

RESEARCH

Open Access

# Automated target tracking and recognition using coupled view and identity manifolds for shape representation

Vijay Venkataraman<sup>1</sup>, Guoliang Fan<sup>1\*</sup>, Liangjiang Yu<sup>1</sup>, Xin Zhang<sup>2</sup>, Weiguang Liu<sup>3</sup> and Joseph P Havlicek<sup>4</sup>

## Abstract

We propose a new couplet of identity and view manifolds for multi-view shape modeling that is applied to automated target tracking and recognition (ATR). The identity manifold captures both inter-class and intra-class variability of target shapes, while a hemispherical view manifold is involved to account for the variability of viewpoints. Combining these two manifolds via a non-linear tensor decomposition gives rise to a new target generative model that can be learned from a small training set. Not only can this model deal with arbitrary view/pose variations by traveling along the view manifold, it can also interpolate the shape of an unknown target along the identity manifold. The proposed model is tested against the recently released SENSAC ATR database and the experimental results validate its efficacy both qualitatively and quantitatively.

**Keywords:** tracking and recognition, shape representation, shape interpolation, manifold learning

## 1 Introduction

Automated target tracking and recognition (ATR) is an important capability in many military and civilian applications. In this work, we mainly focus on tracking and recognition techniques for infrared (IR) imagery, which is a preferred imaging modality for most military applications. A major challenge in vision-based ATR is how to cope with the variations of target appearances due to different viewpoints and underlying 3D structures. Both factors, identity in particular, are usually represented by discrete variables in practical existing ATR algorithms [1-3]. In this paper we will account for both factors in a continuous manner by using view and identity manifolds. Coupling the two manifolds for target representation facilitates the ATR process by allowing us to meaningfully synthesize new target appearances to deal with previously unknown targets as well as both known and unknown targets under previously unseen viewpoints.

Common IR target representations are non-parametric in nature, including templates [1], histograms [4], edge

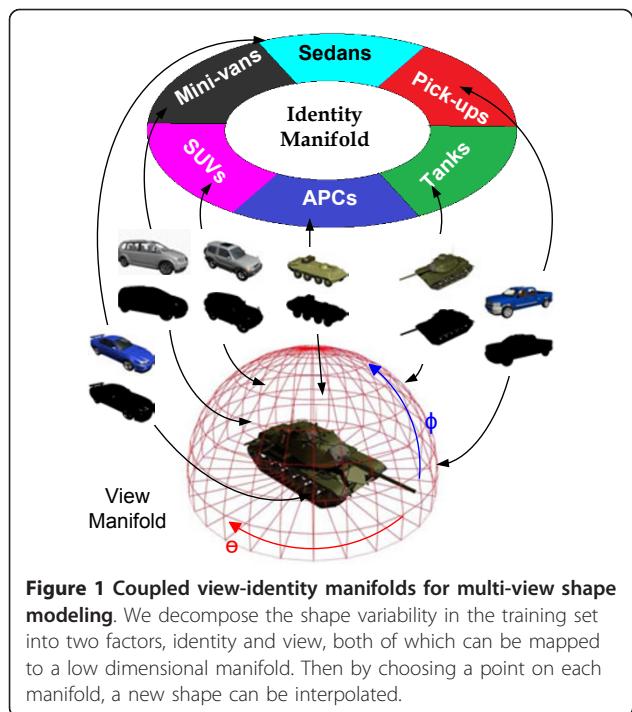
features [5] etc. In [5] the target is represented by intensity and shape features and a self-organizing map is used for classification. Histogram-based representations were shown to be simple yet robust under difficult tracking conditions [4,6], but such representations cannot effectively discriminate among different target types due to the lack of higher order structure. In [7], the shape variability due to different structures and poses is characterized explicitly using a deformable and parametric model that must be optimized for localization and recognition. This method requires high-resolution images where salient edges of a target can be detected, and may not be appropriate for ATR in practical IR imagery. On the other hand, some ATR approaches [8,1,9] depend on the use of multi-view exemplar templates to train a classifier. Such methods normally require a dense set of training views for successful ATR tasks and they are often limited in dealing with unknown targets.

In this work, we propose a new couplet of identity and view manifolds for multi-view shape modeling. As shown in Figure 1, the 1-D *identity manifold* captures both inter-class and intra-class shape variability. The 2-D hemispherical view manifold is used deal with view variations for ground vehicles. We use a nonlinear

\* Correspondence: guoliang.fan@okstate.edu

<sup>1</sup>School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078, USA

Full list of author information is available at the end of the article



**Figure 1** Coupled view-identity manifolds for multi-view shape modeling. We decompose the shape variability in the training set into two factors, identity and view, both of which can be mapped to a low dimensional manifold. Then by choosing a point on each manifold, a new shape can be interpolated.

tensor decomposition technique to integrate these two manifolds into a compact generative model. Because the two variables, view and identity, are continuous in nature and defined along their respective manifolds, the ATR inference can be efficiently implemented by means of a particle filter where tracking and recognition can be accomplished jointly in a seamless fashion. We evaluate this new target model against the ATR database recently released by the Military Sensing Information Analysis Center (SENSIAC) [10] that contains a rich set of IR imagery depicting various military and civilian vehicles. To examine the efficacy of the proposed target model, we develop four ATR algorithms based on different ways of handling the view and identity factors. The experimental results demonstrate the advantages of coupling the view and identity manifolds for shape interpolation, both qualitatively and quantitatively.

The remainder of this paper is organized as follows. In Section 2, we review some related work in the area of 3D object representation. In Section 3, we present our generative model where the identity and view manifolds are discussed in detail. In Section 4, we discuss the implementation of the particle filter based inference algorithm that incorporates the proposed target model for ATR tasks. In Section 5, we report experimental results of target tracking and recognition on both IR sequences from the SENSIAC dataset and some visible-band video sequences, and we also discuss the limitations and possible extensions of the proposed generative model. Finally, we present our conclusions in Section 6.

## 2 Related Work

This section begins with a review of different ways to represent a 3D object and the reasons for our choice of a multi-view silhouette-based method. Then we focus on several existing shape representation methods by examining their ability to parameterize shape variations, the ability to interpolate, and the ease of parameter estimation.

There are two commonly used approaches to represent 3D rigid objects. The first approach suggests a set of representative 2D snapshots [11,12] captured from multiple viewpoints. These snapshots may be represented in the form of simple shape silhouettes, contours, or complex features such as SIFT, HOG, or image patches. The second approach involves an explicit 3D object model [13] where common representations vary from simple polyhedrons to complex 3D meshes. In the first case, unknown views can be interpolated from the given sample set, whereas in the second case, the 3D model is used to match the observed view via 3D-to-2D projection. Accordingly, most object recognition methods can be categorized into one of two groups: those involving 2D multi-view images [14-19] and those supported by explicit 3D models [20-23]. There are also hybrid methods [24] that make use of both the 3D shape and 2D appearances/features.

In this work, we choose to represent a target by its representative 2D views due to two main reasons. First, this is theoretically supported by the psychophysical evidence presented in [25] which suggest that *the human visual system is better described as recognizing objects by 2D view interpolation than by alignment or other methods that rely on object-centered 3D models*. Second, it could be practically cumbersome to store and reference a large collection of detailed 3D models of different target types in a practical ATR system. Moreover, it is worth noting that many robust features (HOG, SIFT) used to represent objects were developed mainly for visible-band images and their use is limited by some factors such as image quality, resolution etc. In IR imagery, the targets are often small and frequently lack sufficient resolution to support robust features. Finally, the IR sensors in the SENSIAC database are static, facilitating target segmentation by background subtraction. Thus the ability to efficiently extract target silhouettes and the simplicity of silhouette-based shape representation motivates us to use the silhouette for multi-view target representation.

There are two related issues for shape representation. One is how to effectively represent the shape variation, and the other is how to infer the underlying shape variables, i.e., view and identity. As pointed out in [26], feature vectors obtained from common shape descriptors, such as shape contexts [27] and moment descriptors

[28], are usually assumed to lie in a Euclidean space to facilitate shape modeling and recognition. However, in many cases the underlying shape space may be better described by a nonlinear low dimensional (LD) manifold that can be learned by nonlinear dimensionality reduction (DR) techniques, where the learned manifold structures are often either target-dependent or view-dependent [29]. Another trend is to explore a shape space where every point represents a plausible shape and a curve between two points in this space represents a deformation path between two shapes. Though this method was shown successful in applications such as action recognition [26] and shape clustering [30], it is difficult to explicitly separate the identity and view factors during shape deformation as is necessary in the context of ATR applications.

This brings us to the point of learning the LD embedding of the latent factors, e.g., view and identity, from the high-dimensional (HD) data, e.g., silhouettes. In an early work [31], PCA was used to find two separate eigenspaces for visual learning of 3D objects, one for the identity and one for the pose. The bilinear models [32] and tensor analysis [33] provide a more systematic multi-factor representation by decomposing HD data into several independent factors. In [34], the view variable is related with the appearance through shape sub-manifolds which have to be learned for each object class. All of these methods are limited to a discrete identity variable where each object is associated with a separate view manifold. Our work draws inspiration from [35] where a non-linear tensor decomposition method is used to learn an identity-independent view manifold for multi-view dynamic motion data. A torus manifold was also proposed in [36,37] for the same purpose that is a product of two circular-shaped manifolds, i.e., the view and pose manifolds. In [36,37,35], the style factor of body shape (i.e., the identity) is a continuous variable defined in a linear space.

