The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

The Application of Intrinsic Variable Preserving Manifold Learning Method to Tracking Multiple People with Occlusion Reasoning

Suiwu Zheng, Hong Qiao *, Bo Zhang and Peng Zhang

Abstract— **Tracking multiple people in crowded and cluttered dynamic scenes is a very difficult task in robotic vision due to the highly frequent occlusion and lack of visibility of objects. In this paper, we present a manifold learning based multiple people tracking approach with occlusion reasoning to solve this problem. In our previous work, a new Intrinsic Variable Preserving Manifold Learning (IVPML) method is proposed, by which the continuity of the intrinsic motion variables for tracking is preserved on a new manifold after dimensionality reduction. In this paper, the IVPML method is extended to be applied to tracking multiple people with occlusion situations. Associated with spatio-temporal continuity of tracking and IVPML method, a novel robust occlusion reasoning method is proposed during the alternations of multiple people. For occlusion recovery, region covariance representation including both spatial and statistic properties of objects are used to detect people after occlusion. The multiple people tracking method has been successfully applied to mobile robotic visual tracking system in several complicated environments. Comparisons and experimental results have shown the effectiveness of the new algorithm in various situations.**

I. INTRODUCTION

Tracking multiple people accurately in dynamic and cluttered environment is a challenging task in robotic vision due to occlusion situation and the changing of illumination. During occlusion, only part of the object is visible, and the foreground blob no longer belongs to a single person. Even worse, a person might be completely occluded by other people and there is no feature available for tracking. Also, during dynamic tracking, the illumination condition may change and result in the change of objects' appearance. In this paper, we focus on these problems to build a robust multiple people tracking algorithm using a single camera.

In literature, there exist many monocular tracking approaches for multiple targets with occlusion in dynamic environments. For example, online sampling and position estimation method is proposed to track multiple people with occlusion. It is assumed that the color and texture of local parts are always similar to the center of objects in sequential frames [1]. In [2], a dynamic layer representation, in which spatial and temporal constraints on shape, motion, and layer appearance are modeled and estimated through a maximum a posteriori (MAP) framework, is proposed to track multiple vehicles undergoing various rigid motions and complex interactions. However, the accurate dynamic layer representation estimation is time-consuming. Meanshift with multiple cues fusion method is used to track multiple people [3]. In this method, multiple cues (e.g., an occlusion grid, the speed, cloth color and the motion trajectory of a person's face in each frame) are combined to detect occlusion and recover from occlusion. However, this algorithm is sensitive to the illumination and color of cloth. In [4], the multiple objects' full contours are tracked and shape priors consisting of shape level sets are used to recover the missing object regions during occlusion. Nevertheless, the contour tracking methods are sensitive to the cluttered background.

Most existent methods mainly use physical and their combined features for people tracking. These physical features may not be general features during the whole tracking process, and they should be adjusted and updated. Also, they are sensitive to external factors, for instance, illumination, occlusion and so on. However, the motion of a human in successive frames is continuous and can be parameterized by only a few intrinsic variables. Hence the appearance of human in tracking can be viewed as a manifold which is generated by theses intrinsic variables. In our previous work, a new Intrinsic Variable Preserving Manifold Learning (IVPML) method is proposed for visual tracking, by which the continuity of the intrinsic motion variables as an intrinsic feature for visual tracking is preserved on a new manifold after dimensionality reduction [5] [6]. In this paper, the IVPML method is extended to be applied to tracking multiple people with occlusion.

Manifold learning methods for dimensionality reduction have been a hot research topic since it was first proposed in *Science*, 2000 [7] [8] [9]. The goal of manifold learning methods is to extract intrinsic degrees of freedom from high-dimensional inputs which lie on or close to a low-dimensional manifold. These methods try to preserve local or global geometric characteristics of data manifolds from the original inputs into their low-dimensional representations.

According to the geometric characteristics which are kept, the recent manifold learning methods can be categorized as local and global approaches. Within local manifold learning approaches, locally linear embedding (LLE) [8] method assumes that each data point can be linearly reconstructed from their neighbors, and preserves local reconstruction weights.

