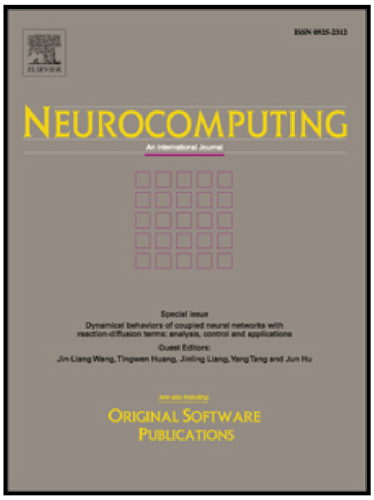
Accepted Manuscript



Fuzzy Logic Approach to Visual Multi-object Tracking

Li Liang-qun , Zhan Xi-yang , Liu Zong-xiang , Xie Wei-xin

PII: S0925-2312(17)31819-2
DOI: 10.1016/j.neucom.2017.11.060
Reference: NEUCOM 19125

To appear in: *Neurocomputing*

Received date: 30 December 2016
Revised date: 8 November 2017
Accepted date: 28 November 2017

Please cite this article as: Li Liang-qun , Zhan Xi-yang , Liu Zong-xiang , Xie Wei-xin , Fuzzy Logic Approach to Visual Multi-object Tracking, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.11.060

Highlights

- A fuzzy logic data association approach is proposed .
- The association probabilities are substituted by the fuzzy membership degrees.
- A track-to-track association approach based on the fuzzy synthetic function is proposed.

Fuzzy Logic Approach to Visual Multi-object Tracking

Li Liang-qun, Zhan Xi-yang, Liu Zong-xiang, Xie Wei-xin

ATR Key Laboratory, Shenzhen University, Shenzhen, Guang Dong 518060,China

Abstract—In this paper, a novel fuzzy logic data association algorithm is proposed for online visual multi-object tracking. Firstly, in the proposed algorithm, in order to incorporate expert experience into the data association for the improvement of performance in multi-object tracking, a fuzzy inference system based on knowledge is designed by using a set of fuzzy if-then rules. Given the error and change of error of motion, shape and appearance models in the last prediction, these rules are used to determine the fuzzy membership degrees that can be used to substitute the association probabilities between the objects and the measurements (or detection responses). Secondly, in order to deal with the fragmented trajectories caused by long-term occlusions, a track-to-track association approach based on the fuzzy synthetic function is proposed, which can effectively stitch track fragments (tracklets). Because of this, the proposed algorithm has the advantage that it does require no assumption of statistical models of measurement noise and of object dynamics. The experiment results on several public data sets show the efficiency and the ability to minimize the number of fragment tracks of the proposed algorithm.

Keywords: Visual Multi-object Tracking; Fuzzy Logic; Data Association; Track Management

1. Introduction

The objective of multi-object tracking is to estimate the current states of objects based on previous visual measurements up to the current time in a video sequence, such as positions, size, identification (ID), etc. It is very important for many computer vision tasks with applications such as automated surveillance, traffic safety, vehicle navigation, human computer interaction and robotics [1], [2], [3]. However, some practical challenges remain to be overcome for implementing this technology, such as object-to-object and object-to-scene occlusions, abrupt object motion, quick changes in lighting conditions, etc. With the development of the detection technology of object, detection-based multi-object tracking methods have been extensively studied [4], [5].

A key problem of the tracking-by-detection approach is the data association between visual measurements and multiple objects. When multiple objects are present, data association is the process of determining which visual measurements should be combined with the existing object trajectories. In order to solve the data association problem, many data association approaches have been proposed in recent decades, which are largely categorized into two types: a global optimization approach and a sampling-based approach. To obtain the global optimal solution or sub-optimal solution, Park *et al.*[1] proposed a binary integer programming formulation for the data association problem which pursues the minimum cost data associations among target measurements via one-to-one, one-to-m, and m-to-one associations. Wang *et al.*[2] proposed an iterative data association and detection update approach by using reliable temporal information to improve the accuracy of human detection, and a global data association to solve the association ambiguities. Yu *et al.*[3] presented a data driven Monte Carlo Markov Chain (MCMC) method to find a global optimal spatio-temporal data association. Milan *et al.* [6] used gradient descent to find strong local minima of complex nonconvex energy that captures image evidence and various physical constraints for tracking. Leibe *et al.* [7] proposed to perform coupled multiple-object detection and tracking by applying the minimum description length principle, formulate it as a QBP, and solve it by expectation maximization (EM)-based method. Benfold and Reid [8] employed MCMC to track multiple heads where motion is exploited to detect false positive (FP) detections. Wu *et al.* [9] combined a sparsity-driven detector with the network-flow data association technique for tracking multiple objects in one or more static cameras. The main advantage of this kind of approach is that it performs the data association problem when a set of the associated measurements are collected, so it generally provides a better solution even in complex scenes, such as long-term occlusions, abrupt motion of objects. However, in order to obtain the globally optimal solution from all time frames of video, all of these approaches require intensively computational load, so it cannot be used to solve the data association problem in real-time tracking system.

In the online multiple object tracking system, the sampling-based data association approaches have been widely employed. In these methods, the posterior probability of the data associations is approximated by a number of possible data associations with certain posterior weights, and some filtering algorithms can be incorporated to improve the performance of tracker. The sampling-based approaches differ in terms of how they deal with data association. Shu *et al.*[5] proposed a multi-object tracking method called Multi-object Tracking with inter-feedback between Detection and Tracking (MTDT). Yoon *et al.*[10] proposed a data association method that effectively exploits structural motion constraints in the presence of large camera motion for multi-object tracking(SCEA). Fagot-Bouquet *et al.*[11] proposed a multi-frame data association step by using a structured sparsity-inducing norm to compute representations more suited to the tracking context(LINF1). Rasmussen *et al.*[12] extended the application of the PDAF and JPDAF from the radar and sonar tracking field to visual tracking by defining measurements suitably and devising a preprocessing step to extract them. Rezatofighi *et al.*[13] revisited the JPDA algorithm (JPDA_m) and proposed an efficient and accurate approximation. Tinne *et al.*[14] developed an online two-level multi-object tracking and detection method using the shape information to help distinguish several interacting targets, and utilizing the JPDA to deal with the association problem between measurements and targets. Hu *et al.*[15] extended the sparse representation framework with multi-feature joint optimization. Breitenstein *et al.*[16] used a pretrained detector and an online-trained classifier to construct the confidence map of objects for the drift problems under occlusions. Shu *et al.*[17] employed deformable part models for describing appearances of objects. Yi *et al.*[18] designed a hierarchical data association tracking (HDAT) framework with branch partition, candidate upgrading and incremental motion pairing inference. In order to solve the track-to-track problem of the fragmented trajectories under long-term occlusions, Bae *et al.*[19,20]] proposed a robust online multi-object tracking method based on tracklet confidence and online discriminative appearance learning (TC_ODAL). Moreover, with the rise of deep learning in recent years, the convolutional neural network (CNN) was also widely applied to the multi-object tracking [21-23]. Milan *et al.*[21] presented a novel approach to online multi-target tracking based on recurrent neural network**s** **(**RNN_LSTM**).** In order to tune the parameters, their approach proposed an end-to-end learning approach for online multi-target tracking. Laura *et al.*[22] proposed a two-stage Siamese convolutional neural network learning based approach to associate detections within the context of pedestrian tracking (SiameseCNN).

The methods mentioned above were mainly based on the Bayesian framework proposed. Since fuzzy theory arises, it has been widely applied in a large range of areas such as information fusion system, nonlinear filtering system and image processing. Kim *et al.*[24] developed a fuzzy object detection algorithm based on target segmentation for tracking a moving object in an image sequence. Yoon *et al.*[25] proposed a visual tracking system by using adaptive fuzzy particle filter framework for complex scenes. Shandiz *et al.* [26] presented a particle filtering approach in which particles are weighted using a fuzzy based color model for object which discriminates between background and foreground elements. Chen and Huang [27] proposed a data association algorithm based on fuzzy-logic called fuzzy data association (FDA) for radar/infrared sensor data fusion. In [28], Ashraf proposed an all-neighbor fuzzy association approach for tracking multiple targets in a cluttered environment. Their approach is similar in form to joint probabilistic data association filter except that the probability weights are replaced by fuzzy weights. Li and Xie [29] proposed a data association algorithm based on intuitionistic fuzzy set for multi-target tracking in a cluttered environment, which reconstructed the joint association probabilities in JPDAF by utilizing the degree of membership of the measurement belonging to the target based on a new intuitionistic fuzzy clustering method. Nuno *et al.*[30] designed a hierarchical fuzzy logic tracking framework for single and multiple visual targets, in which a fuzzy inference engine based on fuzzy rule base was constructed to decide whether the measurement is associate with the object. Thomas *et al.* [31] used an adaptive Gaussian mixture model for background modeling and a fuzzy sequential Monte-Carlo-based tracking algorithm for tracking multiple objects under varying illumination. Paúl *et al.*[32] proposed a fuzzy logic approach based on color, stereo vision for the detecting and tracking of multiple objects. Li Jun *et al.* [33] proposed a frame-by-frame data association algorithm based on intuitionistic fuzzy sets for multiple object tracking. In their algorithm, the association costs between targets and measurements were replaced by the intuitionistic fuzzy membership degrees that can be obtained by a modified intuitionistic fuzzy c-means clustering algorithm based on intuitionistic fuzzy point operator. In this paper, fuzzy logic approach is used to solve the data association problem between the visual measurements and the objects for visual multi-object tracking in a proper way. The goal is to correctly detect appearing and disappearing objects and to obtain a record of the trajectories and sizes of objects over time, maintaining a correct identification for each object throughout.

