http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 13 (13): 2159-2164, 2014
ISSN 1812-5638 / DOI: 10.3923/itj.2014.2159.2164
© 2014 Asian Network for Scientific Information

An Improved Conditional Regression Forests for Facial Feature Points Detection

Xukang Wang, Guanghua Tan and Chunming Gao
College of Information Science and Engineering, Hunan University, Changsha, 410082, China

Abstract: In order to improve the detection accuracy of facial feature points even on deformed and low quality image. This study proposed a new method which combined the constrained local model and conditional regression forests voting. Firstly, the method used the constrained local model to build head pose as global characteristic. This could provide holistic constraint for facial feature points detection. Then conditional regression forests voting was used to train the decision trees which express the relationship between facial image patches and the location of feature points. The resulted decision trees, together with the global face characteristic trained by constrained local model, could be used to fastly and accurately cast votes for the optimal facial feature position. Experimental results show that the algorithm can mark the facial feature points quickly and robustly. The accuracy is improved by 6% in general compared to other methods for facial feature detection.

Key words: Facial feature detection, constrained local model, conditional regression forests voting, statistical shape model

INTRODUCTION

Facial feature detection is a widely-used application in image processing field, especially as a preprocessing step for many applications like human computer interaction or 3D reconstruction, face recognition (Belhumeur *et al*., 2011; Amberg and Vetter, 2011; Valstar *et al*., 2010; Rapp *et al*., 2011). It can be divided into two categories: holistic feature and local feature. Scholars put forward a series of prediction models in this field, one kind of holistic models is based on statistical model which match the shape model with the input image feature to get the prediction results, the typical method includes Active Shape Models (Cootes and Taylor, 1992; Cootes *et al*., 1995; Milborrow and Nicolls, 2008), using a linear Point Distribution Model (PDM) built from alignment training model and the shape model is fit to the results of searching through a proper training. Active Appearance Models (AAM) (Cootes *et al*., 2001; Cootes *et al*., 2002) match is to use an effective parameter updating scheme to combine the relationship between models of shape and texture information and generating the texture characteristics of face region match a linear model to a test image.

Recently, regression forests voting (Cootes *et al*., 2001; Breiman, 2001) is a widely used and efficiently method for facial feature extracting. In this respect, regression forests can learn a mapping from local image or depth patches to a probability over the parameter space, such as the 2D position or orientation rotating. While related Hough forests (Gall *et al*., 2011) has shown that objects (Girshick *et al*., 2011; Fanelli *et al*., 2011) can be effectively located in 2D image by pooling votes from random forests regressors. Valstar *et al*. (2010) have shown that facial feature points can be accurately located using kernel SVM based regressors.

Study present a constrained local model with conditional regression forests voting that accurately detects 2D facial feature points. As subtle deformation can affect shape around feature points and the accuracy of feature points detection. In order to control the effect, study apply the local constrained model framework to face for constraints of facial local situation which can obtain the accurate facial feature point combined with the method of conditional regression forests voting. Since conditional regression forests acquire the probabilities satisfied several conditions from parameter space. The motivation of algorithm is that conditional probabilities are easier to learn since the trees do not have to deal with all facial variation in appearance and shape than the concept of regression forests voting. It also can be used to quickly and accurately cast votes for the optimal facial feature position when combined with a global face characteristics trained by constrained local model.

METHODOLOGY

Here, we will describe the method of constrained local models and random regression forests.

Corresponding Author: Guanghua Tan, School of Information Science and Engineering, Hunan University, Changsha, 410082, People's Republic of China

Constrained local model: The Constrained Local Model (CLM) is an approach for building the points of a global statistical shape model. Here, we only list the key points of approach. One can refer to literature (Cristinacce and Cootes, 2006; Cristinacce and Cootes, 2008) for the details.

In this method, the shape variation model is formulated as:

$$x_i = T(\bar{x}_i + P_i b; t) \quad (1)$$

where, $\bar{x}_i$ is the mean position of the point in a suitable reference frame. P_i is a set of models of variation and T(.;t) applies a global transformation with parameters t.

Actually, this method seek parameters p = {b, t} which minimise:

$$Q(p) = -\log p(b,t \mid I) = -\log p(b) - \alpha \sum_{i=1}^{N} \log p(x_i \mid I) \quad (2)$$

where, the scaling factor α is included to take account of the fact that the conditional probabilities for each point, $p(x_i, |I)$ are not strictly independent and all poses are equally likely, so, p(b, t) = p(b).