This work was supported in part by the Chinese Academy of Science Selective Support of 100 Outstanding Scholarship, in part by the NSFC (National Natural Science Foundation of China)(nos. 60675039, 60505003, 60621001, 60725310 and 90820007), in part by the MOST (The Ministry of Science and Technology of the People's Republic of China) 2007AA04Z228.

Suiwu Zheng and Hong Qiao are with Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China. Suiwu.zheng@ia.ac.cn, Hong.qiao@ia.c.cn

Bo Zhang and Peng Zhang are with the Institute of Applied Mathematics, AMSS, Chinese Academy of Sciences, Beijing 100190, China. b.zhang@amt.ac.cn, zhangpeng@amss.ac.cn

*Corresponding author. Tel.: +86-10-6257-3736. E-mail: hong.qiao@ia.ac.cn

Laplacian Eigenmap (LE) methods [10] preserve the local neighborhood relationship. Locality preserving projections (LPP) [11] is a linearized version of the LE method by assuming that there exists a linear relationship between the high-dimensional inputs and the low-dimensional representations. However, LPP is different from the principal component analysis (PCA) or linear discriminant analysis (LDA) since LPP keeps the local geometric characteristics of a data manifold. Within global manifold learning approaches, the isometric feature mapping (ISOMAP) algorithm [9] preserves the pairwise geodesic distances of high-dimensional data points to their low-dimensional representations. In [12], a unified framework has been proposed to cast most existing manifold learning methods into a common graph model.

In this paper, the IVPML method is extended to multiple people tracking with occlusion reasoning. Different from preserving the local Euclidean distance in the existent manifold learning methods, the new manifold preserves the continuity among intrinsic variables (maybe rotation, scaling and illumination in tracking). As far as we know, no other work has considered tracking multiple noncooperative persons with occlusion by using the space which best keeps the continuity of intrinsic variables and the space is learned through manifold learning. In the proposed method, associated with the spatio-temporal continuity estimation method, people occluded by other people or obstacles is robustly detected. For occlusion recovery, region covariance representation including both spatial and statistic properties of objects and exclusion searching principle are used to detect people after occlusion. The multiple people tracking method has been successfully applied to mobile robotic visual tracking system in complicated environments. Experiments have shown the effectiveness of the new algorithm in various situations.

II. Intrinsic Variable Preserving Manifold Learning for Visual Tracking

In this section, the details of the continuity of the intrinsic variables preserving manifold learning method will be briefly reviewed [5] [6].

A. Building manifolds based on the continuity of intrinsic variables

$\mathcal{X}$ is a given set of high-dimensional data points, which lie on or near to some underlying data manifold $\mathcal{M} = f(\mathcal{U})$, where U denotes the set of intrinsic variables vector and $u^1, u^2, \cdots, u^m$ are various intrinsic variables. Suppose two inputs x_i and x_j from $\mathcal{X}$ are generated as $x_i = f(u_i^1, u_i^2, \cdots, u_i^m)$ and $x_j = f(u_j^1, u_j^2, \cdots, u_j^m)$. If $\{x_i\}_{i=1}^N$ are treated as nodes in a graph, then on the new manifold, adjacency between x_i and x_j is defined as follows.

Definition 2.1: x_i and x_j are **adjacent** if and only if $\left|u_i^k - u_j^k\right| < \varepsilon$, while the other variables $u_i^l = u_j^l, l \neq k$ for any $k \in \{1, 2, ..., m\}$. Here ε is a constant.

Once x_i and x_j become adjacent to each other, they are connected with an edge. Then a connected graph G is built up to form the skeleton of intrinsic manifold structure. The connectivity between x_i and x_j is defined as follows.

Definition 2.2: x_i and x_j are **connected** if there exists a connected path in G between them.

If x_i and x_j are adjacent, the edge length is set to be $\left|u_i^k - u_j^k\right|$ if only the k-th component of u changes. The value of the edge length can also be set manually according to the applications if the value of u is unknown. In general, the variable distance d_{ij} between x_i and x_j is given by

$$d_{ij} = \begin{cases} \text{shortest path length between } x_i \text{ and } x_j & \text{if } x_i \text{ and } x_j \text{ are connected} \\ \infty & \text{otherwise,} \end{cases} \tag{1}$$

where path length equals to the sum of edge lengths along connected path.

