International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.7, No.3 (2014), pp.123-134
http://dx.doi.org/10.14257/ijsip.2014.7.3.11

Automatic head pose estimation with Synchronized sub manifold embedding and Random Regression Forests

Yulian Zhu, Zhimei Xue and Chunyan Li

*College of Education, Hebei Normal University of Science & Technology
Qinhuangdao, Hebei, P.R.China, 066004
Ylzhu_ruby@163.com*

Abstract

Head pose can indicate the eye-gaze direction and face toward which is an important part of human motion estimation and understanding. Due to physical factors of the camera, shooting environment, as well as the appearance change of humanity, the head pose estimation becomes a challenging task. Synchronization sub manifold embedding can find the internal structure of nonlinear data for nonlinear dimensionality reduction and random regression forests can make the nonlinear function mapping for getting the right head pose. In this paper, the advantages of these two algorithms are combined with a method for solving the head pose estimation. Data collection step, the depth data come from the 3D sensor; and training data step, the data is using the local linear structure for label and using a statistical model for synchronization pose samples. Meanwhile the experimental results on a publicly available database prove that the proposed algorithm can achieve state-of-the-art performance while the current estimate has a faster speed and higher robustness when large range of pose changes and outperforms existing.

Keywords: *Synchronized sub manifold embedding, Random regression forests, Head pose estimation*

1. Introduction

Head pose estimation is an important part of human interaction and psychological consciousness and it has wide application in computer vision. Changes of the head pose can be not only use to analyze the human attention in "smart" environment, but also reflect the human mind. Head pose estimation is widely used in real life, such as driving in driver assistance systems, the driver's head pose estimation can play an important security role; in a video conference, head pose estimation can be effective through effective analysis of the attitude of the audience and it can also get the participants' attention level in real-time.

Currently, more and more attention on the applications of the head poses estimation, such as the U.S. VACE projects have done a lot of research in this regard. In these research areas, the head pose estimation is an integral part of the foundation, both human-computer interaction and human interaction, the key is how to use the computer vision to sense the presence of the human, in order to effectively analyze and understand people behavior. It's easy to see that pose estimation has great significance in these areas.

Existing head pose estimation method can be divided into model-based approach and appearance-based approach. Among them, the model-based approach using several facial feature points build a three-dimensional model of human face or other geometric model and the model make the use of head rotating that make the geometry changes to judge the head pose [1-2]. The advantage is easy to implement, high computation efficient, accurate and easy

ISSN: 2005-4254 IJSIP

to understand. However, it may be sensitive to the effects of noise, occlusion, lighting, scale, and other expressions of the image that result it can't extract the required facial feature points stable. Appearance-based approach establishes a direct mapping relationship between the image and the head pose [3, 4]. Related work often focuses on subspace methods [5, 6] and learning methods based on image features [7]. Such method has high robustness and estimation accuracy, but because of it requires a lot of proper training data and accurate image registration, when in practical application it will make a larger workload. Considering combine the advantages of the two methods, Synchronized Sub manifold Embedding (SSE) [8, 9] uses labeled data to calculate a projection matrix that maps range image samples into a lower dimensional representation. The regression is implemented within a random forest framework [10], this projection is optimized for separation of different head poses while minimizing the residual part of non pose related information.

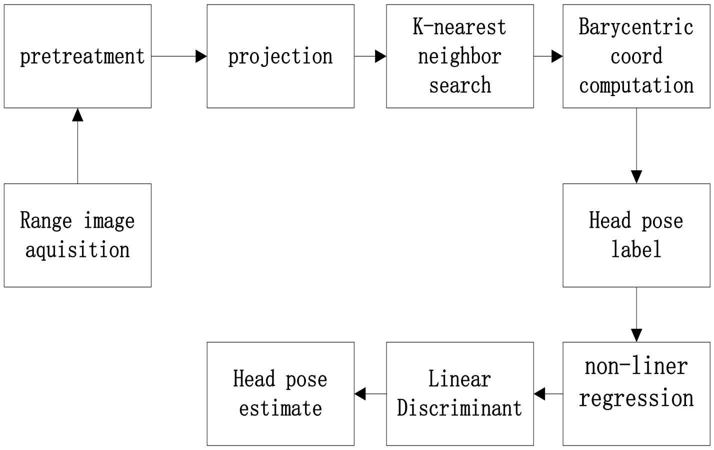


Figure 1. Flowchart of the Head Pose Estimation

2. Related Work

Head pose estimation is a very important field of research in computer vision and pattern recognition [11]. The correlation algorithm can be divided into eight categories, including the appearance-based method, detector array, nonlinear regression, manifold embedding method, elastic model method, geometric method, tracing method and a combine of variety of algorithms.

