# Neighborhood-Preserving Estimation Algorithm for Facial Landmark Points

Yin Liu, Ying Cui, and Zhong Jin

School of Computer Science and Engineering
Nanjing University of Science and Technology, Nanjing, China
seraphever@126.com,
{cuiying,jinzhong}@patternrecognition.cn

**Abstract.** In this paper, we propose a neighborhood-preserving estimation (NPE) algorithm for facial landmark points at arbitrary poses. The proposed NPE algorithm is based on the following assumption: the neighboring structure of the face shapes in non-frontal view is consistent with that in frontal view. A face shape both in frontal and non-frontal view is represented as a linear combination of its neighbors. It is assumed the neighbors both in frontal and non-frontal view are from same persons and with same weights in the combinations. Extensive experiments are conducted on IMM and BU_3DFE to validate the performance of the proposed algorithm.

**Keywords:** neighborhood-preserving estimation, regression model, facial landmark points.

## 1 Introduction

Automatic face annotation plays an essential role in many applications. Accurate face annotation is a premise for virtual generation of face images and face recognition. It is an interesting and challenging problem to develop an automatic annotation method, which can effectively predict the spatial locations of the landmark points for face images under varying poses and expressions.

In [1], modeling images and related visual objects as bags of pixels or sets of vectors was proposed, which made the computation of correspondences for all images simultaneously possible. In [2], the automatic face annotation problem was viewed as an energy-minimizing image coding problem which can be solved by an efficient gradient descent algorithm. [3] and [4] made full use of image sequences to annotate face images automatically. Recently, an approach for automatic annotation of the faces with any arbitrary pose and expression from annotated frontal faces only was put forward by Asthana [5]. The relevant work was extended in [6-8], in which a regression based algorithm to predict the landmarks of unseen images was presented and demonstrated on the examples of automatically annotating face images. The approach based on regression learning drastically simplifies the process of deformable modeling building and decreases the computation.

Manifold learning algorithms are based on the idea that the data points are actually samples from a low-dimensional manifold that is embedded in a high-dimensional

J. Yang, F. Fang, and C. Sun (Eds.): IScIDE 2012, LNCS 7751, pp. 630–638, 2013.

space. Locally Linear Embedding (LLE) [9] is one of the manifold learning algorithms, which attempts to reduce the dimension of the data while preserving the relationships between neighboring data points. Motivated by LLE, we propose the neighborhood-preserving estimation algorithm for facial landmark points.

The rest of the paper is organized as follows. Section 2 introduces regression based learning methods briefly. Section 3 describes the proposed method in detail. In section 4, we exhibit the experimental results and section 5 provides the discussion and conclusion of the paper.

## 2 Regression Based Learning

The core idea of the regression based learning approach is to learn the correspondences of the frontal and non-frontal face images in a regression framework.

Given a set of training set $\{(s_1, s_1^{'}),...(s_m, s_m^{'})\} \subset R^{2n} \times R^{2n}$, where $m$ is the number of the frontal and the corresponding non-frontal images and $s_i, s_i^{'}(i = 1,...,m)$ is the shapes of the $i$-th person for frontal and non-frontal images respectively:

$$s = [x_1, y_1, ..., x_n, y_n]^T \tag{1}$$

where $x_j$,$y_j$ ($j=1, \cdots, n$) is the coordinate of the $j$-th landmark point of the face image. What we desire to do is to minimize the objective function:

$$\sum_{i=1}^{m} \left| s_i^{'} - P(s_i) \right| \tag{2}$$

where $P(s_i)$ means the shape we estimate. The error between the manual and estimating annotation is expected to be as small as possible.

Viewed the shape vector as an integral whole, the correspondences can be posed in a linear framework and the location and local movements of the landmark points are completely dependent, then we'll get the dense linear regression (DLR) approach, in which the regression function can be represented as

$$s^{'} = P(s) = Ws + v \tag{3}$$

where W is the regression matrix, $s^{'}$ is the shape vector estimated in non-frontal view and $v$ is the noise. The problem becomes a ridge regression problem,

$$\min\{\|P - WG\|_2^2 + \lambda \|W\|_2^2\} \tag{4}$$

where $G = [s_1, s_2, ..., s_m]$ and $P = [s_1^{'}, s_2^{'}, ..., s_m^{'}]$, $\lambda > 0$ is a regularization factor and $\|a\|_2^2 = (\sum_i a_i^2)^{1/2}$ represents the L$_2$ norm of a. Eq. (4) can be solved by

$$W = PG^T (GG^T + \lambda I)^{-1} \tag{5}$$

Similar to DLR, if the correspondences are put in the fully sparse representation framework, we'll get the fully sparse linear regression (FSLR) approach. What's more, if the regression function is non-linear, the dense non-linear regression (DNLR) and the fully sparse non-linear regression (FSNLR) are generated. Reference [8] expands on all these methods, so we skip here.