After adjacency relationship is determined, the similarities S_{ij} between two inputs can be estimated as

$$S_{ij} = \begin{cases} \exp\{-d_{ij}\} & \text{if } d_{ij} \neq 0 \\ 0 & \text{if } d_{ij} = \infty. \end{cases} \tag{2}$$

Then variation in intrinsic variables can be reflected in the change of similarities.

B. Intrinsic variable preserving manifold learning with LPP

The next step is to compute a set of low-dimensional representations $\mathcal{Y} = \{y_1, y_2, \ldots, y_N\} \subset \mathbb{R}^m$, which can best preserve the continuity of intrinsic variables, i.e., the similarities defined by equation (2). In this paper, we use LPP [11] to achieve this goal.

In LPP, it is assumed that there exists an $n \times m$ projection matrix U such that for any x_i, $i = 1, 2, \ldots, N$, its low-dimensional representation y_i satisfies $y_i = U^T x_i$. Then pairwise similarities are best preserved by solving the following optimization problem

$$\min_{\{y_i\}} \sum_{ij} \|y_i - y_j\|^2 S_{ij}. \tag{3}$$

Let X and Y be corresponding data matrices whose columns are data vectors. Substitute the projection assumption into equation (3) and add a non-degenerate constraint, equation (3) is transformed into

$$\begin{aligned} \min \quad & Tr(U^T X L X^T U) \\ s.t. \quad & U^T X D X^T U = I_m, \end{aligned} \tag{4}$$

where $L = D - S$ with $S = (S_{ij})$ and D is a diagonal matrix with $D_{ii} = \sum_j S_{ij}, i = 1, 2, \ldots, N$.

The columns of U are the eigenvectors of the following eigenvalue problem corresponding to the 2nd to the $(m+1)$-th smallest eigenvalues:

$$\begin{gathered} X L X^T \hat{u}_i = \lambda X D X^T \hat{u}_i \\ \hat{u}_i^T X D X^T \hat{u}_j = \delta_{ij}, \quad i, j = 1, 2, \ldots, m. \end{gathered} \tag{5}$$

Here $\hat{u}_i$ is the i-th column of U.

After the projection matrix U is computed, the low-dimensional representation y_{new} of a new coming sample x_{new} can be easily obtained as follows:

$$y_{new} = U^T x_{new}. \tag{6}$$

Combining the new manifold structure and manifold learning via LPP together, we form the proposed manifold learning method as Intrinsic Variable Preserving Manifold Learning (IVPML) method.

III. Multiple People Tracking Algorithm with Occlusion Reasoning

In this section, based on the IVPML method, a novel multiple people tracking algorithm with occlusion reasoning is proposed. For multiple people tracking, it is assumed that the motion of each person is smooth with no abrupt changes. In the training process, the continuity of human's vertical and horizontal rotation are used to build the new manifold. Therefore, the distance between object in current frame and previous frames in the low-dimensional space, which represents the continuity of object's motion, is used intuitively and novelly as an important cue for the human tracking and occlusion judgement. The algorithm is introduced in detail as below.

A. Overview of the multiple people tracking algorithm

The diagram of multiple people tracking algorithm with occlusion reasoning is shown in Fig. 1. The tracking algorithm can be divided into three parts: training process, tracking process and occlusion handling process. In training process, the low-dimensional space and the mapping relationship between high-dimensional space and low-dimensional space are learned off-line. In tracking process, people in complicated scene will be detected and tracked in consecutive frames using the learned mapping relationship between high-dimensional image space and learned low-dimensional space. In occlusion handling process, with the spatial-temporal continuity, occlusion will be detected when people is occluded by other people or other obstacles. When occlusion ends, recovery mechanism will detect the object using region covariance feature with exclusion searching principle.

B. Training process

The first step is the collection of training samples. The intrinsic variables of a human head during tracking are considered to be horizontal rotation r_h and vertical rotation r_v, where $r_h \in [0, 2\pi]$ and $r_v \in [-r_{v_0}, r_{v_0}]$. Here, $u = (r_h, r_v)$ is used to denote the intrinsic variables. Then head images with continuously altering u along r_h and r_v are collected. The head images evenly distribute along the r_h axis and the r_v axis.

