

A Supervised Manifold Learning Method

Zuojin Li¹, Weiren Shi¹, Xin Shi¹ and Zhi Zhong²

¹ College of Automation, Chongqing University,
No.174, Shazheng Street, Shapingba District, Chongqing, China
cqulzj@gmail.com

² The Smartech Institute, Room 1808, Anhui Building,
No 6007, Shennan Road, Futian District, Shenzhen, China
zzhong@mae.cuhk.edu.hk

Abstract. The Locally Linear Embedding (LLE) algorithm is an unsupervised nonlinear dimensionality-reduction method, which reports a low recognition rate in classification because it gives no consideration to the label information of sample distribution. In this paper, a classification method of supervised LLE (SLLE) based on Linear Discriminant Analysis (LDA) is proposed. First, samples are classified according to their label values, and low dimensional features of intra-class data are expressed through LLE manifold learning. Then, the base vectors in Fisher subspace of the low dimensional features are generated through LDA learning. This method increases inter-class variation, and decreases the intra-class variation when samples are projected to the Fisher subspace. Hence, the samples of different labels can be recognized, and the recognition rate and robustness of the LLE learning are improved. Experiments on handwritten digit recognition show that the proposed method is featuring high recognition rate.

Keywords: Manifold learning, Locally Linear Embedding, fisher subspace, manifold perception.

1. Introduction

To learn from the mechanism people perceive the world and improve machine intelligence has long been a great challenge facing artificial intelligence. Literature in 1985 [1] showed that the key for human beings to recognize objects is to uncover the inherent law of high-dimensionality data and form our own cognition. In other words, low-dimensionality manifolds of objects formed in human brains enable us to quickly recognize the objects. Another article [2] in Science, 2000, by H.S. Seung, et al. discussed the manifold learning based on cognition and put forward the idea that human beings have the ability to instantaneously perceive the inherent low-dimensionality structure of objects and can thus recognize the corresponding objects under complex surroundings. It proposed the assumption that human perception acts in the form of manifolds. Therefore, computer-based recognition can be inspired

from this assumption, that is, the low dimensional data of objects acquired from LLE manifold learning can reveal their features sufficiently.

The essence of manifold learning is that when the sampling data are in a low-dimensionality manifold space, their dimensionalities can be reduced to show the inherent geometrical distribution rule of the low-dimensionality manifold. In this way, the manifold learning approach can better reveal the substantive characteristics of objects. This non-linear unsupervised learning approach is simple, intuitive and effective. At present, its mainstream methods include: Locally Linear Embedding (LLE) [3], Isotacticity Mapping (ISOMAP) [4], Laplace Characteristic Mapping [5], Local Preserving Projection (LPP) [6], Stochastic Neighbor Embedding (SNE) [7], Charting a Manifold [8], Locally Linear Smoothing [9], etc.

As for classification, the purpose of dimensionality reduction is not only to reduce the dimensionalities of known sample points, but more importantly to discover the inherent manifold of unknown sample points and augment, to the greatest extent, the margin between sample points and classification plane so as to improve its robustness. Consequently, the label information plays an important role in classification. Recently, the supervised LLE manifold learning approach has also experienced some progress [10] [15] and shown favorable results in pattern recognition. For supervised feature extraction, sufficient consideration should be given to the generalization of dimensionality reduction learning, ensuring precise reflection of manifold structure in the embedded space and largest variation between sub-manifolds in the embedded space after dimensionality reduction, so as to increase the margin between different classes and reduce the inner-class data variation as much as possible. The Fisher subspace with LDA algorithm is undoubtedly the best learning approach with the largest intra-class and smallest inner-class variation, and improved separability of data.

This paper aims to improve the LLE manifold learning method with the idea of LDA learning. In the experiment of handwritten digit classification, this method shows a better recognition result compared with unsupervised LLE.

The following section will review the traditional LLE and the main contribution of this paper will be set out in section 3, where the structure and process of the proposed SLLE method is described. Section 4 will display experiments of its application to handwritten digit recognition.

2. Locally Linear Embedding (LLE)

Assume that the given sample data aggregate $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ is a matrix with the size of $D \times N$, where D is the dimensionality of the vector and N is the quantity of vectors or samples. These data are collected in a potential smooth manifold. Then a matrix \mathbf{Y} , with the size of $d \times N$, is generated through LLE algorithm, where, d is the eigenvector of the embedded space

of the corresponding sample. During image processing, each image can be scanned and represented as a column vector.

