Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Quality-Aware Estimation of Facial Landmarks in Video Sequences

Haque, Mohammad Ahsanul; Nasrollahi, Kamal; Moeslund, Thomas B.

Published in:
IEEE Winter Conference on Applications of Computer Vision

DOI (link to publication from Publisher):
10.1109/WACV.2015.96

Publication date:
2015

Document Version
Early version, also known as pre-print

Quality-Aware Estimation of Facial Landmarks in Video Sequences

Mohammad A. Haque, Kamal Nasrollahi, Thomas B. Moeslund
Visual Analysis of People Laboratory, Aalborg University, Denmark
{mah, kn, tbm}@create.aau.dk

Abstract

Face alignment in video is a primitive step for facial image analysis. The accuracy of the alignment greatly depends on the quality of the face image in the video frames and low quality faces are proven to cause erroneous alignment. Thus, this paper proposes a system for quality aware face alignment by using a Supervised Decent Method (SDM) along with a motion based forward extrapolation method. The proposed system first extracts faces from video frames. Then, it employs a face quality assessment technique to measure the face quality. If the face quality is high, the proposed system uses SDM for facial landmark detection. If the face quality is low the proposed system corrects the facial landmarks that are detected by SDM. Depending upon the face velocity in consecutive video frames and face quality measure, two algorithms are proposed for correction of landmarks in low quality faces by using an extrapolation polynomial. Experimental results illustrate the competency of the proposed method while comparing with the state-of-the-art methods including an SDM-based method (from CVPR-2013) and a very recent method (from CVPR-2014) that uses parallel cascade of linear regression (Par-CLR).

1. Introduction

Automatic analysis of facial image plays an important role in many different areas such as surveillance, medical diagnosis, biometrics, and expression recognition [1]. Detecting facial landmarks, also called face alignment, is an essential step in automatic facial image analysis. The accuracy of alignment, i.e., pertinent detection of facial landmarks, affects the performance of the analysis. Face alignment is considered as a mathematical optimization problem and a number of methods were proposed to solve this problem. The Active Appearance Model (AAM) fitting along with its derivatives are some of the early but effective solutions in this area [2]. The AAM fitting works by estimating some parameters of a model which is close enough to the given image. A number of regression based approaches was proposed for the AAM fitting, e.g., a linear regression based fitting [3], a non-linear regressor using boosted learning [4], a boosted ranking model [5], and discriminative approaches [6-7]. Figure 1(a) shows 66 facial landmark points detected in two examples by an AAM fitting algorithm (Fast-Simultaneous Inverse Compositional algorithm (Fast-SIC)) [2].

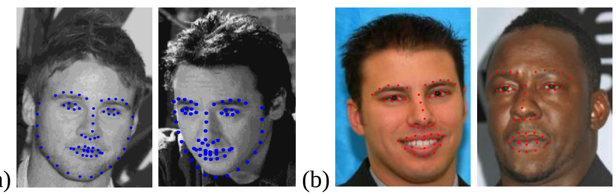
(a) (b)

Figure 1: Examples of facial landmarks detection: (a) Fast-SIC detected 66 points [2] and (b) SDM detected 49 points in four images of the LFPW database [14].

Though the results of regression based AAM fitting methods are good, they are computationally expensive due to the iterative nature of learning the shape and the appearance parameters. To deal with this problem a number of works were done by optimizing least-square functions. For example, Matthews et al. formulated the AAM fitting as a Lukas-Kanade (LK) problem which can be solved using Gauss-Newton optimization [8, 9]. Similar Gauss-Newton or gradient decent based optimization methods for this problem can be found in [10-11]. Standard gradient decent algorithms when applied to AAMs are, however, inefficient in term of computational complexity [2, 12]. Two fast AAM fitting approaches were proposed recently in CVPR [13, 14]. Asthana et al. proposed a Parallel Cascade of Linear Regression (Par-CLR) to detect the landmarks [13]. On the other hand, Xiong et al. developed a Supervised Descent Method (SDM) to minimize a non-linear least square function and employed it for face alignment [14]. Both of these methods in [13, 14] are able to work real-time and showed competent results. Figure 1(b) shows two examples of detecting 49 facial landmark points by the SDM [14].