After collecting head images, the manifold based on the intrinsic variables $u = (r_h, r_v)$ is built and dimensionality reduction is achieved using the IVPML algorithm presented in Section II.

The second step is to construct the adjacency relationship among input samples. Since $r_h \in [0, 2\pi]$ and $r_v \in [-r_{v_0}, r_{v_0}]$, the human head images distribute on a manifold which looks like a cylinder surface. r_h is ordered as $r_{h1}, r_{h2}, ..., r_{hH}$ according to their values from 0 to 2π and r_v is ordered as $r_{v1}, r_{v2}, ..., r_{vV}$ according to their values from $-r_{v_0}$ to r_{v_0}. In this paper, $H = 120$ and $V = 8$.

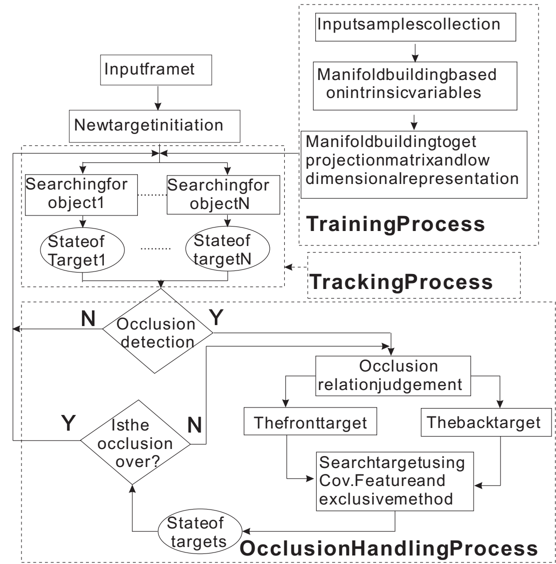


Fig. 1. The flowchart of multiple people tracking algorithm with occlusion reasoning.

It is clear that the head images $I(r_{hi}, r_{vj})$ $, when\ i = 1, 2, \ldots, H,\ j = 1, V$ have three adjacent neighbors, and the others have four adjacent neighbors. Then a connected graph is built and pairwise variable distances $\{d_{ij}\}$ defined in (1) are estimated. Similarities $\{S_{ij}\}$ are obtained along with the computation of $\{d_{ij}\}$.

In the third step, the similarity matrix $S = (S_{ij})$ is used to train the mapping relationship U as described in Section II, and the low-dimensional representations are also obtained.

C. Tracking Process

The key problem in tracking is to determine where the object should be in the next frame according to the current state of the object and the new frame.

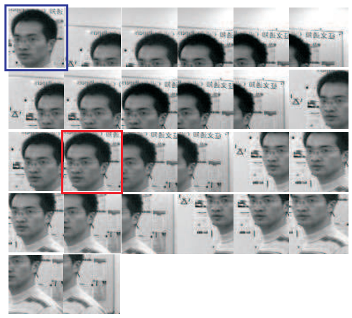
Assume that $(o^t_{i,1}, o^t_{i,2})$ are the coordinates of the center of the detection window for the ith human's head in Frame t. Then in Frame $t+1$, positions of its candidate windows are acquired by sampling around the location $(o^t_{i,1}, o^t_{i,2})$ with equal intervals.

Assume that x^t_i is the detected ith head image from Frame t and that $x^{t+1}_{i,c}$ is a candidate of the ith head image from Frame $t+1$. Then their low-dimensional representations y^t_i and $y^{t+1}_{i,c}$ are computed as follows:

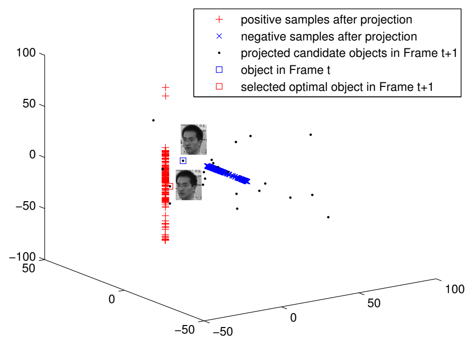
$$\begin{aligned} y^t_i &= U^T x^t_i \\ y^{t+1}_{i,c} &= U^T x^{t+1}_{i,c}. \end{aligned} \tag{7}$$

The proposed tracking algorithm is described as follows.