Current research on appearance based head pose estimation can be roughly divided into three categories. The first category [12] formulates pose estimation as a conventional multiclass pattern recognition problem, and only rough pose information is inferred from these algorithms. The second category takes pose estimation as a regression problem, and nonlinear regression algorithms, *e.g.,* Neural Network [13], are used for learning the mapping from the original appearance features to the pose label. The last category assumes that the pose data lie on or nearly on a low-dimensional manifold, and manifold embedding techniques [14] are utilized for learning a more effective representation for pose estimation. In this work, we address the challenging problem of head pose estimation, instead of multiclass pattern recognition problem as done conventionally, and, hence, the algorithms takes pose estimation as a regression problem from the second category are applicable in our scenario.

Manifold Embedding Methods exploit the fact that even though represented as high dimensional images, head poses should form a lower dimensional manifold. At its best, the dimensionality of this manifold can go down to the number of varied degrees of freedom of the head poses. Most approaches, that are based on this idea use methods like Locally Linear Embedding (LLE), ISOMAP or LaplacianEigenmaps (LE) for a non-linear dimensionality reduction.

Balasubramanian, *et al.*, [15] gain person independence for LLE, ISOMAP and LE by weighting the Euclidean feature distance with the corresponding label (head pose) distance. In contrast Ben Abdelkader [16] shows a direct way to incorporate label distance related information into the objective functions of LLE and LE. Wang, *et al.*, [17] performs two steps of dimensionality reduction, first an unsupervised step consisting of ISOMAP which is followed by a supervised step using linear Local Fisher Discriminant Analysis (LFDA) [18]. The linear mapping obtained by this method also allows the subsequent projection of out-of-sample examples. However, this approach does not tackle the problem of person dependence. Similarly Yan, *et al.*, [9] and Tofighi, *et al.*, [19] use a linear projection based on Multiclass Linear Discriminant Analysis (M-LDA) instead of a non-linear dimension reduction.

Some researchers try to combine more approaches and features to achieve the head pose estimate. Such as the authors of [20] use a combination of the face appearance and a set of specific feature points, which bounds the range of recognizable poses to the ones where both eyes are visible. The approach presented in [21] uses head pose estimation only as a preprocessing step to face recognition, and the reported errors are only calculated on faces belonging to the same people. Breitenstein, *et al.*, [22] proposed a real-time system which can handle large pose variations, partial occlusions, and facial expressions from range images.

The second category takes pose estimation as a regression problem ,and the random forests [23] have become a popular method in computer vision which given their capability to handle large training datasets, high generalization power, fast computation, and ease of implementation. Recent works showed the power of random forests in mapping image features to votes in a generalized Hough space [24] or to real-valued functions [25, 26]. Multiclass random forests have been proposed in [26] for real-time head pose recognition from 2D video data. Fanelli, *et al.*, [27] applies random forest regression and gains a high accuracy for head pose estimation. The random forest is trained on an extensive synthesized head pose database and uses binary features. To the best of our knowledge, we present the approach that uses random regression forests for the task of head pose estimation from range data.

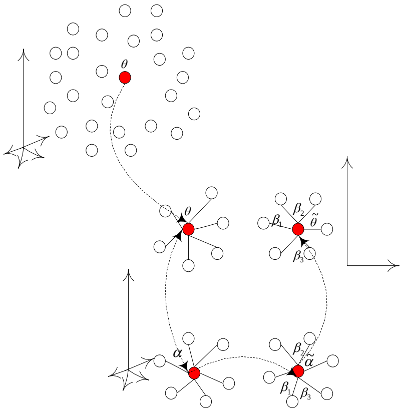


Figure 2. The Process of the Synchronized Submanifold Embedding

3. Head Pose Estimation

3.1. Synchronization

SSE requires discretized pose classes of different orientations which contain one sample from (almost) each subject. For the purpose of high resolution and accuracy, these classes should densely cover all possible head poses. Since databases of natural head movements usually do not meet both requirements, SSE performs a cross-wise synchronization between all training subjects (see Figure 2a). This synchronization step interpolates missing range image samples in a locally linear manner using a simplexization of the subject specific label space. Yan *et al.*, [9] define the k-simplex for an arbitrary sample within the label space as,

$$S_k(\alpha^c) = \left\{ \beta_0\alpha^c + \sum_{j=1}^{k} \beta_j\alpha_j^c : \sum_{j=1}^{k} \beta_j = 1, \beta_j \geqslant 0 \right\}$$