Although the SDM provides good estimates of the facial landmarks, its detection accuracy is suffered by facial image quality measures like resolution (Figure 2, col. 1, row 1), pose (Figure 2, col. 1-3, row 2), brightness (Figure 2, col. 2, row 1), and sharpness (Figure 2, col. 1, row 2). Moreover, the SDM-based face alignment of [14] uses the landmarks of the current frame as the initial points of searching in the next frame in a video, which produces erroneous results when no face is detected in the current frame or when the face is of low quality (Figure 2, col. 3, row 1).

Figure 2: Depiction of bad performance of quality unaware SDM-based method in the alignment of low quality faces in video sequences from *Youtube Celebrities* dataset [14].

When a video acquisition system acquires facial video frames, low quality facial images are very common in many real-world problems [15]. For example, a human face at 5 meters distance from a surveillance camera subtends only about 4x6 pixels on a 640x480 sensor with 130 degrees field of view, which is an insufficient resolution for almost any further analysis [16]. A face region with size 48x64 pixels, 24x32 pixels, or less is not likely to be used for expression recognition due to inadequate information available in the low resolution face [17]. Similar problems are exhibited in facial analysis for high pose variation, very high or very low brightness, and low sharpness value of a facial image [18-20]. When a high quality face image is provided to a face alignment system (e.g., SDM) it detects the landmarks very accurately. However, when the face quality is low, the detected landmark positions are not trustworthy.

To deal with the alignment problem of low quality facial images, a Face Quality Assessment (FQA) system [21] can be employed before running the alignment algorithm. Such a system uses some quality measures to determine whether a face is qualified enough and provides assistance to further analysis by providing the quality rating. Figure 3 shows a typical FQA method which consists of three steps: video frame acquisition, face detection in the video frames, and FQA by measuring face quality metrics. In this paper, we propose a 'quality-aware' estimation method for improved face alignment in video sequence, where facial landmarks in high quality faces are estimated by the SDM and in low quality faces are estimated by a motion based forward extrapolation method.

The rest of the paper is organized as follows: Section 2 presents the proposed approach. Section 3 states the experimental results and finally Section 4 concludes the paper.

Figure 3: Steps of a typical face quality assessment system.

2. The proposed method

The SDM based face alignment system results in erroneous landmarks detection when: (1) a face is detected in a wrong position, and (2) the face quality is low. To deal with these problems, we propose a quality aware system as shown in Figure 4. The steps of the system are described in the following subsections.

2.1. Face detection module

The first step of face alignment in a video sequence is face detection. We employ the well-known Viola and Jones face detection approach [22] for this purpose. This method utilizes the so called Haar-like features in a linear combination of some weak classifiers to classify face and non-face. In order to speed up the detection process an evolutionary pruning method is adopted in classification in order to form strong face/non-face classifier from fewer weak classifiers [23]. Following Figure 4, when a face is detected, the face is passed to the Face Quality Assessment (FQA) module.

2.2. Face quality assessment module

The FQA module is responsible to assess the quality of the extracted faces. Nasrollahi et al. proposed a face quality assessment system in video sequences [21]. Haque et al. utilized face quality assessment while capturing video sequences from an active pan-tilt-zoom camera [18]. FQA was also employed in [24] before constructing facial expression log. All of these previous methods used four quality parameters: out-of-plan face rotation (pose), sharpness, brightness, and resolution. For facial geometry analysis and detection of landmarks all of these quality metrics are critical as discussed in Section 1 (and Figure 2). Thus, we calculate these four quality metrics to assess the face quality. A normalized score is then obtained in the range of *[0:1]* for each quality parameter and a weighted combination of the scores is utilized to generate a single quality score, Q^i, as in [18]. The Q^i which represents the quality of the face in ith frame will be used in the subsequent blocks of the system to make a decision about the method that the proposed system needs to use for the face alignment.