The three major steps of LLE are [3]:

Step 1: searching for K closest adjacency points of each point in the training samples based on Euclidean Distance, which form an adjacency matrix A with $K \times N$ size, where the column vectors represent the distance between the corresponding points to their K adjacency points.

Step 2: reconstructing coefficient W_{ij} for each point i and its adjacency matrices to get the smallest cost function $J(W)$, as in equation 1.

$$J(W) = \sum_{i=1}^N \left\| X_i - \sum_{j=1}^N W_{ij} X_j \right\|^2. \quad (1)$$

To maintain the translation invariance between each point and its adjacency points, the constraint condition for the above equation is

$$\sum_j W_{ij} = 1. \quad (2)$$

It means, in the weight matrices, the sum of elements in each column is 1. If X_i and X_j are not adjacencies, then $W_{ij} = 0$.

Step 3: calculating the low dimensional vector Y_i of the embedded space of each X_i , whose cost function is shown as equation 3:

$$\sigma(Y) = \sum_{i=1}^N \left\| Y_i - \sum_{j=1}^N W_{ij} Y_j \right\|^2 = \text{trace}(Y^T (I - W)^T (I - W) Y) \quad (3)$$

When $\frac{1}{N} \sum_{i=1}^N Y_i Y_i^T = 1$, $\sum_{i=1}^N Y_i = 0$, the cost function has minimum. Y can be obtained through decomposing the eigenvalue of $M = (I - W)^T (I - W)$, that is, d are non-zero minimum eigenvalues of M and their corresponding eigenvectors are $\langle \lambda_i, v_i \rangle, i = 1, \dots, d$, and thus the j th component embedded in the i th point is $\sqrt{\lambda_j} v_{ji}$, where v_{ji} is the i th component of v_j .

3. Supervised Locally Linear Embedding (SLLE)

The process of LLE algorithm can be improved by placing supervision on two steps, so as to limit the searching scope for adjacencies and add supervised learning after LLE dimensionality reduction.

Step 1: intra-class adjacency search (to supersede the first step of LLE algorithm). Let the whole sample set be Ψ , which is then divided into m subsets $\Psi_1, \Psi_2, \dots, \Psi_m$ according to their tag values. And when $\forall i \neq j$, $\Psi = \Psi_1 \cup \Psi_2 \cup \dots \cup \Psi_m$ and $\Psi_i \cap \Psi_j = \emptyset$, each Ψ_i contains the data of the same tag value and the whole sample forms an array of m matrices, each of which consists of the sample data representing the same tag value. For each $\mathbf{x}_i \in \Psi_j$, the K adjacent points are searched within the Ψ_j class. This ensures that the reconstruction of its weight coefficient is exclusively based on the intra-class data, which is beneficial to the classification. In contrast, in LLE algorithm, this step covers all the sample points, leading to a mixed embedding of data from different classes and thus blurring the distinction between classes.

Step 2: Fisher subspace learning. A further supervised learning is thought to be necessary after manifold learning, so as to maximize the inter-class variation and minimize the intra-class variation for increased preciseness of classification. The low dimensional feature \mathbf{Y}_i of the manifold learning is considered as the learning object of Fisher subspace, which can be gained through equation 4 [16]:

$$\mathbf{F} = \arg \max_w \left(\frac{\|\mathbf{W}^T \mathbf{S}_b \mathbf{W}\|}{\|\mathbf{W}^T \mathbf{S}_w \mathbf{W}\|} \right), \quad (4)$$

where, \mathbf{S}_w , the intra-class variation matrices, representing the average variation degree of the elements in each class and \mathbf{S}_b , inter-class variation matrices, showing the variation degree between the centers of different classes, are obtained respectively by equations 5 and 6:

$$\mathbf{S}_w = \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^{N_i} N_i (\mathbf{x}_{ij} - \overline{\Psi}_i)(\mathbf{x}_{ij} - \overline{\Psi}_i)^T \quad (5)$$

