Shape Outlier Detection Using Pose Preserving Dynamic Shape Models

Chan-Su Lee CHANSU@CS.RUTGERS.EDU
Ahmed Elgammal ELGAMMAL@CS.RUTGERS.EDU
Department of Computer Science, Rutgers University, Piscataway, NJ 08854 USA

Abstract

In this paper, we introduce a framework for shape outlier, like carrying object, detection in different people from different views using pose preserving dynamic shape models. We model dynamic human shape deformations in different people using kinematics manifold embedding and decomposition of nonlinear mapping using 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

The shape deformation in human motion contains rich information such as body pose, person identity, and even emotional states of the person. It implies that human shape deformations vary in different body poses, shape styles, and emotional states. Different observation conditions like view and background cause further variations in the observed image or extracted shapes. The shape deformation of human motion like gait causes variations of topology and nonlinear deformations in observed shape sequences. When we have a good generative shape deformation models according to state parameters like body pose, shape style and view, we can solve many problems in human motion analysis, tracking and recognition.

This paper presents 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 in low dimensional space based on motion captured data and invariant to different people and view, is used to model dynamics of shape deformation according to intrinsic body configuration. The entire intrinsic configuration can have 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. By utilizing multilinear analysis for these mappings, 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 of given the 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 shape outlier, like carrying objects, from sequences of silhouette images. The detection of carrying objects is one of the key element in visual surveillance systems (Haritaoglu et al., 1999). In gait challenge problem, the performance of gait recognition decrease dramatically for probe set with briefcase (Sarkar et al., 2005). Our pose-preserving dynamic shape model detects carrying objects as outliers. By removing 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 (Sec. 3.4). Experimental results using CMU Mobo gait database (Gross & Shi, 2001) and our own data set from multiple views show accurate estimation of carrying object with correction of silhouettes from multiple people and multiple view silhouettes with holes (Sec. 4).

Appearing in Proceedings of the Workshop on Machine Learning Algorithms for Surveillance and Event Detection at the 23rd ICML, 2006.

1.1. Related Work

There have been a lot of work on contour tracking from cluttered environment such as active shape models (ASM) (Cootes et al., 1995), active contours (Isard & Blake, 1998), and exemplar-based tracking (Toyama & Blake, 2001). Spatiotemporal models are also used for contour tracking (Baumberg & Hogg, 1996). 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 our previous work (Elgammal & Lee, 2004b), where the motion is represented in a low dimensional nonlinear manifold. Nonlinear manifold learning technique can be used to find intrinsic body configuration space (Wang et al., 2003; Elgammal & Lee, 2004b). However, discovered manifolds are twisted differently according to person styles, views, and other factors like clothes in image sequences (Elgammal & Lee, 2004a). We propose kinematics manifold embedding as an alternative uniform representation of intrinsic body configuration (Sec. 2.1).

Shape models are used for segmentation and tracking of medical image using level sets (Tsai et al., 2003; Paragios, 2003; Leventon et al., 2000). The shape priors also used as constraints in geodesic active contours (Rousson & Paragios, 2002). These shape priors can be used for pose-preserving shape estimation. However, this model does not contain dynamic characteristics of shape deformations in human motion. This paper presents the generative dynamic shape model in multiple view and people using kinematics manifold embedding.

In spite of the importance of carrying objects or outliners 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. Detecting carrying object has been designed to work under a visual surveillance system (Haritaoglu et al., 1999). By analyzing symmetry in silhouette model, they detected carrying object by aperiodic outlier regions. The system is very sensitive to noise of foreground object detection, size of carrying object, and the axis of symmetry which is used to compute asymmetric of shape (BenAbdelkader & Davis, 2002). Amplitude of the shape feature and the location of detected objects are constrained to improve accuracy of carrying object detection (BenAbdelkader & Davis, 2002). Detecting outlier accurately and removing noise and filling hole in extracted silhouette still remains unresolved. This paper proposes gradual detection of outlier, and correction of noise silhouette by hole filling and shape 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 the person shape 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. For the generative model, we need low dimensional representation of body pose b_t invariant to the view and shape style. We need universal representation for body configuration invariant to the variation of observation in different people 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 3 dimensional spaces. We align global transformation in advance in order to model motion only due to body configuration change. Locally linear embedding (LLE) (Roweis & Saul, 2000) is applied to find low dimensional intrinsic representation from the high dimensional data (collection of joint location). The discovered manifold is one-dimensional twisted circular manifold in three-dimensional spaces.

