Robust Facial Landmark Localization Based on Two-Stage Cascaded Pose Regression

Ziye Tong,[1] **Junwei Zhou,**[1,*] **Yanchao Yang,**[1] **Lee-Ming Cheng**[2]
[1]School of Computer Science and Technology, Wuhan University of Technology, Wuhan, PRC
[2]Department of Electronic Engineering, City University of Hong Kong, Hong Kong, PRC
[*]Corresponding Author, Email: junweizhou@msn.com

Abstract

In this paper, we propose a two-stage cascaded pose regression for facial landmark localization under occlusion. In the first stage, a global cascaded pose regression with robust initialization is performed to get localization results for the original face and its mirror image. The localization difference between the original image and the mirror image is used to determine whether the localization of each landmark is reliable, while unreliable localization with a large difference can be adjusted. In the second stage, the global results are divided into four parts, which are further refined by local regressions. Finally, the four refined local results are integrated and adjusted to get the final output.

Introduction

Many studies about facial landmark localization achieved desirable performances(Burgos-Artizzu, Perona, and Dollar 2013). However, it still has obstacles for facials with large variations including pose, expression, especially occlusions.

Cascaded pose regression (CPR) has emerged as one of the most famous methods of facial landmark localization since its superior performance. To localize facial landmarks under occlusion, Burgos-Artizzu et al. proposed the scheme of Robust CPR (RCPR)(Burgos-Artizzu, Perona, and Dollar 2013), which can detect occlusion information and localize the facial landmarks simultaneously. Robust Initialization for CPR (RICPR)(Pan et al. 2018) improved performance by providing texture and pose correlated initial shapes for the testing face. However, these methods including RCPR, DRDA(Zhang et al. 2016), SLPD(Wu, Gou, and Ji 2017), RICPR usually take the entire facial as a whole to make a global regression, while partial occlusions break the structure of the facial and bring obstacles to process the local variations. Moreover, the existed methods directly take the mean of all predictions as the final estimation without evaluating individuals.

TSCPR Architecture

In this paper, we propose a Two-Stage CPR (TSCPR) for facial landmark localization under occlusion. Firstly, we perform a global regression using RICPR to get first-stage localization and occlusion detection results of the original facial and its mirror, respectively. Then, we use the localization difference between the original facial and its mirror to determine whether the landmarks are reliable and adjust the unreliable localization results. In the second stage, the adjusted global results are divided into four parts including left eye, right eye, nose and mouse to learn different regressors, respectively. Then, the four results including local localization and occlusion detection are integrated. Finally, the results are evaluated and adjusted again according to mirror error.

The framework of the proposed TSCPR is illustrated in Figure. 1, which includes the global stage and the local stage. At each stage, we use the mirrorability of face alignment method(MFA) (Yang and Patras 2015) to adjust the unreliable localizations, which gave the clue that the mirror error is strongly correlated to the localization/alignment error. We first describe the design of the global stage and the local stage, then introduce implementation details of the entire model.

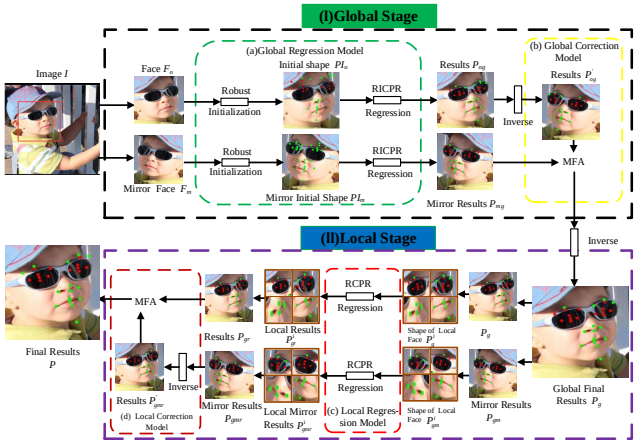


Figure 1: The pipeline of the proposed method TSCPR for facial landmark localization under occlusion.

Regression Model

RCPR divides the face image into 9 zones. At each iteration t, the image features are calculated as $f^t = h^t(I, S^{t-1})$ and

then RCPR trains S_{ta} regressors in each primitive fern regressor R_k^t $(k = 1, .., K)$, where K is the number of primitive fern regressors. Moreover, each regressor can only draw features in 9 pre-defined zones, and the occlusion representing one of the 9 zones can be estimated by the last occlusion estimation state of the image S^{t-1}. Finally, the occlusion presented in the zone is inversely proportional to the weight w_i^k, which is combined with the updates of the regressors δS_i^k to get ΔS_k^t.

Since RCPR is sensitive to initialization, where an improper initialization can severely degrade the performance, RICPR improved RCPR by providing texture and pose correlated initial shapes. In this work, we use RICPR to get the localization and occlusion detection results, as well as the mirror localization and occlusion detection results.

