Visual Object Tracking using Particle Filtering with Dual Manifold Models

YINGHONG XIE1,2 AND CHENGDONG WU1
1College of Information Science and Engineering
Northeastern University
Shenyang, 110819 China
2School of Information Engineering
Shenyang University
Shenyang, 110044 China

Compared with affine transformation, projection transformation represents the process of imaging objects more accurately. This paper proposes a novel object tracking method using particle filtering with dual manifold models. One is the covariance manifold used for the object observation model, and the other is the geometric deformation on SL(3) group, where the rank of projection transformation matrix equals 1, adapted to utilize for object dynamic model. Our main contribution is to utilize both the geometry of SL(3) group and covariance manifolds in developing a general particle filtering-based tracking algorithm. Extensive experiments prove that the proposed method can realize stable and accurate tracking of object with significant geometric deformation, even with illumination changes and when an object is obscured.

Keywords: visual object tracking, SL(3) group, Riemannian manifold, covariance, dual models

1. INTRODUCTION

Visual object tracking is a critical task in many computer vision domains and it is widely used for visual surveillance, human computer interaction, augmented reality, and other applications. Still, robust and accurate tracking of deforming and moving objects is a fundamental and challenging task.

A variety of tracking algorithms have been proposed to overcome dynamic deformation of a target. Recently, use of particle filter to provide a robust tracking framework has attracted interest, as it is not limited to linear systems, nor does it require the noise to be Gaussian. The current density of the state is represented by a set of random samples with associated weights and the new density is computed based on these samples and weights. For example, papers [1-4] utilized particle filter to track deformable targets. They all used affine transform as the parameter model. The six affine parameters were treated as a vector. However, the affine parameters belong to Lie group structure, which does not belong to vector spaces. Some papers [1-5] used the intrinsic geometry of manifold to design the tracking algorithms.

Target representation is one of the major components of a typical visual tracker. Conventional trackers based on histogram [6, 7] have difficulty with tracking deformable objects. Extensive research has been done on this topic. Many trackers [8-10] utilized
covariance matrix as object region descriptor, which has proved reliable and effective for a modest computational cost, as the covariance matrix enables efficient fusion of spatial and statistical features while its dimensionality is small. Because covariance matrix is definite symmetric manifold and belongs to Riemannian manifolds, the similarity between covariance matrices is measured on Riemannian manifolds. [11] represented an object region by the feature covariance matrix, and used Lie group structure of the positive definite matrices to compute the similarity between two covariance matrices, which was proposed by [12]. The method could track objects with moderate pose change; however, it was sensitive to drastic variations in deformation and illumination. More and more method combing covariance matrix with particle filter to gain more robust tracking result has been described. [13] proposed a tracking approach on Riemannian manifolds with incremental covariance tensor learning, combined with the particle filter framework to allow better handling of background clutter, and incrementally learned a low dimensional covariance tensor representation, which resulted in real time performance. [14, 15] proposed an object tracking scheme that exploited the geometrical structure of Riemannian manifold and piecewise geodesic under a Bayesian framework. Covariance matrix combined with two particle filters was used in the tracking strategy. It was applicable to both visual and IR videos.

The performance of an algorithm is closely related to the parameter model that it utilizes. Recently, the parameter models of visual tracking algorithms are largely built on the special Euclidean group SE(3) and the affine group Aff(2) in 3D and 2D visual tracking [16]. The process of object imaging is a non-linear process. There are many imaging models (refer to paper [17]), including translation, similarity, affine, bilinear, and projection. The more complex the imaging model is, the more accurately the imaging process can be reflected. However, affine transformation is an approximation of projection transformation, which is suitable for the work environment in which the camera is far away from the object. So we chose the more complex imaging model (projection transformation) to describe the imaging process of the object.