where α^c is a sample label of subject c, α_1^c, . . . , α_k^c are the k nearest neighbors of α^c and β_0, . . . , β_k are the generalized barycentric coordinates (GBC) which allow addressing arbitrary points within the simplex $S_k(\alpha^c)$. Based on this simplex structure an arbitrary pose sample$\tilde{\alpha}$, that is missing for subject c, can be interpolated. The first step to perform this interpolation consists in solving the following optimization problem:

$$\arg\square\min_{\alpha^c,\beta} \left\| \tilde{\alpha} - \alpha_k^{\beta}(\alpha^c) \right\|^2$$

Thereby $\alpha_k^{\beta}(\alpha^c)$ refers to a certain label (head orientation), that can be addressed within the simplex $S_k(\alpha^c)$ using the GBC vectorβ . For the following it is assumed, that this locally linear reconstruction relationship is transferable between features and pose label space. Based on this assumption, the missing feature sample $\bar{\varepsilon}$, which is associated with the pose label$\tilde{\alpha}$, can be interpolated with

$$\bar{\theta}^c = \beta_0\varepsilon^c + \sum_{j=1}^{k} \beta_j\varepsilon_j^c$$

Therebyε^c, ε_1^c,…ε_k^c are the feature samples that are associated with the pose label samples α^c, α_1^c,…, α_k^c. The described procedure is repeated for all occurring pose labels of all pairs of subjects. It results in head pose classes that are filled up with additional interpolated pose samples where original samples were missing.

3.2. Random Regression Forests

Classification and regression trees are powerful tools capable of mapping complex input spaces into discrete or respectively continuous output spaces. A tree achieves highly non-linear mappings by splitting the original problem into smaller ones, solvable with simple predictors. Such models are stored at the leaves, computed from the annotated data which reached each leaf at train time.

Breiman [23] shows that, while standard decision trees alone suffer from over fitting, a collection of randomly trained trees has high generalization power. Random forests are thus ensembles of trees trained by introducing randomness either in the set of examples provided to each tree, in the set of tests available for optimization at each node, or in both. Figure 3 shows a very simple example of the regression forest used in this work.

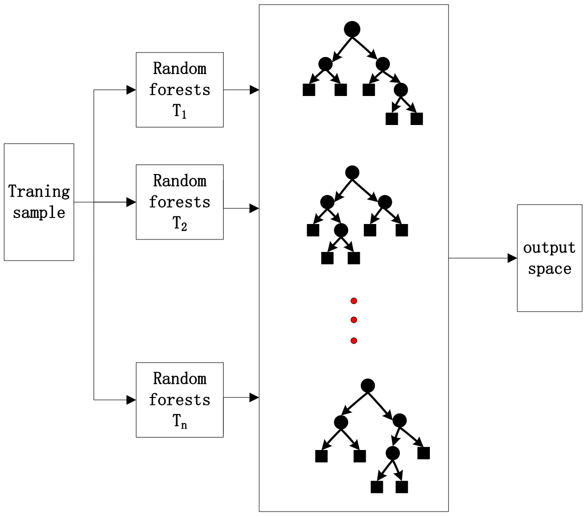


Figure 3. The Decision Procedure of the Random Forest Classifier

3.3. Training

The learning is supervised, *i.e.*, training data is annotated with values in$\mathbf{R}^D$, where D is the dimensionality of the desired output. In our setup, training examples consist of range images of faces annotated with local linear structure. We limit ourselves to the problem of estimating the head pose, thus assume that the head has been already detected in the image. However, a random forest could be trained to jointly estimate the head position in the range image together with the pose, as in [24, 28].

Each tree T in the forest $\mathbf{T} = \{\mathrm{T_t}\}$ is constructed from a set of patches $\{\mathrm{P_i} = (\alpha_i^f, \omega_i)\}$randomly sampled from the training examples. α_i^f are the extracted visual features for a patch of fixed size; in the current setup, we use one to four feature channels, namely depth values and, optionally, the X, Y, and Z values of the geometric normal computed over neighboring, non-border pixels. The real-valued vector$\omega_i = \{\omega_x, \omega_y, \omega_z, \omega_{yaw}, \omega_{pitch}, \omega_{roll}\}$contains the pose parameters associated to each patch. The components ω_x, ω_y, and ω_z represent an offset vector from the point in the range scan falling on the center of the training patch to the nose position in 3D, while $\omega_{yaw}, \omega_{pitch}$, and ω_{roll} are the head rotation angles denoting the head orientation.

