CASCADE SUPPORT VECTOR REGRESSION-BASED FACIAL EXPRESSION-AWARE FACE FRONTALIZATION

Yiming Wang[1], *Hui Yu*[1*], *Junyu Dong*[2], *Muwei Jian*[3], *Honghai Liu*[1]

[1]University of Portsmouth, Portsmouth, United Kingdom
[2]Ocean University of China, Qingdao, China
[3]Shandong University of Finance and Economics, Jinan, China

ABSTRACT

The main aim of face frontalization is to synthesize the frontal facial appearances from non-frontal facial images. How to estimate the frontal face-shape is a crucial but very challenging problem in the frontalization task. Most existing methods use a single shape template to fit in with frontal facial appearances, which will result in a loss of expression-related information. In this work, we present a novel facial expression-aware face frontalization method which directly learns the pair-wise relations between non-frontal face-shape and its frontal counterpart. The support vector regression is explored to train the pair-wise regression model. Considered the pair-wise relationship is non-linear, an appropriate cascade manner is applied to iteratively adjust and optimize the model. With the estimated frontal shape, facial appearances are synthesized through a texture-fitting process formulated by solving a simple optimization problem. The proposed method has been evaluated on a in-the-wild facial expression database. The experimental results shows an outstanding performance of both visual effects of expression recovery and facial expression recognition.

Index Terms— Face frontalization, facial expression-aware, facial expression recognition, support vector regression, facial expression analysis

1. INTRODUCTION

Facial expression recognition (FER) in the wild addresses the unconstrained background and environment of facial images involving the challenging problem of large variations in head pose and occlusions [1]. FER based on 2D/3D view-invariant face-shape-free models have seldom been investigated. Most view-invariant models focus on the problem of face recognition, but not FER. View-invariant face-shape-free FER requires accurate alignment of face shape between non-frontal face and its corresponding frontal face, which is always challenging under various facial expression changes.

*Corresponding author: Hui Yu, at School of Creative Technologies, University of Portsmouth, Portsmouth, United Kingdom. Tel.:(+44)23 92845470 *Email address*: hui.yu@port.ac.uk

The existing view-invariant FER methods commonly focus on view modelling by learning the view-invariant features [2, 3] or training pose-wise classifiers [4, 5, 6, 7]. Most of the models need to be trained per viewpoint/person/expression in order to improve accuracy. Thus, a pose estimation step is always inevitable and the training data must be fairly large. Meanwhile, none of these methods address the problem of occlusions.

An appropriate way to overcome the above problems is to introduce face frontalization. Face frontalization is a comprehensive study which often involves face alignment, face morphing, face synthesis etc. Substantial progress has been made on face recognition [8, 9, 10]. The main purpose of face frontalization is to recover the frontal facial appearances from unconstrained images.

Frontal face-shape estimation is the fundamental step of face frontalization. The existing work in [6, 7, 11, 12] either achieve only person-specific frontalization or fail to rectify textures. The task of texture-fitting is to reconstruct facial appearances according to the frontal face-shape. In [9, 13], piece-wise affine warp [14] is used for texture rectification.In [15], symmetry-based strategy is used to compensate missing part due to out-of-plane head pose.

While frontal face-shape estimation is challenging, hard frontalization has shown its promising advantages. Currently, the approaches of [9] and [15] are the most effective generic (person-independent) face frontalization methods. They use a single 2D/3D reference facial template to fit facial appearances rather than estimating frontal face-shape. Both methods have successfully been tested for face recognition. However, hard frontalization may result in the loss of expression-related cues since all the obtained frontal faces share a same template shape. In order to achieve facial expression-aware face frontalization, a generic frontal face-shape estimation is inevitable.

In this paper, we focus on 2D generic frontalization and come up with an explicit idea of frontal-shape estimation and facial expression-aware face frontalization. Inspired by the success of regression based 2D face alignment and face morphing methods in [10, 16, 17], we proposed a novel

