Particle-filter Multi-target Tracking Algorithm Based on Dynamic Salient Features

Zhang Yan*, Shi Zhi-Guang, Li Ji-Cheng, Yang Wei-Ping

ATR lab, National University of Defense Technology, Changsha, China

Copyright © 2015 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract In order to address the problem of tracking different moving targets in image sequence against a complicated background, this paper presents a particle-filter multi-target tracking algorithm based on their dynamic salient features. By making use of the research findings on visual attention, the algorithm adopts the robust dynamic salient features as a result of combining the gray-scale and details with the motion characteristics of such targets as the state vector of particle filter. The algorithm is highly robust as it contains salient features originating from the low-level features of the targets. Meanwhile the particle filter allows optimized estimation of non-linear and non-Gaussian models. As a consequence, the algorithm is capable of managing traces in tracking different targets and dealing with their appearance, disappearance, mergence, splitting and sheltering by obstacles. Experiments show that this new algorithm enables tracking of multiple targets in complicated image sequence.

Keywords Salience, Target Feature, Particle-filter, Multi-target Tracking

1. Introduction

The research in multi-target tracking plays an important role in a number of military and civil areas. Compared with single-target tracking, it is challenged by state measurement, linkage between different targets and model estimation for each moving target. Since the 1970s, researchers have successively proposed many multi-target tracking algorithms, including nearest-neighbor filter algorithm, joint probability data association (JPDA) filter algorithm, multi-hypothesis tracking filter algorithm and multi-target tracking algorithm based on the theory of random sets [1-4]. Some of the algorithms, however, only select the measurement closest to the predicted position of the targets tracked in respect of statistics as the candidate through which their traces are updated, which is subject to false or lost tracking [5]. Some algorithms suppose that the quantity of targets remain constant during tracking, which is often unrealistic [5,6]. Some may have some limitation in actual applications due to the number of calculations that increases rapidly with the quantity of targets and observations [7]. Others are difficult to offer optimal state solutions in analytic form though they provide a relatively self-contained theoretical system for eliminating multi-target tracking problems by integrating the theory of random sets with Bayes’s theory. In such a situation, therefore, someone advanced a multi-goal tracking algorithm based on particle filter [5, 8, 9], which renders optimal state solutions in analytic form using a group of weighted particles approaching probability hypothesis density (PHD). However, the algorithm has the disadvantage of lack of data relevant modules such that the states of targets are described in a collective form and it is unclear which state corresponds to a specific target. Therefore this algorithm cannot reveal the motion trace of a single target definitely. It is for this reason that the academic community is in great need of new multi-target tracking algorithms.

As we know, particle filter is virtually an open system whose state vector may be set depending on the actual problems to be solved. Many researchers have done a lot in this respect. For example, Yao Jianmin et al solved target tracking to some extent by choosing Gabor wavelet features as the state vector of particle filter [10]; Wang Jian et al achieved single-target tracking of color images against a complicated background [11]; and Yang Tao et al conducted real-time detection and tracking of human head by combining grads and geometric information with particle filter [12]. Even more researchers consider the position information of targets as the state vector of particle filter, and determine their positions through iterative filtering. Based on the above algorithms, the paper presents an algorithm that uses the salient features of targets as the state features in particle filtering process and predicts their positions though iterative filtering while detecting moving targets in single frame using visual saliency maps for data association between predicted and detected positions of targets, managing the trace of each moving target according to the result of such association and thereby tracking several
targets. It has overcome the disadvantages of common algorithms of multi-target tracking such as uncertain motion trajectory of single target, unclear target state, incorrect estimation of target quantity, and uncontrollable computation amount. In addition, because of dynamic salient features resulting from the fusion of diverse low-level target features, the algorithm has high robustness and low image SNR requirement so that it is fully capable of tracking a number of targets and addressing the issues like target appearance, disappearance, mergence, splitting, sheltering by obstacles and multi-target track management.

