Carrying Object Detection Using Pose Preserving Dynamic Shape Models

Chan-Su Lee and Ahmed Elgammal

Department of Computer Science,
Rutgers University, Piscataway, NJ, USA
{chansu, elgammal}@cs.rutgers.edu

Abstract. In this paper, we introduce a framework for carrying object detection in different people from different views using pose preserving dynamic shape models. We model dynamic shape deformations in different people using kinematics manifold embedding and decomposable generative models by kernel map and multilinear analysis. The generative model supports pose-preserving shape reconstruction in different people, views and body poses. Iterative estimation of shape style and view with pose preserving generative model allows estimation of outlier in addition to accurate body pose. The model is also used for hole filling in the background-subtracted silhouettes using mask generated from the best fitting shape model. Experimental results show accurate estimation of carrying objects with hole filling in discrete and continuous view variations.

1 Introduction

This paper presents a new approach for carrying object detection using a dynamic shape model of human motion with decomposition of body pose, shape style and view. To model nonlinear shape deformations by multiple factors, we propose kinematics manifold embedding and kernel mapping in addition to multilinear analysis of collected nonlinear mappings. The kinematics manifold embedding, which represents body configuration invariant to different people and views in low dimensional space based on motion captured data, is used to model dynamics of shape deformation according to intrinsic body configuration. The intrinsic body configuration has one-to-one correspondence with kinematics manifold (Sec. 2.1). Using this kinematics manifold embedding, individual differences of shape deformations can be solely contained in nonlinear mappings between manifold embedding points and observed shapes. By utilizing multilinear analysis for collection of these nonlinear mappings in different people and views, we can achieve decompositions of shape styles and views in addition to the body poses (Sec. 2.2). Iterative estimation of body pose, shape style and view parameters for the given decomposable generative model provides pose preserving, style preserving reconstruction of shape in different view human motion (Sec. 2.3).

The proposed pose preserving, dynamic shape models are used to detect carrying objects from sequences of silhouette images. The detection of carrying objects is one of the key element in visual surveillance systems [7]. The performance of gait recognition is degraded dramatically when people carry objects like briefcases [13]. Our pose-preserving dynamic shape model detects carrying objects as outliers. By removing

F.J. Perales and R.B. Fisher (Eds.): AMDO 2006, LNCS 4069, pp. 315–325, 2006.
© Springer-Verlag Berlin Heidelberg 2006

outliers from extracted shape, we can estimate body pose and other factors accurately in spite of variations of shapes due to carrying objects (Sec. 3.3). Hole filling based on signed distance representation of shape (Sec. 3.1) also helps correcting shapes from inaccurate background subtraction (Sec. 3.2). Iterative procedure of hole filling and outlier detection using pose preserving shape reconstruction achieves gradual hole filling and advance in precision of carrying objects detection in iterations (Sec. 3.4). Experimental results using CMU Mobo gait database [6] and our own dataset from multiple views show accurate estimation of carrying object with correction of silhouettes from multiple people and multiple view silhouettes with holes (Sec. 4).

1.1 Related Work

There have been a lot of work on contour tracking from cluttered environment such as active shape models (ASM) [2], active contours [8], and exemplar-based tracking [15]. However, there are few works to model shape variations in different people and views as a generative model with capturing nonlinear shape deformations. The framework to separate the motion from the style in a generative fashion was introduced in [5] where the motion is represented in a low dimensional nonlinear manifold. Nonlinear manifold learning technique can be used to find intrinsic body configuration space [18,5]. However, discovered manifolds are twisted differently according to person styles, views, and other factors like clothes in image sequences [4]. We propose kinematics manifold embedding as an alternative uniform representation of intrinsic body configuration (Sec. 2.1).

In spite of the importance of carrying objects detection in visual surveillance system, there has been few works focused on carrying objects detection due to difficulties in modeling variations of shape due to carrying objects. By analyzing symmetry in silhouette model, carrying objects can be detected by aperiodic outlier regions [7]. Amplitude of the shape feature and the location of detected objects are constrained in [1] to improve accuracy of carrying object detection. Detecting outlier accurately and removing noise and filling hole in extracted silhouette still remains unresolved.

Shape models are used for segmentation and tracking using level sets [16,11]. Shape priors can be used for pose-preserving shape estimation. However, previous shape prior models like [11] cannot represent dynamic characteristics of shape deformations in human motion. This paper proposes gradual detection of outlier, and correction of noise silhouette by hole filling and outlier removal using pose-preserving dynamic shape model.