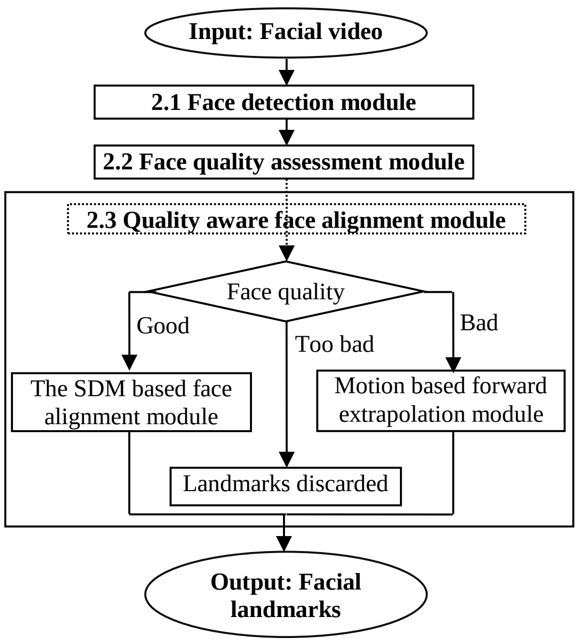

Figure 4: Block diagram of the proposed system of face alignment in video with face quality assessment. The steps of the method are explained in Section 2.

2.3. Quality-aware face alignment module

The SDM method for face alignment uses a set of training samples to learn a mean face shape. This mean shape is used as an initial point for an iterative optimization of a non-linear least square function towards the best estimates of the positions of the landmarks in facial test images. The minimization function can be defined as a function over Δx as:

$$f_{SDM}(x_0 + \Delta x) = \left\| g\big(d(x_0 + \Delta x)\big) - \theta_* \right\|_2^2 \quad (1)$$

where, x_0 is the initial configuration of the landmarks in a facial image, $d(x)$ indexes the landmarks configuration (x) in the image, g is a nonlinear feature extractor, $\theta_* = g(d(x_*))$, and x_* is the configuration of the true landmarks. The Scale Invariant Feature Transform (SIFT) [25] is used as the feature extractor g. In the training images Δx and θ_* are known. By utilizing these known parameters the SDM iteratively learns a sequence of generic descent directions, $\{\delta_n\}$, and a sequence of bias terms, $\{\beta_n\}$, to set the direction towards the true landmark configuration x_* in the minimization process, which are further applied in the testing phase [14]. This is done by:

$$x_n = x_{n-1} + \delta_{n-1}\sigma(x_{n-1}) + \beta_{n-1} \quad (2)$$

where, $\sigma(x_{n-1}) = g(d(x_{n-1}))$ is the feature vector extracted at previous landmark location x_{n-1} and x_n is the new location. The succession of x_n converges to x_* for all images in the training set.

In the proposed approach, following Figure 4, the face region is passed to the SDM based alignment module, if the quality score is greater than an empirical threshold, Q_{Th_high}. Otherwise, the face region is passed to the motion based forward extrapolation module in order to estimate the landmarks or reject the landmark if the quality based on an empirical threshold, Q_{Th_low}, is too bad. To be more precise, we first rewrite the SDM's objective function in (1) for a video sequence as:

$$f_{SDM}(x_0^i + \Delta x^i) = \left\| g\left(d\left(x_0^i + \Delta x^i\right)\right) - \theta_*^i \right\|_2^2 \quad (3)$$

where, the i superscript implies the frame number and the other symbols have similar meaning as (1). Then, we include the information Q^i provided by the FQA system as a prior for the processing of the next frame. This will change the objective function of (3) to:

$$f_{Proposed\ system} = \begin{cases} f_{SDM}(x_0^i + \Delta x^i),\ Q^i > Q_{Th\_high} \\ f_{Est}(x^i),\ Q_{Th\_low} < Q^i < Q_{Th\_high} \\ No\ Landmarks,\ Q^i < Q_{Th\_low} \end{cases} \quad (4)$$

where, $f_{Est}(x^i) = \left\| d(x^i) - d(x_*^i) \right\|_2^2$ minimizes the non-likelihood error between corrected landmarks $d(x^i)$ using the landmark configuration of previous good quality frames and the true landmark configuration $d(x_*^i)$. The working procedure of the SDM based method with the minimization function f_{SDM} and the forward extrapolation method with the minimization function f_{Est} are described in the following subsections, respectively.