The rest of the paper is organized as follows. Section 2 contains the fundamental principles and description of the multi-target particle-filter tracking algorithm based on salient features. Section 3 describes the experimental results. A conclusion is drawn in Section 4.

2. Principles and Description of Particle-filter Multi-target Tracking Algorithm based on Salient Features

2.1. Salient Features and Extraction

2.1.1. Definition of salient features

Visual attention is a process of human vision system in choosing knowledge and filtering visual information based on scenery image salience as well as the relationship between target and scene. Besides the interested targets in a real scenery image, it includes remarkable interference information. Visual attention mechanism can help brain filter out such interference information, and pay attention to interested objects such that the vision perception process is selective. The ability of the visual attention mechanism in filtering and selecting visual information has greatly aroused the researchers’ enthusiasm of research. They have presented a number of visual attention models in the field of machine vision. Visual attention model based on salience was first proposed by Koch and Ullman at first and thereafter a corresponding computation model by Itti and Koch. The latter has three basic information processing modules: several parallel and separable feature maps recording the properties in a number of feature dimensions in different positions in visual field and thereby calculating single feature salience of different feature dimensions in each position; a salient map integrating the salience of different feature maps into a total salient measure and guiding the visual attention process; and a WTA (winner-take-all) network selecting the most salient position dynamically as the attention focus from the salient map.

Therefore, from the theoretical model of Itti and Koch, the so-called salient feature may be defined as a total salient measure resulting from fusion of low-level features. With the salient measure, a visual salient reflecting the difference of salient features is created and a series of salient positions are obtained through inhibition of return and transfer of attention focuses.

By using wavelet analysis, motion estimation and other methods based on the above theory, a total dynamic salient feature can be fused with the extracted visual sensitivity features including grayscale, details and motion of preprocessed achromatic scenery image in multi-scales. Then a two-dimensional salient map reflecting the visual salience in each position can be obtained according to the measurement of the salient features in different image positions. Because the salient feature is the fusion of local contrast difference of multiple visual sensitivity features in several scales, it is characterized by high robustness and anti-jamming capacity as compared with the unitary low-level feature. With reference to Itti and Koch’s visual attention model, the visual salient attention model is proposed. See Fig.1 It consists of three modules: low-level feature extraction module, salient feature fusion module and most-salient position determining module. The realization process of the visual salient attention model is as follows. Firstly, extract diverse low-level features (e.g. gray-scale, details and motion) of visual sensitivity from the scenery image. Thereafter fuse the features to obtain salient features in each image position according to the feature integration theory. Finally form a visual salient map reflecting the salient feature distinction in which the positions compete mutually and the winning position becomes the visual attention focus. The transfer of attention focus is realized by restricting the return of the attention focus.
2.1.2. Extraction of salient features

Visual salience is a local contrast resulting from multiple visual sensitivity features. The more obvious the difference is, the stronger the salience will be, and vice versa. For achromatic infrared or visible image, the features of visual sensitivity include gray-scale, details, motion, etc. Visual sensitivity is generated under the following two premises [5].

(1) For some feature dimensions, the features in the central region are markedly different from surrounding regions. (2) The features of surrounding region have better consistency. If the features in the central and surrounding regions are referred to as “receptive field”, human vision system has the ability of sensing local difference of features through the receptive fields in different sizes, i.e. multi-scale factors of sensing. The common method to extract salient features and form visual salient map is as follows: firstly, set up multi-feature and multi-scale expressions of image; then for each feature channel, extract the local contrast features; and finally, integrate different types of features into a total salient feature and form a visual salient map.