Our work presented in this paper is distinct from that in [36,37,35] primarily in terms of two main original contributions. The first is our couplet of view and identity manifolds for multi-view shape modeling; unlike [36,37,35] where the identity is treated linearly, for the first time we propose a 1D identity manifold to support a continuous nonlinear identity variable. Also, the view and pose manifolds in [36,37,35] have well-defined topologies due to their sequential nature. However, in our IR ATR application the topology of the identity manifold is not clear owing to a lack of understanding of the intrinsic LD structure spanning a diverse set of targets. Finding an appropriate ordering relationship among a set of targets is the key to learning a valid identity manifold for effective shape interpolation. To better support ATR tasks, the view manifold used here

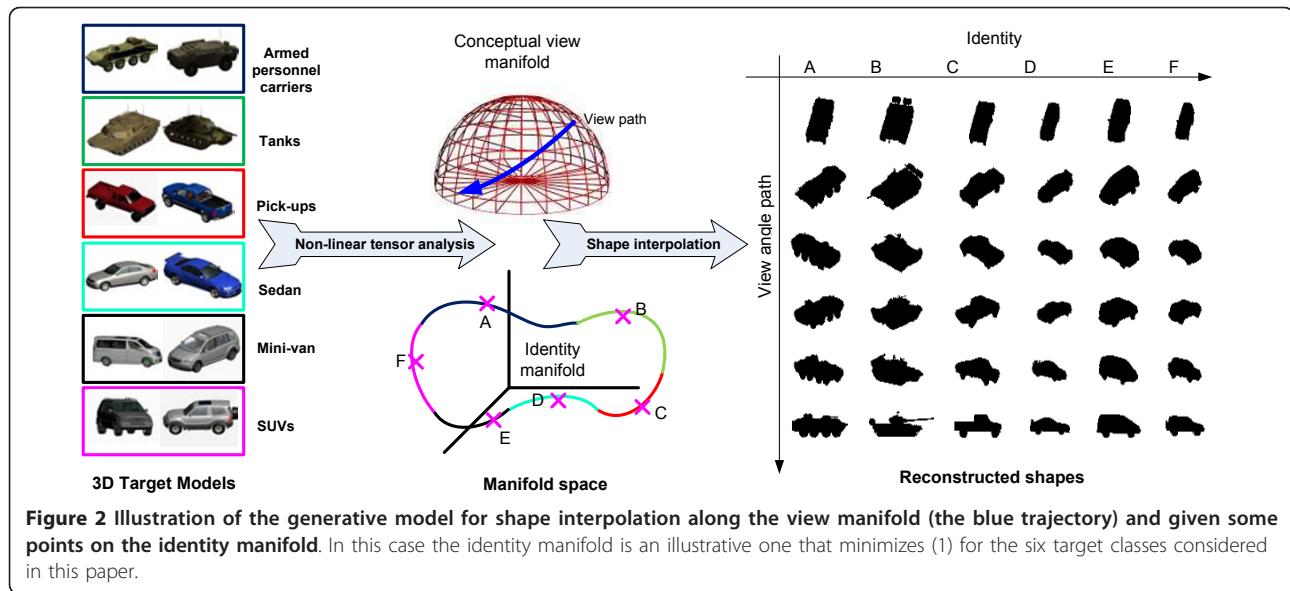
involves both the azimuth and elevation angles, compared with the case of a single variable in [36,37,35]. The second contribution is the development of a particle filter-based ATR approach that integrates the proposed model for shape interpolation and matching. This new approach supports joint tracking and recognition for both known and unknown targets and achieves superior results compared with traditional template-based methods in both IR and visible-band image sequences.

### 3 Target Generative Models

Our generative model is learned using silhouettes from a set of targets of different classes observed from multiple viewpoints. The learning process identifies a mapping from the HD data space to two LD manifolds corresponding to the shape variations represented in terms of view and identity. In the following, we first discuss the identity and view manifolds. Then we present a non-linear tensor decomposition method that integrates the two manifolds into a generative model for multi-view shape modeling, as shown in Figure 2.

#### 3.1 Identity manifold

The identity manifold that plays a central role in our work is intended to capture both inter-class and intra-class shape variability among training targets. In particular, the continuous nature of the proposed identity manifold makes it possible to interpolate valid target shapes between known targets in the training data. There are two important questions to be addressed in order to learn an identity manifold with the desired interpolation capability. The first one is which space this identity manifold should span. In other words, should it be learned from the HD silhouette space or a LD latent space? We expect traversal along the identity manifold to result in gradual shape transition and valid shape interpolation between known targets. This would ideally require the identity manifold to span a space that is devoid of all other factors that contribute to the shape variation. Therefore the identity manifold should be learned in a LD latent space with only the identity factor rather than in the HD data space where the view and identity factors are coupled together. The second important question is how to learn a *semantically valid* identity manifold that supports meaningful shape interpolation for an unknown target. In other words, what kind of constraint should be imposed on the identity manifold to ensure that interpolated shapes correspond to feasible real-world targets? We defer further discussion of the first issue to Section 3.3 and focus here on the second one that involves the determination of an appropriate topology for the identity manifold.



The topology determines the span of a manifold with respect to its connectivity and dimensionality. In this work, we suggest a *1D closed-loop structure* to represent the identity manifold and there are several important considerations to support this seemingly arbitrary but actually practical choice. First, the learning of a higher-dimensional manifold requires a large set of training samples that may not be available for a specific ATR application where only a relatively small candidate pool of possible targets-of-interest is available. Second, this identity manifold is assumed to be closed rather than open, because all targets in our ATR problem are man-made ground vehicles which share some degree of similarity with extreme disparity unlikely. Third, the 1D closed structure would greatly facilitate the inference process for online ATR tasks. As a result, the manifold topology is reduced to a specific *ordering relationship* of training targets along the 1D closed identity manifold. Ideally, we want targets of the same class or those with similar shapes to stay closer on the identity manifold compared with dissimilar ones. Thus we introduce a *class-constrained shortest-closed-path* method to find a unique ordering relationship for the training targets. This method requires a view-independent *distance* or *dissimilarity* measure between two targets. For example, we could use the shape dissimilarity between two 3D target models that can be approximated by the accumulated mean square errors of multi-view silhouettes.

Assume we have a set of training silhouettes from  $N$  target types belonging to one of  $Q$  classes imaged under  $M$  different views. Let  $\mathbf{y}_m^k$  denote the vectorized silhouette of target  $k$  under view  $m$  (after the distance transform [29]) and let  $L_k$  denote its class label,  $L_k \in [1, Q]$

( $Q$  is the number of target classes and each class has multiple target types). Also assume that we have identified a LD identity latent space where the  $k$ 'th target is represented by the vector  $\mathbf{i}^k$ ,  $k \in \{1, \dots, N\}$  ( $N$  is the number of total target types). Let the topology of the manifold spanning the space of  $\{\mathbf{i}^k | k = 1, \dots, N\}$  be denoted by  $\mathbf{T} = [t_1 \ t_2 \ \dots \ t_{N+1}]$  where  $t_i \in [1, N]$ ,  $t_i \neq t_j$  for  $i \neq j$  with the exception of  $t_1 = t_{N+1}$  to enforce a closed-loop structure. Then the class-constrained shortest-closed-path can be written as

$$\mathbf{T}^* = \arg \min_{\mathbf{T}} \sum_{i=1}^N D(\mathbf{i}^{t_i}, \mathbf{i}^{t_{i+1}}), \quad (1)$$

where  $D(\mathbf{i}^u, \mathbf{i}^v)$  is defined as

$$D(\mathbf{i}^u, \mathbf{i}^v) = \sum_{m=1}^M \| \mathbf{y}_m^u - \mathbf{y}_m^v \| + \beta \cdot \varepsilon(L_u, L_v), \quad (2)$$

$$\varepsilon(L_u, L_v) = \begin{cases} 0 & \text{if } L_u = L_v, \\ 1 & \text{otherwise,} \end{cases} \quad (3)$$

where  $\| \cdot \|$  represents the Euclidean distance and  $\beta$  is a constant. The first term in (2) denotes a view independent shape similarity measure between targets  $u$  and  $v$  as it is averaged over all training views. The second term is a penalty term that ensures targets belonging to the same class to be grouped together. The manifold topology  $\mathbf{T}^*$  defined in (1) tends to group targets of similar 3D shapes and/or the same class together, enforcing the best local *semantic smoothness* along the identity manifold, which is essential for a valid shape interpolation between target types.

It is worth mentioning that the identity manifold to be learned according to  $\mathbf{T}^*$  will encompass multiple target classes each of which has several sub-classes. For example, we consider six classes of vehicles in this work each of which includes six sub-class types. Although it is easy to understand the feasibility and necessity of shape interpolation within a class to accommodate intra-class variability, the validity of shape interpolation between two different classes may seem less clear. Actually,  $\mathbf{T}^*$  not only defines the ordering relationship within each class but also the neighboring relationship between two different classes. For example the six classes considered in this paper are ordered as: *Armored Personnel Carriers* (APCs)  $\rightarrow$  *Tanks*  $\rightarrow$  *Pick-up Trucks*  $\rightarrow$  *Sedan Cars*  $\rightarrow$  *Minivans*  $\rightarrow$  *SUVs*  $\rightarrow$  APCs. Although APCs may not look like Tanks or SUVs in general, APCs are indeed located between Tanks and SUVs along the identity manifold according to  $\mathbf{T}^*$ . It occurs because that (1) finds an APC-Tank pair and an APC-SUV pair that have the *least shape dissimilarity* compared with all other pairs. Thus this ordering still supports sensible inter-class shape interpolation, although it may not be as smooth as intra-class interpolation, as will be shown later in the experiments.

### 3.2 Conceptual view manifold

We need a view manifold to accommodate the view-induced shape variability for different targets. A common approach is to use non-linear DR techniques, such as LLE or Laplacian eigenmaps, to find the LD view manifold for each target type [29]. One main drawback of using identity-dependent view manifolds is that they may lie in different latent spaces and have to be aligned together in the same latent space for general multi-view modeling. Therefore, the view manifold here is designed to be a hemisphere that embraces almost all possible viewing angles around a ground vehicle as shown in Figure 1 and is characterized by two parameters: the azimuth and elevation angles  $\Theta = \{\theta, \varphi\}$ . This conceptual manifold provides a unified and intuitive representation of the view space and supports efficient dynamic view estimation.

### 3.3 Non-linear Tensor Decomposition

We extend the non-linear tensor decomposition in [35] to develop the proposed generative model. The key is to find a view-independent space for learning the identity manifold through the commonly-shared conceptual view manifold (the first question raised in Section 3.1).

