Neurocomputing 214 (2016) 881–893

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Robust discriminative regression for facial landmark localization under occlusion

Yanming Wang [a], Jiguang Yue [a], Yanchao Dong [a,*], Zhencheng Hu [b]

[a] School of Electronics and Information, Tongji University, No. 4800 Cao'an Road, Shanghai, China
[b] Department of Computer Science, Kumamoto University, No. 2-40-1 Kurokami Chuo-ku, Kumamoto, Japan

ARTICLE INFO

Article history:
Received 27 April 2015
Received in revised form
25 June 2016
Accepted 14 July 2016
Communicated by Zidong Wang
Available online 21 July 2016

Keywords:
Robust discriminative regression
Facial landmark localization
Discriminative shape regression
Facial alignment

ABSTRACT

Facial landmark localization or facial alignment is a crucial initial step in face analysis. The paper proposes a novel discriminative regression framework called Robust Discriminative Regression (RDR) for facial landmark localization. RDR framework consists of multiple partial feature regressors and a regression tree combination strategy. The proposed method copes with the partial facial landmarks invisible problem together with the optimization problem of multiple outputs combination. The RDR framework can be applied to both raw shape regression and model-based shape parameters regression. In model-based shape parameters regression we propose a two-level regression strategy, the first level is for rigid motion parameter regression and the second one is for non-rigid deformation parameter regression. Experiments on three widely used "face in-the-wild" databases (LFPW, COFW and IBUG) show that the proposed RDR outperforms other state-of-the-art facial landmark localization strategies in raw shape regression especially under partial occlusions or large pose variations. It also shows that the two-level regression strategy within RDR framework could achieve better performance than one-level parameters regression.

1. Introduction

Human face analysis has a wide range of applications in human-computer interaction, such as expression recognition, face identification, driver intent prediction etc. Facial landmark localization or facial alignment is a crucial initial step in face analysis. The accuracy of landmark localization directly influences the performance of the whole application. Thus, an efficient and robust facial landmark localization method is urged, especially under uncontrolled environments with large appearance variations.

In the controlled environment the facial landmark localization can be solved by canonical methods, such as Active Shape Models (ASM) [1], Active Appearance Models (AAM) [2] or Constrained Local Models (CLM) [3]. In the wild environment, the discriminative regression method for facial landmark localization has gained approval due to its good accuracy and real-time performance [4–12]. Discriminative regression may map the extracted features to the raw shape update as in [6–8], or to the shape parameter update in the model-based shape parameter regression as in [10]. Given a test image with an initial shape, Discriminative Shape Regression (DSR) uses a cascade of pre-learned regression functions to update the shape stage by stage. The success of DSR mainly lies on the following properties: 1) it utilizes the shape-indexed features in each stage, which computes the features with respect to the estimated landmarks location of the last stage. This kind of feature mapping methodology copes with the effect of large appearance variations and enhances the robustness and accuracy of regression; 2) gradient boosting framework is adopted in the training process. In each stage, a regression function is learned using the previous estimated shape error and shape-indexed features. Thus the output error in each stage decreases monotonically and converges in 4 to 5 stages [7]; 3) the regression compute a linear combination of training shapes as output, which guarantees a reasonable face shape inherently.

Though the discriminative regression methods make significant progress for facial landmark localization, its performance degrades dramatically when partial landmarks become invisible because of occlusions or large pose variation. Moreover, most of the discriminative regression methods utilize multiple initial shapes and chose the mean or median of the multiple output shapes as the final output, which is sensitive to outliers. Seldom of them try to find the optimal method for combining the multiple output shapes. Several DSR variants have been implemented with different feature mapping functions or regression functions. However, it needs to do cross validation among the feature mapping functions and regression functions to discover the most efficient implementation.

* Corresponding author.
E-mail address: dongyanchao@tongji.edu.cn (Y. Dong).

http://dx.doi.org/10.1016/j.neucom.2016.07.025
0925-2312/

The paper proposes a novel Robust Discriminative Regression (RDR) framework for facial landmark localization to solve the partial occlusion problem and the optimization problem of multiple-output combination. The proposed RDR inherits all the advantages of DSR. The multiple partial feature regressors are utilized to solve the partial occlusion problem. And the multiple-output combination deals with the problem of combining the multi-shape estimates into a single solution. The major contributions of the paper are:

1. The proposed RDR framework handles the partial occlusion problem together with the optimal multiple output combination problem.
2. In model-based shape parameter regression we propose a two-level iteration strategy within RDR framework. The first level is for rigid motion parameter regression and the second one is for non-rigid deformation parameter regression.
3. A thoroughly inspection on different implementations of the basic DSR method is conducted. It is found that DSR implementation with SIFT feature and linear regression obtains the balance of accuracy and computation efficiency.

Experiments on three widely used “face in the wild” databases (LFPW, COFW and IBUG) show that our RDR framework outperforms the state-of-the-art face alignment methods especially under partial occlusions or large pose variations.