2 Pose Preserving Dynamic Shape Models

We can think of the shape of a dynamic object as instances driven from a generative model. Let $y_t \in \mathbb{R}^d$ be the shape of the object at time instance t represented as a point in a d-dimensional space. This instance of the shape is driven from a model in the form

$$y_t = \gamma(b_t; s, v), \tag{1}$$

where the $\gamma(\cdot)$ is a nonlinear mapping function that maps from a representation of the body pose b_t into the observation space given a mapping parameter s, v that characterizes shape style and view variations in a way independent of the configuration. Given

this generative model, we can fully describe observation instance y_t by state parameters b_t, s, and v. In the generative model, we model body pose b_t invariant to the view and shape style. We need a unified representation for body configuration invariant to the variation of observation in different person and in different view. Kinematics manifold embedding is used for intrinsic manifold representation of body configuration b_t.

2.1 Kinematics Manifold Embedding

We find low dimensional representation of kinematics manifold by applying nonlinear dimensionality reduction techniques for motion captured data. We first convert joint angles of motion capture data into joint locations in three dimensional spaces. We align global transformation in advance in order to model motions only due to body configuration change. Locally linear embedding (LLE) [12] is applied to find low dimensional intrinsic representation from the high dimensional data (collections of joint locations). The discovered manifold is one-dimensional twisted circular manifold in three-dimensional spaces.

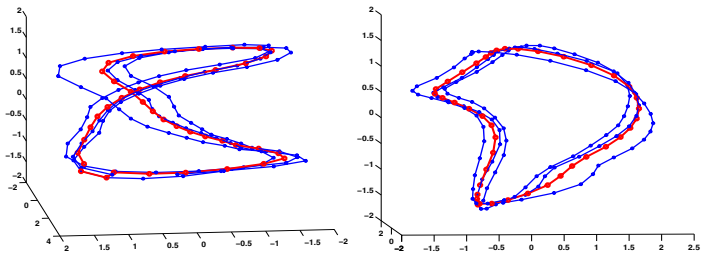

Fig. 1. Kinematics manifold embedding and its mean manifold: two different views in 3D space

The discovered manifold is represented using a one-dimensional parameter by spline fitting. We use multiple cycles to find kinematics intrinsic manifold representation by LLE. For the parametrization of kinematics manifold, we use mean-manifold representation from the multiple cycle manifold. The mean manifold can be found by averaging multiple cycles after detecting cycles by measuring geodesic distance along the manifold. The mean-manifold is parameterized by spline fitting by a one-dimensional parameter $\beta_t \in \mathbb{R}$ and a spline fitting function $g : \mathbb{R} \rightarrow \mathbb{R}^3$ that satisfies $b_t = g(\beta_t)$, which is used to map from the parameter space into the three dimensional embedding space. Fig. 1 shows a low dimensional manifold from multiple cycles motion captured data and their kinematics mean manifold representation.

2.2 Modeling Shape Variations Using Decomposable Generative Models

Individual variations of shape deformations can be discovered in the nonlinear mapping space between the kinematics manifold embedding and the observation in different people. If we have pose-aligned shapes for all the people, then it becomes relatively easy to model shape variations in different people. Similarly, as we have common representation of the body pose, all the differences of the shape deformation can be contained

in the mapping between the embedding points and observation sequences. We employ nonlinear mapping based on empirical kernel map [14] to capture nonlinear deformation in difference body pose. There are three steps to model shape deformations in decomposable nonlinear generative models. Here we focus on walking sequence but the framework can be applicable to other motion analysis in different variation factors.

First, for a given shape deformation sequence, we detect gait cycles and embed collected shape deformation data to the intrinsic manifold. In our case, kinematics manifold is used for embedding in each detected gait cycle. As the kinematics manifold comes from constant speed walking motion captured data, we embed the shape sequence in equally spaced points along the manifold. Second, we learn nonlinear mappings between the kinematics embedding space and shape sequences. According to the representer theorem [9], we can find a nonlinear mapping that minimizes the regularized risk in the following form:

$$f(x) = \sum_{i=1}^{m} \alpha_i k(x_i, x), \tag{2}$$

for given patterns x_i and target values $y_i = f(x_i)$. The solutions lie on the linear span of kernels centered on data points. The theorem shows that any nonlinear mapping is equivalent to a linear projection from a kernel map space. In our case, this kernel map allows modeling of motion sequence with different number of frames as a common linear projection from the kernel map space. The mapping coefficients of the linear projection can be obtained by solving the linear system

$$[y_1^{sv} \cdots y_{N_{sv}}^{sv}] = C^{sv}[\psi(x_1^{sv}) \cdots \psi(x_{N_{sv}}^{sv})]. \tag{3}$$