$$\mathbf{S}_b = \frac{1}{N} \sum_{i=1}^m N_i (\overline{\Psi}_i - \overline{\mathbf{x}})(\overline{\Psi}_i - \overline{\mathbf{x}})^T, \quad (6)$$

where, N_i represents the number of samples in the i th class, \mathbf{x}_{ij} represents the j th column vector (data point) in the i th-class sample matrices, $\overline{\Psi}_i$ represents the average column vector of the i th-class sample matrices, $\overline{\mathbf{x}}$ represents the average vector of the whole sample. Therefore, the vector $\hat{\mathbf{Y}}_i$ of \mathbf{Y}_i gained through Fisher subspace learning is shown as in equation 7:

$$\hat{\mathbf{Y}}_i = \mathbf{F} \mathbf{Y}_i. \quad (7)$$

Concluded from the above steps, this proposed method has three characteristics: (1) in the first adjacency search, data involved are attached with class information, which reduces interference from other data and thus improves the reliability of data features compared with LLE algorithm. Meanwhile, adjacency search can reduce the amount of data storage and calculation; (2) the supervised LLE reflects the topological relationship between intra-class data, avoiding interference from inter-class topology, thus enhances its accuracy; (3) learning of fisher subspace further intensifies the differences in class features among data, so as to realize high recognition rate through low-dimensionality features. These characteristics theoretically justify SLLE's application in engineering.

This proposed SLLE algorithm process can be summarized as figure 1:

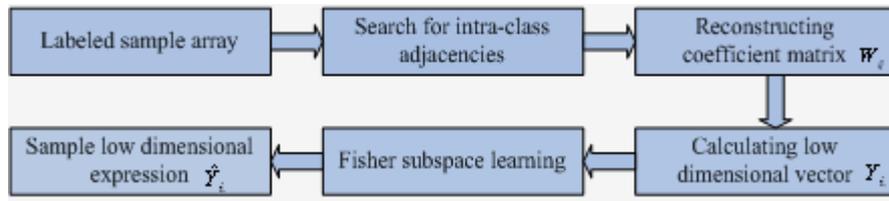


Fig.1. Computational process of SLLE proposed in this paper

4. Experiment

The MNIST handwritings database [17] is used as the experiment sample, which contains grey-scale images of digits from '0' to '9', with pixel of 28×28. Each digit has 4000-6000 samples, containing images taken from different angles. To facilitate processing, each handwritten digit has been scanned and its pixels have been transformed uniformly to a column vector with 784 dimensionalities. In experiment, the same amount of 4000 samples is chosen for each digit as the subject.

4.1. Manifold Distribution Verification for Handwritten Samples

The only difference between images of the same digit is the angle for taking the picture. Each image in one group can be seen as a point in an image space, moving along a curve or a curved surface with the rotation of images. Therefore, the perception invariance (rotation, zoom, and translation etc) of a group of images can be shown as low dimensional manifold. Figure 2 shows randomly chosen 200 samples of the digit '1' from the MNIST database. Figure 3 is the two-dimensionality representation of Figure 2, showing the manifold structure of the sample image distribution.