We build the trees following the random forest framework [23]. At each non-leaf node, starting from the root, a test is selected from a large, randomly generated set of possible binary tests. The binary test at a non-leaf node is defined as$t_{f,x_1,x_2}(I)$:

$$|\mathbf{x}_1|^{\;1} \sum_{q \in x_1} \alpha^f(q) \quad |\mathbf{x}_2|^{\;1} \sum_{q \in x_2} \alpha^f(q) > \lambda$$

Where $\alpha^{\mathbf{f}}$ indicates the feature channel, $\mathbf{x_1}$ and $\mathbf{x_2}$ are two rectangles within the patch boundaries, and λ is a threshold. The test splits the training data into two sets: When a patch satisfies the test it is passed to the right child, otherwise, the patch is sent to the left child. We chose to take the difference between the average values of two rectangular areas (as the authors of [25]) rather than single pixel differences (as in [24]) in order to be less sensitive to noise. Figure 3 shows a patch (marked in red) and the two randomly generated regions $\mathbf{x_1}$ and $\mathbf{x_2}$ as part of a binary test; the arrow indicates the 3D offset vector stretching from the patch center (in red) to the annotated nose location (green).

During the construction of the tree, at each non-leaf node, a pool of binary tests$\{\mathbf{t}^{\mathbf{k}}\}$is generated with random values for f,$\mathbf{x_1}$,$\mathbf{x_2}$, and λ . The set of patches arriving at the node is evaluated by all binary tests in the pool and the test maximizing a predefined measure is assigned to the node. Following [25], we optimize the trees by maximizing the information gain defined as the differential entropy of the set of patches at the parent node P minus the weighted sum of the differential entropies computed at the children $\mathbf{P_L}$ and $\mathbf{P_R}$:

$$\alpha\mathbf{G} = \mathbf{H(P)} \quad (\boldsymbol{\varphi}_{\mathbf{L}}\mathbf{H(P_L)} + \boldsymbol{\varphi}_{\mathbf{R}}\mathbf{H(P_R)})$$

Where $\mathbf{P}_{\mathbf{i}\in\{\mathbf{L,R}\}}$ is the set of patches reaching node and is the ratio between the number of patches in and in its parent node, *i.e.*, $\varphi_{\mathbf{i}} = \frac{\mathbf{P_i}}{\mathbf{P}}$.

We model the vectors ω at each node as realizations of a random variable with a multivariate Gaussian distribution, *i.e.*, $\mathbf{p}(\boldsymbol{\omega}) = \mathbf{N}(\boldsymbol{\omega}, \overline{\boldsymbol{\omega}}, \boldsymbol{\Sigma})$. Therefore, *Eq.*, (2) can be rewritten as:

$$\alpha\mathbf{G} = \log|\boldsymbol{\Sigma}(\mathbf{P})| \quad \sum_{\mathbf{i}\in\{\mathbf{L,R}\}} \boldsymbol{\varphi}_{\mathbf{i}} \log|\boldsymbol{\Sigma}_{\mathbf{i}}(\mathbf{P}_{\mathbf{i}})|$$

Maximizing *Eq.*, (3) favors tests which minimize the determinant of the covariance matrix $\boldsymbol{\Sigma}$, thus decreasing the uncertainty in the votes for the output parameters cast by each patch cluster.

We assume the covariance matrix to be block-diagonal$\Sigma = \begin{pmatrix} \boldsymbol{\Sigma}^{\mathbf{v}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^{\mathbf{a}} \end{pmatrix}$, *i.e.*, we allow covariance only among offset vectors ($\boldsymbol{\Sigma}^{\mathbf{v}}$) and among head rotation angles ($\boldsymbol{\Sigma}^{\mathbf{a}}$), but not between them. *Eq.*, (3) thus becomes:

$$\alpha\mathbf{G} = \log(|\boldsymbol{\Sigma}^{\mathbf{v}}| + |\boldsymbol{\Sigma}^{\mathbf{a}}|) \quad \sum_{\mathbf{i}\in\{\mathbf{L,R}\}} \boldsymbol{\varphi}_{\mathbf{i}} \log(|\boldsymbol{\Sigma}_{\mathbf{i}}^{\mathbf{v}}| + |\boldsymbol{\Sigma}_{\mathbf{i}}^{\mathbf{a}}|)$$

A leaf is created when the maximum depth is reached or a minimum number of patches are left. Each leaf stores the mean of all angles and offset vectors which reached it, together with their covariance, *i.e.*, a multivariate Gaussian distribution.

