P.: 'Improved particle filters for multi-target tracking', *J. Comput. Phys.*, 2012, **231**, pp. 602–611

9 Whiteley, N., Singh, S., Godsill, S.: 'Auxiliary particle implementation of probability hypothesis density filter', *IEEE Trans. Aerosp. Electron. Syst.*, 2010, **46**, pp. 1437–1454

10 Punithakumar, K., Kirubarajan, T., Sinha, A.: 'Multiple-model probability hypothesis density filter for tracking maneuvering targets', *IEEE Trans. Aerosp. Electron. Syst.*, 2008, **44**, pp. 87–98

11 Chorin, A.J., Morzfeld, M., Tu, X.: 'Implicit filters for data assimilation', *Commun. Appl. Math. Comput. Sci.*, 2010, **5**, pp. 221–240

12 Papadakis, N., Bugeau, A.: 'Tracking with occlusions via graph cuts', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011, **33**, pp. 144–157

13 Ottlik, A., Nagel, H.H.: 'Initialization of model-based vehicle tracking in video sequences of inner-city intersections', *Int. J. Comput. Vis.*, 2008, **80**, pp. 211–225

14 Chen, C., Lin, X., Shi, Y.: 'Moving object tracking under varying illumination conditions', *Pattern Recognit. Lett.*, 2006, **27**, pp. 1632–1643

15 Dezert, J., Smarandache, F.: 'Advances and applications of DSmT for information fusion' (American Research Press, Rehoboth , 2004)

16 Tehamova, A., Dezert, J., Semerdjiev, T.: 'Multi-target tracking applications of Dezert-Smarandache theory'. NATO Advanced Study Institute, Albena, Bulgaria, 2005, pp. 66–69.

17 Qiansu, Q.: 'Modified combination rule based on DSmT', *Electron. Meas. Technol.*, 2010, **33**, pp. 53–55

18 Yang, H.-c., Chen, K., Yang, J.-a.: 'Fusion of communication interception information based on DSmT', *Comput. Eng. Appl.*, 2010, **46**, pp. 129–132

19 Fang, Y., Wang, Y., Jin, W.: 'Multiple moving targets tracking research in cluttered scenes', *Procedia Eng.*, 2011, **16**, pp. 54–58

20 Sun, J., Wang, Y.: 'Research on moving object tracking under occlusion conditions'. IEEE Int. Conf. on Management and Service, 2009, pp. 1–4

21 Amlampalam, M.S., Maskell, S., Gordon, N., Clapp, T.: 'A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking', *IEEE Trans. Signal Process.*, 2002, **50**, pp. 174–188

22 Lao, Y., Zhu, J., Zheng, Y.F.: 'Sequential particle generation for visual tracking', *IEEE Trans. Circuits Syst. Video Technol.*, 2009, **19**, pp. 1365–1378

23 Yin, S., Na, J.H., Choi, J.Y., Oh, S.: 'Hierarchical Kalman-particle filter with adaptation to motion changes for object tracking', *Comput. Vis. Image Underst.*, 2011, **115**, pp. 885–900

24 Wang, S.P., Ji, H.B.: 'Adaptive scale appearance model for object tracking', *J. Comput. Aided Des. Comput. Graph.*, 2008, **20**, pp. 1466–1470

25 Dubois, D., Prade, H.: 'Represent and combination of uncertainty with belief functions and possibility measures', *Comput. Intell.*, 1988, **4**, pp. 244–264

26 Levr, E.E., Colot, O.: 'A generic framework for resolving the confliction the combination of belief structures'. Third Int. Conf. on Information Fusion, Paris, France, July 2000