Given motion sequences with N_s shape styles and N_v views, we obtain $N_s \times N_v$ number of mapping coefficients. Third, multi-linear tensor analysis is applied to decompose the gait motion mappings into orthogonal factors. Tensor decomposition is achieved by higher-order singular value decomposition (HOSVD) [17], which is a generalization of SVD. All the coefficient vectors can be arranged in an order-three gait motion coefficient tensor $\mathscr{C}$ with a dimension of $N_s \times N_v \times N_c$, where N_c is the dimension of the mapping coefficients. The coefficient tensor can be decomposed as $\mathscr{C} = \mathscr{A} \times_1 S \times_2 V \times_3 F$ where S is the collection of the orthogonal basis for the shape style subspace. V represents the orthogonal basis of the view space and F represents the basis of the mapping coefficient space. $\mathscr{A}$ is a core tensor which governs the interactions among different mode bases.

The overall generative model can be expressed as

$$y_t = \mathscr{A} \times s \times v \times \psi(b_t). \tag{4}$$

The pose preserving reconstruction problem using this generative model is the estimation of parameter b_t, s, and v at each new frame given shape y^t.

2.3 Pose Preserving Reconstruction

When we know the state of the decomposable generative model, we can synthesize the corresponding dynamic shape. For given body pose parameter, we can reconstruct

best fitting shape by estimating style parameter and view parameter with preserving the body pose. Similarly, when we know body pose parameter and view parameter, we can reconstruct best fitting shape by estimating style parameter with preserving view and body pose. If we want to synthesize new shape at time t for a given shape normalized input y_t, we need to estimate the body pose x_t , the view v, and the shape style s which minimize the reconstruction error

$$E(x_t, v, s) = || \, y_t - \mathscr{A} \times v \times s \times \psi(x_t) \, || \, . \tag{5}$$

We assume that the estimated optimal style can be written as a linear combination of style class vectors in the training model. Therefore, we need to solve for linear regression weights α such that $s^{est} = \sum_{k=1}^{K_s} \alpha_k s^k$ where each s^k is one of the K_s shape style vectors in the training data. Similarly for the view, we need to solve for weights β such that $v^{est} = \sum_{k=1}^{K_v} \beta_k v^k$ where each v^k is one of the K_v view class vectors.

If the shape style and view factors are known, then equation 5 reduces to a nonlinear 1-dimensional search problem for a body pose b_t on the kinematics manifold that minimizes the error. On the other hand, if the body pose and the shape style factor are known, we can obtain view conditional class probabilities $p(v^k|y_t, b_t, s)$ which is proportional to the observation likelihood $p(y_t \mid b_t, s, v^k)$. Such the likelihood can be estimated assuming a Gaussian density centered around $\mathscr{A} \times v^k \times s \times \psi(b_t)$, i.e.,$p(y \mid b_t, s, v^k) \approx \mathscr{N}(\mathscr{C} \times v^k \times s \times \psi(b_t), \Sigma^{v^k})$.

Given view class probabilities we can set the weights to $\beta_k = p(v^k \mid y, b_t, s)$. Similarly, if the body pose and the view factor are known, we can obtain the shape style weights by evaluating the shape factor given each shape style class vector s^k assuming a Gaussian density centered at $C \times v \times s^k \times \psi(b_t)$. An iterative procedure similar to a deterministic annealing where in the beginning the each view and shape style weights are forced to be close to uniform weights to avoid hard decisions about view and shape style classes, is used to estimate x_t, v, s from given input y_t. To achieve this, we use variables, view and style class variances, that are uniform to all classes and are defined as $\Sigma^e = T_v \sigma_v^2 I$ and $\Sigma^s = T_s \sigma_s^2 I$ respectively. The parameters T_v and T_s start with large values and are gradually reduced and in each step and a new configuration estimate is computed.