Given an estimate of the scale and orientation, in order to compute quality of fit $C_i(x_i) = -\log p_i(x_i|I)$ with scanning the global shape model. The objective function is then:

$$Q(p) = -\log p(b) + \alpha \sum_{i=1}^{N} C_i(x_i) \quad (3)$$

The cost function Q(p) can be optimized either using the mean-shift approach advocated by saragih. For each feature point location, algorithm describe the precise location of CLM model tag feature points using conditional regression forests voting which are on the basis of considering global variable.

Random regression forests: Random regression forests have been used for a large number of classification and regression tasks. We outline the training and testing of a random regression forest for detecting facial feature point in 2D images.

Training random forests includes 4 steps (Breiman, 2001):

- Generate a pool of splitting candidates $\phi = \{\theta, \tau\}$
- Divide the set of patches P into two subsets P_L and P_R for each ϕ:

$$P_L(\phi) = \{P / f_\theta(P) < \tau\} \quad (4)$$

$$P_L(\phi) = P / P_L(\phi) \quad (5)$$

- Select the splitting candidate ϕ which maximizes the evaluation function information gain:

$$\phi^* = \arg\max IG(\phi) \quad (6)$$

$$IG(\phi) = H(P) - \sum_{S \in \{L,R\}} \frac{|P_S(\phi)|}{|P|} H(P_S(\phi)) \quad (7)$$

where, H(P) is the defined class uncertainty measure. Selecting a certain split amounts to adding a binary decision node to the tree

- Create leaf l when a maximum depth is reached or the Information Gain (IG) is below a predefined threshold. Otherwise continue recursively for the two subsets $P_L(\phi)$ and $P_R(\phi)$ at first step

In regard to test of random forest regression forests, we initially run a face detection algorithm to find the position and the size of the face. After enlarging the bounding box of the face and rescaling the image to a common size, we densely sample patches $P_i(y_i)$ inside the bounding box, where y_i is the pixel location of the patch P_i.

Considering the Gaussian Kernel K and the bandwidth parameter h, the density estimator for facial feature point n at pixel location x^n can be written as:

$$f(x^n) \propto \sum_i \sum_{l \in L_i} w_l^n K\left(\frac{x^n - (y_i + \bar{d}_l^n)}{h}\right) \phi_n(l) \quad (8)$$

$$\phi_n(l) = \begin{cases} 1, p(c_n \mid l) \geq \alpha \\ 0, \text{otherwise} \end{cases} \quad (9)$$

where, w^n_l is the confidence weight of the leaf l. the factor ϕ_n avoid a bias towards an average face configuration. In order to reduce the influence of votes coming from other parts of the face and to improve the efficiency, we consider only leafs with a class-affiliation higher than α. The facial feature points are obtained by performing mean-shift for each point n.

METHOD OF FACIAL FEATURE POINTS DETECTION

This method begins with selecting image annotated with facial feature points from Labeled Faces Parts in the Wild (LFPW) database, then establishes head pose as the global characteristics with constrained local model and estimate the precise location of facial feature point using conditional regression forests voting. According to the

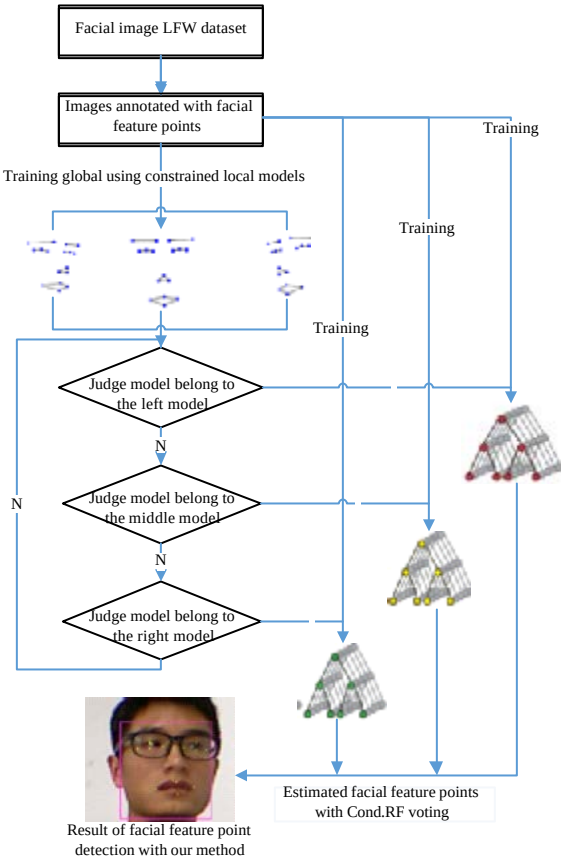

Fig. 1: Flow chart of facial feature points detection

global face properties learned by CLM, the conditional regression forests can easily and quickly learn the relationship between facial image patches and the location of feature points. We can accurately detect facial feature points when combined the two methods. The chart flow of facial feature points detection illustrated by Fig. 1.