2.3.1 SDM-based face alignment module

In the proposed system, we employed a simple deviation of SDM for good quality faces. We initialize the iterative minimization of SDM for landmark configuration in each frame by giving a shape estimate (initial landmark positions) that is obtained from the mean shape of the training images. This is in contrary with [14] which initializes the landmarks in subsequent video frames by using the detected landmarks of the previous frame. This technique of [14] incurs the problem of getting trapped into local maxima, especially in the videos where the faces have high velocity due to video motion [26] among the frames. Thus, instead of using landmarks of the previous frames to initialize the minimization process, we initialize each frame by following the same procedure of using an estimated mean shape from the training data. This provides a more genuine estimate for initial position of landmarks for high-velocity frames. Moreover, as the mean shape initialization is a low-cost computational process involving merely a translation and scaling operation on the pre-trained landmark configuration, this does not incur any undoable computational complexity for real-time operation in video. Besides, the translational and scaling differences are also considered during initialization by scaling the mean shape from training with the size of the present face region.

2.3.2 Motion-based forward extrapolation module

As shown in Figure 2 when face quality is low, the SDM's minimization function f_{SDM} fails to converge in the right place. Furthermore, when a face is detected in a wrong position, the SDM tries to converge in that wrong place. These can be dealt with by considering the temporal stability that is usually present among subsequent facial images of a video. In another words, the alignment problem in low quality and/or wrongly detected facial frames can be addressed by exploiting landmarks positions in the previous good quality faces. This is exactly the point that we have utilized in this paper: the landmarks in low quality faces and/or wrongly detected faces are extrapolated from the landmarks of the previous frames that are of good quality.

When f_{Est} is called for action, the FQA module provides some information such as the quality of the detected face, the displacement of the detected face from the face in the previous frame (the velocity of the face in the frames), the degree of pose in the current face and the amount of pose variation in the two previous faces in the video sequence. In order to calculate $d(x^i)$ for a low quality face in the ith frame of a video, let $d(x^{i-m})$ and $d(x^{i-n})$ denote two landmark configurations of the previous good quality face frames. We define the correction polynomial to minimize $f_{Est}(x^i)$ as:

$$d(x^i) = d(x^{i-n}) + \frac{d(x^{i-m}).-d(x^{i-n})}{x^{i-m}-x^{i-n}} \times (x - x^{i-n}) \qquad (5)$$

where, $x^{i-m} - x^{i-n}$ and $(x - x^{i-n})$ are the velocities of face in the corresponding frames with respect to the previous frames, and $(.-)$ indicates the subtraction operation for each landmarks (49 landmarks) separately.

In the proposed method, we utilize the extrapolation polynomial of (5) in two different ways for two different conditions:

Condition 1) When the face is detected in a wrong position. In this case the face motion shows a larger displacement for the current face than the faces in the previous frames and the face quality shows a sharp change than the face quality in the previous frames. We employed *Algorithm I* in order to extrapolate the landmarks. Two parameters, *Displacement* and *Quality_Change*, are compared with two empirically set thresholds and the landmarks in current frame (*Current_Landmarks*) are calculated from the landmarks in the previous qualified frame (*Prev_Landmarks*).

Condition 2) When the face quality is low. In this case the landmarks detected from the previous high-quality facial images are used to incorporate the motion information of the pose variation in the landmarks to extrapolate the landmarks of the current face. We employed *Algorithm II* in order to extrapolate the landmarks in this case. Two face quality parameters for pose and size, *Qual_Pose* and *Qual_Size* are compared with empirically set *Pose_Threshold* and *Size_Threshold*, and used for conditionally estimating *Current_Landmarks* by using *Prev_Landmarks* and *Displacement*. If the correction is not possible due to the missing of face in previous video frames, the landmarks detected in the current low quality face are also discarded in order to reduce false estimation.