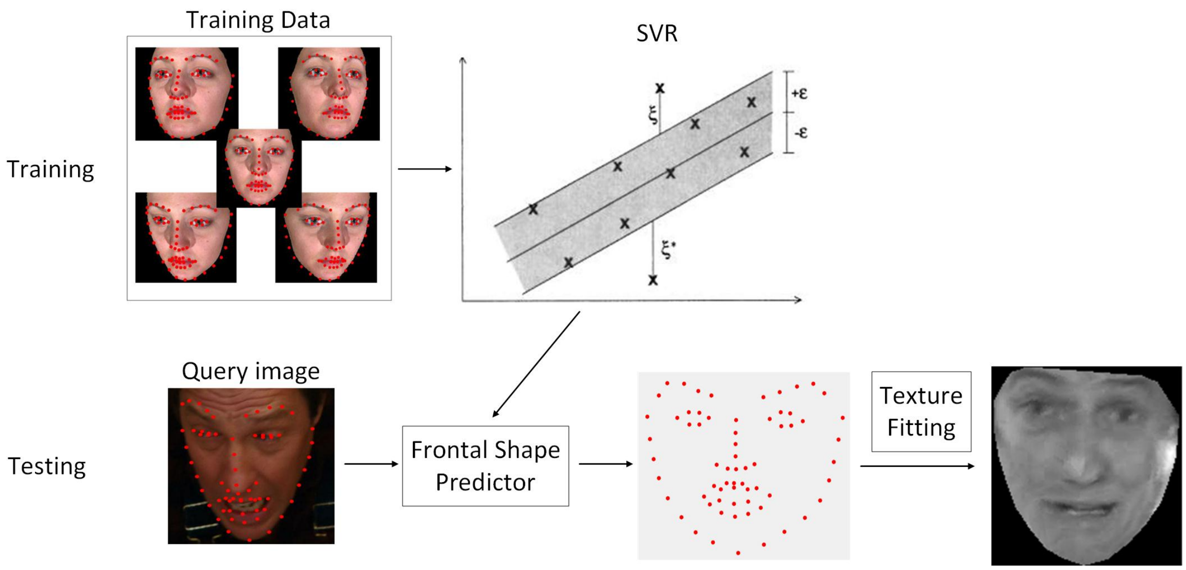

Fig. 1. Main process of proposed method

regression-based method for 2D frontal face-shape estimation. As is shown in Fig 1, a pair of a non-frontal facial image and its corresponding frontal image are collected in order to train the regressor pair-wise. Support vector regression (SVR) is chosen due to its effectiveness and simplicity. Empirically, one step SVR cannot fit well with all shape instances because the changes in poses, expressions and individual characteristics are non-linearly coupled. So we explored a cascade SVR to match with any possible shape features. In each cascade, we train a support vector regressor that best matches the groundtruth. During testing, the input face-shape will go through a sequentially cascade of SVR models and its frontal shape will be obtained.

The frontal face-shape can be viewed as base mash and we synthesize the frontal facial appearance by employing Active Appearance model instantiation [14]. The reconstructed face will be in frontal view and remain detailed expressions. Without using reference template, the facial geometric model for each input image will be unique. The details of facial shape, individual features and facial expressions can be maintained after texture fitting. The main contributions include:

1) We propose a SVR-based framework for facial expression-aware face frontalization which is a new research branch of view-invariant FER. As far as we know, this is the first work in which facial template is totally withdrawn and an explicit idea of 2D frontal face shape estimation is proposed by using regression-based model.

2) Unlike many other view-invariant FER methods which is extremely sensitive to the error of head pose estimation, there is no need for this method to estimate head pose, which effectively avoid the error accumulation.

3) Compared with other generic face frontalization method, the proposed method is able to reconstruct detailed facial expressions.

2. CASCADE SVR-BASED FACE FRONTALIZATION

2.1. Support Vector Regression

SVR is a supervised learning method for regression problem. Given the training data $\{(x^1, y^1), ..., (x^l, y^l)\}$, where $x \in \mathbb{R}^n, y \in \mathbb{R}$. Consider the linear function $f(x) = \langle\omega, x\rangle + b$, the SVR function can be expressed as:

$$\begin{aligned} \min \quad & \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{l}(\xi_i^- + \xi_i^+) \\ s.t. \quad & \begin{cases} y_i - \langle\omega, x_i\rangle - b \leq \epsilon + \xi_i^+ \\ \langle\omega, x_i\rangle + b - y_i \leq \epsilon + \xi_i^- \\ \xi_i^+, \xi_i^- \geq 0 \end{cases} \end{aligned} \tag{1}$$

where $C > 0$ is a constant which make the balance between maximum margin and tolerance ϵ, and ξ is the slack variable which suggests that part of error is tolerated.