Conditional regression forests: While a regression forest aims to model the probability $p(d^n|P)$ given an image patch P, a conditional regression forest models the conditional probability $p(d^n|\omega, P)$ and estimates by:

$$p(d^n \mid P) = \int p(d^n \mid \omega, P)p(w \mid P)d\omega \tag{10}$$

where, ω is an auxiliary parameter that can be estimated from the image. In algorithm, ω corresponds to head pose as the global characteristics that can be estimated by constrained local models.

In order to learn $p(d^n|\omega, P)$, the training set is split into subsets, where the space of the parameter ω is discretized into disjoint sets Ω_i. hence, (10) becomes:

$$p(d^n \mid P) = \sum_i (p(d^n \mid \Omega_i, P)\int_{\omega\in\Omega_i} p(\omega \mid P)d\omega) \tag{11}$$

The conditional probability $p(d^n|\Omega_i, P)$ can be learned by training a full regression forest $T(\Omega_i)$ on each of the training subsets Ω_i. Similarly, the probability $p(\omega|P)$ can be learned by a regression forest on the full training set Ω.

While regression forests average the probabilities over all tree T_t and select T trees from the conditional regression forests $T(\Omega_i)$ based on the estimated probability $p(\omega| P)$. to this end:

$$p(d^n \mid P) = \frac{1}{T}\sum_i \sum_{t=1}^{k_i} p(d^n \mid l_t, \Omega_i(P)) \tag{12}$$

where, $l_t \Omega_i$ is the corresponding leaf for patch P of the tree $T_t \in T(\Omega_i)$.

Global properties estimation: To get the global properties, we learn head pose with the method of CLM. To address this problem, we quantize the image data into 3 subsets that correspond to left profile, front and right profile face since it is difficult to get continuous ground truth head pose data from 2D images. As for facial feature points detection, we rescale the faces based on the face detection result. Since the face bounding box is not always perfect, we train trees which can be used to predict the head pose. We use:

$$\left|Q_{Headpose} = -\log p(b) - \alpha\sum_c p(c|P)\log(c|P)\right) \tag{13}$$

as cost function (3). For facial feature detection, we can choose if c corresponds to the labels of head pose class affiliation. In experiment, we replace the head pose labels by real world angles $\omega \in \{-45,0,+45\}$ representing the yaw angle. We can achieve more robust facial feature estimation with head pose as global properties.

EXPERIENTAL RESULTS

This system runs on a PC with an Intel Core i5 (2.6 GHz) CPU and an ordinary web camera which records 640×480 resolution at 30 fps. The regression forests is trained with the images from LFPW (The labeled face parts in the wild) facial database (Huang *et al*., 2007; Kostinger *et al*., 2011). The images in this database have been collected in the wild and vary in pose, lighting conditions, resolution, quality, expression gender, race,

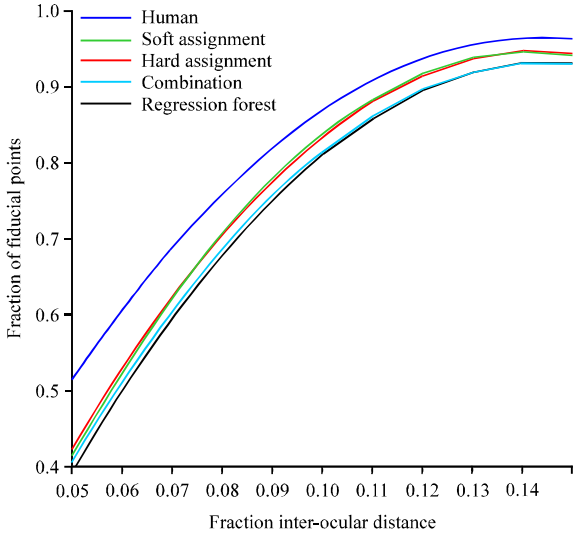

Fig. 2: Performance of soft and hard assignment compared to human performance and to a standard regression forest

occlusion and make-up-13000 faces taken from LFPW database have been annotated with the location of 12 facial feature points.