Algorithm I

```
WRONG_DETECTION_ESTIMATE (Displacement, Quality_Change)
{
IF Displacement > Dis_Threshold AND
        Quality_Change > Qual_Threshold THEN
  Prev_Landmarks = The landmarks of previous qualified face;
  Current_Landmarks = ESTIMATE(Prev_Landmarks,
    Displacement);
  RETURN Current_Landmarks;
END
}
```

Algorithm II

```
LOW_QUALITY_ESTIMATE (Qual_Pose, Qual_Size, Displacement)
{
IF  Qual_Pose  < Pose_Threshold AND
    Qual_Size > Size_Threshold THEN
  Prev_Landmarks = The landmarks of previous qualified face;
  Current_Landmarks = ESTIMATE (Prev_Landmarks,
    Displacement);
  RETURN Current_Landmarks;
END
}
```

```
ESTIMATE (Prev_Landmarks, Displacement) {
  temp_Landmarks =
          Landmarks calculated from (Eq. 5)
  Current_Landmarks = Rotation of
          temp_Landmarks with pose variation
  RETURN Current_Landmarks;
}
```

3. Experimental results

The proposed system was implemented in a combination of VISUAL C++ and MATLAB environments. An implementation of SDM along with its trained direction and bias terms is given in [14]. In order to evaluate the proposed system we used the well-known *Youtube Celebrities* database [27]. However, as a general database created for face recognition research, most of the videos in this database are either single-subject video (as we assume single face from the same subject in a video), or too short to provide enough motion information for erroneous frames, or do not subjected to the problem of low-quality face. However, we managed to select 18 videos containing *2537* frames from this database wherein the problem of low face-quality in alignment are exhibited. To generate ground truth data, we manually aligned the low quality faces in all of these selected videos and compared the performance of the proposed system and state-of-the-art facial alignment systems against this ground truth data.

3.1. Performance evaluation

Figure 5 shows some results of landmarks detection in some good quality face frames of *Youtube Celebrities* database. The results for SDM [14] and the proposed method are similar for these good quality faces. When the face quality is too bad to detect the landmarks, the proposed algorithm discards the erroneous landmarks detected by SDM. Figure 6 shows some example faces from *Youtube Celebrities* database where erroneous landmarks detected by SDM are discarded. Figure 7 illustrates some results of landmarks correction by the proposed method in order to provide the mean of qualitative assessment. The first and third rows of Figure 7 present the results generated by the SDM, and the second and fourth rows present the results generated by our method. From the results (col. 1, 2, 4, row 1-2 of Figure 7) it is observed that when the face detector produces a wrong detection, the proposed approach can detect the error from the face quality and the displacement (face velocity in consecutive frames) parameters, and then extrapolate the landmarks by using the motion information. In the other cases the faces have low resolution problem (col. 3-5, row 1-2), low brightness (col. 6, row 1-2), high pose variation (col. 1-4, row 3-4) and low sharpness (col. 5, row 3-4) in Figure 7. Some faces do not exhibit problem from low-quality, instead they get trapped into local minima in the SDM's minimization process due to poor initialization for high velocity face frames (e.g., col. 6, row 3-4).

Figure 8 shows the relationship between face quality with the SDM and the proposed methods. We used 84 frames of *0450_03_001_bill_clinton* sequence of *Youtube Celebrities* database in order to generate this result. It can be seen that when the face quality is low (in frames 70-80) the detection error is high in the SDM-based method. When the proposed method utilizes the face quality metric along with SDM-based method, the detection error reduces.

Figure 5: Some examples of good quality facial images from *Youtube Celebrities* database for which the proposed method produces similar results to SDM [14].

Figure 6: Some of the alignment results of SDM-based method [14] on the frames of the *Youtube Celebrities* database which are discarded by the proposed method due to excessively low face quality. The first row shows the low-quality face images and the second row shows the landmarks points detected by SDM.

3.2. Performance comparison

Table 1 shows the point to point error in landmark detection by the SDM and the proposed methods compared to the manually generated ground truth of some low quality faces in *0450_03_001_bill_clinton* sequence. From the results it can be seen that when the face quality is low the SDM produces erroneous results, however the proposed method provides better results. The face quality-scores in frames 76 and 77 are very low due to detection of the face in wrong places. However, our landmark correction method, using the motion information, provides better estimates. The result of SDM is a little better than the proposed method in the frame 70. According to our observation this is not because of the slip of the proposed method, instead this is because of trivial difference of few pixels between manual annotation from our perception of true landmarks, and the automatic detection by SDM and the proposed method.

Figure 7: Comparing the landmark correction results of the proposed system (second and fourth row) against the results of the SDM-based method of [12] (first and third row) in some low quality frames of the *Youtube Celebrities* database.