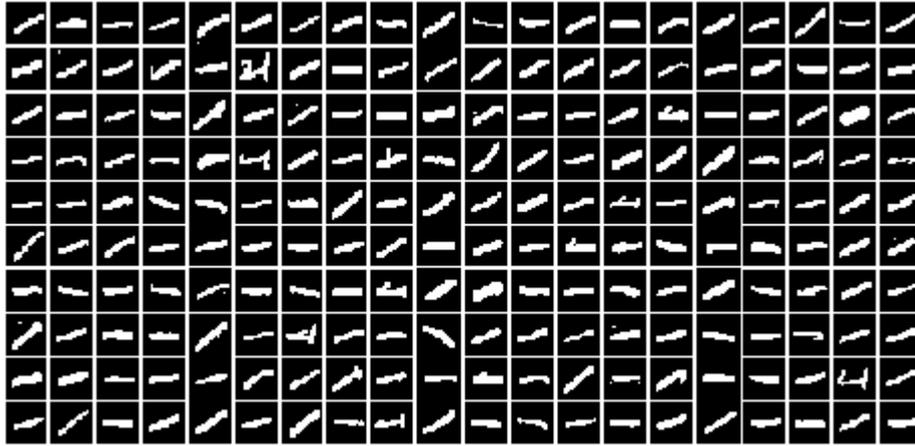


Fig. 2. Samples of 200 images of '1' from MNIST

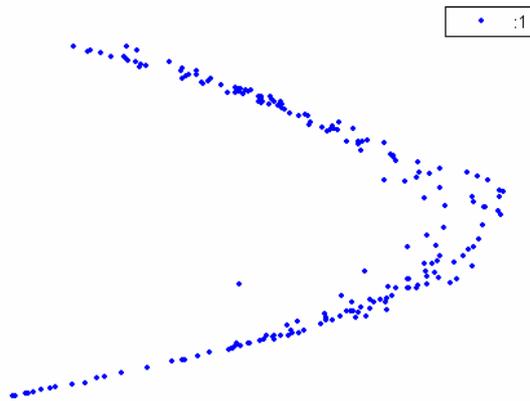


Fig.3. Two-dimensionality expression of digit '1'

4.2. SLLE and LLE Recognition of Handwritten Digits

To compare the impacts of unsupervised and supervised low dimensional eigenvectors on classification results, a classifier is designed based on the same neural network (NN). Limited by computer resources, only 400 samples are chosen for each digit and hence 4000 in total for each experiment. The ideal and actual outputs in each experiment are recorded in detail. As the NN is unstable, each experiment is carried out many times and recorded the average values so as to ensure the authenticity of the results. Table 1 and 2 show the average recognition rates of supervised and unsupervised LLE respectively in 50 experiments on 10-dimensionality eigenvector and with the

A Supervised Manifold Learning Method

same classifier of BP NN. The horizontal lines present the percentages of actual recognition rates, and vertical lines, actual digit inputs or ideal outputs.

Table 1. Average recognition rate (%) of MNIST with SLLE

Output In	'0'	'1'	'2'	'3'	'4'	'5'	'6'	'7'	'8'	'9'
'0'	99.6 875	0	0	0	0.3 125	0	0	0	0	0
'1'	0	99. 375	0.3 125	0	0	0	0	0	0	0.3 125
'2'	0.93 75	0.3 125	98. 125	0.6 25	0	0	0	0	0	0
'3'	0	1.5 625	0.9 375	95. 625	0	0.62 5	0	0.6 25	0.31 25	0
'4'	0.31 25	0.3 125	0	0	97. 5	0.62 5	0.31 25	0.6 25	0	0.3 125
'5'	0.31 25	1.8 75	0	0.3 125	0.3 125	95.3 125	0	0.9 375	0.31 25	0.6 25
'6'	0	0	0	0	0.3 125	0	99.6 875	0	0	0
'7'	0.31 25	0	0.6 25	0.6 25	0.3 125	0.93 75	0.31 25	96. 25	0	0.6 25
'8'	0.62 5	0.3 125	0.9 375	1.2 5	0	0.31 25	0	0	95.9 375	0.6 25
'9'	0	0.3 125	0.3 125	0.6 25	0.6 25	0	0	0.6 25	0.31 25	96. 875

Table 2. Average recognition rate (%) of MNIST with LLE

Output Input	'0'	'1'	'2'	'3'	'4'	'5'	'6'	'7'	'8'	'9'
'0'	95. 625	1.2 5	0.62 5	0	1.2 5	0.93 75	0.31 25	0	0	0
'1'	0	99. 375	0.31 25	0	0	0.31 25	0	0	0	0
'2'	8.7 5	1.5 625	80.9 375	5.9 375	0	0	1.25	0	0.93 75	0
'3'	2.5	9.0 625	7.81 25	73. 75	0.3 125	3.12 5	0	2.18 75	0.93 75	0
'4'	0.6 25	0.3 125	0	0	86. 875	3.43 75	0.31 25	0	0.62 5	4.0 625
'5'	0.9 375	5.6 25	1.25	4.6 875	2.8 125	78.4 375	0	5	0.62 5	0.3 125
'6'	0.9 375	0	0	0	2.8 125	0.31 25	95.9 375	0	0	0
'7'	0.3 125	0.9 375	0.31 25	1.2 5	0.3 125	1.56 25	0.62 5	93.4 375	0	0.6 25
'8'	6.5 625	3.1 25	5.62 5	2.5	0.6 25	0.93 75	0.62 5	0	72.8 125	3.7 5
'9'	1.8 75	2.5	0	0.9 375	0.9 375	1.25	0	0.31 25	9.68 75	78. 125

4.3. Comparison of Recognition Results between SLLE and Other Supervised Manifold Learning Methods

To prove the effectiveness of the proposed SLLE, it was compared with LLE+M and Isomap+M [18] through experiments with the samples in section 4.2. The dimensionality is set as 10 and the three-layer BP NN is used as the classifier, of which the hidden layer is set as 12. The results are shown in Table 3.