Considering that the process of imaging objects is essentially projection transformation [18], and based on our previous work [19], this paper proposes a novel target tracking algorithm on projection transformation group, which is SL(3) group, and builds dual manifold models. One is the covariance manifold using for the object observation model, and the other is the geometric deformation on SL(3) group using for object dynamic model. To our best knowledge, there is little discussion in the literature using SL(3) group to develop particle filter based tracking algorithms. The advantages of the algorithm proposed by this paper are the following: (1) SL(3) group reflects the target imaging process more accurately. (2) Dual models technology takes both the geometry of SL(3) group and covariance manifolds into consideration in developing a general particle filtering-based tracking algorithm. (3) Covariance matrix fuses multiple features and modalities of tracking region. All of these make the tracking performance more robust and accurate.

The rest of this paper is organized as follows. In Section 2 we describe the two manifold models which are SL(3) group and region covariance manifold, and their metric. In Section 3, we build the tracking model based on particle filter, where dynamic and observation models are designed on SL(3) group and covariance manifold respectively. In Section 4, tracking algorithm is proposed. The results of applying this approach to the
target with significant geometric transformation and with obscured object and illumination change are shown in Section 5. Finally, some conclusions are drawn in Section 6.

2. TWO MANIFOLD MODELS

2.1 SL(3) Group and its Lie Algebra

The tools used here come primarily from differential geometry. For more information on these subjects, readers can refer to [20, 21].

A Lie group is a group with the structure of an analytic manifold such that the group operations are analytic, which means the maps

\[ G \times G \to G \]  \quad (X, Y) \to XY,

\[ G \to G \]  \quad X \to X^{-1}  \tag{1}

are analytic [12]. The local neighborhood of any group element \( G \) can be adequately described by its tangent space. The tangent space at the identity element forms its Lie algebra.

The set of nonsingular \( n \times n \) square matrices forms a Lie group where the group product is modeled by matrix multiplication, usually denoted by \( \text{GL}(n, R) \) for the general linear group of the order \( n \), where \( R \) is an \( n \)-dimensional real space. Lie groups are differentiable manifolds on which we can do calculus. Being a sub-group of \( \text{GL}(n, R) \), the special linear group \( \text{SL}(n, R) \) is the space of all real \( n \times n \) matrices \( H \) satisfying \( \det H = 1 \). Its Lie algebra denoted by \( \text{sl}(n, R) \) consists of the real matrices of trace zero.

In our proposed algorithm, we use projective transformation as the parameter model, that is, 3-by-3 matrices where matrices differing only by scalar multiplication represent the same projective transformation. To get rid of this ambiguity, we normalize the matrices to have determinant 1. This group is usually denoted by \( \text{SL}(3) \). The corresponding Lie algebra is \( \text{sl}(3) \). Matrices in this algebra are \( (3 \times 3) \) with a null trace. The exponential map is a homeomorphism between a neighborhood of \( I \in \text{SL}(3) \) and a neighborhood of the null matrix \( 0 \in \text{sl}(3) \) [22, 23].

Let \( A_i \) with \( i \in \{1,2,...,8\} \), be a basis of the Lie algebra \( \text{sl}(3) \) (i.e. the dimension of \( \text{sl}(3) \) is 8). Any matrix \( A \in \text{sl}(3) \) can be written as a linear combination of the matrices \( A_i \): \( A(x) = \sum_{i=1}^{8} x_i A_i \), where \( x = (x_1, x_2,..., x_8) \) is a \( (8 \times 1) \) vector and \( x_i \) is the \( i \)th element of the base field. We use the following \( \text{sl}(3) \) basis matrix:

\[
A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_6 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}
\]
\[
A_7 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_8 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}
\]  \tag{2}
It will be recalled that for matrix groups, the Riemannian distance is defined by the matrix logarithm operation. Given matrix group elements \(X\) and \(Y\) we have

\[
d(X, Y) = \|\log(YX^{-1})\|.
\]  

(3)

### 2.2 Geodesic on \(SL(3)\) Group

\(SO(3)\) and other compact Lie groups have bi-invariant Riemannian metric [24], and the exponential map on Lie group is consistent with the geodesic on the group. Non-compact Lie groups do not have bi-invariant Riemannian metric, so the exponential map on Lie group is not consistent with the geodesic. \(SL(3)\) group is non-compact Lie group. In order to calculate the geodesic on it, a metric structure needs to be defined on \(SL(3)\) group to calculate the new exponential map \(Expp\), which is also known as Riemannian exponential map.