3 Carrying Object Detection

We can detect carrying objects by iterative estimation of outlier using the generative model that can synthesize pose-preserving shapes. In order to achieve better alignment in normalized shape representation, we performed hole filling and outlier removal for the extracted shape iteratively.

3.1 Shape Representation

For consistent representation of shape deformations in variant factors, we normalize silhouette shapes by resizing and re-centering. To be invariant to the distance from camera and different height in each subject, we normalized the extracted silhouette height from background-subtracted silhouettes. In addition, the horizontal center of the shape is re-centered by the center of gravity of silhouette blocks. We use silhouette

blocks whose sizes are larger than a specific threshold value for consistent centering of shape in spite of small incorrect background block due to noise and shadow. we perform normalization after morphological operation and filtering to remove noise spot and small holes.

We parameterize the motion shape contour using signed distance function with limitation of maximum distance for robust shape representation in learning and matching shape contour. Implicit function z(x) at each pixel x such that $z(x) = 0$ on the contour, $z(x) > 0$ inside the contour, and $z(x) < 0$ outside the contour are used, which is typically used in level-set methods [10]. We add threshold values $d_c^{TH_p} - d_c^{TH_n}$ as follows,

$$z(x) = \begin{cases} d_c^{TH_p} & d_c(x) \geq d_c^{TH_p} \\ d_c(x) & x \text{ inside } c \\ 0 & x \text{ on } c \\ -d_c(x) & x \text{ outside } c \\ -d_c^{TH_n} & -d_c(x) \leq -d_c^{TH_n} \end{cases}, \tag{6}$$

where the $d_c(x)$ is the distance to the closest point on the contour c with a positive sign inside the contour and a negative sign outside the contour. We threshold distance value $d_c(x)$ by $d_c^{TH_p}$ and $-d_c^{TH_n}$ as the distance value beyond certain distance does not contain meaningful shape information in similarity measurements. Such representation imposes smoothness on the distance between shapes and robustness to noise and outlier. In addition, by changing threshold value gradually, we can generate mask to represent inside of the shape, which is useful in gradual hole filling. Given such representation, an input shape sequence is points in a d dimensional space, $y_i \in \mathbb{R}^d, i = 1, \cdots, N$ where all the input shapes are normalized and registered and d is the dimensionality of the input shape vector , and N is the number of frame in the sequence.

3.2 Hole Filling

We fill holes in the background-subtracted shape to attain more accurate normalized shape representation. When the foreground color and the background color are the same, most of the background subtracted shape silhouettes have holes inside the extracted shape. This can cause inaccurate description of shape in normalization and in signed distance representation. A Hole can induce misalignment in normalized shape as the hole can cause shifting the center of gravity for the horizontal axis alignment. In addition, holes inside shape result in inaccurate shape description in signed distance representation. So, holes can cause incorrect estimations of the best fitting shape to the given observation.

We utilize *inside shape mask* generated from shape models to fill holes in the original shape. We can generate the mask to represent inside of the shape for estimated style, view, and body pose parameters by threshold in the signed distance representation.

$$h(x)_{hole\,mask} = \begin{cases} 1 & d_c(x) \geq d_c^{TH_{hole}} \\ 0 & \text{otherwise} \end{cases}, \tag{7}$$

where $d_c^{TH_{hole}} \geq 0$ is the threshold value for the inner shape mask for hole filling. If the threshold value is zero, the mask will be the same as the silhouette image generated

by the nonlinear shape model for the given style, view and configuration. As we don't know the exact shape style, view and configuration at the beginning, and as holes can causes misalignments, we start from a large threshold value, which generates a small mask of the inner shape area in order to be robust to any misalignment and inaccurate state estimation. We reduce the threshold value as estimated model parameters get more accurate.

The hole filling operation can be described by $y_{hole\ filling} = z(\texttt{bin}(y) \oplus h(y^{est}))$, where $\oplus$ is logical OR operator to combine extracted foreground silhouette and mask area, $\texttt{bin}(\cdot)$ converts signed distance shape representation into binary representation, and $z(\cdot)$ convert binary representation into signed distance representation with threshold. Fig. 2 shows an initial shape normalized silhouette with holes (a), the best estimated shape model (b) which is generated from the generative model with estimated style and view parameters and body configuration, and the hole mask (c) when $d_c^{TH_{hole}} = 3$, and a new shape after hole filling (d). We improve the best matching shape by excluding mask area in the computation of the similarity measurement for generated samples in searching the best fitting body pose. Re-alignment of the shape and re-computation of the shape representation after hole filling provide better shape description for next iteration.