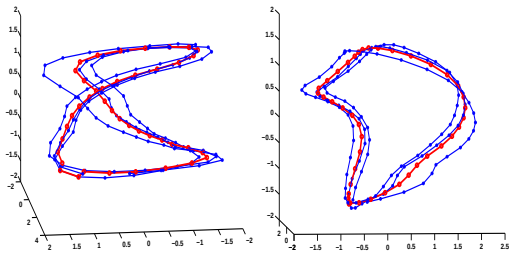

Figure 1. Kinematics manifold embedding and its mean manifold: two different views in 3D space

The manifold is represented using a one-dimensional parameter by spline fitting. In order to find intrinsic manifold representation using nonlinear dimensionality reduction, dense sampling from the manifold points is required. We use multiple cycles to find kinematics intrinsic mani-

fold representation by LLE. For one-dimensional representation of the multiple cycles, we use mean-manifold representation in the parameterizations. 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 kinematics manifold from motion capture data with three walking cycles and their mean manifold representation.

2.2. Decomposing and Modeling Shape Style Space

Individual variations of the shape deformation can be discovered in the nonlinear mapping space between the kinematics manifold embedding and the observation in different people. If we have pose-aligned shape for all the people, then it becomes relatively easy to find shape variations according to the shape style. 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 (Schlkopf & Smola, 2002) to capture nonlinear deformation in difference body pose. There are three steps to model individual shape deformations using nonlinear mapping. We focus on gait, walking sequence. But it can be applicable to other cyclic motion analysis in different people.

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 gait embedding in each detected cycle. As the kinematics manifold comes from constant speed walking motion captured data, we can embed the shape sequence in equally spaced points along the manifold. Second, we learn nonlinear mapping between the kinematics embedding space and shape sequences. According to the representer theorem (Kimeldorf & Wahba, 1971), we can find a mapping that minimizes the regularized risk in the following form for given patterns x_i and target values $y_i = f(x_i)$:

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

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 sequence with N_s shape styles and N_v view, we obtain $N_s \times N_v$ number of mapping coefficients. Third, multi-linear tensor analysis can be used to decompose the gait motion mapping into orthogonal factors. Tensor decomposition can be achieved by higher-order singular value decomposition (HOSVD) (Lathauwer et al., 2000)(Vasilescu & Terzopoulos, 2003), 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 mapping coefficient. 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 configuration parameter b_t, shape style parameter s, and view parameter 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 shapes. For given body pose parameter, we can reconstruct best fitting shape by estimating style 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 b_t , the view v, and the shape style s which minimize the reconstruction error

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

We assume that the estimated optimal style can be written as a linear combination of style vectors in the training data. 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 given each shape style 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 $\Sigma^v = T_v \sigma_v^2 I$ 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 shape. In order to achieve better alignment in normalized shape representation, we performed hole filling and outlier removal for the extracted shape.

3.1. Shape Representation

Background Subtraction: We captured gait sequence on treadmill with multiple camera. Nonparametric kernel density estimation methods (Elgammal et al., 2002) are applied for per-pixel background models assuming static camera and employing local model of intensity.

To learn shape deformations in normal walking without carrying object, we collected walking sequences on treadmill for five people with 11 different views around circle in the same camera height. Fig. 2 (a) shows an example of captured real image. We performed background subtraction after learning statistical model for the background. However, due to lighting around camera and cluttered indoor environments, some of the silhouettes are not good as shown in Fig. 2 (b). To achieve consistent shape representation in different people and hole filling for inaccurate background subtraction, we correct silhouettes used for training.

Normalization of Silhouette Shape: For consistent representation of shape deformation by 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 body 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 specific threshold value for consistent centering of shape in spite of small incorrect background block from noise and shadow. we perform normalization after morphological operation and filtering to remove noise spot and small holes.