Let  $\mathbf{y}_m^k \in \mathbb{R}^d$  be the  $d$ -dimensional, vectorized distance transformed silhouette observation of target  $k$  under view  $m$ , and let  $\Theta_m = [\theta_m, \varphi_m]$ ,  $0 \leq \theta_m \leq 2\pi$ ,  $0 \leq \varphi_m \leq \pi$ , denote the point corresponding to view  $m$  on the LD

view manifold. For each target type  $k$ , we can learn a non-linear mapping between  $\mathbf{y}_m^k$  and the point  $\Theta_m$  using the generalized radial basis function (GRBF) kernel as

$$\mathbf{y}_m^k = \sum_{l=1}^{N_c} w_l^k \kappa(\|\Theta_m - \mathbf{S}_l\|) + [1 \ \Theta_m] b_l, \quad (4)$$

where  $\kappa(\cdot)$  represents the Gaussian kernel,  $\{\mathbf{S}_l | l = 1, \dots, N_c\}$  are  $N_c$  kernel centers that are usually chosen to coincide with the training views on the view manifold,  $w_l^k$  are the target specific weights of each kernel and  $b_l$  is the coefficient of the linear polynomial  $[1 \ \Theta_m]$  term included for regularization. This mapping can be written in matrix form as

$$\mathbf{y}_m^k = \mathbf{B}^k \psi(\Theta_m), \quad (5)$$

where  $\mathbf{B}^k$  is a  $d \times (N_c + 3)$  target dependent linear mapping term composed of the weight terms  $w_l^k$  in (4) and  $\psi(\Theta_m) = [\kappa(\|\Theta_m - \mathbf{S}_1\|), \dots, \kappa(\|\Theta_m - \mathbf{S}_{N_c}\|), 1, \Theta_m]$  is a target independent non-linear kernel mapping. Since  $\psi(\Theta_m)$  is dependent only on the view angle we reason that the identity related information is contained within the term  $\mathbf{B}^k$ . Given  $N$  training targets, we obtain their corresponding mapping functions  $\mathbf{B}^k$  for  $k = \{1, \dots, N\}$  and stack them together to form a tensor  $\mathbf{C} = [\mathbf{B}^1 \ \mathbf{B}^2 \ \dots \ \mathbf{B}^N]$  that contains the information regarding the identity. We can use the high-order singular value decomposition (HOSVD) [38] to determine the basis vectors of the identity space corresponding to the data tensor  $\mathbf{C}$ . The application of HOSVD to  $\mathbf{C}$  results in the following decomposition:

$$\mathbf{C} = \mathbf{A} \times_3 \mathbf{i}^k, \quad (6)$$

where  $\{\mathbf{i}^k \in \mathbb{R}^N | k = 1, \dots, N\}$  are the *identity basis vectors*,  $\mathbf{A}$  is the core tensor with dimensionality  $d \times (N_c + 3) \times N$  that captures the coupling effect between the identity and view factors, and  $\times_j$  denotes mode- $j$  tensor product. Using this decomposition it is possible to reconstruct the training silhouette corresponding to the  $k$ 'th target under each training view according to

$$\mathbf{y}_m^k = \mathbf{A} \times_3 \mathbf{i}^k \times_2 \psi(\Theta_m). \quad (7)$$

This equation supports shape interpolation along the view manifold. This is possible due to the interpolation friendly nature of RBF kernels and the well defined structure of the view manifold. However it cannot be said with certainty that any arbitrary vector  $\mathbf{i} \in \text{span}(\mathbf{i}^1, \dots, \mathbf{i}^N)$  will result in a valid shape interpolation due to the sparse nature of the training set in terms of the identity variation.

To support meaningful shape interpolation, we constrain the identity space to be a 1D structure that includes only those points on a closed B-spline curve connecting the identity basis vectors  $\{i^k | k = 1, \dots, N\}$  according to the manifold topology defined in (1). We refer to this 1D structure as the *identity manifold* denoted by  $\mathcal{M} \subset \mathbb{R}^N$ . Then an arbitrary identity vector  $i \in \mathcal{M}$  would be semantically meaningful due to its proximity to the basis vectors, and should support a valid shape interpolation. Although the identity manifold  $\mathcal{M}$  has an intrinsic 1D closed-loop structure, it is still defined in the tensor space  $\mathbb{R}^N$ . To facilitate the inference process, we introduce an intermediate representation, i.e., a unit circle as an equivalent of  $\mathcal{M}$  parameterized by a single variable. First, we map all identity basis vectors  $\{i^k | k = 1, \dots, N\}$  onto a set of angles uniformly distributed along a unit circle,  $\{\alpha_k = (k - 1) * 2\pi/N | k = 1, \dots, N\}$ .

Then, as shown in Figure 3, for any  $\alpha' \in [0, 2\pi)$  that is between  $\alpha_j$  and  $\alpha_{j+1}$  along the unit circle, we can obtain its corresponding identity vector  $i(\alpha') \in \mathcal{M}$  from two closest basis vectors  $i^j$  and  $i^{j+1}$  via spline interpolation along  $\mathcal{M}$  while maintaining the distance ratio defined below:

$$\frac{|\alpha' - \alpha_j|}{|\alpha' - \alpha_{j+1}|} = \frac{\mathcal{D}(i(\alpha'), i^j | \mathcal{M})}{\mathcal{D}(i(\alpha'), i^{j+1} | \mathcal{M})}, \quad (8)$$

where  $\mathcal{D}(\cdot | \mathcal{M})$  is a distance function defined along  $\mathcal{M}$ . Now (7) can be generalized for shape interpolation as

$$y(\alpha, \Theta) = A \times_3 i(\alpha) \times_2 \psi(\Theta), \quad (9)$$

where  $\alpha \in [0, 2\pi)$  is the identity variable and  $i(\alpha) \in \mathcal{M}$  is its corresponding identity vector along the identity manifold in  $\mathbb{R}^N$ . Thus (9) defines a generative model for multi-view shape modeling that is controlled by two continuous variables  $\alpha$  and  $\Theta$  defined along their own manifolds.

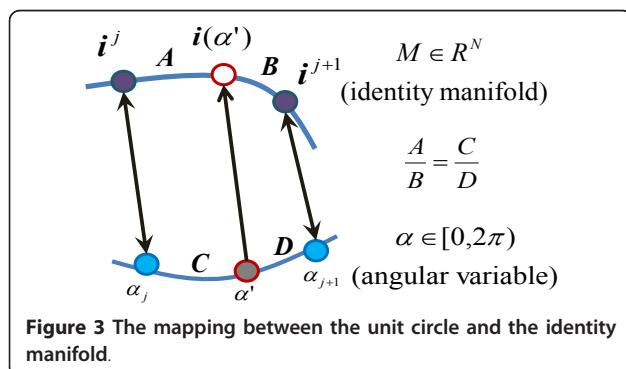


Figure 3 The mapping between the unit circle and the identity manifold.

#### 4 Inference Algorithm

We develop an inference algorithm to sequentially estimate the target state including the 3D position and the identity from a sequence of segmented target silhouettes  $\{z_t | t = 1, \dots, T\}$ . We cast this problem in the probabilistic graphical model shown in Figure 4. Specifically, the state vector  $X_t = [x_t \ y_t \ z_t \ \phi_t \ v_t]$  represents the target's position along the horizon, the elevation, and range directions, the heading direction (with respect to the sensor's optical axis) and the velocity in a 3D coordinate system.  $P_t$  is the camera projection matrix. Considering the fact that the camera in the SENSIAC dataset is static, we set  $P_t = P$ . We let  $\alpha_t \in [0, 2\pi)$  denote the angular identity variable.

In addition to  $\alpha_t$ , the generative model defined in (9) also needs the view parameter  $\Theta$ , which can be computed from  $X_t$  and  $P_t$ , in order to synthesize a target shape  $y_t$ . Target silhouettes used in training the generative model are obtained by imaging a 3D target model at a fixed distance from a virtual camera. Therefore  $y_t$  must be appropriately scaled to account for different imaging ranges. In summary, the synthesized silhouette  $y_t$  is a function of three factors:  $\alpha_t$ ,  $P_t$  and  $X_t$ . Given an observed target silhouette  $z_t$ , the problem of ATR becomes that of sequentially estimating the posterior probability  $p(\alpha_t, X_t | z_t)$ . Due to the nonlinear nature of this inference problem, we resort to the particle filtering approach [39] that requires the dynamics of the two variables  $p(X_t | X_{t-1})$  and  $p(\alpha_t | \alpha_{t-1})$  as well as a likelihood function  $p(z_t | \alpha_t, X_t)$  (the condition on  $P_t$  is ignored due to the assumption of a static camera in this work). Since the targets considered here are all ground vehicles, it is appropriate to employ a simple white noise motion model to represent the dynamics of  $X_t$  according to

$$\begin{cases} \varphi_t = \varphi_{t-1} + w_t^\varphi, \\ v_t = v_{t-1} + w_t^v, \\ x_t = x_{t-1} + v_{t-1} \sin(\varphi_{t-1}) \Delta t + w_t^x, \\ y_t = y_{t-1} + w_t^y, \\ z_t = z_{t-1} + v_{t-1} \cos(\varphi_{t-1}) \Delta t + w_t^z, \end{cases} \quad (10)$$

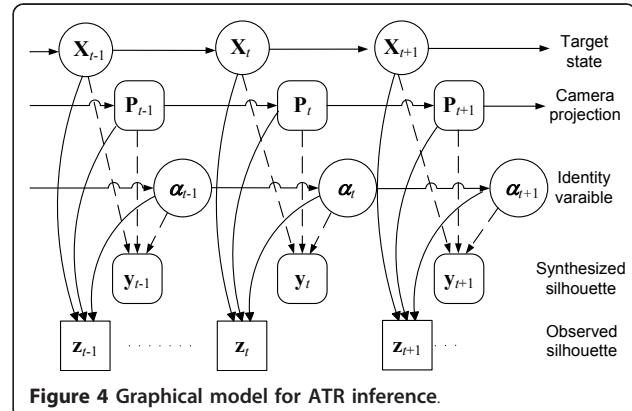


Figure 4 Graphical model for ATR inference.

where  $\Delta t$  is the time interval between two adjacent frames. The process noise associated with the target kinematics is Gaussian, i.e.,  $w_t^\varphi \sim N(0, \sigma_\varphi^2)$ ,  $w_t^x \sim N(0, \sigma_x^2)$ ,  $w_t^y \sim N(0, \sigma_x^2)$ ,  $w_t^y \sim N(0, \sigma_y^2)$ , and  $w_t^z \sim N(0, \sigma_z^2)$ . The Gaussian variances should be chosen to reflect the possible target dynamics and ground conditions. For example, if the candidate pool includes highly maneuvering targets, then large values  $\sigma_\varphi^2$  and  $\sigma_v^2$  are needed while tracking on a rough or uneven ground plane requires larger values  $\sigma_y^2$ .

Although the target identity does not change, the estimated identity value along the identity manifold could vary due to the uncertainty and ambiguity in the observations. We define the dynamics of  $\alpha_t$  to be a simple random walk as

$$\alpha_t = \alpha_{t-1} + w_t^\alpha, \quad (11)$$

where  $w_t^\alpha \sim N(0, \sigma_\alpha^2)$ . This model allows the estimated identity value to evolve along the identity manifold and converge to the correct one during sequential estimation. There are two possible future improvements to make this approach more efficient. One is to add an annealing treatment to reduce  $\sigma_\alpha^2$  over time and the other is to make  $\sigma_\alpha^2$  view-dependent. In other words, the variance can be reduced near the side view when the target is more discriminative and increased near front/rear views when it is more ambiguous.