The main contributions of this paper are summarized as follows:

1) A novel fuzzy logic data association method for online visual multi-object tracking is proposed. In order to deal with the data association problem in complex environment, the error and the change of error of motion, shape and appearance models are used to construct the fuzzy input variables and fuzzy inference system. By using the fuzzy inference system, the association probabilities between measurements and objects are replaced by the fuzzy membership degrees, which can incorporate reasoning based on the fuzzy rule base in the same sense as human reasoning.

2) To cope with the track-to-track association problem of fragmented trajectories caused by frequent occlusions, an adaptive track-to-track association method based on fuzzy synthetical function is proposed for track birth, death and merging. According to the fuzzy synthetical similarity degree obtained, the method can effectively merge the terminated tracks and the new tracks belonging to the same objects, so that the track can maintain a correct identification.

3) Finally, Multi-object tracking experiments on several publicly available challenging video show that the proposed algorithm has a good performance for visual multiple object tracking.

The rest of this paper is organized as follows. Section 2 describes the proposed multi-object tracking algorithm. Experiment results that compare the performances of all algorithms are presented in Section 3. Finally, some conclusions are provided in Section 4.

2. Proposed Multi-Object Tracking Algorithm

Fuzzy logic provides a framework and flexibility to couple human judgment with standard mathematical, and can deal with engineering problems, which are too complex or ill-defined to yield analytical solutions in a simple and robust way. In this section, a novel online multi-object tracking algorithm based on fuzzy logic theory is proposed. The block diagram of the proposed multi-object tracking algorithm is shown in Fig.1. From Fig.1, it shows that the proposed multi-object tracking algorithm consists of the following steps: 1) Fuzzy logic data association; 2) Track state update; 3) Track Management; 4) Model learning and prediction. In the proposed algorithm, we mainly focus on solving the data association problem, while the track states are estimated by using Kalman filtering, and the multi-feature model learning method is similar to [20].

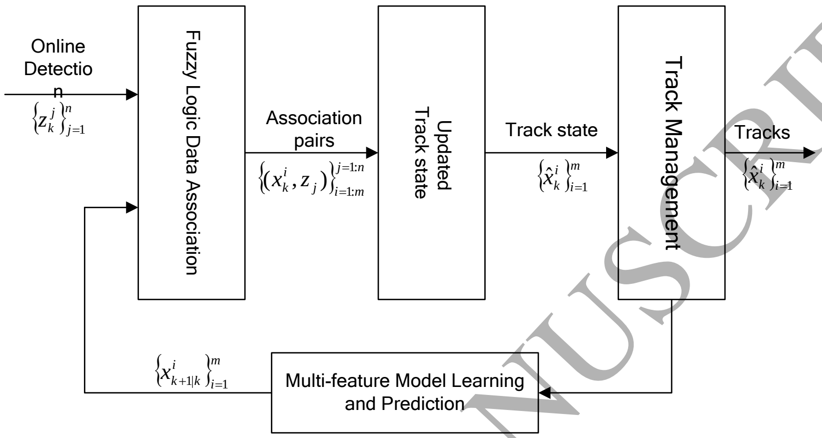


Fig.1. Block diagram of the proposed multi-object tracking algorithm

2.1. Fuzzy Logic Data Association

To calculate the fuzzy association probabilities $\mu_{i,j}(k)$ between the objects $\{\hat{x}_{k-1}^i\}_{i=1}^m$ and measurements $\{z_k^j\}_{j=1}^n$ (or detection responses), a fuzzy inference system (FIS) is designed based on the affinities of motion, shape, RGB color histogram and local binary pattern (LBP). Fig.2 shows the fuzzy inference system scheme, where $E_{i,j}(k)$ and $\Delta E_{i,j}(k)$ denote the prediction errors and change of errors of feature models, respectively. The FIS contains four basic elements: fuzzifier of input variables, fuzzy knowledge-base, fuzzy inference engine and defuzzifier. Therefore, we can add some human knowledge into multi-object tracking system to improve the performance of data association algorithm.

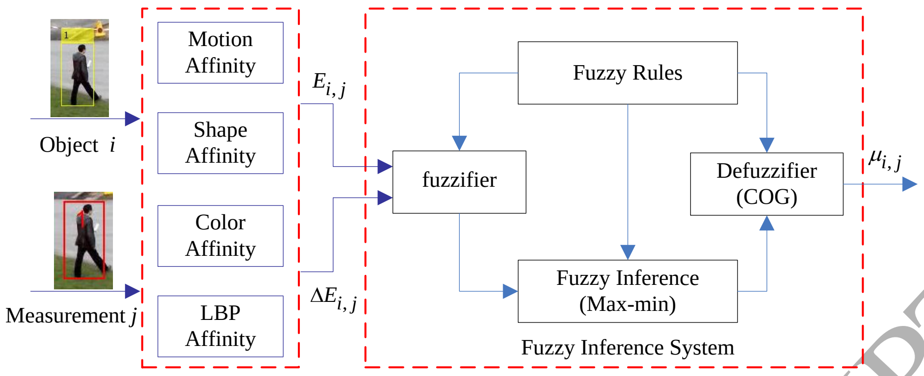


Fig.2. Fuzzy inference system scheme for $\mu_{i,j}$

1) System variables

Suppose that the predicted state of object i at the kth frame is $X=\{\hat{x}_i(k),s_i(k)\}_{i=1}^{m}$, where $\hat{x}_i(k)=(x_i(k),v_{x_i}(k),y_i(k),v_{y_i}(k))$, $s_i(k)=(h_i(k),w_i(k))$. $x_i(k)$ denotes the x-coordinate of the object's prediction position, $y_i(k)$ denotes the y-coordinate of the object's prediction position, $v_{x_i}(k)$ and $v_{y_i}(k)$ denote the corresponding velocities of object i, respectively, $h_i(k)$ and $w_i(k)$ denote the height and width of object i ,respectively.

Given detection responses, we denote the set of all measurements at the frame k as $Z=\{z_j(k),s_j(k)\}_{j=1}^{n}$, where $z_j(k)=(z_x^j(k),z_y^j(k))$, $s_j(k)=(z_h^j(k),z_w^j(k))$. $z_x^j(k)$, $z_y^j(k)$, $z_h^j(k)$ and $z_w^j(k)$ denote the x-coordinate, y-coordinate, height and width of the measurement j ,respectively. The rules of the fuzzy data association approach are expressed in terms of two input variables and an output variable. The input variables $E_{i,j}(k)$ and $\Delta E_{i,j}(k)$ are defined in terms of the prediction errors and change of errors of motion, shape and appearance models. Firstly, the normalized prediction errors of motion and shape models are defined as follows:

$$E_{i,j}^{M}(k)=\begin{cases}1, & If\ \|z_j(k)-\bar{z}_j(k)\|_2>\|z_j(k)-H_k\cdot\hat{x}_i(k-1)\|_2\\ \dfrac{\|z_j(k)-\bar{z}_j(k)\|_2}{\|z_j(k)-H_k\cdot\hat{x}_i(k-1)\|_2}, & If\ \|z_j(k)-\bar{z}_j(k)\|_2<\|z_j(k)-H_k\cdot\hat{x}_i(k-1)\|_2\\ 0, & If\ \|z_j(k)-\bar{z}_j(k)\|_2=\|z_j(k)-H_k\cdot\hat{x}_i(k-1)\|_2\end{cases}\tag{1}$$

$$E_{i,j}^{S}(k)=\begin{cases}1, & If\ |h_i(k)-z_h^j(k)|>3\sigma_s\\ \left|\dfrac{h_i(k)-z_h^j(k)}{3\sigma_s}\right|, & If\ |h_i(k)-z_h^j(k)|<3\sigma_s\\ 0, & If\ |h_i(k)-z_h^j(k)|=3\sigma_s\end{cases}\tag{2}$$

where $\bar{z}_j(k)=H_k\cdot F_k^i\cdot\hat{x}_i(k-1)$ denotes the predicted measurement at time k, $H_k\cdot\hat{x}_i(k-1)$ denotes the predicted measurement at time k-1, F_k^i denotes a 4×4 state transition matrix, H_k^i denotes a 2×4 measurement transition matrix. F_k^i and H_k^i are defined as follows:

$$F_k^i = \begin{bmatrix} 1 & \tau & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \tau & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad H_k^i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{3}$$

where $\tau = 1$ denotes the sampling period.