The common method for defining the metric structure on manifold \(M\) is that an inner product \(\langle \cdot, \cdot \rangle\) is given on the tangent space \(T_pM\) for each point \(p \in M\), which is Riemannian metric, and the length of a tangent vector \(U \in T_pM\) is: \(\|U\| = \langle U, U \rangle^{1/2}\). So we can define exponential map \(Expp\) as follows [24]:

\[
Expp(U) = \exp(-U^T)\exp(U + U^T).
\]  

(4)

### 2.3 Region Covariance Manifold

The method of region covariance feature is that using \(d\)-dimensional covariance matrix describes the features of the region. It represents the variances of each component itself and the correlation between each component. Suppose \(Z\) is a one-dimensional gray image or a three-dimensional color image, and the size of target region \(R\) is \(M*\)N, \(F\) is \(M*N*d\) dimensional features extracted from the image.

\[
F(x, y) = \phi(Z, x, y)
\]  

(5)

where the function \(\phi(\cdot)\) represents the mapping from image \(Z\) to feature image \(F\), it may be the image gray, and the combination of colors, gradients, filter response value, or the combination of several relations.

Let \(\{f_k\}_{k=1, 2, \ldots, n}\) denotes the set of \(d\)-dimensional feature vectors of \(F(x, y)\). The \(d\)-dimensional covariance matrix in the region \(R\) is defined as:

\[
C_x = \frac{1}{MN} \sum_{k=1}^{MN} (f_k - \mu_k)(f_k - \mu_k)^T
\]  

(6)

where \(\mu_k = \frac{1}{MN} \sum_{k=1}^{MN} f_k\), which is mean vector, and the covariance \(C_x\) is a \(d\times d\) dimensional positive symmetric matrix.

It can be concluded according to Eq. (6) that covariance matrix combines the spatial and statistical properties as well as their correlation. Furthermore, the size of \(C_x\) only depends on the dimension of the feature vectors, and has no relation to the size of the
target area. We define the following feature vector:

\[ f_k = (x, y, I_x, I_y, I_{xy}, I_z) \]  

(7)

where \((x, y)\) is the coordinate of image \(I\), \(I_x\) and \(I_y\) denotes gradient value on \(x\)-direction and \(y\)-direction respectively. \(I_{xy}\) denotes the convolution of \(I_x\) and \(I_y\).

The dissimilarity between two regions covariance matrices can be given by the distance between two points of the manifold, considering that those points are the two regions.

The covariance matrix, which is symmetric positive definite matrix, forms a Riemannian manifold. According to [8], we define a Riemannian metric as the follows:

\[ \langle y, z \rangle_x = tr(X^{-1/2}yX^{-1}zX^{-1/2}). \]  

(8)

The exponential map associated with the above Riemannian metric is

\[ \exp_x(y) = X^{1/2} \exp(X^{-1/2}yX^{-1/2})X^{1/2}. \]  

(9)

By Eq. (10), we can obtain the logarithm map

\[ y = \log_x(y) = X^{1/2} \log(X^{-1/2}yX^{-1/2})X^{1/2}. \]  

(10)

Submit Eqs. (10) to (8)

\[ d^2(X, Y) = \|y\|^2_x = \langle y, z \rangle_x \]

\[ = \langle \log_x(Y), \log_x(Y) \rangle_x \]

\[ = tr(\log^2(X^{-1/2}YX^{-1/2})) \]  

(11)

Furthermore, Eq. (11) is equivalent to

\[ d(X, Y) = \sqrt{\sum_{k=1}^{d} \lambda_k (X, Y)}. \]  

(12)

Where \(\lambda_k\) is the generalized eigenvalues of \(X\) and \(Y\).

