Head Pose-Based Conditional Regression Forest for Facial Feature Detection

Liyuan Zhuo, Huawei Pan, Chunming Gao

College of Information Science and Engineering, Hunan University, Changsha, 410082, China

Keywords: LPP; regression forests; facial feature point; global feature; head pose.

Abstract. Multi-angles of facial feature detection is still a challenging research. In this paper, the author proposes a precision head pose estimation method as a condition to improve the performance of regression forests, and decreases the missing rate caused by head deflection. The basic idea is used by locality preserving projection, a kind of manifold learning, and nonlinear regression (LPP+NLR) for getting the global information of pose and label it, then utilize trained conditional regression classifier to identify the feature points in global characteristics. The effectiveness of the proposed facial feature detection algorithm is illustrated in the experiments and the comparison with several recent methods.

Introduction

Facial feature point detection is an important step towards face analysis, crucial for the application of the analysis of facial data, face recognition and reconstruction, accurate and efficient facial feature detection is able to lay a good foundation for these applications. In practice, people's head will appear as a variety of gestures, the influence of the head pose is inevitable, therefore must take head pose account. Multi-angles of facial feature detection is always a challenging research.

Classic shape constraint method of active shape models (ASM) [1] and active appearance model (AAM) [2]. They are generated based on the principal component analysis method of fitting a linear model to obtain the shape changes and changes from the appearance of the face. Recent works attempt to build regression forests that are conditioned with some global or additional properties. Danton [3] propose a regression forests model conditioned on head pose for facial feature detection. They use the method that finely [4] proposed, training the regression forest with 3D face model dataset and getting the position of tip and head rotation angle, of the head pose estimation. Then correspond to the 2D image and divide the training set into subsets according to head pose yaw angle. However, the cost of training and testing is still relatively high, and the results of the experiment are not so good.

To solve this problem, we propose a method for facial feature detection based on head pose and regression forest. The head pose estimation algorithm which combined the locality preserving projection and nonlinear regression [5], the locality preserving projection for dimensionality reduction and nonlinear regression to map the data into linear space. We estimate the head pose with the results that get from nonlinear regression. Experiment shows that the method is able to estimate the head pose fast and robustness. Then treat the head gesture as a global condition, for training the conditional regression forests, which could estimate facial feature points in real-time.

Global Model of head Pose Estimation

Combined the locality preserving projection and nonlinear regression is mainly for classification all the face image, set face image as digital image sampled in high dimensional data space, and represent each head pose with feature in low dimensional manifold [5]. First, we put each face in the database project to a low dimensional space, marked with different position that can distinguish different orientation of the head effectively. Learning the dataset by a nonlinear model, for achieve the best fit with data that already have. When the test data come in, projected it to the model in low dimensional space in the corresponding position. Thus we find the mapping model of the face image in a high dimensional space to a lower dimensional space, generating a low dimensional

manifold structure. Finally, we use linear regression that with high efficiency to analyse these nonlinear mapping data, getting the label of the head pose.

Description of LPP and NLR:

Input: T-training data; t-testing data; I-label for training data; k-parameter for finding the neighbourhood; d-dimension of the low dimensional manifold.

Output: The head pose S of testing data;

S= LPP+NLR (T, t, l, k, d);

Step:

For each sample x_i in T

Using $\| x_i - x_j \|^2 < \varepsilon$ to find k nearest neighbours $x_j (j = 1, 2, ..., k)$ of x_i ;

End For;

Compute the weight matrix W where $W_{ij} = e^{-\frac{\|x_i - x_j\|^2}{t}}$;

Compute the projection matrix, to embed high dimensional samples into the low dimensional manifold;

The problem of solving the projection matrix could be transformed into formula (1) for eigenvalue, corresponding d eigenvectors of eigenvalue constitute the projection matrix E for achieving embed high dimensional samples into the low dimensional manifold

$$XLX^T a = \lambda XDX^T a \tag{1}$$

Where X is a matrix constituted by samples x_i , L indicates Laplace matrix and $L = D - W$. D Is diagonal matrix, and $D_{ii} = \sum_j W_{ji}$.

Normalize the projection matrix E that computed from (3);

For (the maximum distance between the input vector and the initial value σ ,

$$x \rightarrow S = f(x, \sigma)) \tag{2}$$

According to criterion of the least square error of train data;

End for;

Mapped the test samples into a low dimensional that linear and separable space through the projection matrix E , get the data of low dimensional.

Using linear regression for head pose estimation;

For facial feature detection, this paper based on the global property of the head pose that obtained from LPP and NLR, and then trained conditional regression forest to get the precision facial points.