Table 3. Results comparison between three supervised manifold learning methods

Method	Training Time (s)	Recognition rate (%)
Isomap+M	1514.0	92.6
LLE+M	984.0	95.54
SLLE (presented in this paper)	189.6	99.42

As shown in Table 3, the proposed SLLE algorithm can not only save training time but also improve recognition rate in contrast with other methods. On the one hand, the adjacency search mechanism introduced in this method demonstrates obvious advantage over other methods in terms of searching efficiency. On the other hand, fisher subspace greatly increases the efficiency of learning and brings about better results than other methods.

4.4. Impacts of Different Low Dimensional Eigenvectors on Recognition Rate

Experiments in 4.2 are conducted on 10-dimensionality eigenvectors. However, eigenvectors of other low dimensions should also be considered to compare the results of the two methods. Figure 4 shows the relation between the dimension and the average recognition rate gained from 10 experiments. Below 10 dimensions, SLLE reports an obvious advantage over LLE; and to draw the same results, LLE needs more feature information. Another interesting fact shown in Fig.4 is the flatter change trend of recognition rates when eigenvectors reach a certain dimension (i.e. SLLE: 25; LLE: 30), rather than the-higher-the-better result.

A Supervised Manifold Learning Method

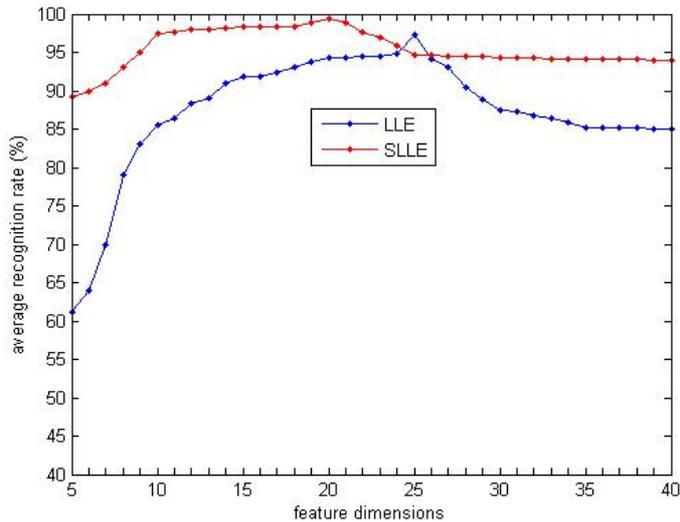


Fig.4. Recognition results of two methods with different feature dimensions

5. Conclusion and Prospect

Manifold learning method has received considerable attention from scholars in the field of visual perception home and abroad. Its ability of non-linear dimensionality reduction and effects of its supervised version in handwritten digit recognition have been proven in experiments.

1) Tables 1 and 2 show that the supervised LLE learning method proposed by this paper is evidently superior to unsupervised LLE method in terms of recognition rate and applicability, especially for lower feature dimensions.

2) This method has shown high recognition rates in its application to handwritten digit, giving theoretically as well as practically valuable reference to handwritten digit recognition.

3) The experiment results shown in Fig.4 approve that manifold learning results let the machine recognition further approach the nature of human recognition. The reason is that human beings can recognize objects as long as there exist some features, and the increment of feature information does not change the recognition of class information.

4) Compared with the methods introduced in literature [17], the supervised LLE method proposed by this paper is closer to the international level. However, as these methods are all unstable in practice, to improve the robustness of this method is worth our further study.

6. Acknowledgement

This work is supported by National Basic Research Program of China (Program 973), No: 2007CB311005-01. I would also like to give my special thanks to the anonymous reviewers of this paper for their contributions to this work.