After introducing dual problem and Lagrangian multipliers. The optimization problem becomes:

$$\max \begin{cases} \dfrac{1}{2}\displaystyle\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle x_i, x_j\rangle \\ -\epsilon\displaystyle\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) \end{cases} \tag{2}$$

$$s.t. \sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*) = 0 \quad and \quad \alpha_i, \alpha_* \in [0, C]$$

where α_i and α_* are the Lagrangian multipliers. The implementation of SVR varies. Quadratic programming is one of commonly used optimization method.

2.2. Frontal face-shape estimation

In this section, we present formulation the problem of SVR training. Given M annotated facial image pairs of non-frontal and corresponding frontal faces, the linear function of SVR can be defined as $\mathbf{x_0} + \Delta\mathrm{x} = \langle \omega, \mathbf{x_0} \rangle + b$, where $\mathbf{x_0}$ and $\mathbf{x}$ represents the shape vector for non-frontal and frontal images, respectively. So $\Delta\mathrm{x} = \mathbf{x} - \mathbf{x_0}$ is known. Consequently, the final function can be expressed as:

$$\Delta\mathrm{x} = \langle \omega, \mathbf{x_0} \rangle + b \tag{3}$$

This equation can be referred to as the linear function of SVR.

Then we introduce the cascade manner of SVR regarding $\Delta\mathrm{x}^i$ representing the obtained $\Delta\mathrm{x}$ in the ith cascade. In each cascade, we revise Eq.3 into:

$$\Delta\mathrm{x}^i = \langle \omega^i, \mathbf{x_0}^{\boldsymbol{i}} \rangle + b^i \tag{4}$$

and train a SVR model using:

$$\max \begin{cases} \dfrac{1}{2} \sum_{j,k=1}^{l} (\alpha_j^i - \alpha_j^{i,*})(\alpha_k^i - \alpha_k^{i,*}) k(x_{0,j}^i, x_{0,k}^i) \\ -\epsilon \sum_{j=1}^{l} (\alpha_j^i + \alpha_j^{i,*}) + \sum_{j=1}^{l} \Delta\mathrm{x}_j^i (\alpha_j^i - \alpha_j^{i,*}) \end{cases} \tag{5}$$

$$s.t. \sum_{j,k=1}^{l} (\alpha_j^i - \alpha_j^{i,*}) = 0 \quad and \quad \alpha_j^i, \alpha_*^i \in [0, C]$$

In the next cascade, we used the learned parameter to compute

$$\Delta\mathrm{x}^i = \sum_{j,k}^{l} (\alpha_j^i - \alpha_j^{i,*}) k(x_j^i, x_k^i) \tag{6}$$

In the new round, $\mathbf{x_0}^{\boldsymbol{i+1}} = \Delta\mathrm{x}^i + \mathbf{x_0}^{\boldsymbol{i}}$, $\Delta\mathrm{x}^{i+1} = \mathbf{x} - \mathbf{x_0}^{\boldsymbol{i+1}}$ and they will be used to train this round SVR model.

With the cascade SVR iteratively moving on, the value of the parameters, C and ϵ, reduces gradually . The algorithm will stop when these parameters and $\Delta\mathrm{x}^i$ turn zero. Empirically, the algorithm converges in 4 or 5 steps.

During testing, the non-frontal facial landmarks should be localized first. There are many existing facial landmark detection methods that has been proved to be effective. With the obtained facial landmarks, frontal face-shape will be estimated using Eq. 6 sequentially.

2.3. Face-texture fitting

Let $I \in \mathbb{R}^{m \times n}$ describe a non-frontal facial image whose landmarks is $\mathbf{x_0}$ and corresponding frontal shape is $\mathbf{x}$. The piece wise affine warp strategy can be directly used to compute the warped image $\boldsymbol{W}(I;p)$ [14].

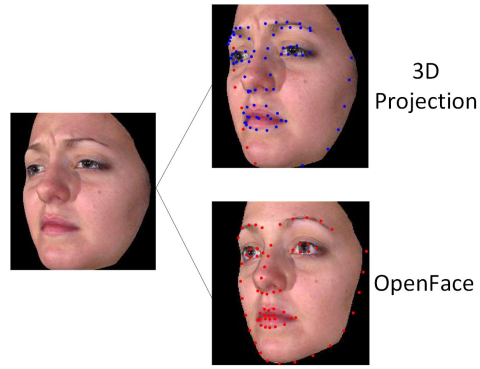

Fig. 2. Two different ways of collecting training data

The frontal facial appearances are synthesized via a linear combination of a set of pre-defined eigen faces:

$$F = Uc \tag{7}$$

where U is eigen face vector and c is parameters. Consequently, the optimization problem can be solved by:

$$\underset{c}{\operatorname{argmax}} \parallel \boldsymbol{W}(I;p) - Uc \parallel^2 \tag{8}$$

With a simple derivation, c can be obtained by $c = U\boldsymbol{W}(I;p)$. The synthesized frontal face can be obtained by Eq. 7.