3. TRACKING MODEL

3.1 Particle Filter

The visual tracking problem is cast as an inference task in a Markov model with hidden state variables. The main idea of particle filter is using a group of weighted particles to describe the posterior probability distribution. We adopt state variables \(S_t\) to describe the projection transformation parameters of the target at the moment \(t\). The aim of target tracking is to estimate the current state \(S_t\) by the image observation sequence \(I_{1:t} = \)
\{I_t, \ldots, I_t\}, according to the Bayesian theorem:

\[ p(S_t | I_{t-1}) = \int p(S_t | S_{t-1}) p(S_{t-1} | I_{t-1}) dS_{t-1}, \]  
\[ \text{(13)} \]

\[ p(S_t | I_t) = \frac{p(I_t | S_t) p(S_t | I_{t-1})}{p(I_t | I_{t-1})}. \]  
\[ \text{(14)} \]

Eq. (13), known as the predictive equation, and represents the probability model of target state transition. Eq. (14) is called update equation. The tracking process is described as follows: Given the probability density \( p(S_{t-1} | I_{t-1}) \) of target state at the moment \( t-1 \), the posterior probability density \( p(S_t | I_{t-1}) \) at moment \( t \) is calculated recursively, according to state transition probability model \( p(S_t | S_{t-1}) \) and observation \( I_t \).

The particle filter is Monte Carlo simulation of the realization of Bayesian timing filter, the particles (sample) are referred to as defining the target state by a set of parameters, multiple discrete particles can denote the distribution of the target attribute, which can carry out target tracking. The idea is that the posterior probability distribution of the unknown state is denoted approximately by random sampling collection (each sample with corresponding weights). If the number of samples is large enough, the distribution of the sample collection can be considered equivalent to posteriori probability distribution.

Let \( \{x_{0,k}, w_i\}_{i=1}^{N} \) denote the particle collection of the posterior probability density function \( p(x_{0,k} | z_k) \), where \( \{x_{0,k}, i=1, \ldots, N\} \) is sample collection with corresponding weights \( \{w_i, i=1, \ldots, N\} \) meeting \( \sum_{i=1}^{N} w_i = 1 \), and \( x_{0,k} = \{x_j, j=0, \ldots, k\} \) denotes all the state collections to the moment \( k \), \( N \) is the number of samples, so posterior probability density at moment \( k \) can be denoted approximately as:

\[ p(x_{0,k} | z_k) \approx \sum_{i=1}^{N} w_i \delta(x_{0,k} - x_{0,k}^i) \]  
\[ \text{(15)} \]

where \( \delta() \) is a Dirac delta function.

### 3.2 Dynamic Model

Dynamic model, also known as state transition model, can describe transition of object state in tracking process. As the object state is constantly changing in the process of tracking, the tracking performance largely depends on the accuracy of the dynamic model. The deformation and location of an object can be represented by projection transformation. The change of object between adjacent frames can be seen as their corresponding points moving on a Riemann manifold. The basic idea of building a dynamic model is to find the dynamic conversion relationship between two adjacent points on the manifold. We make use of tangent vectors of the points on the manifold, which is tangent space, to describe the conversion relationship. So the dynamic model on Riemann manifold and its tangent space are defined as follows:

\[ S_t = S_{t-1} \exp(V_{t-1}) \]  
\[ \text{(16)} \]
where the eight-dimension vector $S_t = [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8]^T$ is defined as projective transformation parameters of object state. $V_t$ represents the speed between $S_{t-1}$ and $S_t$, describing the motion of the target, which is tangent vector along the state $S_{t-1}$ on manifold. And $\mu_t$ is Gaussian white noise. First-order autoregressive model is adopted to represent the change of speed, where $a$ is autoregressive process parameters.