4. Experiments

Our method is evaluated on the publicly available ETH Face Pose Range Image Data Set with provided ground truth. Random forests can be built from large training datasets in reasonable time and are very powerful in learning the most distinctive features for the problem at hand. It has been acquired using a high quality 3D scanner and contains sequences at a frame rate of 28 fps of 10K range images of 20 people who freely turned their head. The experimental environment: Pentium (R) Dual-Core CPU E5400@2.7 GHz, 2.69 GHz, 4 GB RAM. In total, this sums up to more than 104 range images. The continuous head pose annotation comprises yaw and pitch angles in intervals of ±90◦ and ±45◦ respectively. While the images have a resolution of 640*480 pixels the head typically covers an area of 190*250 pixels. The provided ground truth for each image consists of the 3D nose tip coordinates and the coordinates of a vector pointing in the face direction.

Training comprises the Synchronization and Dimension Reduction steps. In order to speed up training, a preceding PCA reduces the initial range image dimensionality by keeping 99% of variance. For the interpolation of missing samples, the Synchronization step uses the 10 closest neighbors within label space. Afterwards the Dimension Reduction reduces the sample dimensionality to 10. The Classification itself is based on the 16 closest neighbors within the lower dimensional representation. In order to ensure that head poses of different sequences are correctly aligned, the learning and classification procedures, is applied to all samples within the pitch range of [-10°--10°] and the yaw range of [-15°--15°].

In the following experiments, we always trained each tree sampling 25 patches from each of 3000 synthetically generated images, while the full ETH database was used for testing.

Table 1. Error's Average Value and Variance

Approach	Dimension				
	3	10	30	60	90
LLE	(8.8,43.9)	(8.2,30.9)	(5.1,15.6)	(4.3,12.1)	(4.5,13.0)
LEA	(20.7,176.9)	(11.7,81.6)	(9.4,62.0)	(7.9,43.4)	(7.9,52.0)
LDA	(21.0,244.7)	(13.1,138.9)	(9.1,59.0)	(8.8,54.4)	(7.5,49.1)
PCA	(18.9,220.5)	(12.7,90.9)	(8.4,41.1)	(8.3,41.2)	(7.0,33.6)
LPP	(22.8,320.9)	(18.1,189.3)	(13.6,105.2)	(11.3,68.0)	(8.6,43.9)
OURS	(10.5,93.2)	(5.6,18.0)	(4.5,13.8)	(4.8,15.0)	(4.4,13.7)

The time of our method mainly on the training, *i.e.,* see Figure 1 is about the non-linear regression, according to the error to minimum the parameters setting directly then we read the results, normalized and dimensionality reduction of 10 human face image, training time is greater than 3h, but based on the input data set processing parameters using random regression forests which reduce the time within 1.5h which can significantly improve the speed of training. After training, namely linear discriminant stages at Figure 1, the predict speed can be achieved in real-time basically. The experiment is on the group of 10 people and predicts rate is stabilized at around 0.5 s and have significantly improved compared to use of Euclidean distance determination, the time on test phase is shown in Table 1. Table 2 shows the percentage of successful estimations on the entire ETH database. Samples are considered to be correctly estimated if their directional angular errors fall below 15°, 10° and 5°respectively. The results of these comparisons indicate that our head pose detection approach that better than the method that in the Table 2.