According to the theory of multi-resolution analysis, any image can be split into slow-varying low frequency part (the principal information of image) and fast-varying high frequency part (detail texture information of image). Based on the above conception, this latest algorithm operates in the following steps: use the wavelet transformation to disassemble the gray-scale and detail features of the image to obtain their feature figures; then by following the principle of image feature matching, extract the target motion features from the image to form motion feature figure using gray projection; and finally, integrate the gray-scale, detail and motion feature figures into total salient features and obtain the corresponding visual salient map. The specific operating procedure of the algorithm is as follows:

At first, set up multi-feature and multi-scale expressions of image. Suppose the input image \( f(x, y) \) is a gray image. Then us Mallat algorithm[1,4] to decompose the image into several levels, compose a gray pyramid \( L_k(s) \) with the low -frequency components at each level in the wavelet decomposition, and combine the three high-frequency components into a detail pyramid \( O_k(s,d) \), where \( s \in [0..N] \) refers to scale, \( N \) gray-scale, and \( d \in \{HL, LH, HH\} \) the orientations of the three details. \( L_k(s) \) and \( O_k(s,d) \) obtained following the above steps form the multi-scale expression of the gray-scale and detail features of the original image.

According to the theory of mechanic vision, the local contrast of the positions in specific regions with certain features can be obtained by means of Center-Surround operator[5,6,7,8,9,18,19] and convolution of the feature map. In shape of DOG function, the Center-Surround operator simulates the manner in which human vision system senses stimulus. See formula (1). However, the convolution is difficult to be performed in engineering practice. By consulting references [5] and replacing the convolution process with the process of “difference between scales”, the algorithm enhances the central excitation region, restraining surrounding regions, strengthens the contrast of the features and achieve transformation from the feature of visual sensitivity to vision stimulation.

\[
\text{DOG}(x, y, \sigma_c, \sigma_s) = \frac{1}{2\pi\sigma_c^2} \exp\left(-\frac{r^2}{2\sigma_c^2}\right) - \frac{1}{2\pi\sigma_s^2} \exp\left(-\frac{r^2}{2\sigma_s^2}\right)
\]

Where, \( r = \sqrt{x^2 + y^2} \), \( \sigma_c \) controls the size of central excitation region, and \( \sigma_s \) controls the size of surrounding restraining regions.

During the realization process of “difference between scales”, the rough scale image \( F(c) \) stands for the image features of surrounding constraining region, and the fine/precise scale image \( F(s) \) represents the image features.
of central excitation region. The “difference between scales” [5, 6] can be defined as
\[ F(c, s) = F(c) \Theta F(s) \] (2)

Where, \( \Theta \) means the following process:
Interpolate \( F(s) \) to the same size of image region as \( F(c) \), then subtract pixel by pixel and obtain its absolute value. Thus, a salient map of gray-scale and detail features may be obtained by following the above definition:
\[ \hat{L}_k(c, s) = L_k(c) \Theta L_k(s) \] (3)
\[ \hat{O}_k(c, s, d) = O_k(c, d) \Theta O_k(s, d) \] (4)

Secondly, in order to obtain a salient map of motion features, the gray-scale of image pixels can be projected in \( x \) and \( y \) directions respectively so that 2 dimensional images can be transformed into one dimensional curve. After that perform operations related to the characteristic curves of adjacent frames to determine the translation estimation between two continuous frames. Based on this estimation, the motion salient map \( \hat{M}_k \) can be determined as follows:
\[ \hat{M}_k(i, j) = |G_k(i, j) - G_{k-m}(i - \hat{dy}, j - \hat{dx})| \] (5)

Where, \( G_k(i, j) \) represents the gray-scale value of pixel \( (i, j) \) in the \( k^{th} \) frame image, and \( \hat{dx} \) and \( \hat{dy} \) mean the translation estimations between the \( k^{th} \) and \( (k + m)^{th} \) frames.

Finally, fuse the salient maps of gray-scale, details, motion, and use the normal function \( N(\cdot) \) to normalize the multiple salient maps. Here we can define a set of weighted factors \( \alpha, \beta, \gamma \), where they can be obtained according different image processing tasks or image features. Here the image processing task aims to detect moving targets or to calculate the motion salient map \( \hat{M}_k \).