1) Initialize the location and size of multiple people's heads using multi-face detection algorithm in Frame $t = 1$. The vector representation of the ith person in image space is x^1_i. Then its low-dimensional projection y^1_i is obtained through (7).
2) For $t = 2, 3, ..., N$, candidate images $x^t_{i,c}$ are sampled around the location in Frame $t-1$. Let $y^t_{i,c}$ be their representations in projected $\mathcal{R}^3$ space and Y^t_i be the



(a) Object and candidates



(b) 3D view of tracking result

Fig. 2. A tracking example of single people. The first image in (a) shows the object in Frame t, while the 14th image indicates the optimal candidate in Frame $t+1$. In (b), object in Frame $t+1$ and all its candidates are projected to the trained feature space. The projection of object in Frame t is marked with blue square. The projection of the optimal candidate which is selected by our searching principle is marked with red square.

set containing all $y_{i,c}^t$. Then with their low-dimensional representations, the optimal window of the ith person is selected according to the following optimal searching principle (see (8)).

$$y_i^t = \arg\min_{y_{i,c}^t \in Y_i^t} \{\varepsilon_{i,c}^t = \sum \|y_{i,c}^t - Y_i^{t-1}\| + \|y_{i,c}^t - y_i^{t-1}\|\}, \tag{8}$$

where Y_i^{t-1} includes the k nearest neighbors of y_i^{t-1} in $\mathcal{R}^3$, and the $\varepsilon_{i,c}^t$ is the distance between the cth candidate in tth frame and the target in $t-1$th frame in the low dimensional space. Then the optimal ith object in Frame t can be obtained based on the location of y_i^t. An illustrative example of the searching result is shown in Fig. 2.

D. Occlusion handling process

For multiple people tracking in cluttered scenes, the occlusion situation often occurs. Therefore, it is very important to build a robust occlusion reasoning method. In this part, a novel occlusion handling mechanism is proposed.

1) occlusion reasoning: In our tracking algorithm, it is assumed that the target is spatio-temporal continuous in consecutive image frames. This assumption is usually addressed in most of the tracking algorithms. The spatio-temporal continuity is: 1) an object cannot be in more than one place at the same time; 2) the motion of object must be continuous (i.e. it cannot instantaneously 'jump' from one place to another). On our new manifold, we aim to preserve the continuity of intrinsic variables in visual tracking (human head's horizontal and vertical rotations). Since the intrinsic variables of human motion are temporal continuous, the continuity of human head's rotation as well as the temporal continuity is preserved in the low-dimensional space. Therefore, the continuity of intrinsic variables is equal to the temporal continuity during tracking process. To utilize the spatio-temporal continuity in occlusion reasoning, multiple people's position relationship and the continuity of intrinsic variables are used to judge and predict the occlusion situation. The mechanism of the occlusion reasoning mechanism can be seen in Fig. 3 and is described in detail as below.

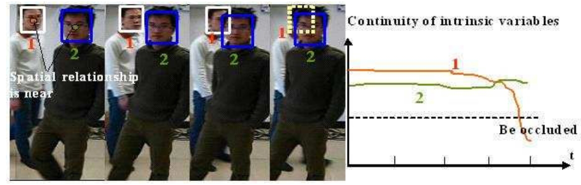


Fig. 3. The occlusion reasoning mechanism.

For convenience, some notations are explained here: x_i^t is the ith target in the tth frame in the ambient space; y_i^t denotes the ith target in the tth frame in the low-dimensional space; $Dx_{i,j}^t$ presents the distance between the ith and jth targets in the jth frame in the ambient space; Dy_i^t is used to define the distance between the ith target in the tth frame and the ith target in previous three frames in the low-dimensional space. The above two kinds of distances in different space are defined as follows.

$$Dx_{i,j}^t = \|x_i^t - x_j^t\|_{L^2}$$
$$Dy_i^t = \frac{1}{3}\sum_{k=1}^{3} \|y_i^{t-k} - y_i^t\|_{L^2}. \tag{9}$$

The L^2 norm of an m-tuple vector $u = (u^1, u^2, ..., u^m) \in \mathbb{R}^m$ is computed as $\|u\|_{L^2} = (\sum_{i=1}^m (u^i)^2)^{1/2}$.