Table 2 shows the comparison between the Par-CLR based method [13], the SDM-based method [14], and the proposed method in normalized point to point error for all 18 selected video sequences from the *Youtube Celebrities* database. These results are depicted in Figure 9. The data for each video in Figure 9 is independent of the data for other videos. Because, we have normalized the point-to-point errors of the frames of a video by the highest value of error in that video. Thus, showing higher error in a video does not necessarily mean that the error of detection in this video is higher than that of the other videos. From the results it is observed that the proposed method outperforms both SDM and Par-CLR when the videos have low quality face-frames. Another significant observation of the experimental results is that the proposed method either improves the detection results or at least maintains the accuracy of SDM and does not worsen the detection error in comparison with both SDM and Par-CLR. Thus, the contribution of the proposed method is significant when high accuracy is expected in landmark-based facial image analysis.

Table 1: Normalized point to point error for both SDM [14] and the proposed method of some low quality faces of *0450_03_001_bill_clinton* sequence of *Youtube Celebrities* database.

Frame number	Face quality	Method	Normalized Error
1 (32)	*0.42*	SDM	*0.586*
		Proposed	*0.008*
2 (69)	*0.31*	SDM	*0.015*
		Proposed	*0.010*
3 (70)	*0.26*	SDM	*0.025*
		Proposed	*0.035*
4 (76)	*0.00*	SDM	*0.995*
		Proposed	*0.284*
5 (77)	*0.00*	SDM	*1.000*
		Proposed	*0.306*

Table 2: Average point to point error of the Par-CLR [13], the SDM [14] and the proposed methods compared to the manually generated ground truth for erroneous frames of 18 experimental videos from the *Youtube Celebrities* database. Higher values indicate higher detection errors.

No.	Sequence name	Pt-pt error		
		Par-CLR	SDM	Proposed
1.	*0450_03_001*	*0.3821*	*0.0315*	***0.0081***
2.	*0051_03_008*	*0.7926*	*0.0786*	***0.0021***
3.	*0049_03_006*	*0.8926*	*0.7506*	***0.3246***
4.	*1304_01_001*	*0.9702*	*0.4145*	***0.2841***
5.	*0009_01_009*	*0.8776*	*0.8369*	***0.3506***
6.	*0033_02_001*	*0.8515*	*0.9000*	***0.4402***
7.	*0054_03_011*	*0.8907*	*1.0000*	***0.6834***
8.	*0079_01_024*	*0.5156*	*0.5060*	***0.0067***
9.	*0162_02_026*	*0.6732*	*0.5000*	***0.3321***
10.	*0182_03_015*	*0.9663*	*0.9515*	***0.5956***
11.	*0193_01_004*	*0.5087*	*0.5064*	***0.0018***
12.	*0458_03_009*	*0.7045*	*0.8898*	***0.4913***
13.	*0492_03_009*	*0.7903*	*0.8973*	***0.3456***
14.	*0518_03_002*	*0.8669*	*0.8096*	***0.1920***
15.	*0532_01_007*	*0.9863*	*0.3348*	***0.0012***
16.	*0606_03_001*	*0.6480*	*0.7661*	***0.1547***
17.	*0795_01_004*	*0.8063*	*0.5039*	***0.0027***
18.	*1744_01_017*	*0.7928*	*0.8552*	***0.1675***
Average error in erroneous frames of 18 videos with 2537 frames in total		*0.7731*	*0.6314*	***0.2318***

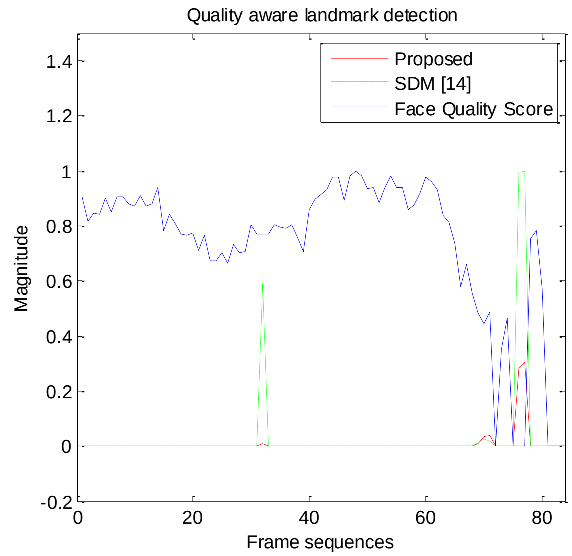

Figure 8: Face quality and normalized point to point error for both the SDM [14] and the proposed method on 84 frames of *0450_03_001_bill_clinton* sequence of *Youtube Celebrities* database.