Consequently, the basic theory of particle filter is: At first indicate a random measure describing the posterior probability distribution \( p(X_{0,k} | Z^k) \) by means of recursion, especially the boundary posterior probability distribution \( p(X_k | Z^k) \), thereby obtaining the optimal estimation of systematic state.

According to experiential distribution, the posterior probability distribution \( p(X_{0,k} | Z^k) \) can be approximated as
\[ p(X_{0,k} | Z^k) \approx \sum_{i=1}^{N} w_i^k \delta(X_{0,k} - X_{0,k}) \] (9)

Where, \( \delta(\cdot) \) is a Dirac function. Then a discrete weighted approximate formula may be used to approach the true posterior probability \( p(X_{0,k} | Z^k) \). So the mathematical expectation of State \( X_k \) is
\[ E\{X_k | Z^k\} = \int X_k p(X_{0,k} | Z^k) dX_k \] (10)

Furthermore, State \( X_k \)'s estimation \( \hat{X}_k \) base on \( Z^k \) may be approached by the following formula:
\[ \hat{X}_k = E\{X_k | Z^k\} \] (11)

According to the Law of great number, formula (11) is
convergent, i.e. \( \mathbb{E}\{X_k \mid Z^k_1\} \stackrel{a.s.}{\longrightarrow} N(0, \text{var}(X_k)) \). In addition, when the posterior variance of \( X_k \) is bounded (i.e. \( \text{var}(X_k) < \infty \)), the central limit theorem holds true,

\[
\mathbb{E}\{X_k\} - E\{X_k\} \sim \frac{1}{\sqrt{N}} N(0, \text{var}(X_k)) \tag{12}
\]

Where, \( \mathbb{E}\{\cdot\} \) means convergence depending on probability distribution [10].

Generally, the flow chart of standard particle filter comprises initialization, time updating, observation updating, re-sampling and other steps. The current filters such as assistant particle filter, regular particle filter and improved particle filter based on genetic algorithm are the transmutations from standard particle filters. The state vector \( X_{0,k} \) may be defined as required depending on the different problems to be solved. It may be defined as the position of a target, certain feature attributes or feature vector characterizing the target, etc.

2.3. Particle-filter Multi-target Tracking Algorithm Based on Dynamic Salient Features

To sum up, integrating the salience formed by grayscale, detail and motion features, the salient feature is characterized by higher robustness and interference resistance than unitary image grayscale feature. If this feature is combined with particle filter and used as the state feature vector of a particle filter, the algorithm is capable of tracking several moving targets by forecasting their positions through iterative filtering while detecting moving targets in single frames using visual salient maps, linking the forecast target positions to detected ones, managing the trace of each target depending on the linkage result. The algorithm combines the robustness of dynamic salient feature with the advantages of particle filter in solving nonlinear and non-Gaussian problems while managing traces according to the forecast and detected positions of targets. Therefore, this algorithm may be used to address the issues including target appearance, disappearance, mergergence, splitting, shedding by obstacles. Based on the concept, the process can be shown as follows,

Step 1: Determine the salient features of the sequence image \( f_k(x, y) \); obtain the corresponding vision salient figure \( S_k(x, y) \), and then detect the moving targets \( Md_{k,j}(x, y) \) which have the salient features of grayscale, details and motion and draw human vision attention from the vision salient figures, where \( l = 1 \cdots L \), \( L \) represents the detected number of targets and \( k \) the number of frames.

Step 2: On the basis of the first frame salient feature figure \( S_1(x, y) \), establish the initial track \( T_1(l) \) and initial particle group \( P_1(l) \) for each detected moving target \( Md_{k,j}(x, y) \), extract the salient feature for each target and use it as the state value for particle filter. When the target number reaches \( L \), the numbers of initial tracks and initial particle groups are both \( L \). Let \( N_p \) be the particle number for each particle group and then initiate each particle group. After the steps of time updating, observation updating and re-sampling independently for every particle group \( k \), the state estimation of each particle group can be obtained.