Appearance is an important cue for the data association in MOT. In order to make the appearance model more robustness, the RGB color feature and the local binary pattern (LBP) feature are used to capture the statistical and texture information of object region. In the RGB color histogram, to satisfy the low-computational cost imposed by real-time processing discrete densities, m-bin histograms will be used. Then, we have

RGB color model of object i: $\hat{H}_c(x_i) = \{H_m(x_i)\}_{m=1,2,\ldots,N}$

RGB color model of measurement j: $\hat{H}_c(z_j) = \{H_n(z_j)\}_{n=1,2,\ldots,N}$

where $H_m(x_i)$ is the number of pixels of object i in the mth color bin, $H_n(z_j)$ is the number of pixels of measurement j in the nth color bin. In order to estimate the similarity between object i and measurement j, we employ the correlation coefficient method to calculate the prediction error of the RGB color histogram.

$$E_{i,j}^{A_c}(k) = \frac{\sum_{m=1}^{N}(H_m(x_i)-\overline{H}_1)(H_m(z_i)-\overline{H}_2)}{\sqrt{\sum_{m=1}^{N}(H_m(x_i)-\overline{H}_1)^2 \sum_{n=1}^{N}(H_n(z_i)-\overline{H}_2)^2}} \tag{4}$$

where

$$\overline{H}_1 = \frac{1}{N}\sum_{m=1}^{N} H_m(x_i), \quad \overline{H}_2 = \frac{1}{N}\sum_{n=1}^{N} H_n(z_j)$$

On the other hand, in order to compute the prediction error of the LBP histogram, the Bhattacharyya coefficient between the LBP histogram of object i and the LBP histogram of measurement *j* is employed

$$\rho\left(\hat{H}_l(x_i),\hat{H}_l(z_j)\right)=\sum_u^N\sqrt{\hat{H}_u(x_i)\hat{H}_u(z_j)} \tag{5}$$

where $\hat{H}_l(x_i)$ and $\hat{H}_l(z_j)$ denote the LBP histogram of object *i* and measurement *j* ,respectively. ρ denotes the Bhattacharyya coefficient between $\hat{H}_l(x_i)$ and $\hat{H}_l(z_j)$. So the prediction error of the LBP can be defined as

$$E_{i,j}^{A_l}(k)=\sqrt{1-\rho\left(\hat{H}_l(x_i),\hat{H}_l(z_j)\right)} \tag{6}$$

and their values lie in [-1,1].

According to the definition of the prediction errors, the change of errors can be defined as

$$\Delta E_{i,j}^m(k)=E_{i,j}^m(k)-E_i^m(k-1),\quad m\in\{M,S,A_c,A_l\} \tag{7}$$

The range of values $[\Delta E_{i,j\,\min}^m(k),\Delta E_{i,j\,\max}^m(k)]$ is dependent on how large the deviation can be between the measurement model and predicted model of the object at the k *th* time. $\Delta E_{i,j\,\min}^m(k)$ can theoretically be zero and $\Delta E_{i,j\,\max}^m(k)$ can be very large. Since different objects have different velocity, appearance and shape, the magnitude of $\Delta E_{i,j\,\max}^m(k)$ may vary from object to object. To design a general fuzzy variable for different objects, the change of errors is normalized as follows:

$$\Delta E_{i,j}^M(k)=\begin{cases}\dfrac{E_{i,j}^M(k)-E_i^M(k-1)}{E_i^M(k-1)}, & if\ \left|E_{i,j}^M(k)-E_i^M(k-1)\right|<\left|E_i^M(k-1)\right|\\ \dfrac{E_{i,j}^M(k)-E_i^M(k-1)}{\left|E_{i,j}^M(k)-E_i^M(k-1)\right|}, & if\ \left|E_{i,j}^M(k)-E_i^M(k-1)\right|>\left|E_i^M(k-1)\right|\\ 0, & if\ E_{i,j}^M(k)=E_i^M(k-1)\end{cases} \tag{8}$$

$$\Delta E_{i,j}^S(k)=\begin{cases}\dfrac{E_{i,j}^S(k)-E_i^S(k-1)}{E_i^M(k-1)}, & if\ \left|E_{i,j}^S(k)-E_i^S(k-1)\right|<\left|E_i^S(k-1)\right|\\ \dfrac{E_{i,j}^S(k)-E_i^S(k-1)}{\left|E_{i,j}^S(k)-E_i^S(k-1)\right|}, & if\ \left|E_{i,j}^S(k)-E_i^S(k-1)\right|>\left|E_i^S(k-1)\right|\\ 0, & if\ E_{i,j}^S(k)=E_i^S(k-1)\end{cases} \tag{9}$$

$$\Delta E_{i,j}^{A_c}(k)=\begin{cases}\dfrac{E_{i,j}^{A_c}(k)-E_i^{A_c}(k-1)}{E_i^{A_c}(k-1)}, & if\ \left|E_{i,j}^{A_c}(k)-E_i^{A_c}(k-1)\right|<\left|E_i^{A_c}(k-1)\right| \\ \dfrac{E_{i,j}^{A_c}(k)-E_i^{A_c}(k-1)}{\left|E_{i,j}^{A_c}(k)-E_i^{A_c}(k-1)\right|}, & if\ \left|E_{i,j}^{A_c}(k)-E_i^{A_c}(k-1)\right|>\left|E_i^{A_c}(k-1)\right| \\ 0, & if\ E_{i,j}^{A_c}(k)=E_i^{A_c}(k-1)\end{cases} \tag{10}$$

and

$$\Delta E_{i,j}^{A_l}(k)=\begin{cases}\dfrac{E_{i,j}^{A_l}(k)-E_i^{A_l}(k-1)}{E_i^{A_l}(k-1)}, & if\ \left|E_{i,j}^{A_l}(k)-E_i^{A_l}(k-1)\right|<\left|E_i^{A_l}(k-1)\right| \\ \dfrac{E_{i,j}^{A_l}(k)-E_i^{A_l}(k-1)}{\left|E_{i,j}^{A_l}(k)-E_i^{A_l}(k-1)\right|}, & if\ \left|E_{i,j}^{A_l}(k)-E_i^{A_l}(k-1)\right|>\left|E_i^{A_l}(k-1)\right| \\ 0, & if\ E_{i,j}^{A_l}(k)=E_i^{A_l}(k-1)\end{cases} \tag{11}$$

From Eq.(8) to Eq.(11), we can see that if $E_{i,j}^m(k)$ is equal to $E_{i,j}^m(k-1)$, the magnitude of $\Delta E_{i,j}^m(k)$ is set to zeros. If $E_{i,j}^m(k)$ is not equal to $E_{i,j}^m(k-1)$, we normalize the change of errors $\Delta E_{i,j}^m(k)$ by using the maximum value between $\left|E_{i,j}^m(k)-E_i^m(k-1)\right|$ and $\left|E_i^m(k-1)\right|$ as a reference standard. As a result, based on the prediction errors { $E_{i,j}^M(k)$, $E_{i,j}^S(k)$, $E_{i,j}^{A_c}(k)$ and $E_{i,j}^{A_l}(k)$ } and the change of errors{ $\Delta E_{i,j}^M(k)$, $\Delta E_{i,j}^S(k)$, $\Delta E_{i,j}^{A_c}(k)$ and $\Delta E_{i,j}^{A_l}(k)$ }, the input variables $E_{i,j}(k)$ and $\Delta E_{i,j}(k)$ can be defined as follows:

$$E_{i,j}(k)=\sqrt{\frac{{E_{i,j}^M}^2(k)+{E_{i,j}^S}^2(k)+{E_{i,j}^{A_c}}^2(k)+{E_{i,j}^{A_l}}^2(k)}{4}} \tag{12}$$

$$\Delta E_{i,j}(k)=\sqrt{\frac{{\Delta E_{i,j}^M}^2(k)+{\Delta E_{i,j}^S}^2(k)+{\Delta E_{i,j}^{A_c}}^2(k)+{\Delta E_{i,j}^{A_l}}^2(k)}{4}} \tag{13}$$

According to the definitions mentioned above, the universe of discourse and the range of values which $E_{i,j}(k)$ and $\Delta E_{i,j}(k)$ each may take is in the interval [0,1] regardless of the object types, object-to-object occlusions and abrupt object motion. The values of $E_{i,j}(k)$ and $\Delta E_{i,j}(k)$ that fall into this range represent crisp values of $E_{i,j}(k)$ and $\Delta E_{i,j}(k)$.

2) Membership Functions (fuzzifier)

The crisp values are mapped into some fuzzy sets defined in the universe of discourse of input and output. Generally, the more numbers of fuzzy sets we made, the more accuracy of output we got. But more numbers of fuzzy sets will increase the computational load of the proposed algorithm. Usually, the number of fuzzy sets will be decided by human experience. In the fuzzy inference system, five fuzzy sets that are labeled in the linguistic terms of zero(ZE), small positive (SP), medium positive (MP),large positive (LP) and very large positive (VP), are specified for each crisp input ($E_{i,j}(k)$and $\Delta E_{i,j}(k)$). Generally, we can employ the triangular and Gaussian membership functions to represent these fuzzy sets. In this paper, the triangular membership functions are employed. These membership functions of each crisp input are shown in Fig.3.

In Fig.3, triangular fuzzy sets whose width is set such that the membership function falls to zero at the center of each adjacent set. The membership functions of the fuzzy sets (such as SP, MP and LP) are defined by triangular functions at equal interval, which ensures that each input data has the same probability falling into each fuzzy set. The boundary sets are left open-ended to ensure that at least one model is applied for all values of U. To the membership functions of each of these fuzzy sets, when the prediction error $E_{i,j}(k)$ or change of error $\Delta E_{i,j}(k)$ is less than or equal to 0.1, we think that the detection response is reliable; when the prediction error $E_{i,j}(k)$ or change of error $\Delta E_{i,j}(k)$ is more than or equal to 0.9, we think that the detection response is unreliable, and the cores are set to be in [0, 0.1] and [0.9, 1], respectively. So we decided that the membership functions of the fuzzy sets ZE and VP are defined by the trapezoidal functions.