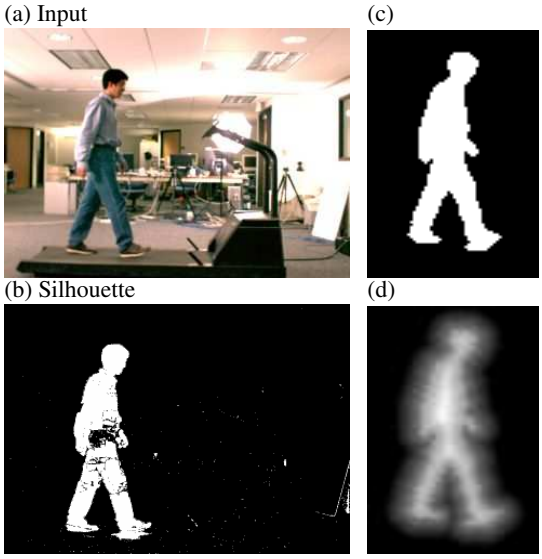

Figure 2. An example of background subtraction and silhouette representation: (a) Input image. (b) Background subtracted image. (c) Normalized silhouette. (d) Signed distance representation.

Shape Representation by Signed Distance Function: We parameterize shape contour using signed distance function with limitation of maximum distance value 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 is used, which is typically used in level-set methods (Osher & Paragios, 2003). We add a threshold value as well

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

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 cer-

tain distance does not contain meaningful shape information in similarity measurements. Such shape representation imposes smoothness on the distance between shapes and robustness to noise and outlier. In addition, by changing threshold values gradually, we can generate mask to represent inside of the shape for hole filling. Given such representation, input shape images are 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, d is the dimensionality of the input space, 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 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 causes inaccurate description of shape in signed distance function. Hence holes inside shape result in inaccurate estimation of the best fitting shape. It can also induce misalignment of shape as the hole can shift center of gravity for the horizontal axis alignment.

From the signed distance representation, we can generate a mask to represent inside of the shape for estimated style, view, and body pose parameters. We can use the mask to fill holes for the original shape. The mask can be generated by thresholds generated from signed distance shape representation like

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

where $d_c^{TH_{hole}} \geq 0$ is the threshold value for inner shape to create mask for hole filling. If the threshold value is zero, the mask will be the same as the silhouette image generated by dynamic shape model given style, view and configuration. As we don't know the exact shape style, view and configuration at the beginning, and the hole causes misalignment, we start from large threshold value, which generates a small mask of inner shape area and robust to misalignment. We reduce the threshold value as estimated model parameters get more accurate.

The hole filling operation can be described by $y_{hole\ filling} = z\left(\texttt{bin}(y) \oplus h(y^{est})\right)$, 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. 3 shows initial shape normalized silhouette with holes (a), the best estimated shape model (b) which is generated from the generative model with style and view estimation and configuration search, and the hole mask (c) when $d_c^{TH_{hole}} = 3$, and new shape after hole filling (d). We can improve the best matching shape by excluding mask area in the computation of similarity measurement for generated samples in searching the best fitting body pose. Re-alignment of shape and re-computation of shape representation after hole filling provide better shape description for next step.

3.3. Carrying Object Detection

Carrying objects are detected by estimating outlier from best matching normal dynamic shape and given input shape. The outlier of a shape silhouette with carrying objects is the mismatching part in input shape compared with 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 for background subtraction, shape misalignment also cause mismatches. For accurate detection of carrying object from outlier, we need to remove other source of outlier such as hole and misalignment in shape. Hole filling and outlier removal are performed iteratively to improve shape representation for better estimation of the matching shape.

We gradually reduce 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 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 thresholds the error distance, we can detect outlier.

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

At the beginning, we start from large $e_c^{TH_{outlier}}$ value and we reduce the value gradually. Whenever we detect outlier, we remove the detected outlier area and perform realignment to reduce misalignment due to the outlier. In Fig. 3, 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 best matching shape.