Step 3: Determine the estimation position \( PF_{k+1,l}(x, y) \) of every moving target in next frame according to the state estimation of each frame’s significant figure. Alternatively, it is expressed as follows

\[
D_{M,P}(l, r) = \{d_{l,r}(Md_{k+1,j}(x, y), PF_{k+1,l}(x, y)), l = 1 \cdots L, r = 1 \cdots L\} \tag{13}
\]

Where,

\[
d_{l,r}(Md_{k+1,j}(x, y), PF_{k+1,l}(x, y)), l = 1 \cdots L, r = 1 \cdots L
\]

means the euro distance between \( Md_{k+1,j}(x, y) \) and \( PF_{k+1,l}(x, y) \). Let \( T_r \) be an associate threshold and then the following formula holds true,

\[
\begin{align*}
D_{M,P}(l, r) &< T_r, & \text{if } Md_{k+1,j}(x, y) \text{ associates with } PF_{k+1,l}(x, y) \\
D_{M,P}(l, r) &\geq T_r & \text{if } Md_{k+1,j}(x, y) \text{ does not associates with } PF_{k+1,l}(x, y)
\end{align*} \tag{14}
\]
According to the association situation, the processing strategy will be different: if the predicted position is associated with the detection results, i.e.  \( D_{M,p}(l,r) < T_r \), maintain the natural target track; if not, that is to say,  \( D_{M,p}(l,r) \geq T_r \), then the following cases requiring special processing will occur:

a. If a new target \( Md_{k+1,l+1}(x,y) \) is detected from the significant image, there is no corresponding track associated with it because there is no estimation position of the corresponding particle filter. Therefore, the new track \( T_{k+1}(L+1) \) and new particle group \( P_{k+1}(L+1) \) shall be set up;

b. If there is no associated salient image to detect target around estimation target position \( PF_{k+1}(x,y) \), it indicates that the target disappears for some reasons. For example, the target is obstructed by obstacles or it stops moving. Here the target track is maintained based on the estimation position, i.e. \( \hat{T}_{k}(l) \rightarrow T_{k+1}(l) \), and the miss-detecting frame number \( F_a \) recorded. If the quantity of missing frames reaches certain threshold \( T_a \) (this threshold can be determined according to prior practice), i.e. \( F_a \geq T_a \), then delete the target track and its corresponding particle group and let \( L = L-1 \);

c. Associate \( n \) ( \( n \geq 2 \)) estimation positions with one detected target from the significant image, i.e. \( D_{M,p}(l,r+i) < T_r, \quad i = 0, \ldots, n-1 \). It indicates that the tracks of \( n \) moving targets will appear cross or incorporate phenomenon. At this point, maintain all the intrinsic target tracks, i.e. \( T_{k}(l) \rightarrow T_{k+1}(l) \), and record frame number \( F_c \) of target merging. If the merging frame number reaches a certain threshold \( T_c \) (the threshold can be determined according to practical instances in advance), i.e. \( F_c \geq T_c \), merge multiple targets into one, delete other unwanted tracks and particle groups, and let \( L = L-n+1 \);

d. Associate one estimation position with \( n \) ( \( n \geq 2 \)) detected targets from the significant images, i.e. \( D_{M,p}(l+i,r) < T_r, \quad i = 0, \ldots, n-1 \). This means that there are moving targets split in the frame. Here the primary track is maintained, alternatively \( T_{k}(l) \rightarrow T_{k+1}(l) \), and distributed to a detected target. Then set up new track \( T_{k+1}(l) \) and particle group \( P_{k+1}(l) \) associated with other detected targets, where \( l = L+i, \quad i = 1, \ldots, n-1 \), and let \( L = L+n-1 \).