Step 1: the judgement of spatial continuity. During the multiple people's tracking process, the spatial relationships among multiple people are calculated through $Dx_{i,j}^t$. If

$$Dx_{i,j}^t < Tx_{nei}, \tag{10}$$

the ith and jth people's locations are close to each other, and they may be occluded by each other. Where Tx_{nei} is the maximal distance threshold of neighborhood in high-dimensional image space. The value of Tx_{nei} is related to the size of the target. In our algorithm, Tx_{nei} is set to be an Gaussian random variable with mean to be 30 and variance to be 4.

If the spatial relationship satisfies the occlusion condition, the occlusion reasoning mechanism will then judge whether the two people interact with each other and which people is being occluded.

Step 2: the estimation of the temporal continuity. The distance Dy_i^t in low-dimensional space represents the change of the intrinsic variables. As mentioned above, the intrinsic variables are continuous along the time axis. If

$$Dy_i^t - Dy_i^{t-1} > Ty_{intr}, \tag{11}$$

the ith human in the scene is occluded by its neighbor human, where Ty_{intr} is the maximum distance threshold of the continuity of intrinsic variable. The value of Ty_{intr} is related to the speed of the motion. In our algorithm, Ty_{intr} is set to be an Gaussian random variable with mean to be 10 and variance to be 2.

2) occlusion recovery: An accurate object detector is needed when the full occlusion ends. In this paper, region covariance representation [13]–[15] which embodies both spatial and statistical properties of objects is used to describe the target, and the exclusion searching method is used to search the target after occlusion.

Region covariance is a newly proposed image feature [13] and has been successfully applied to object detection [13], visual tracking [14] and pedestrian detection [15]. In object detection and tracking, region covariance matrix is used to calculate deviation degree of two or more variables such as color, gradient, coordinate, and filter responses of one pixel. The computation of the region covariance is independent of the size of the image region and this representation has much lower dimensionality than histograms. Furthermore, it is robust against noise and light changes.

For a $W \times H$ region R from an intensity image I, we define a 7-dimensional feature mapping ϕ as

$$\phi(x,y) = \Bigg[x \quad y \quad I(x,y) \quad \left|\frac{\partial I(x,y)}{\partial x}\right| \quad \left|\frac{\partial I(x,y)}{\partial y}\right| \quad \left|\frac{\partial^2 I(x,y)}{\partial x^2}\right| \quad \left|\frac{\partial^2 I(x,y)}{\partial y^2}\right|\Bigg]^T, \tag{12}$$

where $I(x,y)$ is the intensity of pixel(x,y).

Then the region covariance of R is calculated as

$$C_R = \frac{1}{WH-1}\sum_{i=1}^{W}\sum_{j=1}^{H}(\phi(i,j)-\mu)(\phi(i,j)-\mu)^T, \tag{13}$$

where μ is the mean vector of $\phi(x,y)$ in region R.

We use the distance measure proposed in [13] to compute the dissimilarity between two regions based on their region covariance matrices, which is given as follows:

$$\rho(C_1,C_2) = \sqrt{\sum_{i=1}^{7}\ln^2\lambda_i(C_1,C_2)}, \tag{14}$$

where $\lambda_i(C_1,C_2), i = 1,2,\ldots,7$ are the generalized eigenvalues of C_1 and C_2 computed from $\lambda_i C_1 v = C_2 v, i = 1,2,\ldots,7$.

Finally, the dissimilarity between the candidate region and target region is computed as

$$\rho(c,t) = \min_{j\in\{1,2,\ldots,5\}}\left[\sum_{i=1}^{5}\rho(C_i^c,C_i^t) - \rho(C_j^c,C_j^t)\right], \tag{15}$$

where C_i^c and C_i^t stand for the region covariances of the target and candidate respectively and the indices $\{1,2,\ldots,5\}$ represent the whole, the left half, the right half, the top half and the bottom half of the region.