Given the hypotheses on  $\mathbf{X}_t$  and  $\alpha_t$  in the  $t$ th frame as well as  $\mathbf{P}_t$ , the corresponding synthesized shape  $\mathbf{y}_t$  can be created by the generative model (9) followed by a scaling factor reflecting the range  $z_t \in \mathbf{X}_t$ . The likelihood function that measures the similarity between  $\mathbf{y}_t$  and  $\mathbf{z}_t$  is defined as

$$p(\mathbf{z}_t | \alpha_t, \mathbf{X}_t) \propto \exp \left[ -\frac{\|\mathbf{z}_t - \mathbf{y}_t\|^2}{2\sigma^2} \right], \quad (12)$$

**Table 1 Pseudo-code for the particle filter-based ATR algorithm**

- Initialization: Draw  $\mathbf{X}_0^j \sim N(\mathbf{X}_0, 1)$ , and  $\alpha_0^j = \alpha_0 \forall j \in \{1, \dots, N_p\}$ . Here  $\mathbf{X}_0$  and  $\alpha_0$  are the initial kinematic state and identity values, respectively.
- For  $t = 1, \dots, T$  (number of frames)
  1. For  $j = 1, \dots, N_p$  (number of particles)
    - 1.1 Draw samples  $\mathbf{X}_t^j \sim p(\mathbf{X}_t^j | \mathbf{X}_{t-1}^j)$  and  $\alpha_t^j \sim p(\alpha_t^j | \alpha_{t-1}^j)$  as in (10) and (11).
    - 1.2 Compute weights  $w_t^j = p(\mathbf{z}_t | \alpha_t^j, \mathbf{X}_t^j)$  using (12).
  - End
  2. Normalize the weights such that  $\sum_{j=1}^{N_p} w_t^j = 1$ .
  3. Compute the mean estimates of the kinematics and identity  $\hat{\mathbf{X}}_t = \sum_{j=1}^{N_p} w_t^j \mathbf{X}_t^j$  and  $\hat{\alpha}_t = \sum_{j=1}^{N_p} w_t^j \alpha_t^j$ .
  4. Set  $[\alpha_t^j, \mathbf{X}_t^j] = \text{resample}(\alpha_t^j, \mathbf{X}_t^j, w_t^j)$  to increase the effective number of particles [39].
- End

where  $\sigma^2$  controls the sensitivity of shape matching and  $\|\cdot\|^2$  gives the mean square error between the observed and hypothesized shape silhouettes. Pseudo-code for the particle filter-based inference algorithm is given below in Table 1.

## 5 Experimental results

We have developed four particle filter-based ATR algorithms that share the same inference framework shown in Figure 4 and by which we can evaluate the effectiveness of shape interpolation. Method-I uses the proposed target generative model involving both the view and identity manifolds for shape interpolation (i.e., both the identity and view variables are continuous). Method-II applies a simplified version where only the view manifold is involved for shape interpolation (i.e., the identity variable is discrete). Method-III involves shape interpolation along the identity manifold only (i.e., the view variable is discrete). Finally, Method-IV is a traditional template-based method that only uses the training data for shape matching without shape interpolation (i.e., both the view and identity variables are discrete).

We report three major experimental results in the following. First we present the learning of the proposed generative model along with some simulated results of shape interpolation. Then we introduce the SENSAC dataset [10] followed by detailed results on a set of IR sequences of various targets at multiple ranges. We also include three visible-based video sequences for algorithm evaluation, among which two were captured from remote-controlled toy vehicles in a room and one was from a real-world surveillance video. Background subtraction [40] was applied to all testing sequences to obtain the initial target segmentation result in each frame and the distance transform [29] was applied to create the observation sequences that were used for shape matching.

### 5.1 Generative Model Learning

We acquired six 3D CAD models for each of the six target classes (APCs, tanks, pick-ups, cars, minivans, SUVs)

for model learning, as shown in Figure 5. All 3D models were scaled to similar sizes and those in the same class share the same scaling factor. This class-dependent scaling is useful to learn the unified generative model and to estimate the range information in a 3D scene. For each 3D model, we generated a set of silhouettes corresponding to training viewpoints selected on the view manifold. For simplicity, we only considered elevation angles in the range  $0 \leq \varphi < 45^\circ$  and azimuth angles in the range  $0 \leq \theta < 360^\circ$ . Specifically, 150 training viewpoints were selected by setting  $12^\circ$  and  $10^\circ$  intervals along the azimuth and elevation angles, respectively, leading to non-uniformly distributed viewpoints on the view manifold. Ideally, we may need less training views when the elevation angle is large (close to the top-down view) to reduce the redundancy of training data. Our method of selecting training viewpoints is directly related to the kernel parameters set in (4) to ensure that model learning is effective and efficient. After model learning, we evaluated the generative model in terms of its shape interpolation capability through three experiments.

- *Shape interpolation along the view manifold*: We selected one target from each of the six classes and created three interpolated shapes (after thresholding) between three training views, as shown in Figure 6(a). We observe smooth transitions between the interpolated shapes and training shapes, especially around the wheels of the targets.

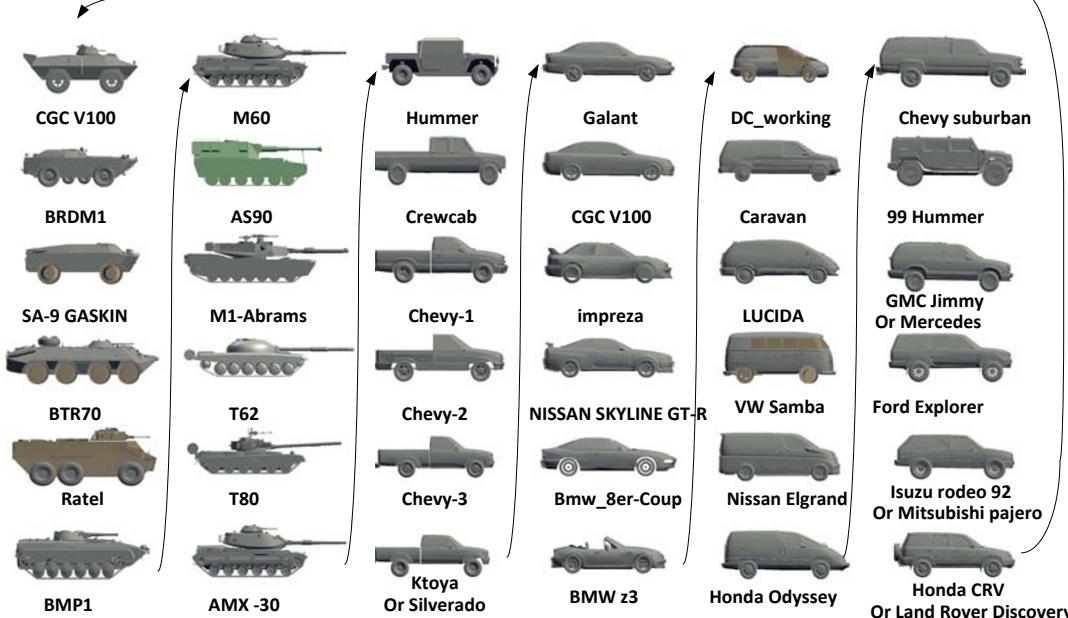
- *Shape interpolation along the identity manifold within the same class*: We generated six interpolated shapes along the identity manifold between three adjacent training targets for each of the six classes, as shown in Figure 6(b). Despite the fact that the three training targets are quite different in terms of their 3D structures, the interpolated shapes *blend* the spatial features from the two adjacent training targets in a natural way.

- *Shape interpolation along the identity manifold between two adjacent classes*: It is also interesting to see the shape interpolation results between two adjacent target classes, as shown in Figure 6(c). Although the series of shape variations may not be as smooth as that in Figure 6(b), the generative model still produces intermediate shapes between two vehicle classes that are realistic looking.

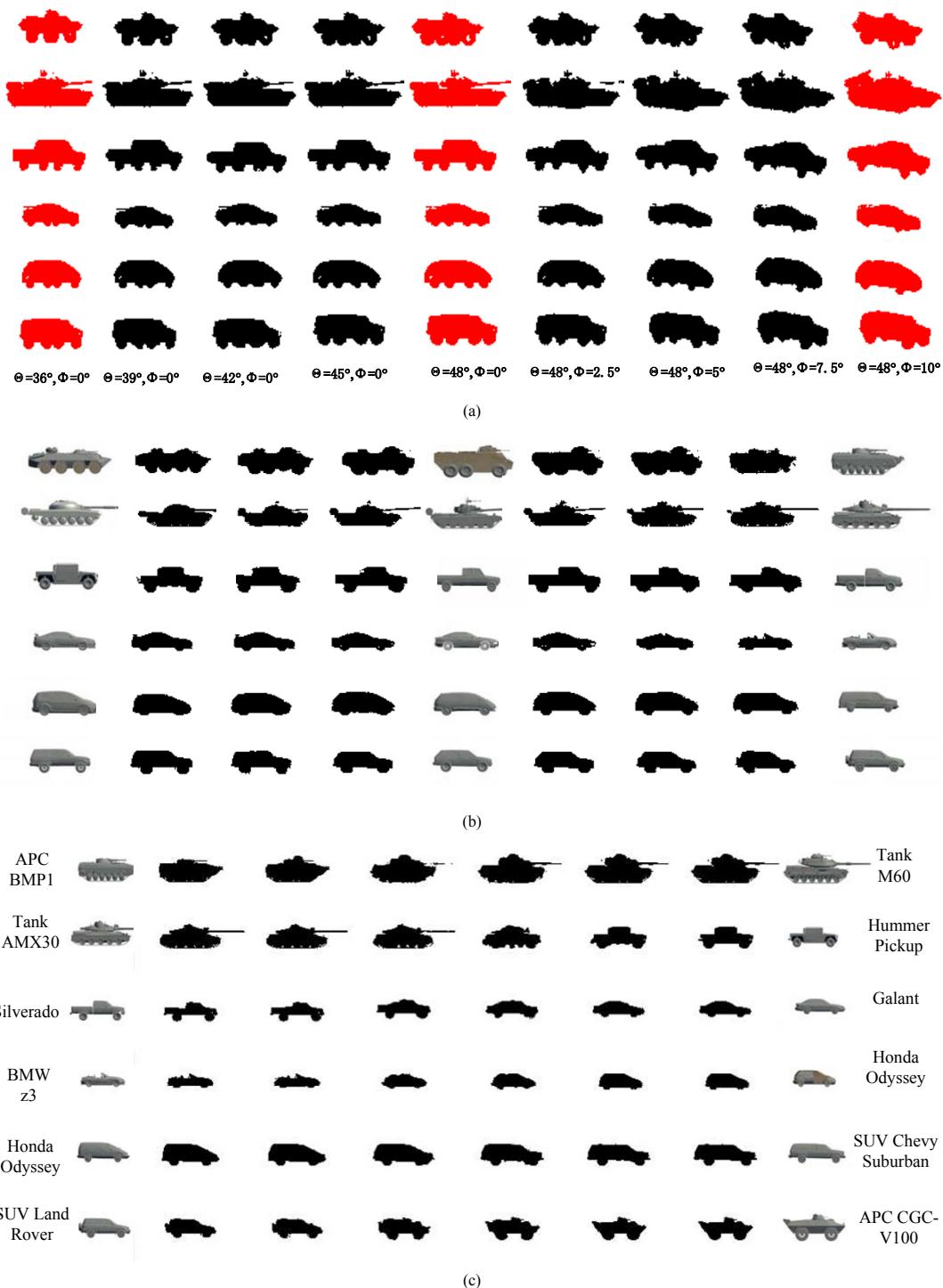
The above results show that the target model supports *semantically meaningful* shape interpolation along the two manifolds, making it possible to handle not only a known target seen from a new view but also an unknown target seen from arbitrary views. Also, the continuous nature of the view and identity variables facilitates the ATR inference process.