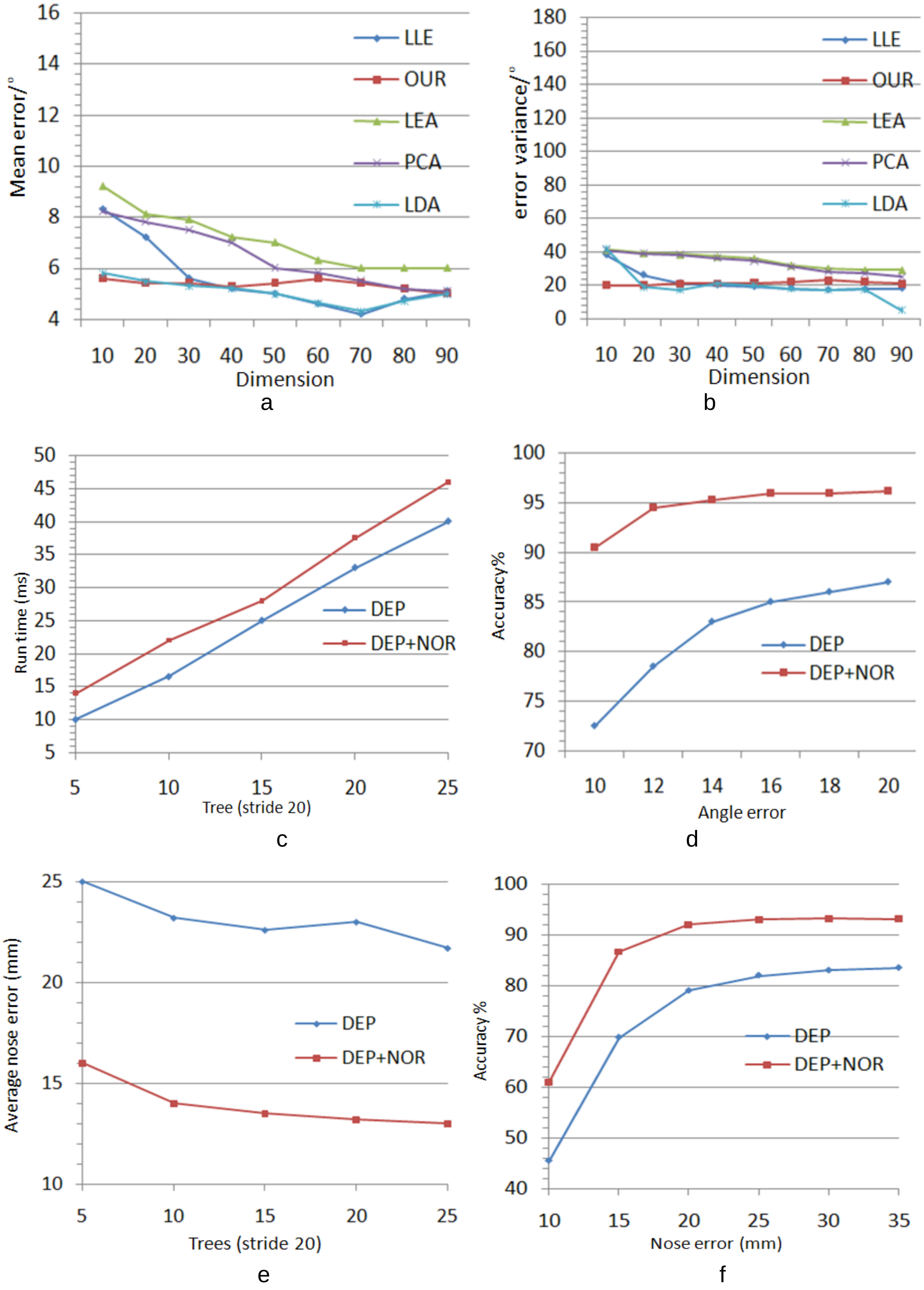
Table 2. Comparison of the Head Pose Accuracy

Method	yaw°	pitch°	direction accuracy%		
			⩽ 15°	⩽ 10°	⩽ 5°
literature [22]	6.1,10.3	4.2,3.9	97.8	80.8	
literature[27]	5.7,15.2	5.1,4.9	95.4	90.4	
OURS	3.0,3.1	2.5, 2.3	99.0	94.5	72.1

Table 1 shows the variance and average error in the experiment and the data in table is average results of 10 groups the head pose estimation values. As seen from Table 1, using our method can find the nonlinear relationship of the data soon, when the dimension reducing turns to 10; the prediction error is around 8°. When the dimension reducing to 30, for several manifold learning methods error variance will eventually reach to 5° and variance is also kept in a small range, less than 20. But the data which directly use dimensionality reduction results of the manifold learning method for linear regression which cannot obtain better perform of the pose estimation, when reduced to 10-dimensional the error is about 10° or more; when down to 30 dimensions, the basic error is around 9°. Meanwhile, the variance of the predicted value is relatively large, more than 40, which reflect the predicted results unstable and poor robustness.

From Figure 4(a) we can be find when the data dimensions down to 30, the prediction of the average error is under 8°, LLE, LDA even reached the error of 4°. Corresponding to Figure 4(b) we can be finding when the data reduced to 10, the variance of the prediction results begin to stabilize, at last the result may all lower than 20. This indicates that use the manifold learning for dimensionality reduction, and even use the traditional dimensionality reduction method, regression networks are better able to map the data, and the high robustness itself has also make the results more stable. We can find that result of traditional PCA and LDA dimensionality reduction method, the effect of PCA dimensionality reduction is bad, the average error of the LDA can achieve some satisfactory levels which are equivalent to the manifold learning methods, but the variance is relatively large, especially after dimension reduction the dimension within 30. Figure 6(c) plots the average runtime of the trees; the plot in Figure 6(d) shows the percentage of correctly estimated images as a success threshold set for the angular error, while in Figure 6(f) the thresholds are defined on the nose localization error. Figure 6(e) show the average error in the nose localization task, plotted as a function of the number of trees when the stride is fixed.