**Step 4:** Reliably track all moving targets appearing in the scenery image.

### 3. Experimental Results

According to algorithm principles described above, we have verified the feasibility and effectiveness of the particle-filter multi-target tracking algorithm based on salient features proposed herein through the following experiments:

#### 3.1. Experiment 1

By following the algorithm proposed herein, static salient features are extracted from a group of traffic surveillance image sequence diagrams, which are in size of \( 768 \times 576 \) and available from ftp://pets.rdg.ac.uk.

As shown in Fig.2, from Fig.2 (a1) to Fig.2 (a5) represent the 5th, 32nd, 45th, 89th and 147th frames in the sequence diagrams. The group of sequence diagrams contains a number of targets in different moving model. From Fig.2 (b-1) to Fig.2 (b-5) are the resultant static saliency maps of the respective frames obtained by extracting static salient features. The comparison of the two groups of pictures indicates that simulation of human’s visual attention mechanism in consideration of the static information including of the gray-scale, details and motion of the targets results in significant change in the images and enhances the areas drawing visual attention while suppressing others.

Thus the following features may be identified in Fig. 2 (b-1) to Fig. 2 (b-5):

- Target 1 and 2 are most eye-catching;
- Several typical targets including Target 3 and 4 are relatively remarkable with less saliency than Target 1 and 2;
- Target 5, 6 and 7 are even less salient than Target 1;
- New targets – Target 8, Target 9 and Target 10 and 11 appear in the 32nd, 45th and 147th frames respectively. However, Target 8 and 9 are obscure in the static saliency maps due to indistinct changes in the gray-scale and details of the targets and background.
- In the 147th frame, Target 3 and 4 no longer move but are still significant in the static saliency maps because of the distinct change in gray-scale and details.

In the 32nd, 45th and 89th frames, thanks to its relatively high distinctiveness of change in margin details, Target 4 is always visible in the saliency maps in the cause of partial obstruction – complete obstruction – partial obstruction by a sign board.

As shown in Fig.2 (b1) to Fig.2 (b5), only using the static salient feature, target 1 is the best salient target, but target 1 is not the moving target. So if we only use the static salient feature, we might not correctly recognize the whole moving targets.

This paper has used 5th, 32nd, 45th, 89th and 147th frames of the sequence diagrams, because these five frames are the best typical images. In these images, the phenomena that targets are disappearing and being sheltered and reappearing have been found. So we select these five frames to show the performance of the algorithm.
3.2. Experiment 2

From Fig.3 (a-1) to Fig.3 (a-5) are the resultant motion saliency maps of the 5th, 32nd, 45th, 89th and 147th frames obtained using corrected gray projection algorithm. In Fig.3 (a-1) to Fig.3 (a-5), it can be seen that:

- Target 1 and 2 that were highly significant in the static salient pictures are not very notable in the motion saliency maps as they are not motion-characteristic.
- Owing to their motion characteristics, Target 3, 4, 5, 6 and 7 are relatively distinct in the motion saliency maps.
- All the new targets appearing in the 32nd, 89th and 147th frames including Target 8, 9, 10 and 11 are relatively notable in the motion saliency maps.
- The saliency of Target 3 in the 89th frame drops as it moves no longer. Target 4 also does not move any longer in the 147th frame. For that reason Target 3 and 4 are almost invisible in this frame.
- In Fig. 3 (a1) to Fig. 3 (a5), it may be seen that the moving areas caused by the minor motion of camera or other factors, e.g. Interference Point 1 and 2 and other interference areas, emerge in the motion saliency maps. However, they are of no saliency completely in the static saliency maps.

As the resultant picture fusing the static and motion saliency maps, Fig.3 (b-1) to Fig.3 (b-5) are referred to as “dynamic saliency feature extracting diagram fusing motion characteristics”. In the group of fusion pictures, the targets with both static and motion saliency are characterized by high saliency as follows:

- Target 1 and 2, which are salient in the static saliency maps, almost disappear in Fig.2 (b1) to Fig.2 (b5) because they have no motion characteristics.
- Target 3 and 4 disappear in the 89th and 147th frames respectively because they slowly stop motion;
- Interference areas such as Interference Point 1 and 2 are almost invisible in Fig. 3 (b1) to Fig. 3 (b5) because their saliency is very low in the static saliency maps.