3. EXPERIMENT

The performance of the proposed method has been validated in two tasks: 1) visual effects of frontal face reconstruction 2) FER in the wild.

3.1. Prepare for training data

Binghamton University 3D Facial Expression (BU3DFE) is a static 3D facial expression database which include 100 subjects with 2500 3D facial expression models. The training data are captured by rendering 2D images using 3D models. Images are captured at 7 pan angles (-45°, -30°, -15°, 0°, 15 °, 30°, 45°) and 5 tilt angles (-30°, -15°, 0°, 15 °, 30°), which results in totally 35 different viewpoints. Each training instance includes the position landmark points in one of the 34 non-frontal rotations and the corresponding points in frontal pose.

BU3DFE provide the 3D position of 83 landmarks for each 3D facial model. When the 3D landmark points were projected to 2D plane, there would be misalignments especially when there was large out-of-plane rotation. As is shown in figure 2, the red points of 3D projection are obviously misaligned. So we use OpenFace to automatically detect landmark points. OpenFace [19] is a very simple and effective tool

Table 1. Recognition rate (%) of different methods on SFEW databse

	Angry	Disgust	Fear	Happy	Neutral	Sadness	Surprise	Total
Baseline	23.00	13.00	13.90	29.00	23.00	17.00	13.50	18.90
[6]	25.89	**28.24**	17.17	42.98	14.00	33.33	10.99	24.70
[18]	24.11	14.12	20.20	50.00	23.00	23.23	21.98	26.14
Proposed	**40.18**	25.88	**48.48**	**55.26**	**37.00**	**36.36**	**37.36**	**40.71**

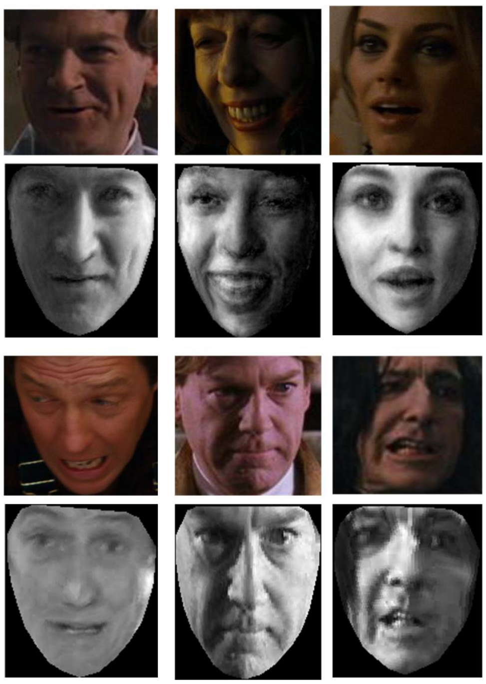

Fig. 3. Examples of face frontalization

for facial landmark detection. Most landmarks can be well detected using this software. Misaligned points were manually revised.

3.2. Frontalization and recognition

The shape model can be trained on the rendered 2D images from BU3DFE. For the two parameters of SVR, C is set by 10, 1, 0.1, 0.01 at the four iterations and ϵ is fixed to 0.

In order to evaluate the performance on the unconstrained images, we use another database for testing. Statistical Facial Expression in the Wild (SFEW) [20] is a spontaneous facial expression database. It contains 700 images captured from movies labelled by seven categories: six universal emotions and neutral.

From Figure 3 we can see the outstanding performance of face frontalization. Whatever pan rotation or tilt rotation can be well recovered to the frontal view. Meanwhile, the facial expression related cues are maintained.

For FER, there is a standard evaluation protocol provided by the authors of SFEW. The evaluation is strictly person-independent. In this experiment, Local Binary Pattern (LBP) and Support Vector Machine (SVM) are used for feature extraction and emotion classification, respectively. In table 1, the methods of [6] and [18] are the state-of-the-art approaches. It is obvious that our method outperforms the others. The overall recognition rate of the proposed method is 10% higher than the others, which suggests a considerable improvement. Based on this result we can conclude that facial expression-aware face frontalization can achieve promising result for FER in the wild. In this comparison, we did not mention deep learn because our work focus on small sample learning task which is quite different from deep learning. Meanwhile, deep learning methods must use a large volume of external training data, which do not totally comply with the evaluation protocol of SFEW.