During the training process, some parameters are fixed according to the empirical observation: The maximum depth of the tree is 20, each mode can randomly generate 3000 splitting candidates and 30 thresholds. Every tree is selected based on the random image collection of 2000 training.

Experiments compared three methods of selecting T trees for the conditional regression forest: Soft assignment, hard assignment and combination of both soft and hard assignment. Soft assignment selects trees conditional to the estimated global features probability while hard assignment selects only trees for the global features with the highest probability. An additional naive approach (combination) randomly selects the trees without taking the estimated global feature into account. Figure 2 shows the results: Soft and hard assignment performs similarly and both methods outperform the standard regression method which is close to the performance of the Human.

Regression forests provide two parameters to balance the running time and accuracy, the number of trees to be loaded and the sampling stride, as shown in Fig. 3. The figure indicates that a higher number of trees and a lower stride improve the accuracy at the cost of a higher average computation time.

In order to evaluate the accuracy of the system, we performed a ten-fold cross validation experiment on the same database. The method is compared with two state of the art methods, Valstar and Everingham

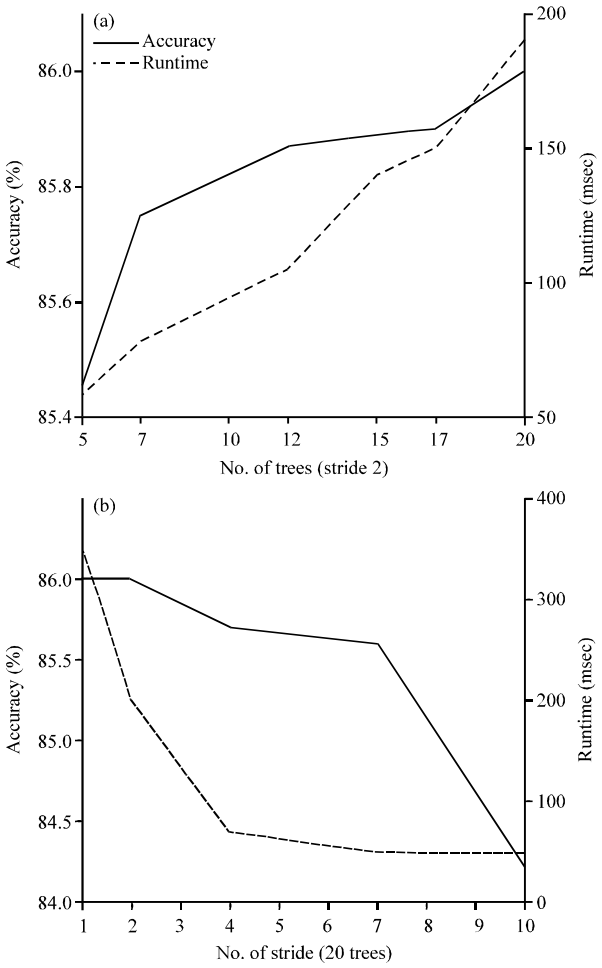

Fig. 3(a-b): (a) Trade-off between runtime and accuracy depend on No. of trees and (b) Trade-off between runtime and accuracy depend on stride

Table 1: Accuracy and mean error of facial feature points compared to the method of Valstar

Facial feature	Cond. RF Accuracy (%)	Cond. RF Mean error	Valstar Mean error
Lower outer lib	75.1	0.0973	-
Upper outer lib	86.6	0.0740	-
Right eye up	88.2	0.0865	0.1108
Left eye up	90.4	0.0692	0.1185
Mouth right	80.4	0.0787	0.1441
Mouth left	81.9	0.0798	0.1167
Right eye right	86.2	0.0786	0.1061
Right eye left	93.1	0.0665	0.0973
Left eye right	93.5	0.0667	0.1007
Left eye left	87.8	0.0692	0.1261

separately. The accuracy of each method is illustrated in Fig. 4. From this figure we can see that our method clearly outperforms both competitors with respect to accuracy.