The above method can be further detailed as follows: A particle filter is applied on the Riemannian manifold to generate candidate points $S_j$, $j = 1, \ldots, N$. Let $S_{t-1}^j$ be the previous manifold points at $t-1$ and $V_{t-1}^j$ be the corresponding velocity vector that connects $(S_{t-1}^j, S_t^j)$ where $S_t^j$ is on the end point of the geodesic starting from $S_{t-1}^j$.

Tracking algorithm does not require the analytical expression of priori probability density, but sampling from priori probability density.

### 3.3 Observation Model

In the tracking process, the observation data are needed to correct the newly predicted state at every moment. The probability of the sample is the real target and is estimated by measuring the similarity between the observation data and the model.

Let $p(I_t | S_t)$ be the observation of $I_t$ under the state $S_t$, we can build the observation model as follows:

$$p(I_t | S_t) \propto \exp(-\lambda \|d^2 (C_t, C_{S_t})\|).$$

Where $C_t$ represents the region covariance features of the image model, and $C_{S_t}$ denotes the region covariance features of the target image under the geometry transformation $S_t$.

The likelihood is then assigned as the weights of particles,

$$w_t' = \exp(-\lambda \|d^2 (C_t, C_{S_t})\|).$$

These weights are then normalized by

$$w_t' = w_t' / \sum_{j=1}^{N} w_j'.$$

### 4. TRACKING ALGORITHM

The dynamic model is achieved by computing the velocity vector $V_t$ on Lie algebra under the first-order autoregressive model, then mapping it back to the Lie group space originated from the previous tracked manifold point $S_{t-1}$. And the dynamic vector is defined by the parameter on $SL(3)$ group, the candidate $S_t'$ is computed by Brownian motion model, however, $C_{S_t'}$ is the corresponding feature covariance matrix of $S_t'$, and utilized for computing the weight value of each candidate dynamic vector.

Then object appearance model on Riemannian manifold is estimated using the
YINGHONG XIE AND CHENGDONG WU

tracked object region covariance matrix \( C_t \). The complete tracking algorithm is designed
as follows:

**Input:** \( \{S'_i, i = 1, 2, ..., N\} \) and \( \{V'_i, i = 1, 2, ..., N\} \).

**Output:** the estimated target state \( S_t \) and the covariance \( C_t \) of tracked target region \( t = 2, 3, ... \).

**Step 1:** Initiate \( w_i^t = 1 / N \).

**Step 2:** According to Eq. (17), generate \( N \) samples of \( V'_i \).

**Step 3:** Draw \( \{S'_i, i = 1, 2, ..., N\} \), According to Eq. (16).

**Step 4:** Compute the covariance feature matrix \( C'_{si} \) corresponding to the dynamic vector \( S'_i \).

**Step 5:** Calculate the weight value \( w_i^t, i = 1, 2, ..., N \), according to Eqs. (19) and (20).

**Step 6:** Compute the weighted Lie group mean:

\[
C_t = \exp(\frac{1}{N} \sum_{i=1}^{N} \log(w_i^t C'_{si}))
\]

**Step 7:** Output \( C_t \), which is the target state. Set \( t = t + 1 \), go to step 1.

5. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed tracking algorithm, we performed two groups of experiments. One is for verifying the tracking performance of target experiencing geometric deformation, which can prove the effectiveness of adopting SL(3) group. The other is for verifying the tracking robustness against obscured and illumination change. The testing videos in the first group of experiments, sequences 1 and 2, can be found at the website: http://cv.snu.ac.kr/jhkwon/tracking/index.htm. Moreover, the testing videos in the second group of experiments, sequences 3 and 4, can be found at the website: http://www.cs.toronto.edu/~dross/ivt/.

5.1 Experiments for Geometric Deformation

For the first group of experiments, several video sequences with the targets undergoing large geometric deformation are adopted. The bounding box in the first frame is manually marked.