Correction Model

Before introducing the correction model, we first verify whether we can evaluate the reliability of the prediction. After regression, we have got orignial localization P_{og} and and its mirror localization P_{mg} as shown in Figure. 1. Then, we invert P_{og} to P'_{og},

$$x'^i_{og} = w - x^i_{og}, y'^i_{og} = y^i_{og}, v'^i_{og} = v^i_{og}, (i = 1, .., n), \quad (1)$$

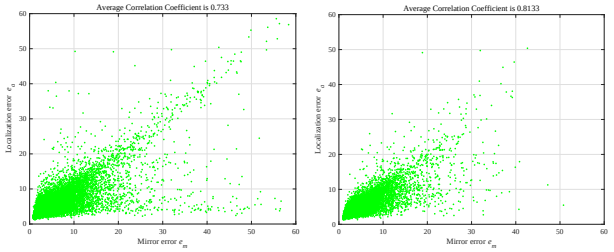
where x^i_{og}, y^i_{og} and v^i_{og} are the prediction in x-coordinate, y-coordinate and occlusion state of i-th landmark of P_{og}. The values of x'^i_{og}, y'^i_{og} v'^i_{og} are obtained by inverting P_{og}, while w is the width of the image and n is the number of the facial landmarks. The localization error is calculated as:

$$e_a = \sqrt{(x^i_{og} - x^i_o)^2 + (y^i_{og} - y^i_o)^2}. \quad (2)$$

where x^i_o, y^i_o and v^i_o are the ground truth in x-coordinate, y-coordinate and occlusion state of i-th landmark of the P_{og}, respectively. The mirror error is calculated as:

$$e_m = \sqrt{(x^i_{mg} - x'^i_{og})^2 + (y^i_{mg} - y'^i_{og})^2}. \quad (3)$$

where x^i_{mg}, y^i_{mg} and v^i_{mg} are the x-coordinate prediction, y-coordinate prediction and occlusion state prediction of i-th landmark P^i_{mg}, respectively.



(a) (b)

Figure 2: (a): The global average correlation of e_m and e_a. (b): The local average correlation of e_m and e_a.

As shown in Figure 2, the correlation between e_m and e_a demonstrates that the mirror error e_m is strongly correlated to the localization error e_a. A larger mirror error usually leads to a larger localization error. Therefore, we can adjust the unreliable predictions using the mirror error. In this paper, we use MFA (Yang and Patras 2015) to evaluate the reliability of the regression results.

Since MFA gets the mirror error for the whole face's landmarks, which results in a large error of the reliability assessment. To solve the problem, we compute the mirror error specifically for each single landmark. Moreover, we use the distance between localization result and mirror localization result to quantify mirror error instead of using the distance between them in x-coordinate, which significantly improves the accuracy of reliability evaluation.

Experimentals and Results

As shown in Table 1, we compared TSCPR with several state-of-the-art methods and two-stage RCPR (TRCPR) on the COFW dataset using the Normalized Mean Error (NME) and the failure rate, where error above 0.1 will be taken as a failure. The results show that NME is 6.34 $\times 10^{-2}$ and the accuracy of occlusion detection is 80/57.1% precision/recall, which are comparable to the state-of-the-arts. Since the proposed method is usually independent of facial landmark localization, it has the potential to be extended and applied to other algorithms.

Table 1: Comparison of facial landmark localization and occlusion detection on the COFW dataset

Methods	Error NME	Failure ($\times 10^{-2}$)	Occlusion Precision/Recall
RCPR	8.01	20	80/42%
TRCPR	7.6	16.1	80/44.32%
RICPR	6.64	11	80/54.6%
DRDA	6.46	-	80/54.4%
SLPD	6.40	-	80/44.3%
TSCPR	**6.34**	**9.8**	**80/57.1%**
Human	5.6	-	-

Acknowledgments

This work is supported by the National Natural Science Foundation of China (61601337).

References

Burgos-Artizzu, X. P.; Perona, P.; and Dollar, P. 2013. Robust face landmark estimation under occlusion. In *ICCV*, 1513–1520.

Pan, Y.; Zhou, J.; Gao, Y.; Xiong, S.; and Yang, Y. 2018. Robust facial landmark localization based on texture and pose correlated initialization. *arxiv [online]*.

Wu, Y.; Gou, C.; and Ji, Q. 2017. Simultaneous facial landmark detection, pose and deformation estimation under facial occlusion. In *CVPR*, 5719–5728.

Yang, H., and Patras, I. 2015. Mirror, mirror on the wall, tell me, is the error small? In *CVPR*, 4685–4693.

Zhang, J.; Kan, M.; Shan, S.; and Chen, X. 2016. Occlusion-free face alignment: Deep regression networks coupled with de-corrupt autoencoders. In *CVPR*, 3428–3437.