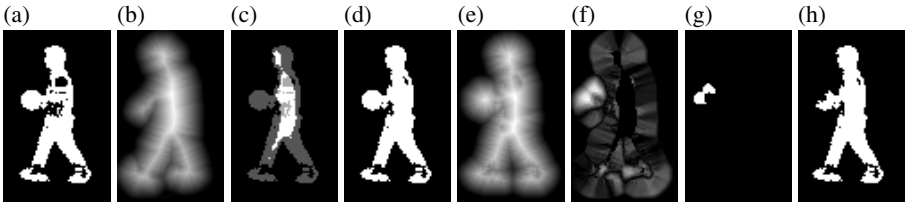

Fig. 2. Hole filling using mask from the best fitting model : (a) Initial normalized shape with holes. (b) The best matching shape. (c) Overlapping with initial silhouette and mask from the best matching shape. (d) New shape with hole filling. (e) A normalized shape for outlier detection. (f) Euclidian distance error. (g) Detected outliers. (h) A new shape after outlier removal.

3.3 Carrying Object Detection

Carrying objects are detected by estimating outliers from the best matching normal dynamic shape from the given input shape. Outliers of a shape silhouette with carrying objects are mismatching parts in input shape compared with the best matching normal walking shape. Carrying objects are the major source of mismatching when we compare with normal walking shape even though other factors such as inaccurate shape extraction, shape misalignment can also cause mismatches. For accurate detection of carrying objects from outliers, we need to remove other source of outlier such as holes and misalignment in shapes. Hole filling and outlier removal are performed iteratively to improve shape representation for better estimation of the matching shape.

We gradually reduce the threshold value for outlier detection to get more precise estimation of outlier progressively. The mismatching error $e(x)$ is measured by Euclidian distance between signed distance input shape and best matching shape generated from the dynamic shape model,

$$e_c(x) = ||z_c(x) - z_c^{est}(x)|| \,. \tag{8}$$

The error $e(x)$ increases linearly as the outlier goes away from the matching shape contour due to signed distance representation. By threshold the error distance, we can detect outliers.

$$O(x)_{outliermask} = \begin{cases} 1 & e_c(x) \geq e_c^{TH_{outlier}} \\ 0 & \text{otherwise} \end{cases}, \tag{9}$$

At the beginning, we start from large $e_c^{TH_{outlier}}$ value. We reduce the threshold value gradually. Whenever we detect outliers, we remove the detected outlier areas and perform realignment to reduce misalignment due to the outliers. In Fig. 2, for given signed distance input shape (e), we measure mismatching error (f) by comparing with best matching shape (b). Outlier is detected (g) with given threshold value $e_c^{TH_{outlier}} = 5$, and new shape for next iteration is generated by removing outlier (h). This outlier detection and removal procedure is combined with hole filling as both of them help accurate alignment of shape and estimation of the best matching shape.

3.4 Iterative Estimation of Outliers with Hole Filling

An iterative gradual estimation of outliers, hole filling and outlier removal is performed by threshold value control. The threshold value for hole filling and the threshold value for outlier detection need to be decreased to get more precise in the outlier detection and hole filling in each iteration. In addition, we control the number of samples to search body pose for estimated view and shape style. At the initial stage, as we don't know accurate shape style and view, we use small number of samples along the equally distant manifold points. As the estimation progress, we increase accuracy of body pose estmation with increased number of samples. We summarize the iterative estimation as follows:

Input: image shape y_b, estimated view v, estimated style s, core tensor $\mathcal{A}$
Initialization: – initialize sample num N_{sp}, $d_c^{TH_{hole}}$, $e_c^{TH_{outlier}}$
Iterate: – Generate N_{sp} samples y_i^{sp} $b_i, i = 1, \cdots, N_{sp}$
- Coefficient $C = \mathcal{A} \times s \times v$
- embedding $b_i = g(\beta_i)$, $\beta_i = \frac{i}{M_{sp}}$

– Generate hole filling mask $h_i = h(y_i^{sp})$
– Update input with hole filling $y_{hole\ filling} = z(\texttt{bin}(y) \oplus h_i(y^{est}))$
– Estimate best fitting shape with hole filling mask: 1-D search for y^{est} that minimizes $E(b_i) = ||y_{hole\ filling} - h_i(C\psi(b_i))||$
– Compute outlier error $e_c(x) = ||y_{hole\ filling} - y^{est}(x)||$
– Estimate outlier $o_{outlier}(x) = e_c(x) \geq e_c^{TH_{outlier}}$
Update: – reduce $d_c^{TH_{hole}}$, $e_c^{TH_{outlier}}$
– increase N_{sp}

Based on the best matching shape, we compute outliers from the initial source after re-centering initial source.