## 5.2 Tests on the SENSIAC database

The SENSIAC ATR database contains a large collection of visible and midwave IR (MWIR) imagery of six military and two civilian vehicles (Figure 7). The vehicles



**Figure 5** All 36 3D CAD models used for learning. There are six models for each target type (from left to right: APCs, tanks, pick-ups, cars, minivans and SUVs, ordered according to the manifold topology determined by (1)) and shown by arrowed lines.



**Figure 6 Shape interpolation along the view and identity manifolds for six target classes.** (a) Shape interpolation along the view manifold: the shapes of the first, middle and last columns are training cases that are adjacent on the view manifold, while the others are interpolated. The first and second training shapes are  $12^\circ$  apart along the azimuth angle, and the second and third ones are  $10^\circ$  apart in the elevation angle. (b) Shape interpolation along the identity manifold: the shapes of the first, middle and last columns are training cases that are adjacent on the identity manifold, while the others are interpolated. (c) Shape interpolation between two adjacent target classes along the identity manifold.



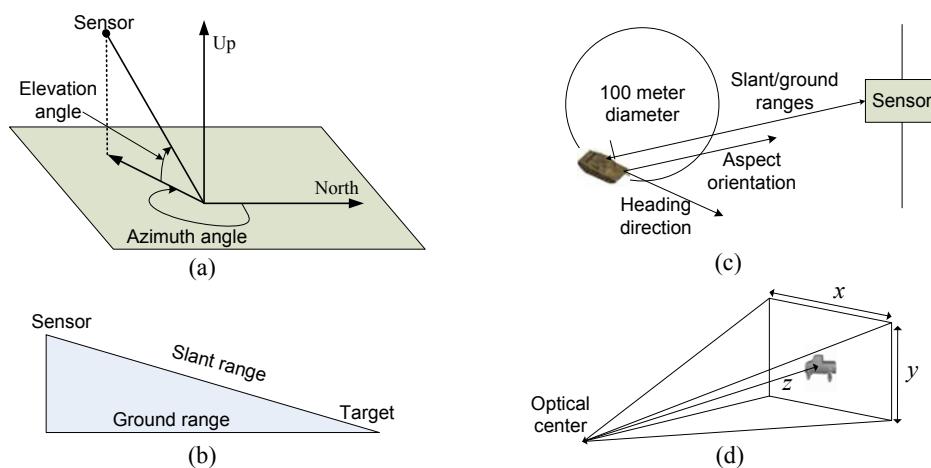
**Figure 7** The eight vehicles of the SENSIAC dataset used in algorithm evaluation.

were driven along a continuous circle marked on the ground with a diameter of 100 meters (m). They were imaged at a frame rate of 30 Hz for one minute from distances of 1,000 m to 5,000 m (with 500 m increment) during both day and night conditions. In the four ATR algorithms, we set  $\sigma_\phi^2 = 0.1$ ,  $\sigma_v^2 = 1$ ,  $\sigma_x^2 = 0.1$ ,  $\sigma_z^2 = 1$ ,  $\sigma_z^2 = 1$  in (10) and  $\sigma_\alpha^2 = 0.01$  in (11). We chose 48 night time IR sequences of eight vehicles at six ranges (1000 m, 1500 m, 2000 m, 2500 m, 3000 m, and 3500 m). Each sequence has approximately 1000 frames. Additionally, the SENSIAC database includes a rich set of meta data for each frame of every sequence. This information includes the true north offsets of the sensor (in azimuth and elevation, Figure 8(a)), the target type, the target speed, the range and slant ranges from the sensor to the target (Figure 8(b)), the pixel location of the target centroid, heading direction with respect to true north, and aspect orientation of the vehicle (Figure 8(c)). Furthermore, we defined a sensor-centered 3D

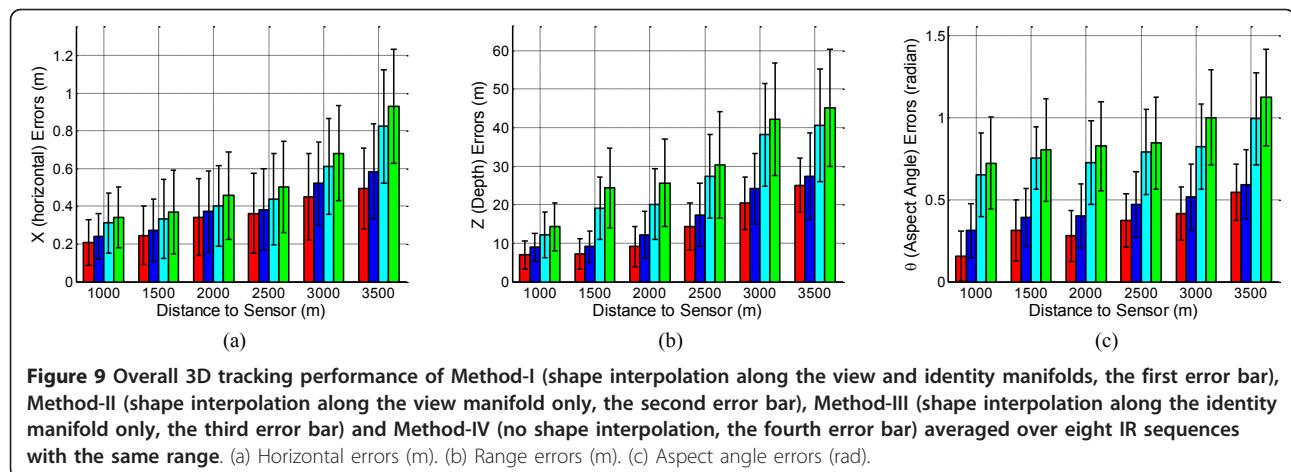
world coordinate system (Figure 8(d)) and developed a pinhole camera calibration technique to obtain the ground-truth 3D position of the target in each frame. The tracking performance is evaluated based on the errors in the estimated 3D position and aspect orientation.

### 5.2.1 Tracking Evaluation

We computed the errors in estimated 3D target positions along the  $x$  (horizon) and  $z$  (range) axes as shown in Figure 8(d), as well as of the aspect orientation of the target (Figure 8(c)). All tracking trials were initialized by the ground truth data in the first frame. The overall tracking performance averaged over eight targets with the same range is shown in Figure 9. All four algorithms achieved comparable errors of less than one meter along the horizon direction, with Method-I delivering performance gains of 10%, 20% - 40%, and 30% - 50% over Methods-II, III and IV, respectively. Method-I also outperforms the other three methods on the range and aspect estimation with over 10% - 50% and 20% - 80%



**Figure 8** Spatial geometry of the sensor and the target in the SENSIAC data. (a) The sensor orientation in a world coordinate system; (b) The slant and ground ranges between the sensor and target (side view); (c) The aspect orientation and the heading direction (top-down view); (d) Sensor-centered 3D coordinate system used for algorithm evaluation.

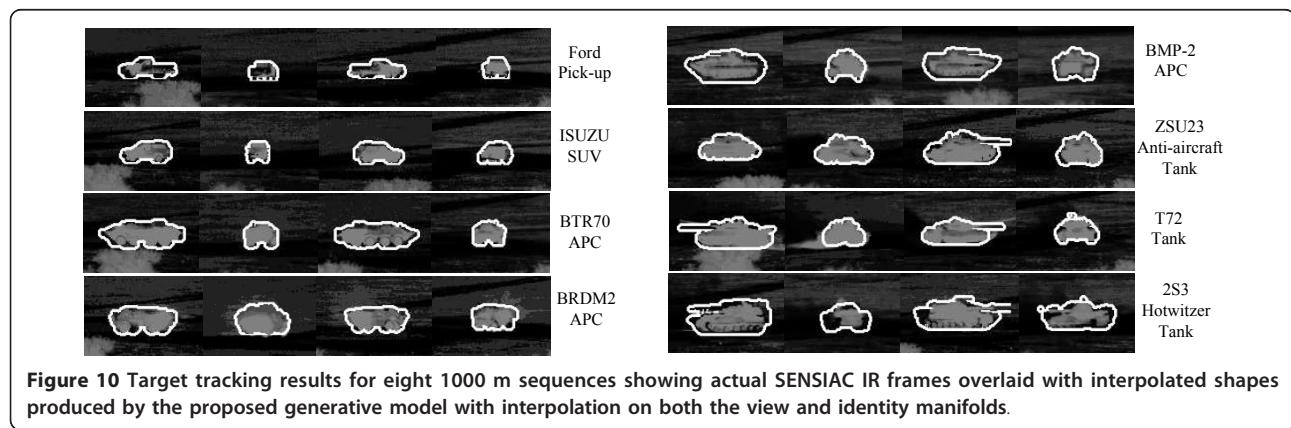


improvements. These results show that shape interpolation along the view manifold is more important than that along the identity manifold and that using both of them yields the best tracking performance. Even at a range of 3500 m, the averaged horizontal/depth/aspect errors of Method-I are only 0.5 m, 25 m, and 0.5 rad ( $28.7^\circ$ ), compared to the Method-IV errors of 0.9 m, 45 m, and 1.1 rad ( $63.1^\circ$ ). We also present some tracking results for Method-I against four 1000 m sequences in Figure 10, where the interpolated shapes are overlaid on the target according to the estimated 3D position and aspect angle as well as the given camera model. All of these results demonstrate the general usefulness of the generative model in interpolating target shapes along the view and identity manifolds for realistic ATR tasks.