## 3 Neighborhood-Preserving Estimation

In this section, we take full account of the neighboring information of the testing face shapes, and based on the hypotheses that the the neighborhood structure in the frontal and non-frontal view are consistent, we propose the neighborhood-preserving estimation algorithm to predict the facial shape vectors.

### 3.1 Motivation

As a new unsupervised learning method, manifold learning is capturing increasing interests of researchers in the field of machine learning. Manifold learning is a popular recent approach to non-linear dimensionality reduction, which maps the data into a lower-dimensional space and preserves the information in the original space simultaneously.

As one of the manifold learning methods, LLE is based on simple geometric intuition [9]: each data point and its neighbors are lie on or close to a locally linear patch of the manifold, so the data point can be represented as a linear combination of its neighbors. What's more, when the original data point is mapped into a certain high-dimensional space, the corresponding data point can also be reconstructed with its corresponding neighbors in the high-dimensional space and the reconstruction weights are invariant to the mapping.

LLE preserves the neighborhood structure of the data sets between the observed and the mapping space. Regression based learning methods consider the shape of the face as a whole and the goal is to find the regression function between the face shapes of the frontal and non-frontal view. Here we look the face shape as a data point in 2n space. Since regression methods look for some relationship to map the data point in frontal view into the data point in non-frontal view, then inspired by LLE, we have a similar assumption: the neighborhood structures of the face shapes are invariant to the head pose changing. That is to say, given a testing face shape in frontal view and its neighbors, the corresponding counterpart in non-frontal view preserve the same neighboring information. The regression weights computed from the frontal view are the same to those of the non-frontal view. Besides, the neighboring face shapes in frontal and non-frontal view are from the same subject.

### 3.2 NPE Algorithm

In order to illustrate our idea clearly and intuitively, we conduct a small experiment. We collect a number of annotated face images, and then choose a frontal shape as the testing sample and the other shapes in frontal and non-frontal views as the training samples. Firstly, we compute the distance between the testing sample and the training samples in frontal view and find ten nearest neighbors of the testing sample. Then we reconstruct the testing sample from its neighbors with linear coefficients. Finally, we use the linear coefficients and the shapes of chosen neighbors in non-frontal view to estimate the testing shapes in non-frontal view. Fig.1 shows the scattered dots distribution of the testing and the training shapes, and the estimated shape as well. From Fig.1 (a) and (b), we can find that the shape of the testing sample is close to the shapes of its neighbors in both frontal and non-frontal view. The estimated shape for the testing sample in non-frontal view is shown in Fig.1 (c). It can be observed that the estimated shape is similar to the manually annotated shape of the testing sample. So, the proposed method can estimate the facial landmark points in non-frontal view.

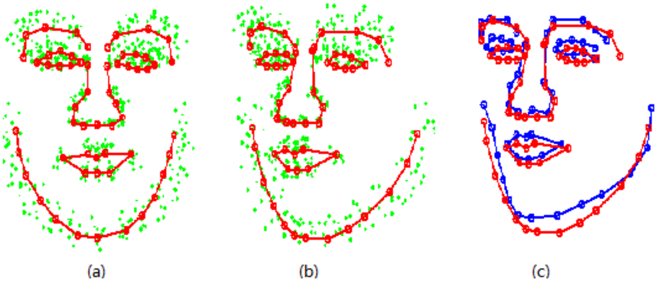


**Fig. 1.** A small experiment conducted with neighborhood-preserving estimation method: red and green dots represent the landmark points in testing and training shapes, blue dots represent the estimating landmarks, (a) testing shape and the shapes of the ten nearest samples for testing sample in frontal view; (b) the shapes of the testing sample and its ten nearest neighbors in non-frontal view; (c) the estimated shape with our method