Compared with the Gaussian fuzzy sets, the triangular fuzzy sets require more less computational load. Since only fuzzy rules with nonzero weighting need to be computed, from Fig.3, we can see that at most two fuzzy sets are required at each step. If it is necessary to increase the coverage of the model range by adding more linguistic terms, the proposed algorithm still only uses two fuzzy sets at each step. Hence good coverage of the model range can be achieved by increasing the number of linguistic terms without a correspondingly large increase in computational load. On the other hand, since

Gaussian fuzzy sets have a much wider range of support than triangular fuzzy sets, the number of fuzzy sets operating is generally increased when these are employed, which will lead to the increasing of computational load.

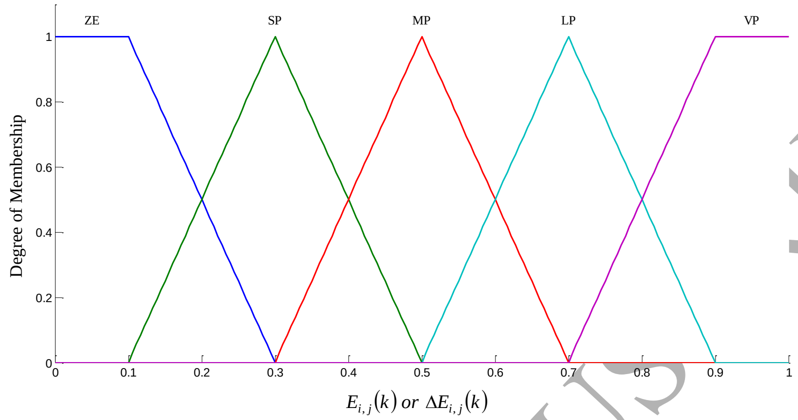


Fig.3 Membership functions for the fuzzy sets of $E_{i,j}(k)$ and $\Delta E_{i,j}(k)$

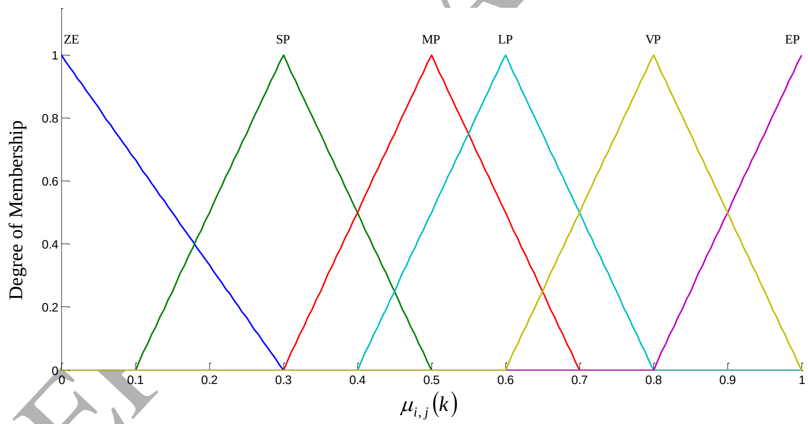


Fig.4 Membership functions for the fuzzy sets of $\mu_{i,j}(k)$

Moreover, unlike that of the input data, the output data have six fuzzy sets labeled in the linguistic terms of ZE, SP, MP, LP, VP and extremely large positive (EP). The membership functions of the fuzzy sets are defined by the triangular functions and the core of these fuzzy sets are not equally spaced. The membership functions of output are shown in Fig. 4.

3) Fuzzy rules

According to the input and output variables defined above, the fuzzy rules can be expressed by employing fuzzy IF-THEN rules as follows:

Rule 1:IF $E_{i,j}(k)$ is ZE AND $\Delta E_{i,j}(k)$ is ZE THEN $\mu_{i,j}(k)$ is EP

Rule 2:IF $E_{i,j}(k)$ is ZE AND $\Delta E_{i,j}(k)$ is SP THEN $\mu_{i,j}(k)$ is VP

Rule 3:IF $E_{i,j}(k)$ is ZE AND $\Delta E_{i,j}(k)$ is MP THEN $\mu_{i,j}(k)$ is LP ……

Because of the complexity of environment and the inaccuracy of object detector, there are some confliction and ambiguity in data association. For example, one object may be associated with multiple measurements; the objects may appear and disappear at any time and in any place. In order to deal with these problems, each fuzzy rule should exhibit the following characteristics: when the object abrupt motions or occlusions in appearance, the prediction error $E_{i,j}(k)$ is also abrupt larger. Hence, under the condition of fixed the change error $\Delta E_{i,j}(k)$, the fuzzy association probability $\mu_{i,j}(k)$ should decrease as the prediction error $E_{i,j}(k)$ increases. Fixed prediction error $E_{i,j}(k)$, the fuzzy association probability $\mu_{i,j}(k)$ should decrease as the change error $\Delta E_{i,j}(k)$ increases.

Table1. Fuzzy rule base for $\mu_{i,j}(k)$

$\mu_{i,j}(k)$		$E_{i,j}(k)$				
		ZE	SP	MP	LP	VP
$\Delta E_{i,j}(k)$	ZE	EP	VP	LP	SP	SP
	SP	VP	LP	MP	SP	SP
	MP	LP	MP	MP	SP	ZE
	LP	MP	MP	SP	SP	ZE
	VP	SP	SP	ZE	ZE	ZE

Tables 1 show the fuzzy rules of fuzzy inference system for the fuzzy association probability $\mu_{i,j}(k)$ .The output of the weight membership functions which are used to evaluate the crisp outputs through the Max–Min compositional rule of inference technique and center of gravity (COG) defuzzification.

2.2. Fuzzy Track-to-Track Association

Despite efforts to detect partially/completely occluded objects, missed detections/false measurements are still inevitable. Fuzzy logic approach can solve the data association problem of object tracking in case of no detection responses and false measurements in a short period. However, in a long-time occlusions or high probability of missed detection environment, when the object state is not updated by any measurements for a long time, the track of object is very difficult to be maintained. In order to deal with these problems, we propose a track-to-track association method based on fuzzy synthetic function to stitch track fragments (tracklets) before and after occlusions. Fig.5 shows the flow diagram of the proposed approach. In the proposed approach, a fuzzy synthetic affinity measure is designed to match the track fragments so that occlusions are solved by filling gaps after stitching track fragments.

Step1: Construction of the affinity matrix based on fuzzy synthetic function

Firstly, in order to associate the terminated track i with the new track j (or new measurement), we define sets of the terminated tracks and new tracks as $\left\{\hat{O}_{l_{0,i}}^{i}(k)\right\}_{i=1}^{n_{t_0}}$ and $\left\{N_{l_{1,j}}^{j}(k)\right\}_{j=1}^{n_{t_1}}$ at time k, where n_{t_0} and n_{t_1} denote the number of the elements of each set, respectively. $l_{0,i}$ denotes the terminated time of the track i. $l_{1,j}$ denotes the first time of the new track j, if $l_{1,j}=k$, it means the new track $N_{l_{1,j}}^{j}(k)$ is a new measurement, which is not associated with any existing tracks.

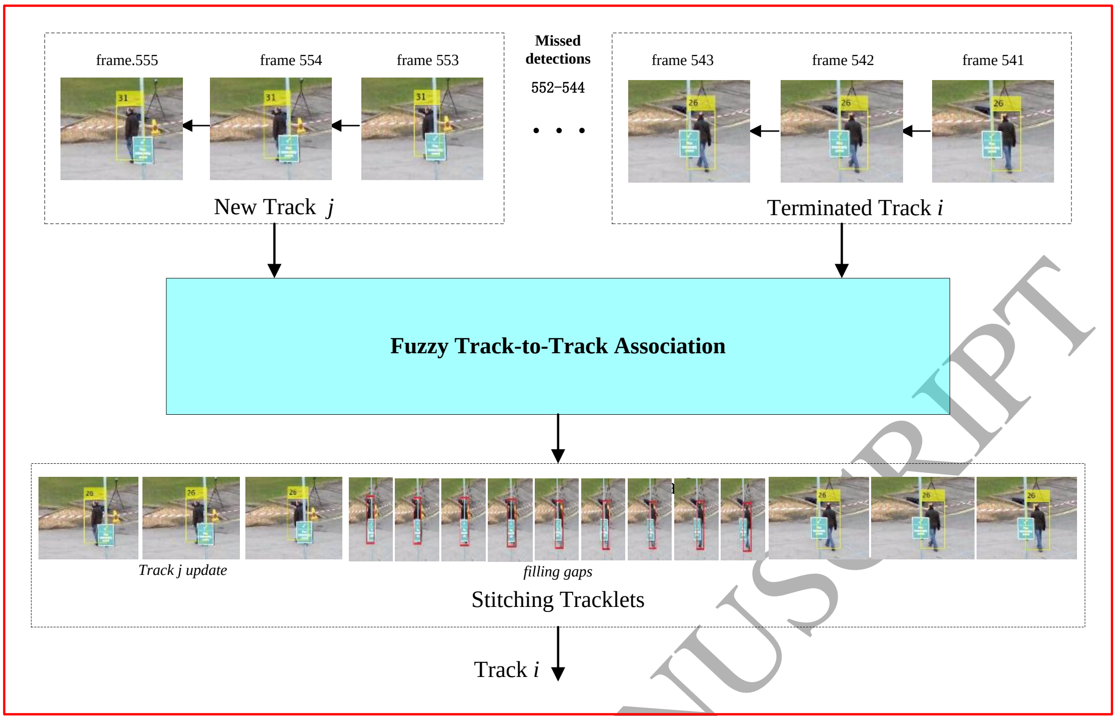


Fig.5 Flow diagram of fuzzy track-to-track association approach

Suppose $U_k = \{\mu_{ij}\}$ is a $n_{t_0} \times n_{t_1}$ similarity matrix, μ_{ij} is the fuzzy synthetic affinity degree between the terminated track i with the new track j. Several object features are used to compute the similarity degree μ_{ij}, such as motion, shape, appearance and velocity features. According to the definitions of $E^{M}_{i,j}(k)$, $E^{S}_{i,j}(k)$, $E^{A_c}_{i,j}(k)$ and $E^{A_l}_{i,j}(k)$, the affinities of motion, shape, color histogram, LBP and velocity are defined as follows:

$$f_M(i,j) = \exp\left(-\frac{E^{M}_{i,j}(k)^2}{2\sigma_m^2}\right) \tag{11}$$

$$f_S(i,j) = \exp\left(-\frac{E^{S}_{i,j}(k)^2}{2\sigma_s^2}\right) \tag{12}$$

$$f_{A_c}(i,j) = \exp\left(-\frac{E^{A_c}_{i,j}(k)^2}{2\sigma_{a_c}^2}\right) \tag{13}$$

$$f_{A_l}(i,j) = \exp\left(-\frac{E_{i,j}^{A_l}(k)^2}{2\sigma_{a_l}^2}\right) \tag{13}$$

$$f_V(i,j) = \exp\left(-\frac{\arctan((y_j - y_i)/(x_j - x_i)) - \arctan(v_{y_i}/v_{x_i})}{2\sigma_v^2}\right) \tag{14}$$

where $\sigma_m^2, \sigma_s^2, \sigma_{a_c}^2, \sigma_{a_l}^2$ and σ_v^2 denote the variances of object motion, shape, color histogram, LBP and velocity respectively, and are set to $\sigma_m^2 = 3, \sigma_s^2 = 3, \sigma_{a_c}^2 = \sigma_{a_l}^2 = 3, \sigma_v^2 = 100$. x_i and y_i denote the x-coordinate and y-coordinate of terminated track i, x_j and y_j denote the x-coordinate and y-coordinate of new track j. v_{x_i} and v_{y_i} denote the velocities along the x-coordinate and y-coordinate of terminated track i.

According to the affinities of motion, shape, appearance and velocity, the affinity vector $\Lambda_k(i,j)$ is defined as follows:

$$\Lambda_k(i,j) = (\bar{d}_M^k(i,j), \bar{d}_S^k(i,j), \bar{d}_{A_c}^k(i,j), \bar{d}_{A_l}^k(i,j), \bar{d}_V^k(i,j)) \tag{15}$$

where $\Lambda_k(i,j) \in [0,1]^5$, $\bar{d}_M^k(i,j)$, $\bar{d}_S^k(i,j)$, $\bar{d}_{A_c}^k(i,j)$, $\bar{d}_{A_l}^k(i,j)$ and $\bar{d}_V^k(i,j)$ denote the average affinities of the motion, shape, color histogram, LBP and velocity between the terminated track i and the new track j at time k, and are defined as follows, respectively.

$$\bar{d}_M^k(i,j) = \frac{1}{\kappa} \cdot \sum_{m=l_{0,i}-3}^{l_{0,i}} \sum_{n=l_{1,j}}^{k} f_M^{m,n}(i,j) \tag{16}$$

$$\bar{d}_S^k(i,j) = \frac{1}{\kappa} \cdot \sum_{m=l_{0,i}-3}^{l_{0,i}} \sum_{n=l_{1,j}}^{k} f_S^{m,n}(i,j) \tag{17}$$

$$\bar{d}_{A_c}^k(i,j) = \frac{1}{\kappa} \cdot \sum_{m=l_{0,i}-3}^{l_{0,i}} \sum_{n=l_{1,j}}^{k} f_{A_c}^{m,n}(i,j) \tag{18}$$

$$\bar{d}_{A_l}^k(i,j) = \frac{1}{\kappa} \cdot \sum_{m=l_{0,i}-3}^{l_{0,i}} \sum_{n=l_{1,j}}^{k} f_{A_l}^{m,n}(i,j) \tag{18}$$

$$\bar{d}_V^k(i,j) = \frac{1}{\kappa} \cdot \sum_{m=l_{0,i}-3}^{l_{0,i}} \sum_{n=l_{1,j}}^{k} f_V^{m,n}(i,j) \tag{19}$$

$$\kappa = k - l_{1,j} + 3$$

where $f_M^{m,n}(i,j)$, $f_S^{m,n}(i,j)$, $f_{A_c}^{m,n}(i,j)$, $f_{A_l}^{m,n}(i,j)$ and $f_V^{m,n}(i,j)$ denote the affinities of the motion, shape, color histogram, LBP and velocity between the terminated track i at time m and the new track j at time n, respectively.

A fuzzy affinity model based on fuzzy synthetic function[34] is developed to measure the matching degree between the terminated tracks and new tracks, with the aim of giving a high similarity score when any new track is associated with a terminated track, and is defined as follows:

$$S_k(\Lambda_k(i,j)) = \frac{1}{2}\left[\underset{l\in\{M,S,A_c,A_l,V\}}{\wedge} \bar{d}_l^k(i,j) + \underset{l\in\{M,S,A_c,A_l,V\}}{\vee} \bar{d}_l^k(i,j) \right] \tag{20}$$

where $\wedge$ is the minimum operator, $\vee$ is the maximum operator. According to the affinity model, the fuzzy synthetic affinity degree μ_{ij} can be defined as

$$\mu_{ij} = S_k(\Lambda_k(i,j)) = S_k((\bar{d}_M^k(i,j), \bar{d}_S^k(i,j), \bar{d}_{A_c}^k(i,j), \bar{d}_{A_l}^k(i,j), \bar{d}_V^k(i,j),)) \tag{21}$$

Step2: Rules of track-to-track association

According to the fuzzy affinity matrix U, because of the complexity of environment, there are some confliction and ambiguity in track-to-track association. For example, one measurement may be associated with multiple targets. In order to associate the terminated track i with the new track j, the maximum association degree and the track quality $m_{ij^*}(k)$ are defined as

$$\mu_{ij^*} = \max_{j\in\{1,2,\ldots,n_{t_1}\}} S_k(\Lambda_k(i,j)), \quad i \in \{1,2,\ldots,n_{t_0}\} \tag{22}$$

$$m_{ij^*}(k) = \begin{cases} m_{ij^*}(k-1)+1, & \text{if track } i \text{ is associated with } j^* \\ m_{ij^*}(k-1)-1, & \text{otherwise} \end{cases} \tag{23}$$

- If $\mu_{ij^*} \geq \varepsilon$, the terminated track i is associated with the new track j^*, and then the new track j^* cannot be associated any other terminated tracks. ε is a confidence degree threshold, and $0.5 \le \varepsilon \le 1$.

- If $m_{ij^*}(k) \ge 3$, the terminated track i and the new track j^* are considered as the same object, and stitch the track i and the track j^* by using the linear interpolation method to fill gaps and allocate the ID of track i into the stitching track. Finally, the terminated track i is recovered the existing track i .

2.3. Overall Visual Multi-object Tracking

According to the derived results mentioned above, the structure of the proposed online visual multi-object tracking algorithm can be summarized as follows. For the sake of clarity, the summary is given for a single forward run of the algorithm.

Algorithm: Online multi-target tracking based on fuzzy Logic Data Association(FLDA)

1. Initialization

Applying the offline-trained detector based on aggregated channel features [35] to obtain the measurements (detection responses) $\{z_k^j\}_{k=1:T}^{j=1:n}$ at the current frame.

2. State estimation

For k=1,2,…

- Data association: Find the best association pairs $\{(\hat{x}_{k-1}^i, z_k^j)\}$between the objects $\{\hat{x}_{k-1}^i\}_{i=1}^m$ and measurements $\{z_k^j\}_{j=1}^n$ by using the fuzzy logic data association algorithm described in Section 2.1.
- State update: Update the object states with the associated measurements by using the Kalman filtering.
- Track Management:

  a) Associated the terminated tracks $\{\hat{O}_{l_{0,i}}^i(k)\}_{i=1}^{n_{t_0}}$ with the new tracks $\{N_{l_{1,j}}^j(k)\}_{j=1}^{n_{t_1}}$ by using the fuzzy track-to-track association approach.

  b) Track initialization and termination as shown in Section 2.2.

