Face Quality Assessment for Face Verification in Video

M. Nikitin[1], V. Konushin[2], A. Konushin[1]
[1]Lomonosov Moscow State University, [2]Video Analysis Technologies, LLC
mnikitin@graphics.cs.msu.ru, vadim@tevian.ru, ktosh@graphics.cs.msu.ru

Abstract

Performance of biometric systems depends on quality of acquired biometric samples. Low sample quality is the main reason for matching errors in biometric systems and may be the principal weakness of some implementations. Therefore, when a biometric system obtains a sequence of person images from a surveillance camera, the quality of the different face images has to be evaluated before performing any analysis on the face of a person. In this paper, we propose an approach for face image quality assessment, which is based on four facial features including facial symmetry, sharpness, quality of illumination and the image resolution. To produce overall face quality score we perform weighted fusion of facial features with automatically tuned weights. Experimental evaluation of the proposed method has demonstrated its high accuracy and efficiency.

Keywords: *face quality assessment, facial symmetry, video surveillance.*

1. INTRODUCTION

When a person is observed by a surveillance camera, a sequence of images of that person is captured. Most of these images are useless due to problems like not facing the camera, motion blur, poor illumination and too small size of the region of interest in that image. For most biometric applications considering some (one or two) of the best images is sufficient to obtain accurate results. Therefore, there is a need for a mechanism, which can choose the best image from a sequence in terms of quality. This is called Quality Assessment. Automatic face quality assessment (FQA) can be used to monitor image quality for different applications such as face logging, video-based face classification [6] and identification [7].

Fig. 1 shows a framework of a face identification system using face quality assessment component. Face images are preprocessed and their quality is evaluated. Low quality images are discarded and only images with acceptable qualities are received for recognition. This allows to significantly accelerate matching speed in cases of large gallery. Moreover, in [13] it was shown that using quality assessment component in video-based face identification system can greatly improve its performance. Also, the quality score may be useful in image-based face recognition. For example, images of different qualities can be processed in different ways or high quality image may be necessarily required for reliable matching.

In different works related to FQA, different quality metrics were analyzed. X. Gao et al.[4] tried to standardize the quality of face images by facial symmetry based methods. Wong et al.[13] developed a method for simultaneous handling issues such as pose variations, cast shadows and blurriness, which quantifies the similarity of a given face to a probabilistic face model, representing an 'ideal' face, via patch-base local analysis. Fourney et al [3] proposed a quality fusion approach to combine head pose, sharpness, human skin presence, resolution, and two illumination measurements. Nasrollahi and Moeslund[8] proposed a similar quality assessment method, by using out-of-plane rotation, sharpness, brightness, and image resolution qualities.

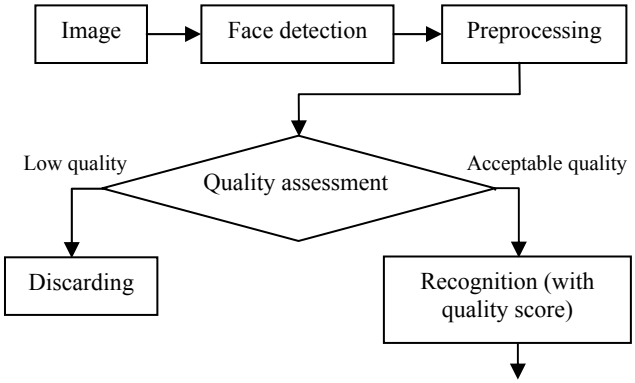

Fig. A framework of face recognition system with quality assessment component

In order to obtain an overall image quality score, it is necessary to combine all the scores of the used quality metrics, measuring different quality parameters. There are various methods of quality metrics fusion. Some works do it by thresholding each quality metric and counting the number of satisfied metrics. Others perform weighted averaging. The significant drawback of many existing works is that all thresholds and weights are obtained experimentally [3, 8].

In this paper, we propose a face quality assessment method, which is based on four facial features including facial symmetry, sharpness, illumination quality and face size. In order to combine our facial quality features we perform weighted averaging. For automatic weights tuning we adapted Ozay's et al [9] face recognition match score based technique.

The majority of existing pose estimation based FQA methods use an analysis of gradient image in order to locate left and right sides and face's axis of symmetry. Such method is not stable when subjects are wearing glasses, or when faces are not upright. To avoid this problem we perform facial symmetry evaluation based on feature points from facial features detector. In addition, using facial points allowed us to conduct sharpness analysis more accurately. In practice modern facial features detectors[5,10] only take about one millisecond per image and that is why they can be used as a part of FQA algorithm without notable performance loss.