To improve the speed and reduce the false-positive rate during the detection of occluded people, the occlusion recovery mechanism uses an exclusion searching principle to reduce the searching region. For example, the region of the target which occludes the other person can be excluded. Also using the property of spatial continuity, the searching area focuses on the neighborhood of the occluding target. By using this method, the detection is real-time during occlusion and has low false detecting rate.

IV. Experiments and Analysis

To test the performance of the proposed algorithm, first the algorithm is compared with some other popular methods in different conditions. Then the IPVML tracking algorithm is embedded into a mobile robot to track multiple people under complicated environment. In the training step, 960 training samples of single person's head are used. The algorithm is implemented using C/C++ on the Windows XP system of robotic platform with P4-3.2G CPU and 1G RAM. The resolution of the image used for tracking is 320×240. The average speed of the algorithm is about 12f/s under various conditions, and it meets the real-time requirement.

A. Comparison with other tracking methods

Incremental Learning Principal Component Analysis (ILPCA) tracking method [16] learns a low-dimensional subspace representation online to adapt the changes in the appearance of the target. It is a popular method to track appearance changing object. Meanshift based multiple cues fusion tracking method [3] is used to track multiple people with full occlusion in many applications. In this part, our IVPML tracking method is compared with these two popular methods under different situations. In Fig. 4, the results show that our proposed tracking algorithm is capable to track a single person under severe occlusion, but the ILPCA tracking method failed in this situation. In Fig. 5, the three algorithm are used to track two people with alternations in the corridor. The results show that our new tracking algorithm is robust to detect the occlusion and recover from full occlusion. ILPCA method lost the target when full occlusion appears because it can not deal with total occlusion situation. The meanshift based multiple cues fusion tracking method is also failed because it is sensitive to the illumination change in the environment. In Fig. 6, it is also validated that the proposed method outperforms meanshift based multiple cues fusion method when the illumination and background change greatly.

B. IVPML based multiple people tracking experiments in complicated environment

In this part, the proposed tracking method is embedded into a mobile robot to track multiple people in the hall of a building with much more complicated illumination

Fig. 4. Comparison results with ILPCA tracking method (in the 2nd row) under severe partial occlusion.

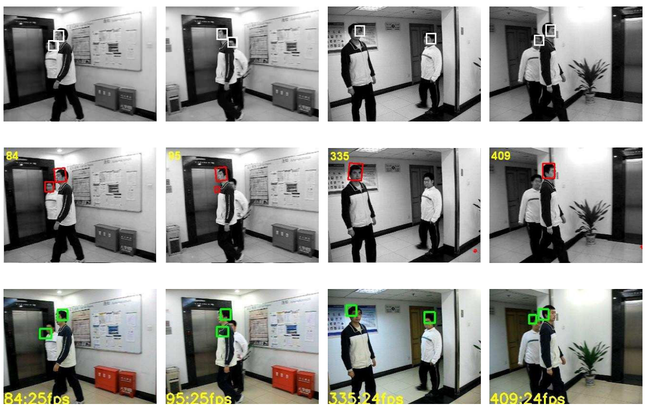


Fig. 5. Comparison results with ILPCA (in the 2nd row) and Meanshift based multiple cues fusion (in the 3rd row) tracking methods on our mobile robotic system in the corridor.

and background conditions. In Fig. 7, two persons were tracked in the hall of a building. The main difficulties in this environment are the obvious change of illumination in different orientations and the similar color with human's head in the background. The tracking results show that the new multiple people tracking algorithm has successfully captured the head during the whole process despite the alternations of people.

V. CONCLUSION

In this paper, a novel and practical multiple people tracking framework based on Intrinsic Variable Preserving Manifold Learning (IVPML) has been proposed. The new algorithm is mainly based on the distance in low dimensional space which preserves the continuity of motion intrinsic variables. Different conditions in visual tracking have been addressed, including full rotation and full occlusion of the object. Experimental results have validated the effectiveness of the proposed algorithm.