Compared to human performance, experiment acquire accuracy of some facial feature point in Table 1. Figure 5

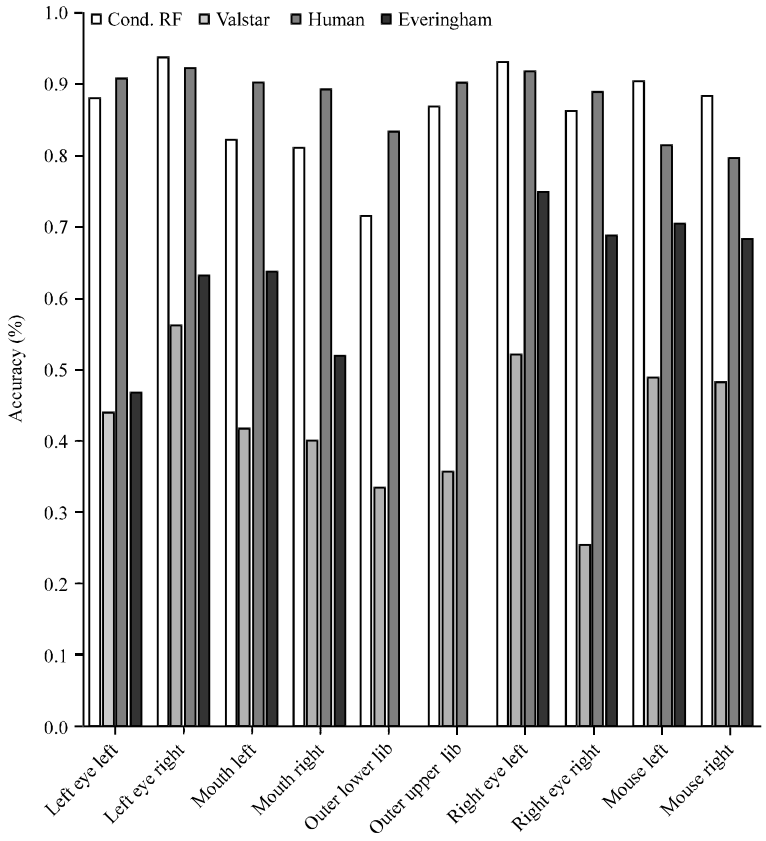

Fig. 4: Accuracy of facial feature points detection using conditional regression forests comparison to other methods

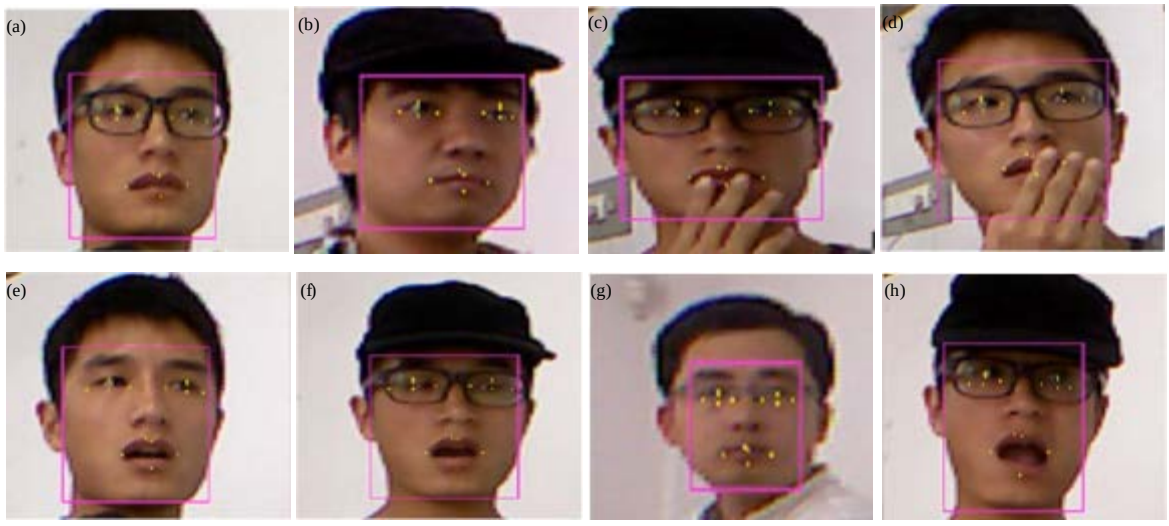

Fig. 5(a-h): Qualitative facial feature points detection results on some images from researchers and me in my labor. (a) Common expression, (b) Covered with hat, (c) Covered with hat, glasses and hand, (d) Covered with glasses and hand, (e) With opening mouth, (f) Covered with hat, glasses and open mouth, (g) Error in detection eyes feature points and (h) Error in detection feature points with surprised expression

shows some qualitative results. From this figure, we can see that the most difficult point to detect is the lower lip and eyes.