We compare the proposed algorithm with the algorithm proposed by [8] which was covariance tracking using model update based on Lie group (CTML), and the algorithm proposed by [15] which was Bayesian online Learning on Riemannian manifolds using a dual model with applications to video object tracking (BDMT). The algorithms were implemented in Matlab2007 running on an Intel Core-2 2.93GHz processor with 2GB memory. The performance of the three algorithms has been compared with the same experimental configuration. For the BDMT and the proposed algorithm, the parameters of the particle filter are set as the same, and the sample number \( N = 300 \). It performs 25 times repeatedly for each set of image sequences. Performance of the three algorithms is compared in two ways. On the one hand, for each frame, The Euclidean distance between the coordinates of the centers of tracked target region and the ground-truth target region is computed. Let the coordinates of the centers of tracked target region at moment \( t \) be \((x'_t, y'_t)\), and the coordinates of the centers of the ground-truth target region at moment \( t \) be \((x_t, y_t)\) (marked manually), then Euclidean distance at moment \( t \) is defined as:
In the first sequence, the target has experienced significant geometric deformation through the whole 271 frames, while each frame with the width of 320 pixels and the height of 240 pixels, and the initial size of template is 42*42. In the proposed algorithm, the state noise of 8-dimensional projection transformation in Eq. (15) is set as [0.05;0.002; 0.02;0.05;5;0.0008;0.0008], and in BDMT algorithm, the state noise of 6-dimensional affine transformation is set as [0.05;0.002;0.002;0.05;5;5]. For BDMT and the proposed algorithm, the initial rotation is both -12.5. The experimental results of the three algorithms are shown in Fig. 1 respectively. For a better visualization, Fig. 1 shows the five tracking results of the frames 3,171,196,269, and the final frame 271. From Fig. 1, we can conclude that the tracking result of CTML algorithm almost failed, for the tracking region is always fixed in the same size and lack of online learning during the tracking process. However the target, that is the cube, has experienced tremendous scale and pose change. In Figs. 1 (b) and (c), we can see that the size of the tracking window changes with the scale change of the target. The tracking performance is similar. But according to the frame 171 and frame 196, it is obvious that the proposed method has much more accurate results than BDMT. Fig. 3 (a) shows the results of the Euclidean distance between the tracked region and the ground truth target region of each frame. From it we can see that the distance of CTML algorithm is much larger than the other two, which is consistent with the tracking result shown in Fig. 1. The distance of BDMT is always larger than our proposed algorithm. The reason is that the proposed algorithm adopts SL(3) group for state modeling, which more accurately reflects the projection process than the affine group.

$$\text{DisError}_t = \sqrt{(x'_t - x_t)^2 + (y'_t - y_t)^2}. \quad (21)$$
In the second sequence, the target (vase) has also experienced large geometric deformation through the whole 316 frames, while each frame with the width of 320 pixels and the height of 240 pixels. The initial size of template is 50*58. In the proposed algorithm, the state noise of 8-dimensional projection transformation in Eq. (15) is set as \([0.04;0.002;0.002;0.04;4;4;0.001;0.001]\), and in BDMT algorithm, the state noise of 6-dimensional affine transformation is set as \([0.04;0.002;0.002;0.04;4;4]\). For BDMT and the proposed algorithm, the initial rotation equals to 0. From Fig. 2 we can see that the experimental results are also similar to the first sequence. From Fig. 3 (b), the distance of CTML algorithm and BDMT algorithm are much larger than the proposed algorithm as well.
5.2 Experiments for Illumination Change and Obscured

We have done experiments for many video sequences with illumination change or obscured. All the parameters for particle filter are as the same as the first group. All of these experimental results show that the proposed method can realize stable and accurate tracking. Because of the length limit of this paper, the tracking results of only three representative sequences are given.