Given an annotated frontal face image, we estimate its face shape $t'$ in non-frontal view. Firstly, we calculate the distance between the testing face shape and the training face shapes to find $K$ nearest neighbors. Thus, we can use the $K$ nearest neighbors to represent the testing face shape as

$$t = w_1 s_1 + \ldots + w_K s_K = SW \tag{6}$$

where $s_i$ ($i$=1,2, ⋯, $K$) are the *ith* nearest neighbor in frontal view and we can rewrite them into $S$=[$s_1$, ⋯, $s_K$], and all the linear coefficients $w_i$ ($i$=1,2, ⋯, $K$) can be constructed as the reconstruction weight vector $W$=[$w_1$, ⋯, $w_K$]$^T$. And W can be acquired by using

$$W = (S^T S + \alpha I)^{-1} S^T t \tag{7}$$

where $\alpha$ is a non-negative constant and $I$ is an unit. The reconstruction weight vector implies the $K$ nearest neighbors' contribution to representing the testing shape in frontal and non-frontal view. Lastly, we predict the face shape with W and the $K$ nearest neighbors in non-frontal view. The consequence is

$$t^{'} = S^{'}W = \sum_{i=1}^{K} w_i s_i^{'} \qquad (8)$$

where $s_i^{'}(i=1,...,K)$ is the face shape of the $i$th person in non-frontal view and $t^{'}$ is the shape vector predicted by NPE approach. Since the shapes of the subjects in non-frontal and frontal views keep the same neighboring structure, $s_i^{'}$ and $s_i$ are the face shapes corresponding to the same person.

The complete framework of our algorithm is summarized in Algorithm I.

**Algorithm** I (Neighborhood-Preserving Estimation Algorithm)

```
Input: An annotated frontal image for testing, the fixed pose rotation
Output: An estimated face shape with the determined pose rotation
    Step 1: Find $K$ nearest neighbors for the testing sample in frontal view;
    Step 2: Use Eq. (7) to compute the reconstruction weight vector;
    Step 3: Use Eq. (8) to estimate the face shape $t'$.
```

# 4 Experiments

## 4.1 Experiments on IMM

We conduct experiments on the IMM database [10]. IMM has a total of 240 images across 40 persons and each of them is manually annotated with 58 landmark points. Fig.2 (a) represents one example with the annotated landmark points in IMM.

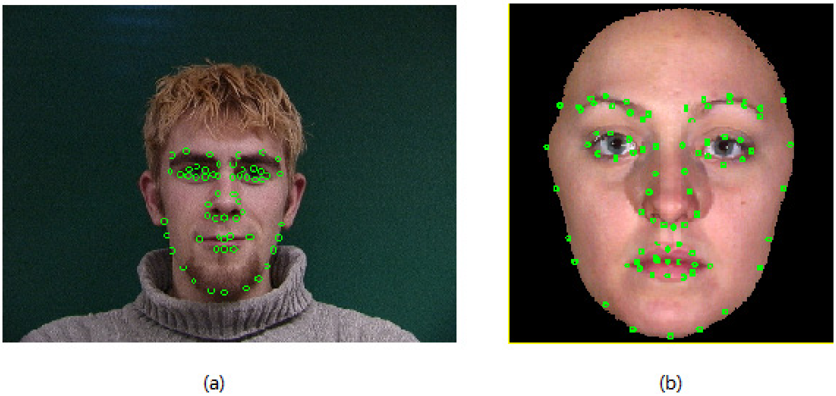


**Fig. 2.** Example faces with manually landmark points in IMM and BU_3DFE: (a) one sample with 58 landmark points in IMM; (b) one sample with 83 landmark points in BU_3DFE

To demonstrate the performance of the proposed method, we estimate the shape vectors for the face images, and compute the error between the manually annotated and the estimating landmark points.

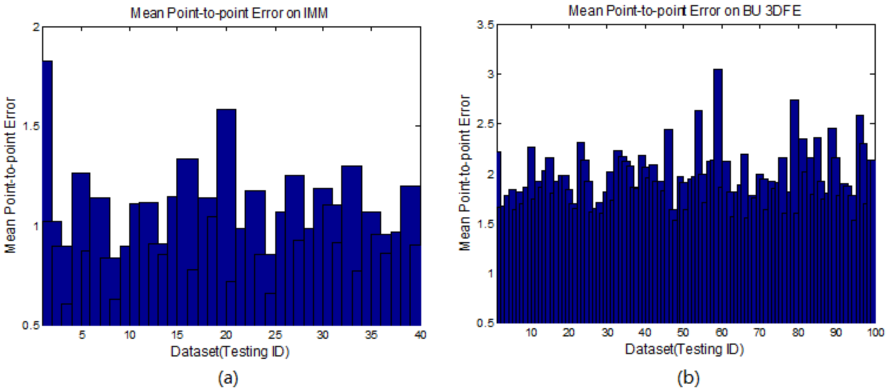
We adopt the leave-one-out cross validation scheme throughout the experiment. Each time we choose one sample in the database as the testing sample and the others are used for training. In order to evaluate the performance of the proposed approach, we calculate the mean point-to-point error between the manually annotated and the estimated landmark points for every image.