CONCLUSION

Study present an improved conditional regression forests for facial feature detection. The experiments show that this method can give robust, real-time detection results compared favorably to related techniques. The further work is to model other properties such as glasses or facial hair which may cause some problems in facial feature detection. Besides,we also plan to try the method of conditional regression forests in other relevant computer vision applications.

ACKNOWLEDGMENTS

This study is supported by projects of Human Province Science and Technology Plans (2013SK3310): Format research and demonstration for "red culture" experience ecotourist in shao shan and the Fundamental Research Funds for the Central Universities.

REFERENCES

Amberg, T. and B. Vetter, 2011. Optimal landmark detection using shape models and branch and bound. Proceedings of the International Conference on Computer Vision, June 2011, Barcelona, Spain.

Belhumeur, P.N., D.W. Jacobs, D. Kriegman and N. Kumar, 2011. Localizing parts of faces using a consensus of exemplars. Proceedings of the International Conference on Computer Vision and Pattern Recognition, June 20-25, 2011, Providence, RI., pp: 545-552.

Breiman, L., 2001. Random forests. Mach. Learn. J., 45: 5-32.

Cootes, T.F. and C.J. Taylor, 1992. Active shape models-smart snakes. Proceedings of the British Machine Vision Conference, September 22-24, 1992, Leeds, UK.

Cootes, T.F., C.J. Taylor, D.H. Cooper and J. Graham, 1995. Active shape models-their training and applications. Comput. Vis. Image Understanding, 61: 38-59.

Cootes, T.F, G.J. Edwards and C.J. Taylor, 2001. Active appearance models. Trans. Pattern Anal. Machine Intell., 23: 681-685.

Cootes, T.F., G.V. Wheeler, K.N. Walker and C.J. Taylor, 2002. View-based active appearance models. Image Vision Comput., 20: 657-664.

Cristinacce, D. and T. Cootes, 2006. Feature detection and tracking with constrained local models. Proceedings of the British Machine Vision Conference, September 4-7, 2006, Edinburgh, UK.

Cristinacce, D. and T. Cootes, 2008. Automatic feature localisation with constrained local models. Pattern Recog., 41: 3054-3067.

Fanelli, G., J. Gall and L. van Gool, 2011. Real time head pose estimation with random regression forest. Proceedings of the Conference on Computer Vision and Pattern Recognition, June 20-25, 2011, Providence, RI., pp: 617-624.

Gall, J., A. Yao, N. Razavi, L. van Gool and V. Lempit-Sky, 2011. Hough forests for object detection, tracking and action recognition. Patt. Anal. Machine Intelli. Trans., 33: 2188-2202.

Girshick, R., J. Shotton, P. Kohli, A. Criminisi and A. Fitzgibbon, 2011. Efficient regression of general-activity human poses from depth images. Proceedings of the International Conference on Computer Vision, November 6-13, 2011, Beijing, China.

Huang, G.B., M. Ramesh, T. Berg and E. Learned-Miller, 2007. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report. University of Massachusetts, Amherst, pp: 1-14. http://tamaraberg.com/papers/Huang_eccv2008-lfw.pdf

Kostinger, M., P. Wohlhart, P.M. Roth and H. Bischof, 2011. Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. Proceedings of the International Conference on Computer Vision Workshops, November 6-13, 2011, Barcelona, pp: 2144-2151.

Milborrow, S. and F. Nicolls, 2008. Locating Facial Features with an Extended Active Shape Model. In: Computer Vision ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part IV, Forsyth, D., P. Torr and A. Zisserman (Eds.). Vol. 5305. Springer, Berlin, Heidelberg, ISBN: 978-3-540-88692-1, pp: 504-513.

Rapp, V., T. Senechal, K. Bailly and L. Prevost, 2011. Multiple kernel learning SVM and statistical validation for facial landmark detection. Proceedings of the International Conference on Automatic Face and Gesture Recognition, March 21-25, 2011, Santa Barbara, CA., pp: 265-271.

Valstar, M., B. Martinez, X. Binefa and M. Pantic, 2010. Facial point detection using boosted regression and graph models. Proceedings of the International Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, CA., pp: 2029-2736.