End For

2.4 Discussion

1) In the proposed algorithm, in order to adaptively compute the association probability based on fuzzy rules, fuzzy logic has been incorporated into the data association of multi-object tracking. These rules determine the association probability $\mu_{i,j}(k)$ according to the magnitudes of the last prediction errors and change of errors of object feature models. To a certain extent they reflect the likelihood of, say, the tracking algorithm encountering object-to-object occlusions, abrupt object motion, quick changes in lighting conditions. Whenever more than one possible situation needs to be considered simultaneously, more than one rule is fired. The degrees at which these rules are fired are combined quantitatively to determine the final values of $\mu_{i,j}(k)$.

2) Unlike the approaches in [19,20], to link the fragmented track caused by long-term occlusions, they employed the track existence probability to construct the cost matrix, and determine the optimal association pairs by minimizing the total cost of the cost matrix using the Hungarian algorithm. In the proposed algorithm, an adaptive fuzzy track management method is proposed for track birth, death and merging, and a fuzzy synthetic affinity measure is designed to match the track fragments so that occlusions are solved by filling gaps after stitching track fragments. For this reason, the proposed algorithm can reduce the number of fragmented tracks when measurements of occluded objects are missed or inaccurate.

3) Track Initialization and Termination. In the online visual multi-object tracking problem, random number of measurements is received due to detection uncertainty or false alarms, and the objects may appear and disappear in any place and at any time. To track multi-object visual tracking, the track initialization and termination are very important. Consequently, several rules are defined as follows: **a)** If the measurement is not associated with any existing tracks or new tracks, this measurement will be considered as a new track and initialized by using the Kalman filtering. **b)** When the new track is consistently associated in $T_{\_init}$ frames, the new track is marked as an existing track. **c)** If the existing track is not associated with any measurement in subsequent $T_{\_term}$ frames, it is considered as the terminated track. **d)** If the terminated track is not associated with any measurement more than λ frames,

it is considered as the complete (ended) track.

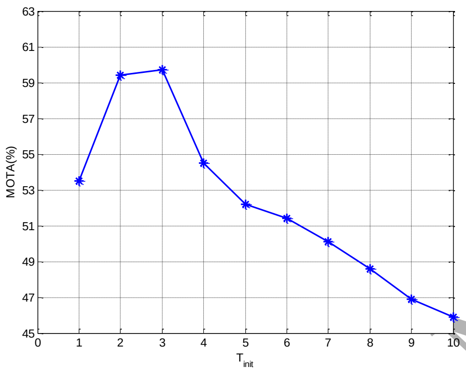
3. Experiment Results

3.1. Datasets and evaluation metrics

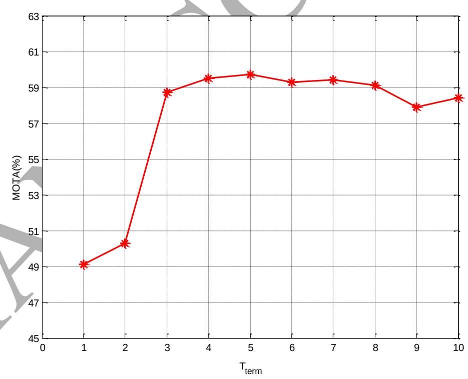
1) Datasets and detections: For the performance evaluation, we use some popular video sequences, including PETS09-S2L1, TUD-Stadtmitte, TUD-Crossing, PETS09-S2L2, ETH-Jelmoli and AVG-TownCentre. PETS09-S2L1and TUD-Stadtmitte datasets are the training sets, another four datasets are the test sets. In all datasets, the ETH-Jelmoli sequence of street scene taken by a moving camera was selected, other datasets shown from a stationary camera. The PETS.S2L2 dataset is much more difficult than other four datasets since the crowded density of this scenario is higher and long-time occlusions are caused by object interactions. In all experiments, we used the public detections given by Aggregate Channel Features (ACF) pedestrian detector [35].

2) Evaluation metrics: Comparisons with several state-of-the-art algorithms were made. Both qualitative analysis and quantitative evaluations were presented to show the effectiveness of our algorithms. The widely used CLEAR MOT metrics was employed to evaluate the proposed algorithm, including the Multi-object tracking accuracy (MOTA) that combines three kinds of errors. It is defined as $MOTA = 1 - (\sum_t (FP_t + FN_t + IDS_t))/(\sum_t m_t)$, where FPs denotes the number of false positives, FNs denotes the number of missed targets, IDS denotes the number of identify switches, m_t denotes the total objects, respectively, at frame t. The Multi-object tracking precision (MOTP) is average of the bounding box overlap over all tracked targets. We report the most tracked trajectories (MT) and lost trajectories (ML). Here Fragments (FG) denote the number of times that a ground truth trajectory is interrupted.

3) System parameters: In all experiments, a Kalman filtering based on constant velocity model is used to predict and update the pedestrian states in video sequences. All system parameters have been found experimentally. From an extensive evaluation, we find that most parameters do not affect the tracking performances of the proposed algorithm except for the number of frames $T_{\_init}$ for the new track initialization and the number of frames $T_{\_term}$ for the track termination. To evaluate the impact of the parameters $T_{\_init}$ and $T_{\_term}$ on the performance of object tracking, we manually tune them on training datasets (PETS09-S2L1 and TUD-Stadtmitte), and then fix them for all test datasets (TUD-Crossing, PETS09-S2L2, ETH-Jelmoli and AVG-TownCentre). Fig.6 show the multi-object tracking accuracy (MOTA) for different $T_{\_init}$ and $T_{\_term}$ values. From Fig.6(a), we can see that the proposed algorithm has the best performance for MOTA when the parameter T_{init} equals to 3 or the parameter T_{term} equals to 5. Moreover, the confidence degree threshold ε is set to 0.6, the threshold λ in section 2.4 is set to 60.



(a). MOTA for different T_{init} values



(b). MOTA for different T_{term} values

Fig.6 the impact of parameters T_{init} and T_{term} on the tracking performances

3.2. Tracking results of the proposed algorithm

To evaluate the performance of the proposed algorithm, Fig.7 and Fig.8 show the tracking results of the proposed algorithm without track management and with it on the PETS.S2L1 data set. To simplify, the proposed algorithm without track management is called FLDA, otherwise FLDA-TM. Comparing Fig.7 with Fig.8, it shows that the FLDA-TM algorithm outperforms the FLDA algorithm. From Fig.7, in frames 102,112,120 and 148, due to the existence of missed detections, the human at the up-left corner of picture is tracked with different identifications by the FLDA algorithm (identifies as 10, 12,14,19, respectively), whereas the FLDA-TM algorithm can track him successfully. The main reason is that with the track-to-track association method, the FLDA-TM algorithm can stitch track fragments (tracklets) caused by occlusions of object interaction or missed detections.

Fig.7. Results of the FLDA algorithm

Fig.8. Results of the FLDA-TM algorithm

Moreover, the evaluation results of the FLDA and FLDA-TM algorithms on two datasets are shown in Table 2. As expected, the performance of the FLDA-TM algorithm is better than that of the FLDA algorithm in term of all metrics. Particularly, the improvement percentage of the number of identify switches (IDS) metric exceeds 30% for all data sets.

Table 2. Performance comparison of the FLDA and FLDA-TM algorithms

Sequence	Method	MOTA (%↑)	MOTP (%↑)	MT (%↑)	ML (%↓)	FP(↓)	FN(↓)	IDS(↓)	FG(↓)
PETS.S2L1	FLDA	86.5	65.4	94.7	0	270	290	11	58
	FLDA-TM	87.2	65.8	94.7	0	237	280	7	56
TUD-Stadtmitte	FLDA	56.5	65.2	36	0	84	110	14	21
	FLDA-TM	57.7	65.5	39	0	73	98	8	18

3.3. Performance comparison with other approaches

In this section, in order to compare the performance of the proposed algorithm with other multiple object tracking algorithms, we chose seven reported state-of-the-art approaches with available corresponding publications for the same sequences on the **MOT Challenge**, including Structural constraint event aggregation for multi-object tracking (**SCEA**)[10], multi-frame data association step by using a structured sparsity-inducing norm (**LINF1**)[11], **JPDA_m**[13], online multi-object tracking method based on tracklet confidence and the online discriminative appearance learning (**TC_ODAL**)[19], multi-target tracking based on recurrent neural networks (**RNN_LSTM**)**[21]**, Siamese convolutional neural network learning based approach **(SiameseCNN)[22]**, multi-object tracking with quadruplet convolutional neural networks (**QuadMOT**)[23]. Table 3 shows the comparison results for all metrics on four sequences individually.

Firstly, the tracking results for the TUD-Crossing data set are compared. This data set is widely used in multi-object tracking literatures. Because the human density is low, though the dataset includes some nonlinear motion of objects and some proximity objects, all algorithms can track these objects with a high tracking accuracy. Fig.9 shows the results of the FLDA-TM algorithm. The comparison results are shown in Table 3. From Table 3, we can see that the FLDA-TM algorithm obtained better performance in terms of ML and FN. However, the QuadMOT[23] algorithm outperforms other algorithms in most metrics, such as MOTA, MOTP, ML,FP, etc.

For the PETS.S2L2 sequences, the human density is higher than the TUD-Crossing, and the objects are frequently occluded. The comparison results are shown in Table 3. As can be seen, the FLDA-TM algorithm exhibited significant improvements on the performance, and the FLDA-TM algorithm obtained the best performance in terms of MOTA, MT and FN scores. Particularly, the MOTA score is improved 11.2% and 22.1% compared with the QuadMOT algorithm and the SCEA algorithm. In addition, the proposed algorithm achieves the second best performance on the ML metric, and the IDS score outperforms the QuadMOT algorithm and the SCEA algorithm. Some visual tracking results of the FLDA-TM algorithm on the PETS.S2L2 sequence are shown in the second row of Fig.9.