As a consequence, it can be shown that after fusing treatment the moving targets in the pictures drawing human’s visual attention are enhanced while other areas are obviously suppressed.
Figure 3. The contrast of the motion saliency maps and dynamic saliency maps which fusing the static and motion saliency maps.
3.3. Experiment 3

Fig. 4 reveals the experimental results of tracking multiple targets by following the particle-filter multi-target tracking algorithm based on dynamic salient features as proposed in this article. In Fig. 4, it can be seen that the algorithm may be used to track multiple targets against complicated background and to effectively deal with the problems of new target appearance, disappearance, mergence, splitting and obstruction by obstacles.

For example, new targets including Target 6, 7 and 8 appear in Fig 4 (b), (d) and (e) respectively. In addition to tracking the existing targets reliably, therefore, the proposed algorithm establishes motion traces and particle swarms concerning the new targets, and tracks them. In Fig. 4(d) and (e), Target 3 and target 4 no longer move, and tracking four frames, the algorithm removes the motion traces of the two targets and the corresponding particle swarm without further tracking. Target 4 is hidden by obstruction in Fig. 4 (b), (c) and (d), and split up into two targets in Fig. 4(c), one small and the other big. However, the algorithm still well tracks Target 4. In Fig. 4(e), Target 7 is split up into two moving targets. The algorithm can not only reliably track Target 7 before splitting, but also properly trace Target 11 originating from Target 7 depending on the mergence and splitting of the target.

Fig. 4(f) is the motion position and trace tracking diagram of 170 frames of individual moving targets. The positions of the rectangles in the diagram are the starting points of the targets while those of triangles end points. It indicates that:
- Target 4 is obstructed by obstacles during movement and therefore its trace is temporarily interrupted.
- Target 7 is split up into two targets midway. One of the targets is numbered as “7” and continues to move while the other new target starts its new trace.

Target 9, 10 and 11 appear later while Target 3 and 4 disappear early, which is obviously shown by their traces.
4. Conclusions

In conclusion, the dynamic salient feature presented in the paper simulates the features of human visual system for observation with superior capability in detecting a number of moving targets. Such dynamic salient feature is combined with open particle filter framework as the state vector of particle filtering in order to achieve filtering iteration and target tracking. The experimental results show that the proposed algorithm can make full use of the robust performance of salient feature and take advantage of the capability of particle filter in solving nonlinear and non-Gaussian problems. In addition, it can be used to address the problems of target appearance, disappearance, mergence, splitting, obstruction by obstacles and thereby track several targets reliably.

5. Innovation of the Paper

The main innovation of the paper is reflected in the following three aspects:

1. The paper presents a method for extracting dynamic salient features by taking advantage of low-level features of image including grayscale, detail and motion and uses the features to achieve multi-target detection depending on the properties of gray-level images.

2. The dynamic salient feature is combined with particle filtering as the state vector of particle filter and a particle-filter multi-target tracking algorithm based on dynamic salient features presented, thereby making it possible to predict the positions of moving targets.

3. The data correlation of filter-predicted positions with saliency-detected position of targets allows the trace management of several moving targets and eliminates the difficulties in terms of target appearance, disappearance, mergence, splitting and obstruction by obstacles that frequently emerge in multi-target tracking against a complicated background.

REFERENCES

[1] Singer.R.A, Stein.J.J. An Optimal Tracking Filter for Processing Sensor Data of Imprecisely Determined Origin in Surveillance Systems[A]. Proc. of IEEE Conference on Decision and Control[C], Miami Beach, 1971: 171-175.

[2] Mahler R.. A Theoretical Foundation for the Stein-Winter Probability Hypothesis Density(PHD) Multi-target Tracking Approach[J]. Proc. of the MSS National Symposium on Sensor and Data Fusion, 2000, 1(3): 99-118.