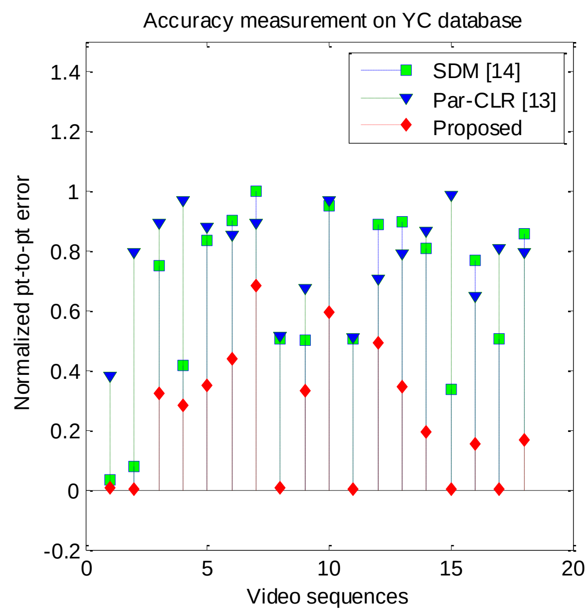

Figure 9: Average point to point error of the SDM [14], Par-CLR [13] and the proposed methods compared to the manually generated ground truth for erroneous frames of 18 experimental videos from the *Youtube Celebrities* database. Detection error is normalized for each video separately.

4. Conclusions

This paper investigated the problem of detecting facial landmarks in low-quality faces of videos. As the SDM-based face alignment system exhibits the problems of misalignment due to low face-quality and initialization by the landmarks of the previous frame for the current frame, we proposed a quality aware method for improved face alignment, where high quality faces were estimated by SDM and low quality faces were estimated by motion based forward extrapolation. The method utilized the quality of the detected face, the displacement of the detected face from the face in the previous frame (the velocity of the face in the consecutive face-frames), the degree of pose in the current face and the amount of pose variation in previous faces in the video sequence in order to extrapolate the landmarks in the face of the current video frame. As the proposed method improves the landmarks detection results in erroneous (low quality) frames and does not worsen the detection error (in high quality frames) while comparing against state-of-the-art approaches, the contribution of the proposed method is noteworthy for facial image analysis. In the future, we will investigate the performance of the proposed face alignment system in facial expression recognition systems and facial image based health monitoring.

References

[1] S.Z. Li, and A.K. Jain, Handbook of Face Recognition, 2nd Edition, Springer, 2011.

[2] G. Tzimiropoulos, and M. Pantic, Optimization Problems for Fast AAM Fitting In-The-Wild, Proc. of the Int. Conf. on Computer Vision (ICCV), 2013.

[3] T. Cootes, G. Edwards, and C. Taylor, Active Appearance Models, IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI), vol. 23, no. 6, pp. 681-685, 2001.

[4] J. Saragih, and R. Gocke, A Nonlinear Discriminative Approach to AAM Fitting, Proc. of the Int. Conf. on Computer Vision (ICCV), 2007.

[5] H. Wu, Z. Liu, and G. Doretto, Face Alignment via Boosted Ranking Model, Proc. of the Int. Conf. on Computer Vision (ICCV), 2008.

[6] X. Liu, Generic Face Alignment using Boosted Appearance Model, Proc. of the Int. Conf. on Computer Vision and Pattern Recognition (CVPR), 2007.

[7] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, Robust Discriminative Response Map Fitting with Constrained Local Models, Proc. of the Int. Conf. on Computer Vision and Pattern Recognition (CVPR), 2013.

[8] I. Matthews, and S. Baker, Active Appearance Models Revisited, International Journal of Computer Vision (IJCV), 60(2):135-164, 2004.