3.4. Iterative Estimation of Outlier with Hole Filling

Iterative estimations of outlier, hole filling, outlier removal, and estimation of shape style, view and configuration are performed with threshold value control. The threshold

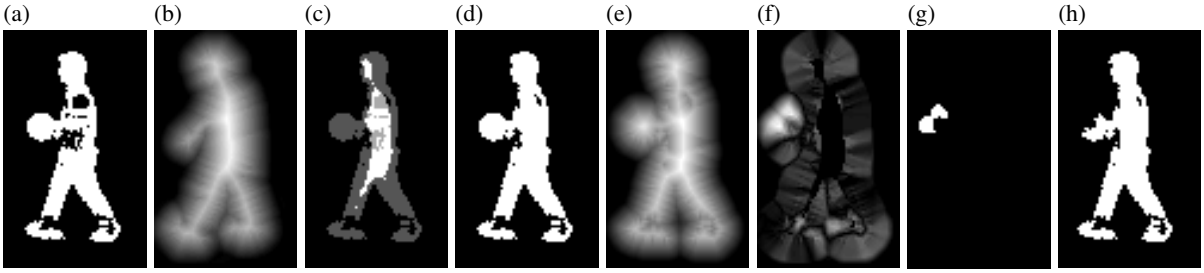

Figure 3. Hole filling using mask from best fitting model : (a) Initial normalized shape with hole. (b) Best matching shape from generative model. (c) Overlapping with initial silhouette and mask from best matching shape by threshold. (d) New shape with reduced hole. (e) Initial normalized shape for outlier detection with signed distance representation. (f) Euclidian distance between best matching model from the generative model and input shape with signed distance representation. (g) Detected outlier with threshold value $e(x) \geq 5$. (h) New shape after removing outlier.

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 estimation with increased number of samples. We summarize the iterative estimation as follows:

Input: image shape y_i, estimated view v, estimated style s, core tensor $\mathscr{A}$

Initialization:
- initialize sample num N_{sp}
- initialize $d_c^{TH_{hole}}$, $e_c^{TH_{outlier}}$

Iterate:
- Generate N_{sp} samples y_i^{sp} $b_i, i = 1, \cdots, N_{sp}$
  - Coefficient $C = \mathscr{A} \times s \times v$
  - 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 the outlier 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 data set for multiple view gait sequence. Robust outlier detection in spite of hole 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 (Gross & Shi, 2001). The database provides silhouette sequence extracted based on one background image. The background subtracted silhouettes in most of the sequences have holes. 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. 4 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 poses after iterative estimations of view and shape style with hole filling and outlier removal (b)(f). Fig. 4 (c)(g) show examples of generated masks during iteration for hole filling. Fig. 4 (d) (h) show detected outlier after iteration. In Fig. 4 (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 (BenAbdelkader & Davis, 2002) can be used to exclude such a outlier from detected carrying objects.

4.2. Carrying Object Detection with Continuous View Variations

We collected 4 people with 7 different views to learn the pose preserving shape model of normal walking for detection of carrying object in continuous view variations. In order to achieve reasonable multiple view interpolation,

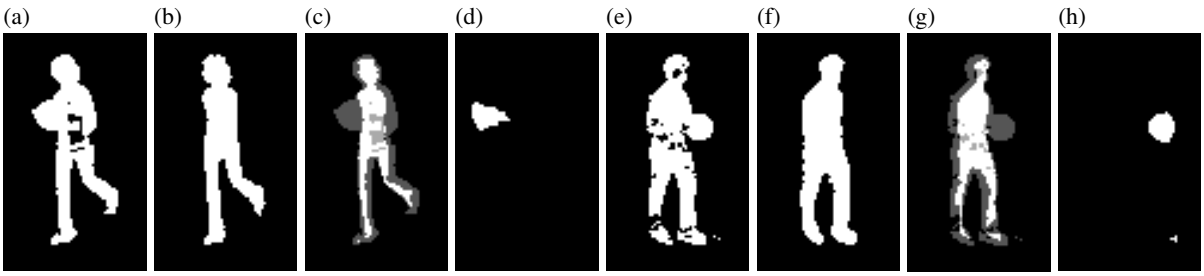

Figure 4. Outlier detection in different view: (a) Initial normalized shape for outlier detection. (b) The best fitting model from the generative model. (c) Overlapping initial input and hole filling mask at the last iteration. (d) Detected outlier. (e) (f) (g) (h) : Another view in different person

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. 5 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 row shows normalized shape after background subtraction. We used the nonparametric kernel density estimation method for per-pixel background models (Elgammal et al., 2002). 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 outlier. Most of the dominant outlier comes from the carrying object.