### 5.2.2 Recognition Evaluation

As mentioned before, the ID closed-loop identity manifold learned from the tensor coefficient space can be mapped into a unit circle to ease the inference process. The identity variable then becomes an angular one  $\alpha \in [0, 2\pi]$ . Correspondingly, the six target classes, i.e., tanks, APCs, SUVs, picks-ups, minivans and cars, can be represented by six angular sections along the circularly

shaped identity manifold (as shown in Figure 1). Since the target type is estimated frame by frame during tracking, we define the overall recognition accuracy as the percentage of frames where the target is correctly classified in terms of the six classes. Also, it is interesting to check the two best-matched training targets for a given sequence that can be found along the identity manifold. The overall recognition results of the four methods for 48 sequences are shown in Table 2, where the accuracy of Tanks is averaged over the T72, ZSU23, and 2S3 target types and that of the APCs is averaged over those of BTR70, BMP2, and BRDM2 target types. Overall, Method-I outperforms the other three methods, again showing the usefulness of shape interpolation along both of the two manifolds. The improvements of Method-I are more significant for long-range sequences when the targets are small and shape interpolation is more important for correct recognition. The reason that recognition accuracies are below 80% for tanks and APCs at long ranges ( $\geq 2500$  m) is mainly because of small target sizes and poor segmentation results, as shown in Figure 11, which shows the targets in the original IR sequences as well as the segmented silhouettes.

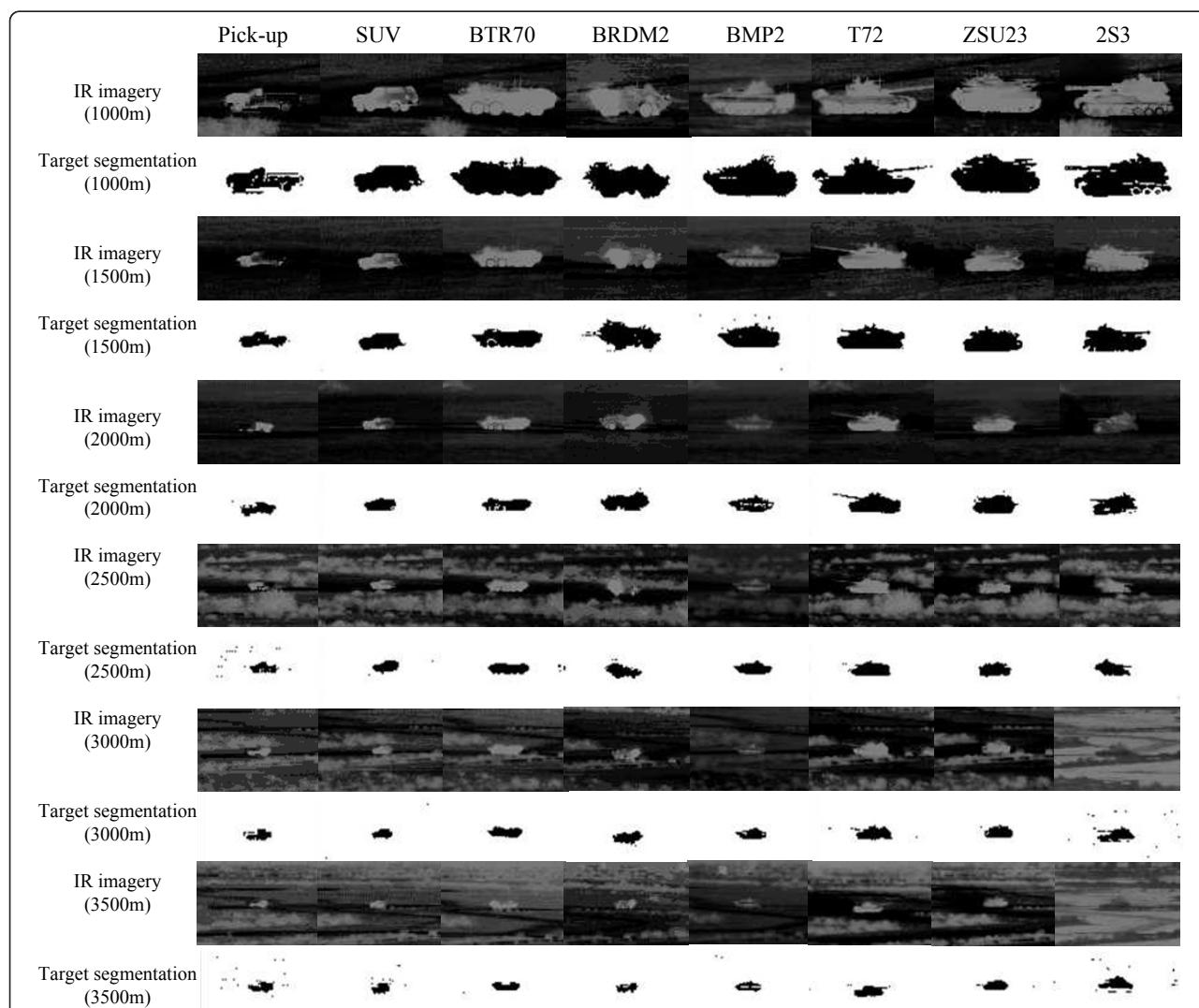


**Table 2 Overall recognition accuracies (%) of four methods (Method-I/Method-II/Method-III/Method-IV) against 48 SENSIAC sequences**

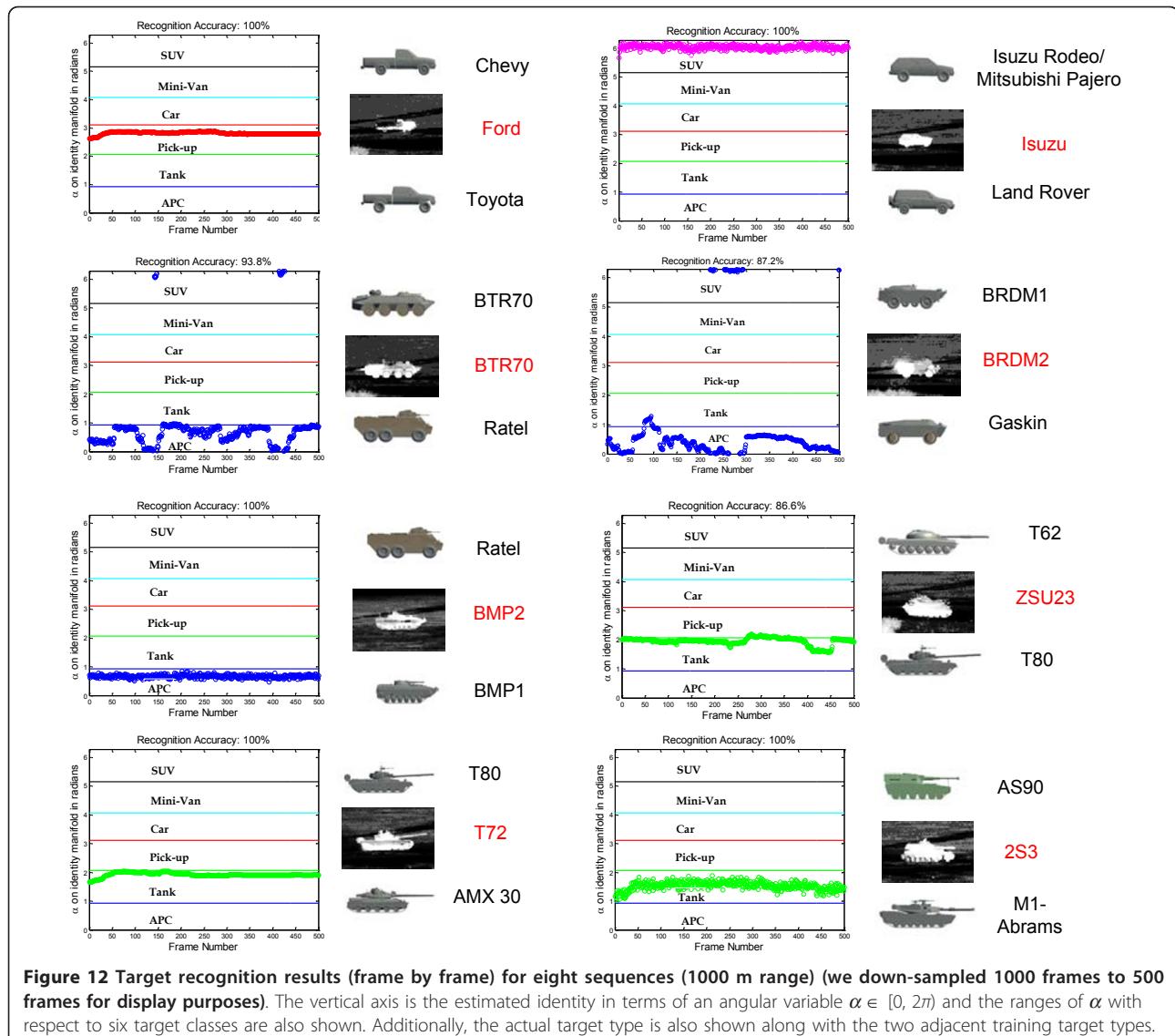
Targets/ranges	Tanks	APCs	SUV	Pick-up
1000 m	96/94/91/90	94/92/89/88	100/100/99/99	100/100/100/99
1500 m	93/91/88/86	88/86/85/82	100/99/98/98	100/100/100/98
2000 m	86/83/82/81	85/83/80/80	98/96/96/95	97/96/97/95
2500 m	78/73/72/69	76/72/71/70	92/90/89/86	90/88/88/86
3000 m	70/65/62/60	72/69/66/65	86/84/82/79	82/80/79/77
3500 m	68/62/58/57	70/65/64/62	78/76/75/70	73/72/70/65

We used a simple morphological opening operation to clean up the segmentation results. However, when the targets are small, morphological opening has to be moderate to ensure the target shapes are well preserved, which also results in noisier segmentations.

More details on the recognition results of Method-I for the eight 1000 m sequences are shown in Figure 12, which shows not only the frame-by-frame target recognition results but also the two best-matched training targets. In most frames, the estimated identity values are in the



**Figure 11** Snapshots of the original IR sequences of eight targets at six ranges along with the segmented targets.



**Figure 12** Target recognition results (frame by frame) for eight sequences (1000 m range) (we down-sampled 1000 frames to 500 frames for display purposes). The vertical axis is the estimated identity in terms of an angular variable  $\alpha \in [0, 2\pi]$  and the ranges of  $\alpha$  with respect to six target classes are also shown. Additionally, the actual target type is also shown along with the two adjacent training target types.