[3] Mahler R.. Multi-target Moments and Their Application to Multi-target Tracking[A]. Proc. of Workshop on Estimation, Tracking and Fusion: A tribute to Yaakov Bar-Shalom[C], Monterey, 2001: 134-166.

[4] Mahler R.. Multi-target Bayes-Filtering via First-order Multi-target Moments[J]. IEEE Trans.on Aerospace and Electronic Systems, 2003, 39(4): 1152-1178.

[5] J.Vermaak, S.J.Godsill, P.Perez. Monte Carlo Filtering for Multi-Target Tracking and Data Association[A]. IEEE Trans. on Aerospace and Electronic Systems, 2005, 41 (1):309-332.

[6] Y.Bar-Shalom, Fortman.T.E. Tracking and data association[A]. New York Academic Press inc[C], San Diego, 1988: 26-31.

[7] Reid D.B. An Algorithm for Tracking Multiple Targets[J]. IEEE Trans. on Automatic Control, 1979, 24(12): 843-854.

[8] Vo B., Singh S., Doucet A.. Sequential Monte Carlo Implementation of the PHD Filter for Multi-target Tracking[A]. Proc. of International Conference on Information Fusion[C], Fairborn, USA, 2003: 792-799.

[9] Vo B., Ma W.-K., Singh S.. Localizing an Unknown Time-varying Number of Speakers: A Bayesian Random Finite Set Approach[A]. Proc. of IEEE International Conference on Acoustics, Speech Signal Processing[C], Philadelphia, 2005, 4: 1073-1076.

[10] YAO Jian-min, SUN Jun-xi, SUN Zhong-sen, SONG Jian-zhong. Wavelet Feature Tracking Method Based on Particle Filter[J]. Computer Simulation, 2006, 23(1): 94-96.
[11] WANG Jian, ZHANG Gui-lin. Real-time Tracking of Moving Target in The Complex Background[J]. Computer & Digital Engineering, 2005, 33(11): 18-21.

[12] YANG Tao, LI Jing, PAN Quan. Head Tracking Algorithm Based on Particle Filter under Complex Environment[J]. Computer Engineering and Design, 2005, 26(2): 354-356.

[13] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Trans, 1998, Pattern Analysis and Machine Intelligence, 20(11): 1254-1259.

[14] [14] ZHANG Zhi-long. Feature Extraction and Recognition Important in Remote Sensing Imagery. Chang Sha: National University of Defense Technology, 2005.

[15] Itti L, Koch C, Niebur E. Rapid Biologically Inspired Scene Classification Using Features Share with Visual Attention [J]. IEEE Trans. Pattern Analysis and Machine Intelligence, 2007, 29(2): 300-312.

[16] Itti L, Koch C. Computational modeling of visual attention [J]. Nature Review Neuroscience, 2001, 2(3): 194-230.

[17] Itti L, Koch C. Feature combination strategies for saliency-based visual attention systems [J]. Journal of Electronic Imaging, 2001, 10(1): 161-169.

[18] Stentiford FWM. An evolutionary programming approach to the simulation of visual attention [A]. Proc. of the IEEE Congress on Evolutionary Computation[C]. Seoul, 2001: 851-858.

[19] Chen Shuo, Wu Chengdong, Chen Dongyue. Rapid scene registration method based on visual saliency. Journal of Image and Graphics, 2011, 16(7):1241-1247.

[20] Yu Zhiruing, Wang Shuozhong, Zhang Xinpeng, Liu Tingting. Image Saliency Detection Based on Contrast Features and Local Sharpness. Journal of Applied Sciences-Electronics and Information Engineering, 2010, 28 (1):25-31.

[21] Achanta R, Hemami S, Estrada F. Frequency tuned salient region detection[C]. Proceedings of IEEE International Conference On Computer Vision and Pattern Recognition. Florida, USA, 2009:1597-1604.