Fig.3 (a) shows the error distributions over IMM obtained from NPE. It is clear that NPE approach can be used to estimate facial landmark points under varying poses. The error between the automatic and the manually annotated landmark points are less than 2 pixels.

For the sake of showing a visual result, we give an example of the estimated facial landmark points on IMM, which are shown in Fig.4 (a). It is noted that the estimated landmark points are accurate under changing head poses. And the result of the fourth image is good as well, which means NPE can predict accurate shapes under changing illuminations. But the location of the landmark points in the fifth image in Fig.4 (a) is not as precise as that under the other conditions.

### 4.2 Experiments on BU_3DFE

The aforementioned NPE approach provides a satisfactory result on IMM, so NPE is able to estimate the facial landmark points under varying poses. If we regard the neutral expression as the frontal view and the expressions, such as happy, as the non-frontal view, we'll have the similar thought under changing expressions. The neighboring information of the shapes with the neutral and the other expressions keeps the same. To validate the thought, we make an experiment on BU_3DFE database [11], which contains 2500 images at seven expressions in frontal view and each of them is annotated with 83 landmark points. Fig.2 (b) represents an example with the annotated landmarks in BU_3DFE.



(a)

(b)

**Fig. 3.** Estimated annotation results for IMM and BU_3DFE with NPE: (a) mean point-to-point errors on IMM database; (b) mean point-to-point errors on BU_3DFE database

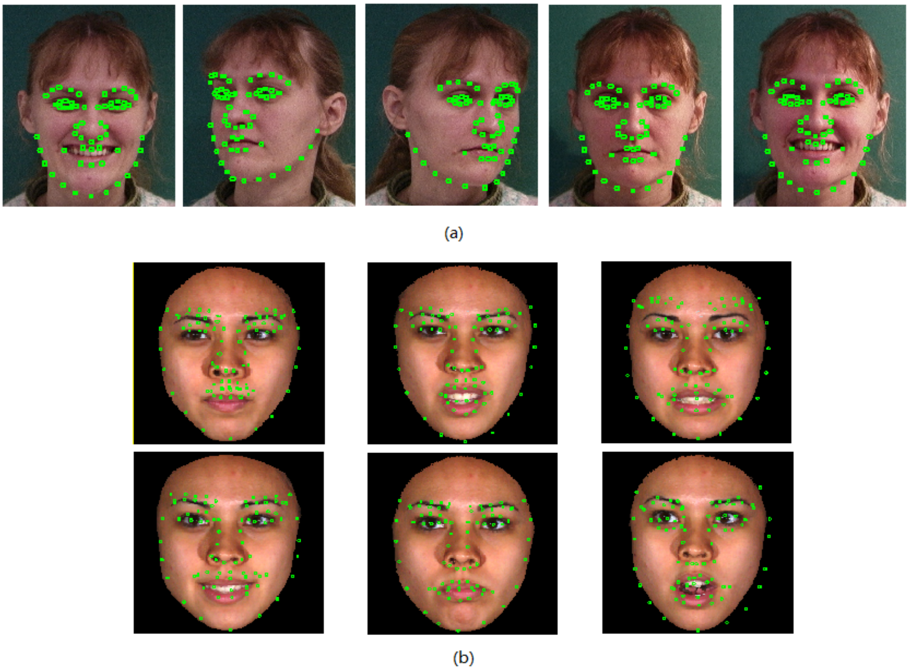


**Fig. 4.** Sample estimated results via NPE from IMM and BU_3DFE: (a) the shapes of the sample estimated in IMM, every image represent full frontal face with happy expression, face rotated approx 30 degrees to the person's right and left, full frontal face spot light added at the person's left side and full frontal face with arbitrary expression respectively; (b) the shapes of the sample estimated in BU_3DFE, each image represents the result with a certain expression (angry, disgust, fear, happy, sad and surprise)

Fig.3 (b) shows the the error distribution over BU_3DFE. It is noted that the errors between the estimated and the manually annotated landmark points are less than 3.5 pixels. So, NPE can estimate shapes at varying poses well, but the errors are a litter bigger when estimating the shapes with arbitrary expressions.