By the contrast we can find that our method has better stability. Meanwhile, the sample in the experiment is random and the differences of the facial feature also have an impact on the result, but our method still maintain the stability and hold relatively low error and volatility. Through analysis and comparison above we can find that the proposed method which combined with the advantages of manifold and nonlinear regression to estimate the head pose, while ensuring accuracy it still has fast speed and high robustness.



a

b

c

d

e

f

Figure 4. Results of the Head Pose Estimate

5. Conclusions

The algorithm uses the Synchronized Submanifold Embedding to find the inherent characteristics of non-linear data structure, combined with the Random Regression Forests nonlinear regression method, the face image is mapped to linearly separable low-dimensional and finally we use linear regression to get the head pose estimation. From the experimental results we find that the proposed method not only can ensure the accuracy of prediction but also can significantly reduce the running time, and has a high robustness. It is worth noting that we use PCA to reduce the initial range image dimensionality, so that we can get more satisfactory results. In the Synchronization step we use the 10 closest neighbors within label space for the missing samples which also improve the results. Based on the above, in particular synchronization submanifold which can obtain more accurate result.

During the experiment we can find that add label information for the face image, which can make the dimensionality reduction data have a good performance, while the outcome of the judgment also improved. Next, we will create a low-dimensional space nearest neighbor candidates, so that a training set can offer better uniform distribution training samples which cover more head pose.

References

[1] J. G. Wang and S. Eric, "EM Enhancement of 3D Head Pose Estimated by Point at Infinity", Image and Vision Computing, vol. 25, no. 12, **(2007)**, pp. 1864-1874.

[2] Y. Ebisawa, "Head Pose Detection with One Camera Based on Pupil and Nostril Detection Technique", Proceedings of the IEEE International Conference on Virtual Environments, Human-computer Interfaces, and Measurement Systems. VECIMS, **(2008)**, pp. 172-177.

[3] M. C. Erik and M. T. Mohan, "Head Pose Estimation in Computer Vision:A Survey", IEEE Trans on Pattern Analysis and Machine Intelligence, vol. 31, no. 4, **(2009)**, pp. 607-626.

[4] L. Zhao, G. Pingali and I. Carlbom, "Rea-l time Head Orientation Estimation Using Neural Networks", Proceedings of the International Conference on Image Processing. ICIP, **(2002)**, pp. 297-300.

[5] B. Raytchev, YodaI and K. Sakaue, "Head Pose Estimation by Nonlinear Manifold Learning", Proceedings of the International Conference on Pattern Recognition. ICPR, **(2004)**, pp. 462-466.

[6] J. W. Wu and M. M. Trivedi, "A Two-stage Head Pose Estimation Framework and Evaluation", Pattern Recognition, vol. 41, no. 3, **(2008)**, pp. 1138-1158.

[7] L. Kun, L. Yu-pin and Y. Shi-yuan, "Static Head Pose Estimation Under Different Illumination", Computer Engineering, vol. 34, no. 10, **(2008)**, pp. 16-18.

[8] S. Yan, H. Wang, Y. Fu, X. Yan and T. S. Huang, "Synchronized submanifold embedding for person-independent pose estimation and beyond", IEEE transactions on image processing, vol. 18, no. 1, **(2009)**, pp. 202–210.

[9] S. Yan, Z. Zhang, Y. Fu, Y. Hu, J. Tu and T. Huang, "Learning a Person-Independent Representation for Precise 3D Pose Estimation", Multimodal Technologies for Perception of Humans, Computer Science, Berlin, Heidelberg, vol. 28, no. 4625, **(2008)**, pp. 297–306.

[10] A. Criminisi, J. Shotton, D. Robertson and E. Konukoglu, "Regression forests for efficient anatomy detection and localization in ct studies", Medical Computer Vision Workshop, **(2010)**.

[11] E. Murphy-Chutorian and M. Trivedi, "Head pose estimation in computer vision: a survey", IEEE Trans Pattern Analysis and Machine Intelligence, vol. 31, no. 4, **(2009)**, pp. 607-326.

[12] S. Li, X. Lu, X. Hou, X. Peng and Q. Cheng, "Learning multiview face subspaces and facial pose estimation using independent component analysis", IEEE Trans. Image Process, vol. 14, no. 6, **(2005)**, pp. 705–712.