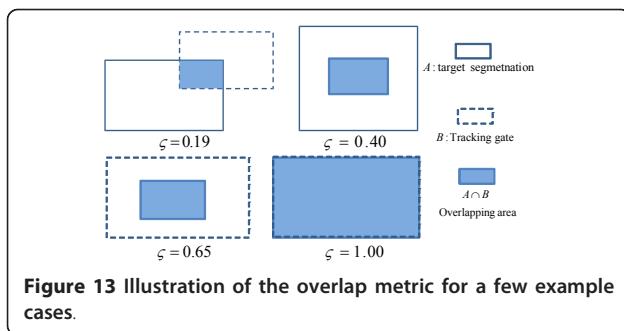
correct region of class and misclassification usually occurs around the front/rear views when the target is not very distinguishable. Interestingly, the two best matches for the BTR70 and ISUZU-SUV sequences include the exact correct target model. Also, the best matches for the other sequences include a similar target model. For example, BMP1, T72, BRDM1, and AS90 are among the two best matches for BMP2, T80, BRDM2 and 2S3, respectively.<sup>1</sup> We do not have 3D models for the Ford pick-up and the ZSU23 in our training set, but their best matches (Chevy/Toyota pick-ups and T62/T80 tanks) still resemble the actual targets in the SENSIAC sequences.

### 5.3 Results on Visible-band Sequences

We also tested the four ATR methods on three visible-band video sequences. Two of them (the car and the

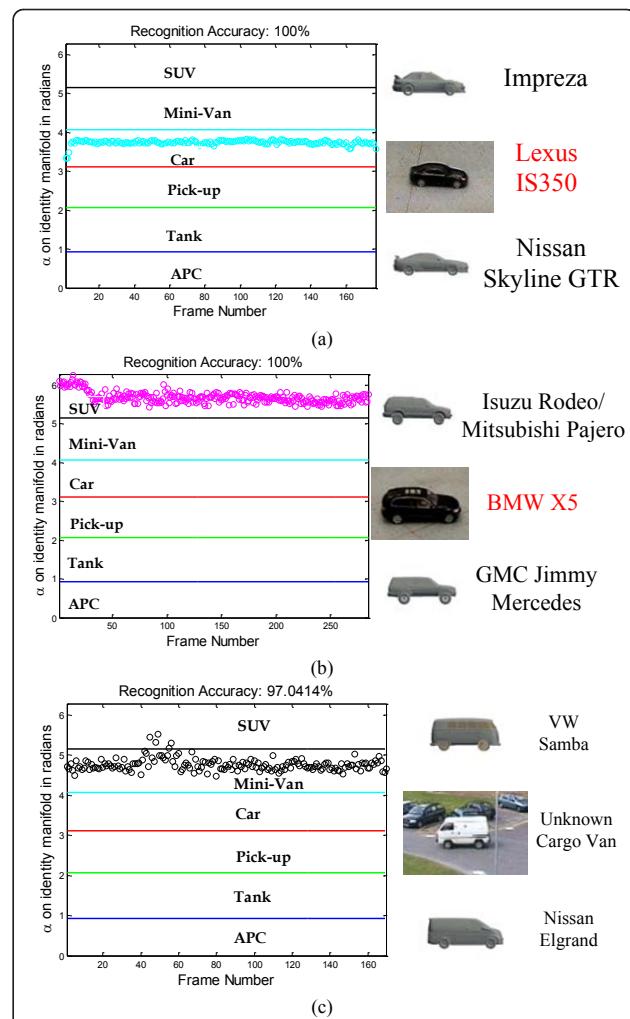
SUV) were captured indoors using a remote controlled toy vehicle where both the target pose and 3D position were estimated by making use of the camera calibration information, and one was a real-world surveillance video (the cargo van) for which camera calibration is not available and only pose estimation was performed from the normalized silhouette sequences. To compare the four methods, we used an overlap metric [41] to quantify the overlap between the interpolated shapes and the segmented target. Let  $A$  and  $B$  represent the tracking gate and the ground-truth bounding box respectively. Then the overlap ratio  $\zeta$  is defined as

$$\zeta = \frac{\#(A \cap B) \times 2}{\#(A) + \#(B)}, \quad (13)$$

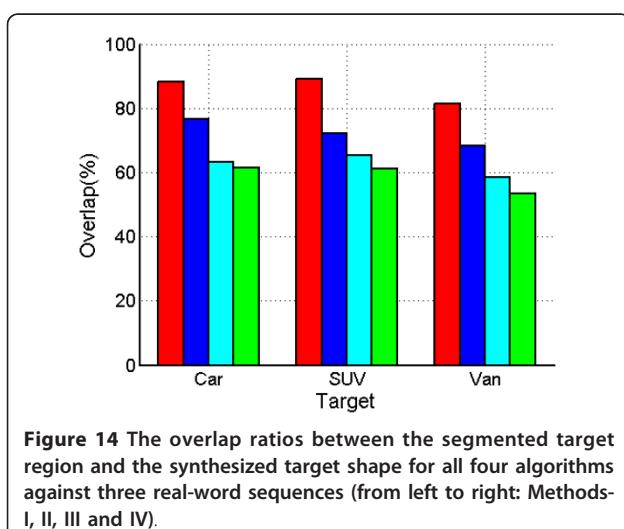


where # is the number of pixels. A larger  $\zeta$  ratio implies a better tracking performance, as shown in Figure 13. The overlap ratios of all four methods on three visible-band sequences are shown in Figure 14. It is clearly seen that Method-I is again superior to other three methods.

We now focus on the recognition results of Method-I, as shown in Figure 15. Although the three targets are previously unknown, the recognition accuracy is still 100% for the first two sequences and close to 100% (97%) for the last one. Note that the two best matches do indeed resemble the unknown target for each sequence. In particular, the cargo van in the third sequence is very different from all training models of minivan. Yet the two best matches, VW Samba and Nissan Elgrand, give a reasonable approximation. Detailed tracking results of the three sequences are shown in Figure 16. Although target segmentation results are not ideal in many frames, especially for the *cargo van*, the estimated pose trajectories along the view manifold are still smooth and represent the actual pose variation of the target during the sequence. Moreover, the interpolated shapes match reasonably well with the segmented targets, indicating the correct estimation for both the view and identity variables.

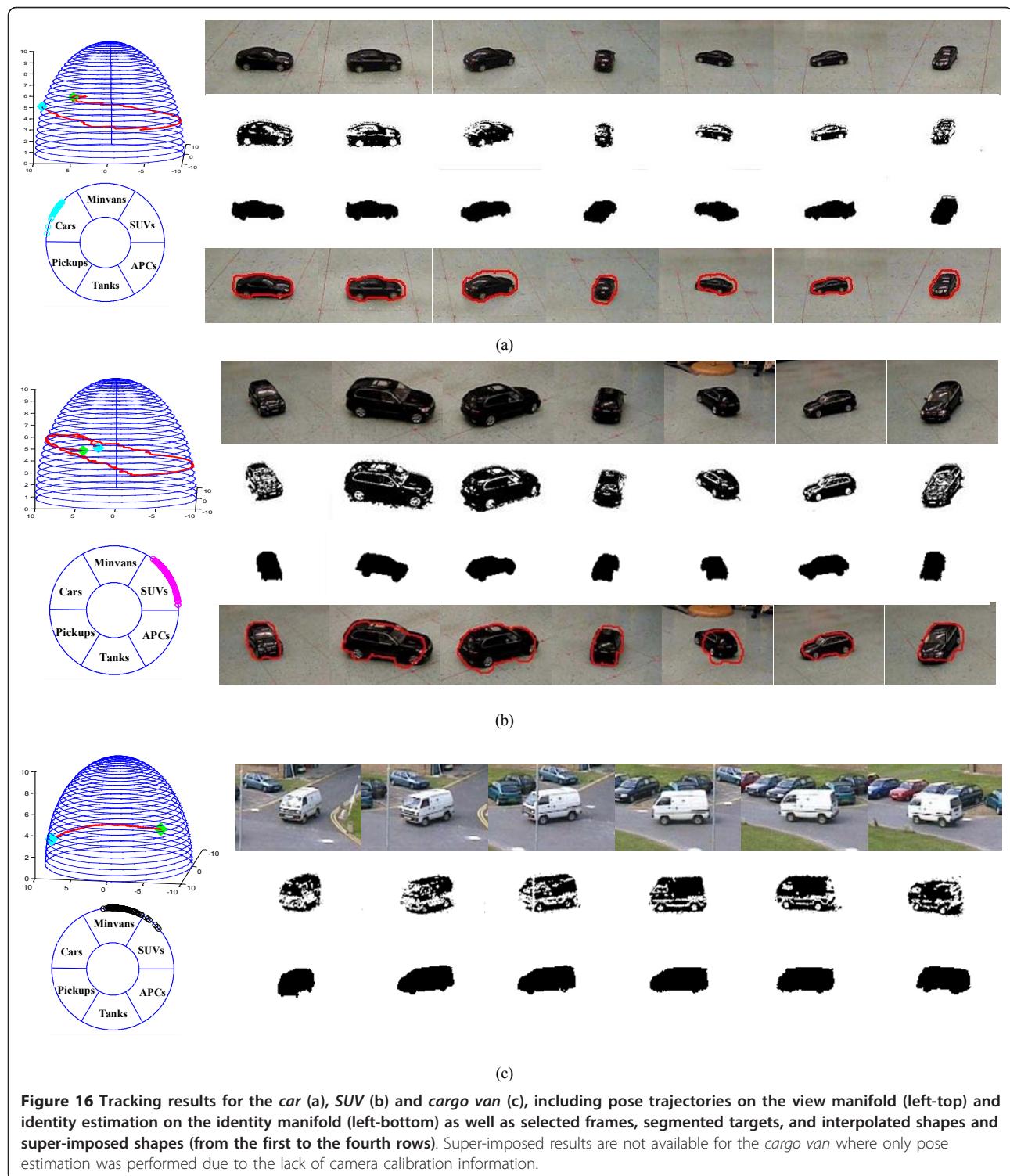


**Figure 15** Recognition results for the car (a), the SUV (b) and the cargo van (c) along with the two best matched training models.



#### 5.4 Discussion and Limitations

Although these results are promising, we still consider this work preliminary for three main reasons. First, the computational complexity of the proposed algorithm (Method-I) is relatively high due to the shape interpolation using the generative model. Our experimental results are based on a non-optimized Matlab implementation. Shape interpolation requires approximately 0.03s on a PC i7 computer (without parallel computation), and the inference with 200 particles requires about 6.9s per frame. Fast implementation is still needed to support real-time processing. Second, we use a silhouette-based shape representation that requires target segmentation. The background subtraction used here assumes that the camera platform is not moving. In the case of a moving camera platform, the initial target segmentation could become a challenging issue. Third, we did not



consider the issue of occlusion that has to be accounted for in any practical ATR system. The silhouette is a global feature that could be sensitive to occlusion. An extension to other more salient and robust features such as SIFT and HOG would increase the applicability of

the proposed method for real-world applications. Nevertheless, our main contribution is a new shape-based target model where, for the first time, both the view and identity variables are continuous and defined along their own respective manifolds.