From Fig.4 (b) , we can find that the estimated results with disgust, happy and surprise expressions are better than the other expressions. The estimated landmark points of eyebrows and eyes shift upward with the fear expression.

Table 1 tabulates the mean point-to-point error of the regression based learning methods and neighborhood-preserving estimation method over IMM and BU_3DFE. We can see that the proposed approach outperforms the regression methods in IMM. However the results on BU_3DFE dataset are not as good as the results on IMM. The error computed by NPE is only lower than DLR. To some extent, NPE is more suitable for estimating landmark points at arbitrary poses, but compared with regression based learning methods, its ability of estimating the landmark points under varying expressions is not good enough.

**Table 1.** Mean estimation errors obtained for the IMM and BU_3DFE

| Mean Error | DLR | FSLR | DNLR | FSNLR | NPE |
|---|---|---|---|---|---|
| IMM | 1.9895 | 1.4519 | 1.7362 | 1.5889 | 1.0195 |
| BU_3DFE | 2.3685 | 1.5883 | 1.7098 | 1.5731 | 1.9381 |

## 5 Discussion and Conclusion

An approach based on neighborhood-preserving has been proposed for simplifying the Active Appearance Model (AAM) fitting process. NPE makes use of the neighboring information of the testing face shape, which uses local linear regression to reflect the global non-linear information. The experiments mentioned above indicate that NPE can estimate the face shapes with varying poses well, while the results on BU_3DFE show NPE is not very suitable to estimate the face shapes with changing expressions. NPE views the face shapes as data points, which implies NPE does not take the local information of the face landmark points into account. The face shapes with varying expressions may be influenced greatly with the local landmark points, so we get poor relatively performance on BU_3DFE.

Future work will focus on exploring an algorithm that can model the subtle changes of the landmark points at arbitrary expressions and studying an approach which not only use the neighboring structure of the testing face shape, but also take the local information of the face shapes to get more accuracy results.

**Acknowledgments.** This work is partially supported by the Natural Science Foundation of China under grant No. 60973098, 61005005.

## References

1. Jebara, T.: Images as Bags of Pixels. In: Proc. IEEE International Conference on Computer Vision, New York, vol. 1, pp. 265–272 (2003)
2. Baker, S., Matthews, I., Schneider, J.: Automatic Construction of Active Appearance Models as an Image Coding Problem. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(10), 1380–1384 (2004)
3. Walker, K., Cootes, T., Taylor, C.: Automatically Building Appearance Models from Image Sequences Using Salient Features. Image and Visual Computing 20(5-6), 435–440 (2002)
4. Saragih, J., Goecke, R.: Learning Active Appearance Models from Image Sequences. In: HCSNet Workshop on the Use of Vision in HCI (VisHCI 2006), Conferences in Research and Practice in Information Technology (CRPIT), Canberra, pp. 51–60 (2006)

5. Asthana, A., Khwaja, A., Goecke, R.: Automatic Frontal Face Annotation and AAM Building for Arbitrary Expressions from a Single Frontal Image Only. In: IEEE International Conference on Image Processing, Cairo, pp. 2445–2448 (2009)
6. Asthana, A., Goecke, R., Quadrianto, N., Gedeon, T.: Learning Based Automatic Face Annotation for Arbitrary Poses and Expressions from Frontal Images Only. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1635–1642 (2009)
7. Asthana, A., Sanderson, C., Gedeon, T., Goecke, R.: Learning-based Face Synthesis for Pose-Robust Recognition from Single Image. In: Proceedings of the British Machine Vision Conference, Dublin, pp. 2–4 (2009)
8. Asthana, A., Lucey, S., Goecke, R.: Regression Based Automatic Face Annotation for Deformable Model Building. Pattern Recognition 44(10-11), 2598–2613 (2011)
9. Roweis, S.T., Saul, L.K.: Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science 290(5500), 2323–2326 (2000)
10. Nordstrom, M.M., Larsen, M., Sierakowski, J., Stegmann, M.B.: The IMM Face Database-An annotated Dataset of 240 Face Images. Informatics and Modeling, Technical University of Denmark, DTU (2004)
11. Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M.J.: A 3D Facial Expression Database for Facial Behavior research. In: The 7th International Conference on Automatic Face and Gesture Recognition (FG 2006), IEEE Computer Society TC PAMI, Southampton, pp. 211–216 (2006)