4. CONCLUSIONS

In this paper, we have presented a novel regression-based frontal face-shape estimation method for facial expression-aware face frontalization and applied it to FER in the wild. As far as we know, this is the first work that can achieves both face frontalization and facial expression recovery. The experiment on unconstrained images shows impressive visual effects of the synthesized faces and a significant improvement in facial expression recognition for the data in the wild.

Acknowledgment

This work was supported by the EPSRC project, Real-Time 4D Facial Sensing and Modelling (EP/N025849/1) and Royal Academy of Engineering Newton Research Collaboration Programme - NRCP1516/1/74.

5. REFERENCES

[1] M. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, “The first facial expression recognition and analysis challenge,” in *IEEE International Conference and*

Workshops on Automatic Face and Gesture Recognition, pp. 921–926, 2011.

[2] U. Tariq, J. Yang, and T. S. Huang, "Multi-view facial expression recognition analysis with generic sparse coding feature," in *Procceedings of European Conference on Computer Vision*, pp. 578–588, 2012.

[3] U. Tariq, J. Yang, and T. S. Huang, "Supervised supervector encoding for facial expression recognition," *Pattern Recognition Letters*, vol. 46, pp. 89–95, 2014.

[4] S. Moore and R. Bowden, "Local binary patterns for multi-view facial expression recognition," *Computer Vision and Image Understanding*, vol. 115, no. 4, pp. 541–558, 2011.

[5] N. Hesse, T. Gehrig, H. Gao, and H. K. Ekenel, "Multi-view facial expression recognition using local appearance features," *International Journal of Computer Vision*, vol. 83, no. 2, pp. 178–194, 2011.

[6] S. Eleftheriadis, O. Rudovic, and M. Pantic, "Discriminative shared gaussian processes for multiview and view-invariant facial expression recognition," *IEEE Transaction on Image Processing*, vol. 24, no. 1, pp. 189–204, 2015.

[7] O. Rudovic, M. Pantic, and I. Patras, "Coupled gaussian processes for pose-invariant facial expression recognition," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1357–1369, 2013.

[8] G. Tzimiropoulos and M. Pantic, "Fast algorithms for fitting active appearance models to unconstrained images," *International Journal of Computer Vision*, pp. 1–17, 2016.

[9] C. Sagonas, Y. Panagakis, S. Zafeiriou, and M. Pantic, "Robust statistical face frontalization," in *Proceedings of IEEE International Conference on Computer Vision*, pp. 3871–3879, 2015.

[10] A. Akshay, S. Lucey, and R. Goecke, "Regression based automatic face annotation for deformable model building," *Pattern Recognition*, vol. 44, no. 10, pp. 2598–2613, 2011.

[11] L. A. Jeni, J. F. Cohn, and T. Kanade, "Dense 3d face alignment from 2d videos in real-time," in *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, pp. 1–8, 2015.

[12] J. Roth, Y. Tong, and X. Liu, "Unconstrained 3d face reconstruction," in *IEEE International Conference Computer Vision Workshops*, pp. 2606–2615, 2015.

[13] H. T. Ho and R. Chellappa, "Pose-invariant face recognition using markov random fields," *IEEE Transaction on Image Processing*, vol. 22, no. 4, pp. 1573–1584, 2013.

[14] I. Metthews and S. Baker, "Active appearance model revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.

[15] T. Hassner, S. Harel, E. Paz, and R. Enbar, "Effective face frontalization in unconstrained images," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4295–4304, 2015.

[16] X. Xiong and F. D. la Torre, "Supervised descent method and its applications to face alignment," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 532–539, 2013.

[17] G. Trigeorgis, P. Snape, M. A. Nicolaou, E. Antonakos, and S. Zafeiriou., "Mnemonic descent method: A recurrent process applied for end-to-end face alignment," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[18] M. Liu, S. Li, and X. Chen, "Au-aware deep networks for facial expression recognition," in *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, pp. 1–6, 2013.

[19] T. Baltusaitis, P. Robinson, and L. Morency, "Openface: an open source facial behavior analysis toolkit," in *IEEE Winter Conference on Applications of Computer Vision*, pp. 1–10, 2016.

[20] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, "Static facial expression analysis in tough conditions: Data evaluation protocol and benchmark," in *IEEE International Conference Computer Vision Workshops*, pp. 2106–2112, 2011.