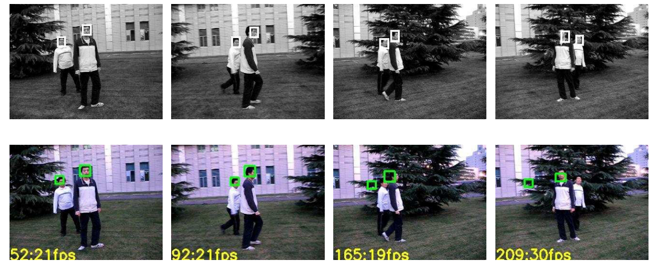


Fig. 6. Comparison results with Meanshift based multiple cues fusion method in the outdoor environment.

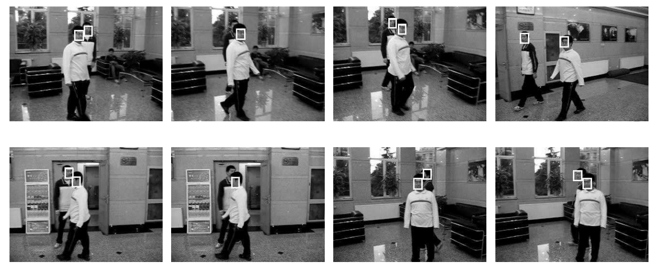


Fig. 7. Results of multiple people tracking with alternations in the hall.

REFERENCES

[1] L. Zhu, J. Zhou , J.Y. Song, "Tracking multiple objects through occlusion with online sampling and position estimation", *Pattern Recognition*, vol. 41, No. 8, pp. 2447-2460, Aug. 2008.
[2] T. Hai, H.S. Sawhney, R. Kumar, "Object Tracking with Bayesian Estimation of Dynamic Layer Representations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.24, No.1, pp75-89, Jan. 2002.
[3] C. Lerdsudwichai, M. Abdel-Mottaleb and A.-N. Ansari, "Tracking multiple people with recovery from partial and total occlusion", *Pattern Recognition*, vol. 38, no. 7, pp. 1059-1070, Jul. 2005.
[4] A. Yilmaz, X. Li, M. Shah, "Contour-Based Object Tracking with Occlusion Handling in Video Acquired Using Mobile Cameras", *IEEE Transactions on Pattern Analysis and Machine Intelligence* vol. 26, no. 11, pp. 1531-1536, Nov. 2004.
[5] H. Qiao, P. Zhang, B. Zhang, S. Zheng, "Tracking Feature Extraction Based on Manifold Learning Framework", accepted by *Journal of Experimental and Theoretical Artificial Intelligence*, 2009.
[6] H. Qiao, P. Zhang, B. Zhang, S. Zheng, "Learning An Intrinsic Variable Preserving Manifold for Dynamic Visual Tracking", *submitted to IEEE Transaction on Systems, Man and Cybernetics - Part B*, 2009.
[7] H.S. Seung and D.D. Lee, "The manifold ways of perception," *Science*, vol. 290, no. 5500, pp. 2268-2269, Dec. 2000.
[8] S.T. Roweis and L.K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323-2326, Dec. 2000.
[9] J.B. Tenenbaum, V. de Silva and J.C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319-2323, Dec. 2000.
[10] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*. vol. 15, no. 6, pp. 1373-1396, Jun. 2003.
[11] X.F. He, S.C. Yan, Y.X. Hu, P. Niyogi and H.J. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328-340, Mar. 2005.
[12] S.C. Yan, D. Xu, B.Y. Zhang, H.J. Zhang, Q. Yang and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40-51, Jan. 2007.
[13] O. Tuzel, F. Porikli, and P. Meer, "Region Covariance: A Fast Descriptor for Detection and Classification," *Proc. Ninth European Conf. Computer Vision (ECCV 06)*, vol. 2, pp. 589-600, 2006.
[14] F. Porikli, O. Tuzel, and P. Meer, "Covariance Tracking Using Model Update Based on Lie Algebra," Proc. *IEEE Conf. Computer Vision and Pattern Recognition (CVPR 06)*, vol. 1, pp. 728-735, 2006.
[15] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian Detection via Classification on Riemannian Manifolds," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1713-1727, 2008.
[16] D.A.Ross, J.Lim-M.-H.Yang, J.Lim and R.-S.Lin, "Incremental Learning for Robust Visual Tracking," *Internatiolal Journal of Computer Vision*