Firstly, the video sequences 3 and 4 are illumination changing sequences. In video sequence 3, the size of each frame is 320*240 [25], the initial size of template is 104*110. The tracking results for sequence 3 and sequence 4 are shown in Figs. 4 and 5 respectively, from which we can see that our algorithm can track the targets accurately.

Secondly, in Fig. 6, the tracking results for the video sequence 5 captured from a hand-held camera are given. The two people represent a situation of group tracking where one or more objects move together in a sequence. Furthermore, the target has experienced severe full occlusion in the tracking process. The size of each frame is 1920 *1080, the initial size of template is 135*442. From Fig. 6, we can see from frame 198, the target is gradually obscured until it is completely obscured. When the target re-emerges, our algorithm finds it again immediately, and tracks the target well, even after the two people meet together.

The reasons why the proposed method can track the object with illumination change are that it adopts covariance matrix manifold to describe object features, and the gray information of an image is insensitive to illumination changes.
The reasons why the proposed method can track the transiently obscured object well are that particle filter algorithm has multiple hypothesis features, so it can re-capture the transiently lost object. Furthermore, the proposed method can store the object template before the object is obscured. When the transient obscured object appears again, the method can re-capture the object by matching with the stored template.

In summary, we can conclude that the proposed algorithm, based on SL(3) group using particle filtering with dual models, is able to provide more stable and accurate performance when tracking the targets undergoing significant geometric deformation. Furthermore, when the target undergoes drastic illumination change or is severely obscured, it can still realize stable and accurate tracking.

5.3 Experimental Analysis

The computational complexity of particle filter algorithm is related to the quantity of particles adopted. Smaller quantity of particles leads to lower robustness of the algorithm, but, larger quantity of particles leads to more computational time. The computational complexity of the proposed method is \( O(N^4) \), where \( N \) is the quantity of particles.

The average computational time per frame of our proposed method is about 0.8 seconds with 300 particles. Although the computational time indicates that the current speed of our tracker is far from real time, we expect further speedup, up to rates of 20 fps, which is possible.

The proposed algorithm consists of two components.

(1) The proposed particle filtering based tracking algorithm explicitly takes the geometry of SL(3) Lie group into consideration in deriving the dynamic equation on Lie group, which is our main contribution and the dominant factor for improving the tracking performance.

(2) Region covariance manifold fuses multiple features and modalities of tracking region, which is adopted for building an observation model, so that the tracking result is more robust to illumination changes.

Applying the above two manifold models, the proposed algorithm makes the tracking performance more robust and accurate.

6. CONCLUSION

We have proposed a visual tracking algorithm with dual manifold models. One is the covariance manifold using for the object observation model, and the other is the geometric deformation on SL(3) group using for the object dynamic model. The distinct advantage of the proposed method is explicitly taking both the geometry of SL(3) group and covariance manifolds into consideration in developing a general particle filtering-based tracking algorithm, which has improved the tracking performance. Theoretic analysis and experimental results have shown that the proposed algorithm can realize robust tracking of the target not only with significant geometric deformation but also with illumination changes and severe obscured. The proposed dual models probabilistic tracker is much more suitable for multi-target which is our ongoing work.
REFERENCES


16. C. Choi and H. I. Christensen, “Robust 3D visual tracking using particle filtering on the SE(3) group,” in Proceedings of IEEE International Conference on Robotics and

**Yinghong Xie (谢英红)** received her B.Sc. degree in 1999 from Shenyang Jianzhu University, received her M.Sc. degree in 2005 from Northeastern University. Now she is a doctoral student at College of Information Engineering, Northeastern University, and works in Shenyang University. Her main research interests include video image processing, pattern recognition.

**Chengdong Wu (吴成东)** received M.S. degree in Automatic Control from Tsinghua University in 1988. He received Ph.D. degree in Automatic Control from Northeastern University in 1994. Now he is currently a Professor in College of Information Science and Engineering, Northeastern University. His current research interests include image processing, wireless sensor networks, home automation and robot.