4 Experimental Results

We evaluated our method using two gait-database. One is from CMU Mobo data set and the other is our own dataset with multiple view gait sequences. Robust outlier detection in spite of holes in the silhouette images was shown clearly in CMU database. We collected our own data set to show carrying object detection in continuous view variations.

4.1 Carrying Ball Detection from Multiple Views

The CMU Mobo database contains 25 subjects with 6 different views walking on the treadmill to study human locomotion as a biometric [6]. The database provides silhouette sequences extracted using one background image. Most of the sequences have holes in the background subtracted silhouette sequences. We collected $12(=4\times3)$ cycles to learn dynamic shape models with view and style variations from normal slow walking sequences of 4 subjects with 3 different views. For the training sequences, we corrected holes manually. Fig. 3 shows detected carrying objects in two different views from different people. The initial normalized shape has holes with a carrying ball (a)(e). Still the best fitting shape models recover correct body pose after iterative estimations of view and shape style with hole filling and outlier removal (b)(f). Fig. 3 (c)(g) show examples of generated masks during iteration for hole filling. Fig. 3 (d) (h) show detected outlier after iteration. In Fig. 3 (h), the outlier in bottom right corner comes from the inaccurate background subtraction outside the subject, which cannot be managed by hole filling. The verification routine based on temporal characteristics of the outlier similar to [1] can be used to exclude such a outlier from detected carrying objects.

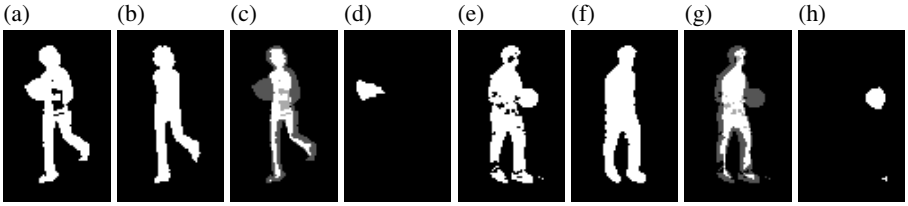

Fig. 3. Outlier detection in different view: (a) Initial normalized shape for outlier detection. (b) The best fitting model from the generative model. (c) Overlapping initial input and hole filling mask at the last iteration. (d) Detected outlier. (e) (f) (g) (h) : Another view in different person.

4.2 Carrying Object Detection with Continuous View Variations

We collected 4 people with 7 different views to learn the nonlinear decomposable dynamic shape model of normal walking for detection of carrying objects in continuous view variations. In order to achieve reasonable multiple view interpolation, we captured normal gait sequence on the treadmill with the same height camera position in our lab. The test sequence is captured separately in outdoor using commercial camcorder. Fig. 4 shows an example sequence of carrying object detection in continuous change of walking direction. The first row shows original input images from the camcorder. The second