5. Conclusions

We presented shape outlier detection using pose preserving dynamic shape model especially for carrying objects detection for given silhouette images. The proposed model may be used for shadow detection and abnormal body pose detection, which is important human motion recognition and event detection. In the carrying object detection, the signed distance representation of shape helps robust matching in spite of small misalignment and holes. To enhance the accuracy of alignment and matching, we preformed hole filling and outlier detection iteratively with threshold change. In our experiment, we controlled threshold for gradual discriminative estimation of outlier and holes. More expert knowledge, such as temporal constraints in body pose transition, periodic characteristics in body limb can be used to enhance outlier detection and identification. Experimental results from CMU Mobo data set show accurate detection of outlier in multiple fixed views. We also showed the estimation of outlier 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 and abnormal motion detection.

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

References

Baumberg, A., & Hogg, D. (1996). Generating spatiotemporal models from examples. *Image and Vision Computing*, *14*, 525–532.

BenAbdelkader, C., & Davis, L. S. (2002). Detection of people carrying objects: A motion-based recognition approach. *Proc. of FGR* (pp. 378–383).

Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active shape models: Their training and applications. *CVIU*, *61*, 38–59.

Elgammal, A., Harwood, D., & Davis, L. (2002). Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. *IEEE Proceedings*, *90*, 1151–1163.

Elgammal, A., & Lee, C.-S. (2004a). Inferring 3d body pose from silhouettes using activity manifold learning. *Proc. CVPR* (pp. 681–688).

Elgammal, A., & Lee, C.-S. (2004b). Separating style and content on a nonlinear manifold. *Proc. CVPR* (pp. 478–485).

Gross, R., & Shi, J. (2001). *The cmu motion of body (mobo) database* (Technical Report TR-01-18). Carnegie Mellon University.

Haritaoglu, I., Cutler, R., Harwood, D., & Davis, L. S. (1999). Packpack: Detection of people carrying objects using silhouettes. *Proc. of ICCV* (pp. 102–107).

Isard, M., & Blake, A. (1998). Condensation–conditional density propagation for visual tracking. *Int.J.Computer Vision*, *29*, 5–28.