7. References

1. Paul J. Besl , Ramesh C. Jain, "Three-Dimensional Object Recognition," ACM Computing Surveys, March, vol.17(1):75-145 ,(1985)
2. H.Sebastian Seung, Daniel D.Lee,"The Manifold Ways of Perception," Science, vol, 290, 2268 – 2269,(2000)
3. S. Roweis and L. Saul, "Nonlinear Dimensionality Reduction by locally linear embedding," Science, vol. 290, 2323–2326.(2000)
4. Joshua B. Tenenbaum, Vin de Silva, John C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," Science, vol. 290, 2319–2323. (2000)
5. Belkin M , Niyogi P. "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," Neural Computation , 15(6) , 1373 -1396.(2003)
6. He XF , Yan SC , Hu YX , et al . "Face Recognition Using Laplacianfaces," IEEE Trans on PAMI , 27 (3) , 328 -340.(2005)
7. Hinton G, Roweis S. "Stochastic Neighbor Embedding," Neural Information Proceeding Systems : Natural and Synthetic. Vancouver , Canada , (2002)
8. Brans MM. "Charting a Manifold," Neural Information Proceeding Systems Natural and Synthetic. Vancouver , Canada , (2002)
9. Park J H , Zhang ZY, Zha HY, et al . "Local linear smoothing for nonlinear manifold learning," IEEE Computer Society Conference on CVPR. Washington DC , 452 — 459.(2004)
10. De Ridder D, Kouropteva O, Okun O, Pietikainen M, Duin RPW. "Supervised locally linear embedding," Aartificial neural networks and neural information processing - ICAN/ICONIP Vol, 2714, 333-341. (2003)
11. Kouropteva O, Okun O, Pietikainen M. "Supervised locally linear embedding algorithm for pattern recognition," Pattern Recognition and Image Analysis ,Proceedings Vol, 2652, 386-394. (2003)
12. Belkin M, Niyogi P. "Semi-Supervised Learning on Riemannian Manifolds," Machine Learning, 56(1),209–239,(2004)
13. Liang D, Yang J, Zheng ZL, Chang YC. "A facial expression recognition system based on supervised locally linear embedding," Pattern Recognition Letters Vol. 26 , (15): 2374-2389,(2005)
14. Tenenhaus A,Giron A , Viennet E , Bera M , Saporta G and Fertil B. "Kernel logistic PLS: A tool for supervised nonlinear dimensionality reduction and binary classification," Computational Statistics & Data Analysis Vol. 51 (9): 4083-4100,(2007)
15. Wang M, Yang H, Xu ZH and Chou KC "SLLE for predicting membrane protein types," Journal of Theoretical Biology Vol: 232 (1), 7-15,(2005)
16. Fisher R A. "The use of multiple measurements in taxonomic problems," [J]. Annals of Eugenics, (part П) ,179~188. (1936)
17. <http://yann.lecun.com/exdb/mnist/>

18. Xin geng, zhan dechuan, zhou zhihua. Supervised nonlinear dimensionality reduction for visualization and classification. IEEE Trans on SMCB, 2005,35(6):1098-1107

Zuojin Li received the MSc in College of Communication Engineering from Chongqing University, Chongqing, China, in 2007. Currently, he is working toward the Ph.D. degree in the College of Automation from Chongqing University, Chongqing, China. His current research interests include visual perception, image processing, pattern recognition and intelligence system.

Weiren Shi received BSc in industrial automation instrumentation from Chongqing University, China. He is currently a professor and tutor of a Ph.D. student in the College of Automation, and the dean of TT&C institute, Chongqing University. His current research interests include information control and intelligence system, image processing and computer vision, wireless sensor network and embedded system. He is a member of the IEEE.

Xin Shi Ph. D. candidate at Waseda University and teacher at Chongqing University. His research interest covers optimization systems, evolutionary computing systems, information gathering platform, intelligent system and image processing.

Zhi Zhong received MSc in electronics engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China, in 2004. From 2004 to 2005, he was studying in Shanghai Jiaotong University for PhD, Shanghai, China, in 2005. He is a research assistant in the Shenzhen Smartech Institute currently. His research interests include computer vision, intelligent surveillance system, and robotics

Received: December 26, 2008; Accepted: May 23, 2009.