The arrangement of the paper is as follows: Section 2 reviews the existing methods for facial landmark localization; Section 3 describes of our Robust Discriminative Regression; Section 4 presents the experimental validation of the proposed RDR framework for facial landmark localization based on three “face in the wild” datasets; Finally, Section 5 makes a conclusion and gives the future work to do.

2. Related work

Active Shape Models (ASM) [1] and Active Appearance Models (AAM) [2] are the canonical methods for facial landmark localization. These kinds of model-based methods have been widely studied in the past decades. In ASM, Procrustes Analysis [13] is first applied to the training shapes to remove similarity transformations, and then Principal Component Analysis (PCA) is utilized to obtain a shape model, which is defined by mean shape and shape eigenvectors. The boundary information is often used to fit a new image in ASM. AAM is an extension of ASM, which contains both the shape statistical model and appearance statistical model. The appearance statistical model is obtained by warping each sample image using a triangulation algorithm to a mean shape, then the pixel information is sampled from the entire region covered by the mean shape, and PCA is utilized to obtain the mean appearance and the eigenvectors. The fitting of AAM can be viewed as an optimization process and Gauss-Newton iterative algorithm is often adopted to solve this problem. Some variants have been proposed in the past years to improve the performance of ASM and AAM [14–17]. However, most of fitting algorithms are sensitive to pose and illumination changes. The initial location in the test image also has a significant impact on the accuracy of ASM or AAM, which makes the algorithm less robustness. Explicit geometrical constraints [18,19] have been used to tackle pose variations. Sukno et al. [18] proposed a projective ASM model for facial landmark detection. Based on the assumption that most facial features lie approximately on the same plane, the projective transformation matrix can be obtained firstly, then the ASM model constructed with frontal-view images can be directly applied to other viewpoints with the obtained projective transformation matrix. Fan et al. [19] introduced a novel projective invariant feature, named characteristic number (CN), which unifies the collinearity, cross ratio, and geometrical characteristics. With these constraints, the facial landmark can be obtained by standard gradient descent.

Constrained Local Models (CLM) [3] utilizes the same shape model as ASM and AAM, excepting that the appearance model in CLM is constructed using local facial landmark features instead of the entire face region features. In addition, the appearance model in CLM is used to generate likely feature templates instead of approximating the image pixels directly. When fitting to a test image, a response image is generated by the joint model and the current parameter values. Then Nelder-Mead simplex algorithm [20] is applied to adjust the shape parameters in order to maximize the sum of response at each landmark in the response image. This optimization process continues until the parameters converged. Wang et al. [21] proposed an enforcing convexity strategy at each local patch response surface to optimize a global warp updating in an efficient way. This local patch response was obtained by linear SVM and the parametric vector was updated using convex quadratic curve fitting method. Lucey et al. [22] utilized a simplified optimization criterion to fit the CLM, which divided the problem of finding a single complex registration/warp displacement into finding multiple simple warp displacement. The simplified optimization problem can be solved in parallel. Saragih et al. [23] pointed that the ambiguity of landmark detections in CLM could be reduced by posing a constraint on joint motion. The non-parametric mean-shift approach was applied to all landmarks simultaneously to impose a global prior over the joint motion.

Discriminative Shape Regression (DSR) has been widely used for facial landmark localization in recent years due to its accurate and real-time performance. The feature mapping and regressor learning are the two main stages of DSR. These two stages are iterated in a gradient boosting framework until converged. The feature mapping function may be global features such as pixel-difference and Haar-like features on the entire face region [5,6,9,24–26], or may be local features such as SIFT and local binary features with respect to the landmarks [7,8,10]. The regression function can be Ferns regression [5,6], regression trees [9,24], regession forest [25] or linear regression [7,8,10]. Cao et al. [5] utilized a correlation-based feature selection method to better explore the huge feature space in a short time and generate good candidate ferns. Kazemi et al. [9] introduced an exponential prior over the distance between the pixels used in a split to courage closer pixel pairs to be chosen. This prior could reduce the prediction error in testing dataset. Burgos-Artizzu et al. [6] proposed a smart restarts in regression and introduced a probabilistic method for occlusion detection. However, this method needs the training database contains visibility information for each landmark. Ren et al. [8] first learnt a local feature mapping function (regression random forest) to generate local binary features for each landmark, and then a global linear regressor was learned using the concatenate local features. Asthana et al. [10] proposed an incremental linear regressor for face alignment but it cannot run in real-time. A low rank driven facial landmark regression method was proposed by Deng et al. [12] to improve the performance of landmark detection. Firstly, the face was frontalized by low-rank matrix recovery algorithm, and then the sparse shape constrained cascade regression model was proposed to simultaneously suppress the ambiguity in local features and outlier caused by occlusion. A shape dictionary was constructed in the training process to code the face shape however, when the number of training data is very large, it is infeasible to simply stack all shapes into the data matrix since sparse shape composition could not handle them efficiently. Ge et al. [27] proposed a joint local regressors learning method for face alignment, different regression parameters were