Conditional Regression Model

Base on random regression forest[6], for each type of head pose training conditional regression forest, and then select the corresponding trees from the training to estimate facial feature points. Conditional regression forests aim to learn several conditional probabilities over the parameter space instead. Training the forests after classified by head pose respectively, so that trees do without dealing with all face changes in appearance and shape. During testing, select the corresponding conditional forests via head pose.

We build the trees following the random forest framework [7]. Each tree T in the forest $T = \{T_F\}$ is built from a different, randomly selected, set of training images. For each image, we randomly extract a set of patches $\{P_i = (I_i, D_i)\}$. Where I_i is defined by multiple channels $I_i = (I_i^1, I_i^2, ..., I_i^f)$, represents the appearance, $D_i = (d_i^1, d_i^2, ..., d_i^N)$ represents the set of offsets to each facial feature point. The test splits the set of patches P into two subsets: When a patch satisfies the test it is passed to the right child, otherwise, the patch is sent to the left child. The binary test at a non-leaf node is defined as $t_{f, R_1, R_2, \tau}(I)$:

$$| R_1 |^{-1} \sum_{q \in R_1} I^a(q) - | R_2 |^{-1} \sum_{q \in R_2} I^a(q) > \tau, \tag{3}$$

Where I^a indicates the feature channel $a \in \{1, 2, ..., f\}$ R_1 and R_2 are two rectangles within the patch boundaries, and τ is a threshold, f, R_1, R_2, τ are generated with random values.

A regression forest aims to model the probability $p(d^n | P)$, given an image patch P and the conditional probability $p(d^n | S, P)$ modelled by a conditional regression forest, then estimates by

$$p(d^n | P) = \int p(d^n | S, P) p(S | P) dS \tag{4}$$

Where S a global property is that consistent with the head pose, estimated from the image, as described in Section 2.

In order to learn $p(d^n | S, P)$, we split the training set into subsets, and the space of the parameter S is discretized into respective sets S_{pose}. To this end, we quantize the training data into 5 subsets that correspond to 'left profile', 'left', 'front', 'right, and 'right profile' faces. Thus (3) could be

$$p(d^n | P) = \sum_i \left(p(d^n | S_{pose}, P) \int_{S \in S_{pose}} p(S | P) dS \right) \tag{5}$$

Where $p(d^n | S_{pose})$ indicates the conditional probability that can be learned by training a full regression forest $T(S_{pose})$. While the probability $p(\omega | P)$ can be learned by a regression forest on the full training set Ω.

Experiments

The head pose estimation is trained on the Facepix [8] database, contains 30 different identities colour image. We normalized each image in our experiment, constitute 32×32 dimensional data. Linear dimensionality reduction methods commonly used principal component analysis (PCA) and linear discriminant analysis (LDA) [9]. The experiment select local embedding (LEA) and locally linear embedding (LLE) [10]. Fig. 1 shows our method compares with methods mentioned above. The experiment selects random samples, different data of facial feature affects the result. The LPP with NLR method can maintain stability, holding low error and volatility. As the above analysis show that the LPP+NLR method combined the advantages of manifold learning and nonlinear regression, guarantee either the accuracy or the high robustness.
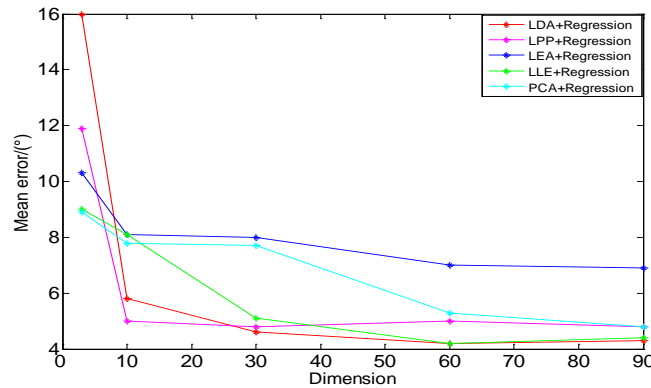

Fig. 1: Mean error of varies manifold learning with nonlinear regression.

Regression forests are trained on the professional facial database on the challenging Labelled Face Parts in the Wild (LFPW) [11], which contain large variations in the imaging conditions. We annotated 13000 faces taken from LFPW database with the location of 14 facial feature points.