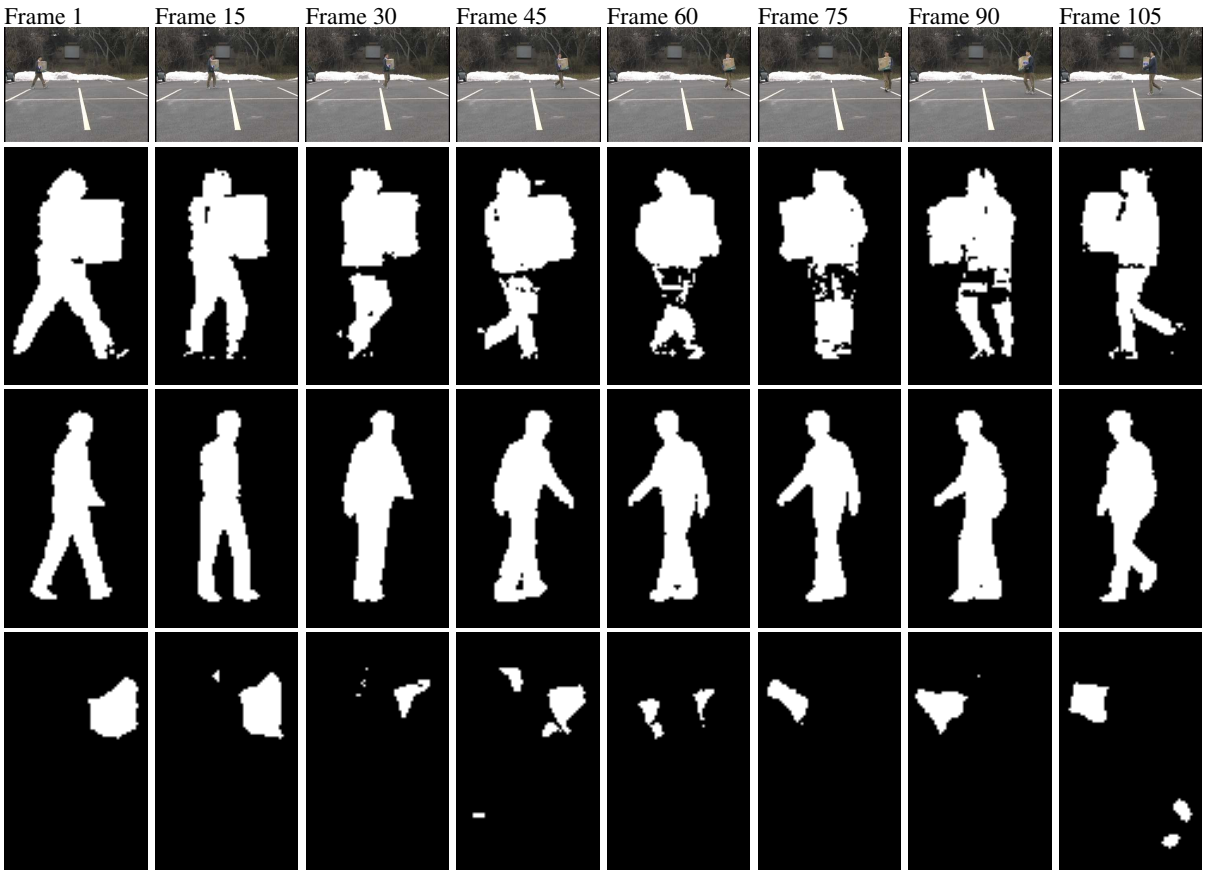

Figure 5. Outlier detection in continuous view variations: First row: Input image. Second row: Extracted silhouette shape. Third row: Best matching shape. Fourth row: Detected carrying object

Kimeldorf, G., & Wahba, G. (1971). Some results on tchebycheffian spline functions. *J. Math. Anal. Applic.*, *33*, 82–95.

Lathauwer, L. D., de Moor, B., & Vandewalle, J. (2000). A multilinear singular value decomposiiton. *SIAM Journal On Matrix Analysis and Applications*, *21*, 1253–1278.

Leventon, M. E., Grimson, W. E., & Faugeras, O. (2000). Statistical shape influence in geodesic active contours. *Proc. of CVPR* (pp. 1316–1323).

Osher, S., & Paragios, N. (2003). *Geometric level set methods*. Springer.

Paragios, N. (2003). A level set approach for shape-driven segmentation and tracking of the left ventticle. *IEEE Trans. on Medical Imaging*, *22*.

Rousson, M., & Paragios, N. (2002). Shape priors for level set representations. *Proc. ECCV, LNCS 2351* (pp. 78–92).

Roweis, S., & Saul, L. (2000). Nonlinear dimensionality reduction by locally linar embedding. *Science*, *290*, 2323–2326.

Sarkar, S., Phillips, P. J., Liu, Z., Vega, I. R., Grother, P., & Bowyer, K. W. (2005). The humanid gait challenge problem: Data sets, performance, and analysis. *IEEE Trans. PAMI*, *27*, 162–177.

Schlkopf, B., & Smola, A. (2002). *Learning with kernels: Support vector machines, regularization, optimization and beyond*. MIT Press.

Toyama, K., & Blake, A. (2001). Probabilistic tracking in a metric space. *ICCV* (pp. 50–59).

Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., Fan, A., & Grimson, W. E. (2003). A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Trans. on Medical Imaging*, *22*.

Vasilescu, M. A. O., & Terzopoulos, D. (2003). Multilinear subspace analysis of image ensembles. *Proc. of CVPR*.

Wang, Q., Xu, G., & Ai, H. (2003). Learning object intrinsic structure for robust visual tracking. *CVPR* (pp. 227–233).