## 6 Conclusion and Future Work

We have presented a new shape-based generative model that incorporates two continuous manifolds for multi-view target modeling. Specifically, the identity manifold was proposed to capture both inter-class and intra-class shape variability among different target types. The hemispherical view manifold is designed to reflect nearly all possible viewpoints. A particle filter-based ATR algorithm was presented that adopts the new target model for joint tracking and recognition. The experiments on both IR and visible-based video sequences show the advantages of shape interpolation along both the view and identity manifolds.

However, the current work only considers the silhouette-based shape for target representation that may not be sufficiently distinctive in some challenging cases. This work could be extended to other more salient and robust features thereby making the proposed model more promising for real-world applications. Another issue that needs further research is the structure and dimensionality of the identity manifold. In some sense, the 1D identity manifold used here is a practical simplification where a small set of training models (e.g., six models for each of the six classes, totally 36 in this work) is used for learning the generative model. It is possible we can learn a 2D or even 3D identity manifold for more generalized target modeling given sufficient training data. However, there will be two major challenges in going to a higher dimension space. One is how to learn an appropriate manifold topology in 2D or 3D, which is much harder than the 1D learning we considered here. The other is how to infer the identity variable effectively in a 2D or 3D identity manifold. There should be a balanced consideration of both the complexity and efficiency when using the couplet of view and identity manifolds for real-world ATR applications.

## Notes

<sup>1</sup>Both 1 AS90 and 2S3 are self-propelled howitzers.

## Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions that helped us improve this paper. This work was supported in part by the U.S. Army Research Laboratory and the U.S. Army Research Office under grants W911NF-04-1-0221 and W911NF-08-1-0293, the National Science Foundation under Grant IIS-0347613, and an OHRS award (HR09-030) from the Oklahoma Center for the Advancement of Science and Technology.

## Author details

<sup>1</sup>School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078, USA <sup>2</sup>School of Electronics and Information Engineering South China University of Technology, China <sup>3</sup>College of Computer Science, Zhongyuan Univ. of Technology, China <sup>4</sup>School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK 73019 USA

## Competing interests

The authors declare that they have no competing interests.

Received: 31 May 2011 Accepted: 7 December 2011

Published: 7 December 2011

## References

1. X Mei, SK Zhou, H Wu, Integrated Detection, Tracking and Recognition for IR Video-Based Vehicle Classification, in *Proc IEEE International Conference on Acoustics, Speech and Signal Processing* (2006)
2. MI Miller, U Grenander, JA Osullivan, DL Snyder, Automatic target recognition organized via jump-diffusion algorithms. *IEEE Trans Image Processing*. **6**, 157–174 (1997). doi:10.1109/83.552104
3. V Venkataraman, X Fan, G Fan, Integrated Target Tracking and Recognition using Joint Appearance-Motion Generative Models, in *Proc IEEE Workshop on Object Tracking and Classification Beyond Visible Spectrum (OTCBVS08)* in conjunction with CVPR08 (2008)
4. V Venkataraman, G Fan, X Fan, Target Tracking with Online Feature Selection in FLIR Imagery, in *Proc IEEE Workshop on Object Tracking and Classification Beyond Visible Spectrum (OTCBVS07)* in conjunction with CVPR07 (2007)
5. J Shaik, K Iftekharuddin, Automated tracking and classification of infrared images, in *Proc International Joint Conference on Neural Networks* (2003)
6. V Venkataraman, G Fan, X Fan, J Havlicek, Appearance Learning by Adaptive Kalman Filters for FLIR Tracking, in *Proc IEEE Workshop on Object Tracking and Classification Beyond Visible Spectrum (OTCBVS09)* in conjunction with CVPR09 (2009)
7. Z Zhang, W Dong, K Huang, T Tan, EDA Approach for Model Based Localization and Recognition of Vehicles. *Proc IEEE International Conference on Computer Vision and Pattern Recognition* (2007)
8. L Chan, N Nasrabadi, Modular wavelet-based vector quantization for automatic target recognition, in *Proc International Conference on Multisensor Fusion and Integration for Intelligent Systems* (1996)
9. L Wang, S Der, N Nasrabadi, Automatic target recognition using a feature-decomposition and data-decomposition modular neural network. *IEEE Trans Image Processing*. **7**(8), 1113–1121 (1998). doi:10.1109/83.704305
10. Military Sensing Information Analysis Center (SENSIAC) [Https://www.sensiac.org/](https://www.sensiac.org/) (2008)
11. T Poggio, S Edelman, A network that learns to recognize three-dimensional objects. *Nature*. **343**, 263–266 (1990). doi:10.1038/343263a0
12. S Ullman, R Basri, Recognition by Linear Combinations of Models. *IEEE Trans Pattern Analysis and Machine Intelligence*. **13**, 992–1006 (1991). doi:10.1109/34.99234
13. S Ullman, An Approach to Object Recognition: Aligning Pictorial Descriptions. *Cognition*. **32**, 193–254 (1989). doi:10.1016/0010-0277(89)90036-X
14. S Khan, H Cheng, D Matthes, H Sawhney, 3D model based vehicle classification in aerial imagery, in *Proc IEEE International Conf on Computer Vision and Pattern Recognition* (2010)
15. A Kushal, C Schmid, J Ponce, Flexible object models for category-level 3D object recognition, in *Proc IEEE International Conference on Computer Vision and Pattern Recognition* (2007)
16. H Su, M Sun, L Fei-Fei, S Savarese, Learning a dense multiview representation for detection, viewpoint classification and synthesis of object categories, in *Proc IEEE International Conference on Computer Vision* (2009)
17. O Ozcanli, A Tamrakar, B Kimia, Augmenting shape with appearance in vehicle category recognition, in *Proc IEEE International Conference on Computer Vision and Pattern Recognition* (2006)
18. S Savarese, L Fei-Fei, Multi-view Object Categorization and Pose Estimation, in *Computer Vision, Volume 285 of Studies in Computational Intelligence, Springer* (2010)
19. A Toshevand, A Makadia, K Daniilidis, Shape-based object recognition in videos using 3D synthetic object models, in *Proc IEEE International Conference on Computer Vision and Pattern Recognition* (2009)
20. J Lou, T Tan, W Hu, H Yang, S Maybank, 3-D model-based vehicle tracking. *IEEE Trans Image Processing*. **14**, 1561–1569 (2005)
21. M Leotta, J Mundy, Predicting high resolution image edges with a generic, adaptive, 3-D vehicle model, in *Proc IEEE International Conference on Computer Vision and Pattern Recognition* (2009)

22. R Sandhu, S Dambreville, A Yezzi, T A, Non-rigid 2D-3D pose estimation and 2D image segmentation, in *Proc IEEE International Conference on Computer Vision and Pattern Recognition* (2009)
23. Y Tsin, Y Gene, V Ramesh, Explicit 3D modeling for vehicle monitoring in non-overlapping cameras, in *Proc IEEE International Conference on Advanced Video and Signal based Surveillance* (2009)
24. J Liebelt, C Schmid, Multi-view object class detection with a 3D geometric model, in *Proc IEEE International Conference on Computer Vision and Pattern Recognition* (2010)
25. H Bülthoff, S Edelman, Psychophysical support for a 2D view interpolation theory of object recognition. *Proc of the National Academy of Science.* **89**, 60–64 (1992). doi:10.1073/pnas.89.1.60
26. M Abdelkader, W Abd-Almageed, A Srivastava, R Chellappa, Silhouette-based gesture and action recognition via modeling trajectories on Riemannian shape manifolds. *Computer Vision and Image Understanding.* **115**(3), 439–455 (2011). doi:10.1016/j.cviu.2010.10.006
27. S Belongie, J Malik, J Puzicha, Shape matching and object recognition using shape contexts. *IEEE Trans Pattern Analysis and Machine Intelligence.* **24**(4), 509–522 (2002). doi:10.1109/34.993558
28. M Hu, Visual pattern recognition by moment invariants. *IRE Trans Information Theory.* **8**(2), 179–187 (1962). doi:10.1109/TIT.1962.1057692
29. A Elgammal, CS Lee, Separating style and content on a non-linear manifold, in *Proc IEEE International Conference on Computer Vision and Pattern Recognition* (2004)
30. A Srivastava, S Joshi, W Mio, X Liu, Statistical shape analysis: clustering, learning, and testing. *IEEE Trans Pattern Analysis and Machine Intelligence.* **27**(4), 590–602 (2005)
31. H Murase, S Nayar, Visual learning and recognition of 3D objects from appearance. *International Journal of Computer Vision.* **14**, 5–24 (1995). doi:10.1007/BF01421486
32. J Tenenbaum, WT Freeman, Separating style and content with bilinear models. *Neural Computation.* **12**, 1247–1283 (2000). doi:10.1162/089976600300015349
33. MAO Vasilescu, D Terzopoulos, Multilinear analysis of image ensembles: Tensorfaces, in *Proc IEEE European Conference on Computer Vision* (2002)
34. C Gosch, K Fundana, A Heyden, C Schnörr, View point tracking of rigid objects based on shape sub-manifolds, in *Proc European Conference on Computer Vision* (2008)
35. C Lee, A Elgammal, Modeling View and Posture Manifolds for Tracking, in *Proc IEEE International Conference on Computer Vision* (2007)
36. A Elgammal, CS Lee, Tracking people on torus. *IEEE Trans on Pattern Analysis and Machine Intelligence.* **31**, 520–538 (2009)
37. C Lee, A Elgammal, Simultaneous Inference of View and Body Pose using Torus Manifolds, in *Proc IEEE Int'l Conference on Pattern Recognition* (2006)
38. MAO Vasilescu, D Terzopoulos, Multilinear image analysis for facial recognition, in *Proc IEEE International Conference on Pattern Recognition* (2002)
39. S Arulampalam, S Maskell, N Gordon, T Clapp, A Tutorial on Particle Filters for Online Non-linear/Non-Gaussian Bayesian Tracking. *IEEE Trans Signal Processing.* **50**(2), 174–188 (2002). doi:10.1109/78.978374
40. Z Zivkovic, F van der Heijden, Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters.* **27**, 773–780 (2006). doi:10.1016/j.patrec.2005.11.005
41. K She, G Bebis, H Gu, R Miller, Vehicle Tracking Using On-Fusion of Color and Shape Features, in *Proc IEEE International Conference on Intelligent Transportation Systems* (2004)

doi:10.1186/1687-6180-2011-124

**Cite this article as:** Venkataraman et al.: Automated target tracking and recognition using coupled view and identity manifolds for shape representation. *EURASIP Journal on Advances in Signal Processing* 2011 **2011**:124.

## Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)