For the AVG-TownCentre sequences, the FLDA-TM algorithm achieves the best performance in MOTA compared with other tracking approaches, and the MOTA scores has been greatly improved over 18.2% compared to the second best algorithm (QuadMOT). Moreover, the FLDA-TM algorithm achieves the second best performance among all tracking algorithms in terms of MOTP, ML and FP scores. As discussed above, the main reason is FLDA-TM's ability to robustly maintain object's identities by using the fuzzy track-to-track association resulting in higher MOTA and lower IDS. Finally, some visual tracking results of the FLDA-TM algorithm are shown in the final row of Fig.9.

For the ETH-Jelmoli sequences, the comparison results of all algorithms are shown in Table 4. Unfortunately, From Table 3, it can see that the FLDA-TM algorithm has lower performance on most metrics than the SCEA, LINF1, JPDA_m, SiameseCNN and QuadMOT algorithms. This is because the ETH-Jelmoli sequence of street scene taken by a moving camera, which make a continuous and rapid changes in the background of the image sequence, and lead to the degradation of the association probabilities between the measurements and objects. On the other hand, the FLDA-TM algorithm achieves the best performance in terms of FN scores. Some visual tracking results are shown in the last row Fig.9.

Table 3. Performance Comparison of Tracking Results on Four Different Datasets

Sequence	Method	MOTA (%↑)	MOTP (%↑)	MT (%↑)	ML (%↓)	FP(↓)	FN(↓)	IDS(↓)	FG(↓)
TUD-Crossing	SCEA	57.0	**74.2**	23.1	23.1	15	442	17	30
	LINF1	64.2	73.2	38.5	15.4	**12**	375	**8**	22
	JPDA_m	60.9	68.4	30.8	23.1	44	385	**2**	26
	TC_ODAL	55.8	72.8	23.1	**7.7**	110	360	17	35
	RNN_LSTM	57.2	71.7	30.8	15.4	81	348	43	49
	SiameseCNN	**73.7**	73.0	**69.2**	15.4	85	**197**	**8**	**10**
	QuadMOT	**72.1**	**74.5**	**46.2**	**0**	**5**	288	15	**21**
	FLDA-TM	64.2	72.8	5	**0**	80	**275**	26	33
PETS09-S2L2	SCEA	44.6	69.3	7.1	14.3	**536**	4633	175	289
	LINF1	37.7	70.2	2.4	26.2	**393**	5497	**114**	**202**
	JPDA_m	37.6	65.9	11.9	19.0	1016	4858	**139**	**260**
	TC_ODAL	30.2	69.2	2.4	19.0	1074	5375	284	499
	RNN_LSTM	38.3	**71.6**	9.5	14.3	1016	4611	320	417
	SiameseCNN	34.5	69.7	7.1	19.0	672	5364	282	424
	QuadMOT	**49.0**	**72.6**	**16.7**	**7.1**	686	**3947**	285	380
	FLDA-TM	**54.5**	67.4	**18.6**	**9.3**	720	**3801**	162	336
AVG-TownCentre	SCEA	29.3	69.6	**15.0**	42.9	738	4226	88	233
	LINF1	11.1	**70.4**	2.2	81.0	**106**	6241	**5**	34
	JPDA_m	18.3	66.8	4.4	63.3	**258**	5561	**23**	**108**
	TC_ODAL	1.6	69.0	0.0	71.7	899	6111	26	114
	RNN_LSTM	13.4	68.8	3.5	41.2	1206	4682	299	414
	SiameseCNN	19.3	69.0	4.4	44.7	698	4927	142	289
	QuadMOT	**30.8**	69.8	**18.1**	**31.4**	1191	**3643**	111	409

	FLDA-TM	**36.4**	**70.0**	14.9	**38.6**	483	**3971**	89	214
ETH-Jelmoli	SCEA	40.4	72.6	15.6	33.3	312	1179	22	50
	LINF1	36.5	**73.5**	15.6	48.9	**169**	1430	**11**	**35**
	JPDA_m	35.6	71.5	6.7	40.0	**145**	1482	**8**	**31**
	TC_ODAL	31.2	72.0	13.3	33.3	495	1227	23	84
	RNN_LSTM	34.8	73.3	**17.8**	**28.9**	314	1280	59	86
	SiameseCNN	**42.3**	72.8	**24.4**	31.1	315	1119	30	61
	QuadMOT	**42.3**	**75.6**	**17.8**	**28.9**	379	**1065**	19	50
	FLDA-TM	35.4	71.1	16	35	602	**1008**	57	71

Notes: The red and purple colors indicate the best and the second best performing tracker on each metric.

Fig.9. Results of the FLDA-TM algorithm on TUD-Crossing(the first row), PETS.S2L2 dataset (the second row),TownCentre dataset(the third row) and ETH-Jelmoli (the fourth row)

Finally, in order to demonstrate the performance of the FLDA-TM algorithm, Table 4 shows the performance comparison for the FLDA-TM algorithm and other seven competitive algorithms on the test set of the 2DMOT2015 challenge. Although we only rely on the provided detections and some simple fuzzy rules, our approach shows a very competitive performance. It can see that the FLDA-TM algorithm achieves the best performance in MT and FM compared with other tracking approaches.

The reason is the same as the discussion above. Moreover, the FLDA-TM algorithm achieves the second best performance among all tracking algorithms in terms of MOTA, MOTP and ML scores.

Table 4 Results for the proposed algorithm on the test set of the 2DMOT2015

Method	MOTA (%↑)	MOTP (%↑)	MT (%↑)	ML (%↓)	FP(↓)	FN(↓)	IDS(↓)	FG(↓)
SCEA	29.1	71.1	8.9	47.3	6060	36912	604	1182
LINF1	24.5	71.3	5.5	64.6	5864	40207	298	744
JPDA_m	23.8	68.2	5	58.1	6373	40084	365	869
TC_ODAL	15.1	70.5	3.2	55.8	12970	38538	637	1716
RNN_LSTM	19	71	5.5	45.6	11578	36706	1490	2081
SiameseCNN	29	71.2	8.5	48.4	5160	37798	639	1316
QuadMOT	33.8	73.4	12.9	36.9	7898	32061	703	1430
FLDA-TM	30.4	71.8	14.2	40	7642	30123	598	1211

Based on the above discussion, we can see that the proposed algorithm can effectively deal with the object-to-object occlusions and abrupt object motion problems of video sequences. The advantages of the FLDA-TM algorithm are that it can incorporate the latest measurement information and experience knowledge into the fuzzy inference system by fuzzy rules to compute the association probabilities, and whenever more than one possible situation needs to be considered simultaneously. Meanwhile, to deal with the fragmented tracks caused by long-term occlusions, a track-to-track association method based on the fuzzy synthetic function is used to stitch track fragments (tracklets) before and after occlusions. However, we can also see that when the video scene was taken by a moving camera, the performance of the FLDA-TM algorithm will degrade compared to other algorithms, such as the SCEA, LINF1, JPDA_m, SiameseCNN and QuadMOT algorithms. To this problem, one solution is to modify the fuzzy rules of fuzzy inference system to adapt to changes in object features under the moving scene.

3.4. Analysis of Computational load of the FLDA-TM algorithm

The proposed algorithm is currently realized using MATLAB and implemented on a PC with 3.60 GHz CPU and 4GB RAM. Most computational time is spent on the fuzzy data association, which depends on the number of objects and detections. Overall, the average frame processing time of the

FLDA-TMalgorithm is 0.474s for PETS.S2L1 dataset, 0.697s for TUD-Stadtmitte dataset, 0.559s/f for TUD-Crossing dataset, 2.727s/f for PETS.S2L2 dataset, 3.127s/f for AVG-TownCentre dataset and 0.613s/f for ETH-Jelmoli dataset, respectively. Because the fuzzy inference system costs most processing time, to reduce the computational time of the FLDA-TM algorithm, the fuzzy rules can be optimized by the incorporation of expert knowledge. Moreover, if the FLDA-TM algorithm is implemented in C++ with GPU acceleration and code optimization, the computational time can be reduced.The real-time implementation of the proposed algorithm is a topic for future research.

4. Conclusion

In this paper, we proposed a new fuzzy logic data association approach for visual multi-object tracking. By incorporating fuzzy logic into multi-object tracking system, the association probabilities are allowed to be adjusted dynamically based on the conclusions of a set of fuzzy rules. In order to design the fuzzy rules, we defined two fuzzy inputs by using the prediction error and change of error of object features. So the proposed algorithm can quickly adjust the association probabilities between objects and measurements in response to long-term occlusions, changes in lighting conditions and abrupt object motion, etc. Moreover, fuzzy inference is helpful to simultaneously take into consideration several different or even conflicting situations to make better decisions. In order to tackle the association problem of fragmented tracks caused by long-term occlusions, we proposed a track-to-track approach based fuzzy synthetic function. For this reason, the proposed algorithm can reduce the number of fragmented tracks when measurements of occluded objects are not available or inaccurate. Finally, experimental results show the effectiveness and robustness of the proposed algorithm.