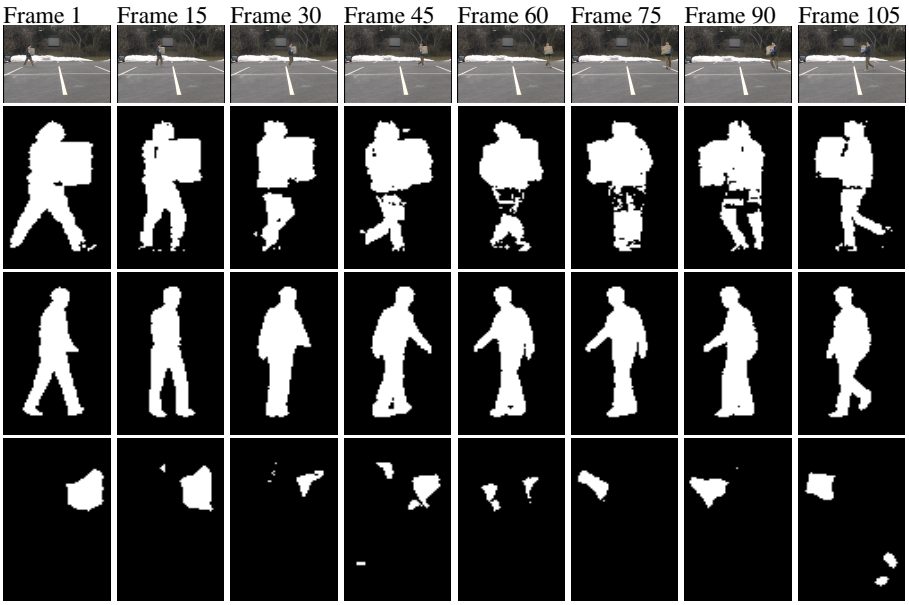

Fig. 4. Outlier detection in continuous view variations: First row: Input image. Second row: Extracted silhouette shape. Third row: Best matching shape. Fourth row: Detected carrying objects.

row shows normalized shape after background subtraction. We used the nonparametric kernel density estimation method for per-pixel background models, which is proposed in [3]. The third row shows best matching shape estimated after hole filling and outlier removal using dynamic shape models with multiple views. The fourth row shows detected outliers. Most of the dominant outliers come from the carrying objects.

5 Conclusions

We presented a new framework for carrying object detection from given silhouette images based on pose preserving dynamic shape model. The signed distance representation of shape helps robust matching in spite of small misalignment and hole. To enhance the accuracy of alignment and matching, we preformed hole filling and outlier detection iteratively with threshold control. Experimental results from CMU Mobo data set show accurate detection of outliers in multiple fixed views. We also showed the estimation of outliers in continuous view variations from our collected data set. The removal of outlier or carrying object will be useful for gait recognition as it helps recovering high quality original silhouette, which is important in gait recognition. We plan to apply the proposed method to test gait recognition with carrying objects.

Acknowledgement. This research is partially funded by NSF award IIS-0328991.

References

1. C. BenAbdelkader and L. S. Davis. Detection of people carrying objects: A motion-based recognition approach. In *Proc. of FGR*, pages 378–383, 2002.
2. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models: Their training and applications. *CVIU*, 61(1):38–59, 1995.
3. A. Elgammal, D. Harwood, and L. Davis. Background and foreground modeling using non-parametric kernel density estimation for visual surveillance. *IEEE Proceedings*, 90(7):1151–1163, 2002.
4. A. Elgammal and C.-S. Lee. Inferring 3d body pose from silhouettes using activity manifold learning. In *Proc. CVPR*, volume 2, pages 681–688, 2004.
5. A. Elgammal and C.-S. Lee. Separating style and content on a nonlinear manifold. In *Proc. CVPR*, volume 1, pages 478–485, 2004.
6. R. Gross and J. Shi. The cmu motion of body (mobo) database. Technical Report TR-01-18, Carnegie Mellon University, 2001.
7. I. Haritaoglu, R. Cutler, D. Harwood, and L. S. Davis. Packpack: Detection of people carrying objects using silhouettes. In *Proc. of ICCV*, pages 102–107, 1999.
8. M. Isard and A. Blake. Condensation–conditional density propagation for visual tracking. *Int.J.Computer Vision*, 29(1):5–28, 1998.
9. G. Kimeldorf and G. Wahba. Some results on tchebycheffian spline functions. *J. Math. Anal. Applic.*, 33:82–95, 1971.
10. S. Osher and N. Paragios. *Geometric Level Set Methods*. Springer, 2003.
11. M. Rousson and N. Paragios. Shape priors for level set representations. In *Proc. ECCV, LNCS 2351*, pages 78–92, 2002.
12. S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linar embedding. *Science*, 290(5500):2323–2326, 2000.
13. S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *IEEE Trans. PAMI*, 27(2):162–177, 2005.
14. B. Schlkopf and A. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press, 2002.
15. K. Toyama and A. Blake. Probabilistic tracking in a metric space. In *ICCV*, pages 50–59, 2001.
16. A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, and W. E. Grimson. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Trans. on Medical Imaging*, 22(2), 2003.
17. M. A. O. Vasilescu and D. Terzopoulos. Multilinear subspace analysis of image ensembles. In *Proc. of CVPR*, 2003.
18. Q. Wang, G. Xu, and H. Ai. Learning object intrinsic structure for robust visual tracking. In *CVPR*, volume 2, pages 227–233, 2003.