[13] L. Brown and Y. Tian, "Comparative study of coarse head pose estimation, in Proc.", IEEE Workshop on Motion and Video Computing, **(2002)**, pp. 125–130.

[14] Y. Fu and T. Huang, "Graph embedded analysis for head pose estimation", 7th Int. Conf. Automatic Face and Gesture Recognition, **(2006)**, pp. 3–8.

[15] V. N. Balasubramanian, S. Krishna and S. Panchanathan, "Person-independent head pose estimation using biased manifold embedding", EURASIP J. Adv. Signal Process, **(2008)**.

[16] C. BenAbdelkader, "Robust Head Pose Estimation Using Supervised Manifold", In ECCV, Lecture Notesin Computer Science, Berlin, Heidelberg, vol. 6316, **(2010)**, pp. 518–531.

[17] X. Wang, X. Huang, J. Gao and R. Yang, "Illumination and Person-Insensitive Head Pose Estimation Using Distance Metric Learning", ECCV, Computer Science, Berlin, Heidelberg, **(2008)**, pp. 624–637.

[18] M. Sugiyama, "Local Fisher discriminant analysis for supervised dimensionality reduction", In Proceedings of the 23rd international conference on Machine learning, ICML, New York, NY, USA, **(2006)**, pp. 905–912.

[19] M. Tofighi, H. Kalbkhani, M. G. Shayesteh and M. Ghasemzadeh, "Robust Head Pose Estimation Using Contourlet Transform", Technical report, Department of Electrical Engineering, Urmia University, Urmia, Iran, **(2012)** April.

[20] J. Whitehill and J. R. Movellan, "A discriminative approach to frame-by-frame head pose tracking", In Aut.Face and GesturesRec., **(2008)**.

[21] A. Mian, M. Bennamoun and R. Owens, "Automatic 3d face detection, normalization and recognition", In 3DPVT, **(2006)**.

[22] M. D. Breitenstein, D. Kuettel, T. Weise, L. Van Gool and H. Pfister, "Real-time face pose estimation from single range images", In CVPR, **(2008)**.

[23] L. Breiman, "Random forests", Machine Learning, vol. 45, no. 1, **(2001)**, pp. 5–32.

[24] J. Gall, A. Yao, N. Razavi, L. Van Gool and V. Lempitsky, "Hough forests for object detection, tracking, and action recognition", TPAMI, **(2011)**.

[25] A. Criminisi, J. Shotton, D. Robertson and E. Konukoglu, "Regression forests for efficient anatomy detection and localization in ct studies", In Medical Computer Vision Workshop, **(2010)**.

[26] C. Huang, X. Ding and C. Fang, "Head pose estimation based on random forests for multiclass classification", In ICPR, **(2010)**.

[27] G. Fanelli, J. Gall and L. Van Gool, "Real time head pose estimation with random regression forests", In CVPR.IEEE Computer Society, **(2011)**.

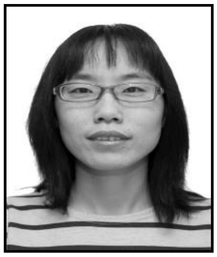
[28] R. Okada, "Discriminative generalized hough transform for object dectection", In ICCV, **(2009)**.

[29] P. Paysan, R. Knothe, B. Amberg, S. Romdhani and T. Vetter, "A 3d face model for pose and illumination invariant face recognition", In Advanced Video and Signal based Surveillance, **(2009)**.

Authors

Yulian Zhu, she received her bachelor's degree of Physics Education in Datong University (the original Yanbei Normal College), Datong, Shanxi. (2001) and master's degree of Educational Technology in Hebei University ,Baoding, Hebei (2004) .Now she is a lecturer in Hebei Normal University of Science & Technology, Qinhuangdao, Hebei. Her major fields of study are instructional design, media teaching theory and distance education.

Zhimei Xue, she received her bachelor's degree of education technology in Shanxi Normal University, Linfen, Shanxi (2002).and master's degree of education in Hebei university, Baoding, Hebei(2005), Now she is a lecturer in Hebei Normal University of Science & Technology, Qinhuangdao, Hebei. Her major fields of study are instruction design, distance education and construction of digital teaching resources.



Chunyan Li, she received her bachelor's degree of science in Northeast Normal University, Changchun, Jilin. (2003) and master's degree of education in Northeast Normal University (2005) , Now she is a lecturer in Hebei Normal University of Science & Technology, Qinhuangdao, Hebei. Her major fields of study are Instructional technology, and distance education.