Regression forests provide two parameters to balance the runtime and accuracy, namely the sampling stride, as shown in Fig.2. The stride parameter is crucial for real-time performance: a stride greater or equal to 3 already allows for over 10 fps (i.e., below 100 ms for one frame) at a marginal loss in accuracy.
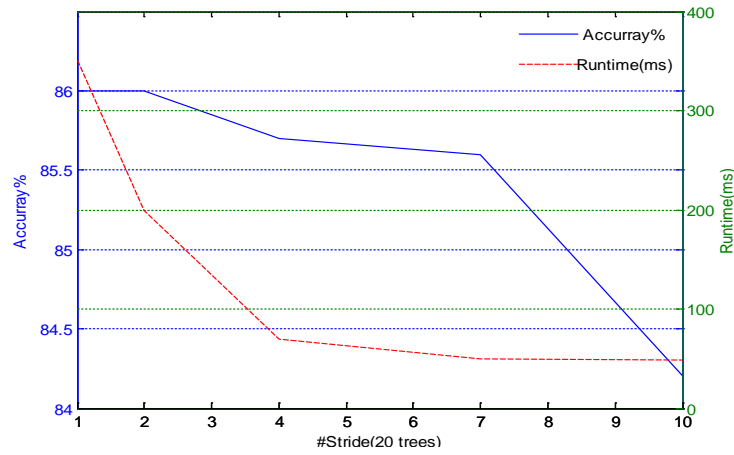
Fig. 2: Trade-off between runtime and accuracy with stride.

In order to accurately assess the performance of our system, we performed a ten-fold cross validation experiment. We compare our results to two state-of-the-art methods [12-13]. Our method clearly outperforms competitors with respect to accuracy. The error for facial feature point is given in Fig. 3.
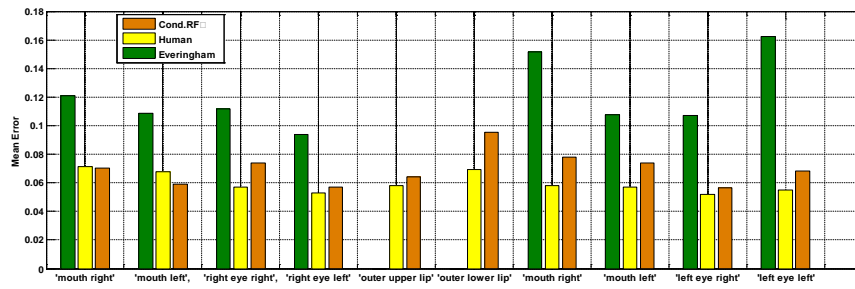

Figure 3.Comparison with other methods of mean error.

Fig. 4: Images of real-time facial feature detection.

Fig. 4 shows the facial feature points images that captured in real-time during experiments. The results prove that our method holds robust with all kinds of angles, and remains good effects for some occlusion.

Summary

In this paper, we present a real-time algorithm for facial feature detection based the improved conditional regression forests. Specifically, we introduce a face shape prior model to capture the face shape patterns under varying head pose based on LPP+NLR. We then extend the frontal face prior model by a conditional random regression forests under pose variation. In the future, we will perform further validation of our methods to model other properties like sunglasses or facial hair that still cause some problems as well.

References

[1] T. F. Cootes, C. J. Taylor, D. H. Cooper. Active shape models their training and application. Computer Vision and Image Understanding, 61(1):38–59, 1995.

[2] I. Matthews and S. Baker. Active appearance models revisited. Int. J. Compute. Vision, 60 (2): 135–164, 2004.

[3] Dantone, M., Gall, J.: Real-time facial feature detection using conditional regression forests. In: CVPR. (2012)

[4]G. Fanelli, T. Weise, J. Gall, and L. Van Gool. Real time head pose estimation from consumer depth cameras. DAGM, 2011.

[5] CHEN Shuming, CHEN Duan-sheng. Head pose estimation based on improved LPP [J]. Computer Engineering and Design, 32 (12): 4218-4222, 2011.

[6] 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.

[7] G. Fanelli, T. Weise, J. Gall, and L. Van Goal. Real time head pose estimation from consumer depth cameras. DAGM, 2011.

[8] Black J A, Gargesha M, ET al. A framework for performance evaluation of face recognition algorithms [C] / / Pros. SPIE. Internet Multimedia Management Systems III. Boston, USA: SPICE, 2002: 163-174.

[9] Jung Dy, Lee SM. Fault detection method with PCA and LDA and its application to induction motor. J. Cent. South Univ. Technol. (2010) 17: 1238-1242;

[10] FuY, Huang TS. Locally linear embedded Eigen space analysis, IFP-TR[R]. Illinois at Urbana-Champaign, 2005.

[11] G. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, University of Massachusetts, Amherst, 2007.

[12] M. Everingham, J. Sivic, and A. Zisserman. Hello! My name is... buffy - automatic naming of characters in TV video. In BMVC, 2006.

[13] M. Val star, B. Martinez, X. Benefit, and M. Pantic. Facial point detection using boosted regression and graph models. In CVPR, pages 2729–2736, 2010.