[9] S. Lucey, R. Navarathna, A. Ashraf, and S. Sridharan, Fourier Lucas-Kanade Algorithm, IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI), vol. 35, no. 6, pp. 1383-1396, 2013.

[10] F. De la Torre, and M.H. Nguyen, Parameterized Kernel Principal Component Analysis: Theory and Applications to Supervised and Unsupervised Image Alignment, Proc. of the Int. Conf. on Computer Vision (ICCV), 2008.

[11] K.T.A. Moustafa, F. De la Torre, and F.P. Ferrie, Pareto Discriminant Analysis, Proc. of the Int. Conf. on Computer Vision and Pattern Recognition (CVPR), 2010.

[12] L. Unzueta, W. Pimenta, J. Goenetxea, L.P. Santos, and F. Dornaika, Efficient Generic Face Model Fitting to Images and Video, Image and Vision Computing, vol. 32, no. 5, pp. 321-334, 2014.

[13] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, Incremental Face Alignment in the Wild, Proc. of the Int. Conf. on Computer Vision and Pattern Recognition (CVPR), 2014.

[14] X. Xiong, and F.D.L. Torre, Supervised Descent Method and Its Application to Face Alignment, Proc. of the Int. Conf. on Computer Vision and Pattern Recognition (CVPR), 2013.

[15] K. Nasrollahi, and T.B. Moeslund, Complete face logs for video sequences using face quality measures, IET Signal Processing, vol. 3, no. 4. pp. 289-300, 2009.

[16] S.J.D. Prince, J. Elder, Y. Hou, M. Sizinstev, and E Olevskiy, Towards Face Recognition at a Distance, Proc. of the IET Conf. on Crime and Security, pp. 570-575, 2006.

[17] Y. Tian, T. Kanade, and J.F. Cohn, Facial Expression Recognition, Handbook of Face Recognition, Chapter 19, pp. 487-519, Springer, 2011.

[18] M.A. Haque, K. Nasrollahi, and T.B. Moeslund, Real-Time Acquisition of High Quality Face Sequences from an Active Pan-Tilt-Zoom Camera, Proc. of the Int. Conf. on Advanced Video and Surveillance Systems (AVSS), pp. 1-6, 2013.

[19] K. Nasrollahi, and T.B. Moeslund, Extracting a Good Quality Frontal Face Image From a Low Resolution Video Sequence, IEEE Trans. On Circuits and Systems for Video Technology, vol. 21, no. 10, pp. 1353-1362, 2011.

[20] I.V.D. Linde, and T. Watson, A combinatorial study of pose effects in unfamiliar face recognition, Vision Research, vol. 50, pp. 522-533, 2010.

[21] K. Nasrollahi, and T.B. Moeslund, Face Quality Assessment System in Video Sequences, 1st European Workshop on Biometrics and Identity Management, Springer-LNCS, vol. 5372, pp. 10-18, 2008.

[22] P. Viola, and M. Jones, Robust real-time face detection, Proc. of the 8th IEEE Int. Conf. on Computer Vision, pp. 747, 2001.

[23] J. Jun-Su, and K. Jong-Hwan, Fast and Robust Face Detection using Evolutionary Pruning, IEEE Trans. On Evolutionary Computation, vol. 12, no. 5, pp. 562-571, 2008.

[24] M.A. Haque, K. Nasrollahi, and T.B. Moeslund, Constructing Facial Expression Log from Video Sequences using Face Quality Assessment, Proc. of the Int. Conf. on Computer Vision Theory and Applications (VISAPP), pp. 1-8, 2014.

[25] D.G. Lowe, Object recognition from local scale-invariant features, Proc. of the Int. Conf. on Computer Vision (ICCV), pp. 1150-1157, 1999.

[26] Y. Feng et al., Exploiting Temporal Stability and Low-Rank Structure for Motion Capture Dara Refinement, Information Sciences, vol. 277, no. 1, pp. 777-793, 2014.

[27] L. Wolf, T. Hassner, and I. Maoz, Face Recognition in Unconstrained Videos with Matched Background Similarity, Proc. of the Int. Conf. on Computer Vision and Pattern Recognition (CVPR), 2011.