Though the experiment results are encouraging, there are still many aspects need to be further improved. Our future work will focus on studying the robustness of the proposed algorithm with respect to changes in the membership functions and the rules, and designing suitable fuzzy sets and fuzzy rules by incorporating expert experiences into fuzzy inference system to improve the performance of multi-object tracking.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (61773267), Science & Technology Program of Shenzhen (JCYJ20170302145519524).

REFERENCES

[1] Chiwoo Park, Taylor J. Woehl, James E. Evans, Nigel D.Browning, Minimum Cost Multi-Way Data Association for Optimizing Multitarget Tracking of Interacting Objects, IEEE Transactions on Pattern Analysis and Machine Intelligence. 37(3) (2015)611-624.

[2] Lu Wang, Nelson Hon Ching Yung, Lisheng Xu, Multiple-Human Tracking by Iterative Data Association and Detection Update, IEEE Transactions on Intelligent Transportation Systems. 15(5)(2014)1886-1899.

[3] K. Okuma, A. Taleghani, N. de Freitas, J. J. Little, D. G. Lowe, A boosted particle filter: Multitarget detection and tracking, in Proc.8th ECCV, 2004, pp. 28-39.

[4] Tao Zhou, Harish Bhaskar, Fanghui Liu, Jie Yang, Ping C, Online learning and joint optimization of combined spatial-temporal models for robust visual tracking, Neurocomputing. 226(2)(2017)221-237.

[5] Shu Tian, Fei Yuan, Gui-Song Xia, Multi-object tracking with inter-feedback between detection and tracking, Neurocomputing. 171(C)(2016) 768-780.

[6] A. Milan, S. Roth, K. Schindler, Continuous energy minimization for multi-target tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence. 36(1)(2014)58-72.

[7] B. Leibe, K. Schindler, N. Cornelis, L. Van Gool, Coupled object detection and tracking from static cameras and moving vehicles, IEEE Transactions on Pattern Analysis and Machine Intelligence. 30(10)(2008)1683-1698.

[8] B. Benfold, I. Reid, Stable multi-target tracking in real-time surveillance video, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 3457-3464.

[9] Zheng Wu, Margrit Betke, Global optimization for coupled detection and data association in multiple object tracking, Computer Vision and Image Understanding. 143(C) (2016) 25-37.

[10] J. Yoon, C. Lee, M. Yang, K. Yoon, Online Multi-object Tracking via Structural Constraint Event Aggregation, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). 2016, pp.1392-1400

[11] L. Fagot-Bouquet, R. Audigier, Y. Dhome, F. Lerasle, Improving Multi-Frame Data Association with Sparse Representations for Robust Near-Online Multi-Object Tracking, European Conference on Computer Vision (ECCV). 2016, pp.774-790

[12] Christopher Rasmussen, Gregory D. Hager, Probabilistic Data Association Methods for Tracking Complex Visual Objects, IEEE Transactions on Intelligent Transportation Systems. 23(6)(2001)560-576.

[13] H. Rezatofighi, A. Milan, Z. Zhang, Q. Shi, A. Dick, I. Reid. Joint Probabilistic Data Association Revisited, In ICCV. 2015, pp.3047-3055.

[14] Tinne De Laet, Herman Bruyninckx, Joris De Schutter, Shape-Based Online Multitarget Tracking and Detection for Targets Causing Multiple Measurements: Variational Bayesian Clustering and Lossless Data Association, IEEE Transactions on Pattern Analysis and Machine Intelligence. 33(12)(2011)2477-2491.

[15] Weiming Hu, Wei Li, Xiaoqin Zhang, Stephen Maybank, Single and Multiple Object Tracking Using a Multi-Feature Joint Sparse Representation, IEEE Transactions on Pattern Analysis and Machine Intelligence. 37(4)(2015)816-813.

[16] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, L. Van Gool, Online multiperson tracking-by-detection from a single, uncalibrated camera, IEEE Transactions on Pattern Analysis and Machine Intelligence. 33(9)(2011)1820–1833.

[17] G. Shu, A. Dehghan, O. Oreifej, E. Hand, M. Shah, Part-based multiple-person tracking with partial occlusion handling, in Proc. IEEE CVPR. 2012, pp. 1815–1821.

[18] Yang Yi, Zengwei Mo, Jie-wen Tan, A novel hierarchical data association with dynamic viewpoint model for multiple targets tracking, J. Vis. Commun. Image R. 34(1) (2016) 37-49.

[19] S. Bae, K. Yoon, Robust Online Multi-Object Tracking based on Tracklet Confidence and Online Discriminative Appearance Learning, IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2014, pp.1218-1225.

[20] Seung-Hwan Bae, Kuk-Jin Yoon, Robust Online Multiobject Tracking With Data Association and Track Management, IEEE Transactions on Image Processing. 23(7)(2014)2820-2833.

[21] J. Son, M. Baek, M. Cho, B. Han, Multi-Object Tracking with Quadruplet Convolutional Neural Networks, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017,pp.5620-5629.

[22] A. Milan, S. Rezatofighi, A. Dick, I. Reid, K. Schindler, Online Multi-Target Tracking using Recurrent Neural Networks, In AAAI. 2017,pp.1-9.

[23] Laura Leal-Taixé, Cristian Canton-Ferrer, Konrad Schindler, Learning by Tracking: Siamese CNN for Robust Target Association. Deep Vision Workshop (CVPR). Las Vegas (Nevada, USA), 2016, pp.33-40.

[24] Byung-Gyu Kim, Dong-Jo Park, Novel target segmentation and tracking based on fuzzy membership distribution for vision-based target tracking system, Image and Vision Computing. 24(12)(2006)1319-1331.

[25] Changyong Yoon, Minkyu Cheon, Mignon Park, Object tracking from image sequences using adaptive models in fuzzy particle filter, Information Sciences. 253(12)(2013)74-99.

[26] H. Shandiz, S.Mirhassani, B.Yousefi,M. Fatemi, Fuzzy based foreground background discrimination for probabilistic color based object tracking, International Journal of Computer Science and Network Security. 10 (1)(2010) 120–125.

[27] Y.M.Chen, H.C.Huang, Fuzzy logic approach to multisensor data association, Mathematics and Computers in Simulation.52 (6) (2000) 399-412.

[28] Ashraf M.Aziz, A novel all-neighbor fuzzy association approach for multitarget tracking in a cluttered environment, Signal Processing. 91(8)(2011)2001–2015.

[29] Li Liang-qun, Xie Wei-xin, Intuitionistic Fuzzy Joint Probabilistic Data Association Filter and its application to Multitarget Tracking, Signal Processing. 96(3)(2014)433-444.

[30] Nuno Vieira Lopes, Pedro Couto, AranzazuJurio, Pedro Melo-Pinto, Hierarchical fuzzy logic based approach for object tracking, Knowledge-Based Systems. 54 (1) (2013)255–268.

[31] Vinu Thomas, Ajoy Kumar Ray, Fuzzy Particle Filter for Video Surveillance, IEEE Transactions on Fuzzy systems.19(5)(2011)937-945.

[32] RuiPaúl, Eugenio Aguirre, Miguel García-Silvente, Rafael Muñoz-Salinas, A new fuzzy based algorithm for solving stereo vagueness in detecting and tracking people, International Journal of Approximate Reasoning. 53(4)(2012)693-708.

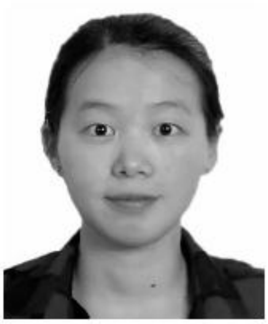
[33] LI Jun, XIE Wei-xin, LI Liang-qun, Online Multiple Target Tracking by Intuitionistic Fuzzy Data Association, International Journal of Fuzzy Systems. 19(2)(2017) 355–366.

[34] He You, Wang Guohong, Guan xin, Information Fusion Theory with Applications, Publishing House of Electronics Industry, 2010.

[35] P. Dollár, R. Appel, S. Belongie, Fast feature pyramids for object detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(8)(2014)1532-1545.

Resumes of Authors:

Li Liang-qun was born in 1979. He is an associate professor of information engineering, Shenzhen University. Address:2-603 Box, School of Information Engineering, Shenzhen University, shenzhen, Guangdong, 518060, P.R.CHINA,E-mail:lqli@szu.edu.cn,Tel:+86-755-26732055.



Zhan Xi-yang was born in 1993. She is a master student of school of information engineering, Shenzhen University. She research interests include multi-sensor information fusion, target tracking, etc. 2150730452@email.szu.edu.cn.

Liu Zong-xiang was born in 1965. He is a professor of School of information engineering, Shenzhen University. His research interests include multi-sensor information fusion, multiple target tracking ,and Random set theory, etc. E-mail: liuzx@szu.edu.cn.

Xie Wei-xin was born in 1941. He is a professor and doctoral tutor of School of information engineering, Shenzhen University. His research interests include fuzzy information processing, multi-sensor information fusion, computer vision, etc. Address:2-603 Box, School of Information Engineering, Shenzhen University, shenzhen, Guangdong, 518060, P.R.CHINA,E-mail: wxxie@szu.edu.cn, Tel:+86-755-26534375.