There are two main contributions in this paper. The first contribution is that we proposed a new facial symmetry estimation method, which is based on the analysis of face local features. The second contribution is that we adapted weights tuning technique of low-level features for our high-level facial features.

2. PROPOSED ALGORITHM

When determining the quality of a face image I, we consider four facial features, weighted to varying degrees of importance. These features are facial symmetry, sharpness, illumination and resolution. The following sections describe how each of these features can be measured, and how they contribute to the overall quality score of an image.

2.1. Facial symmetry

The pose variations and illumination unevenness are two main issues that cause serious performance degradation for the most existing biometric systems, because wide variations in pose and illumination direction can hide most of the useful face image features. We propose to use facial symmetry to assess quality degradations caused by improper facial pose and non-frontal lighting.

The symmetry may be analyzed using some local image features, e.g., the raw image pixel values, or locally-filtered pixel values. When a local filter is chosen properly, it provides a better basis for computing facial symmetry. The degree of similarity between image features at the corresponding left-right pixel locations provides local measures of symmetry. If the face image is strictly left-right symmetric, the similarity scores should all take the maximum value.

We propose (see Fig. 2) to use similarity score between histograms H^L and H^R of facial local features as local measures of symmetry, because local feature histograms are more stable relative to slight misalignments than the local features itself. We use FaceSDK library [2] to find the locations of facial features. One of the compared patches (left or right) is horizontally mirrored before histogram computation. To find similarity score of two normalized histograms we use histogram intersection distance, which can be calculated as follows:

$$d_{int}(l) = \sum_{i} \min(H_i^L(l), H_i^R(l)), \quad (1)$$

where l denotes l-th pair of symmetric facial points. The larger the intersection distance value is, the larger the left-right symmetric of the face image is, and the larger the image quality is in some aspects.

Facial symmetry should be measured based on pose-sensitive image local features. In this paper, the Histograms of Oriented Gradients (HoG)[1] are used for this purpose.

The facial symmetry score S_1 is calculated as the mean value of all the histogram distances:

$$S_1 = Symmetry(I) = \frac{1}{N}\sum_{i=1}^{N} d_{int}(l), \quad (2)$$

where N is the number of pairs of symmetric facial points. The larger the $Symmetry(I)$ value, the more the face is symmetric.

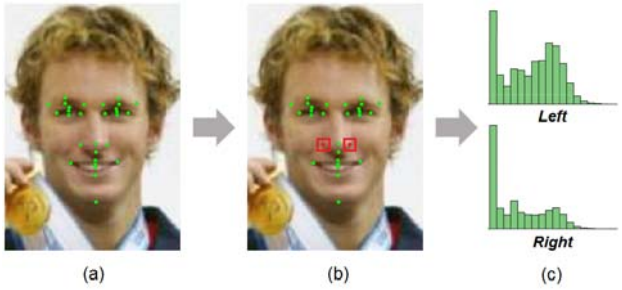

Fig. 2. (a) Face image with detected local facial features. (b) Used areas of symmetric points for local feature histogram computing. (c) Comparing left-right histograms of oriented gradients H^L and H^R

2.2. Sharpness

Since in real world applications the objects are moving in front of the camera, it is possible that the captured image is affected by motion blur, so defining a sharpness feature can be useful for FQA. The sharpness of a face image refers to the degree of clarity in both coarse and fine details in the face region. Well-focused images, which have a better sharpness compared to blurring images, should get a higher score for this feature.

We use a modified discrete Laplace operator to estimate image sharpness:

$$L(I) = \left|\frac{\partial^2 I}{\partial x^2}\right| + \left|\frac{\partial^2 I}{\partial y^2}\right|. \quad (3)$$

The Laplace operator is an example of a second order method of image spatial filtering. It is particular good at finding the fine details in an image. Any feature with a sharp discontinuity will be enhanced by a Laplace operator. The discrete second derivatives can be computed as convolution with the following kernels: $(1,-2,1)$ and $(1,-2,1)^T$.

We perform sharpness estimation strictly inside the facial area (see Fig. 3). For this purpose, we construct a mask using facial points found by face features detector. The mask is constructed in such a way that there would not be background pixels on masked image for any allowable face pose. The image sharpness score S_2 is calculated as the averaged Laplace operator response by masked image. Such scheme allows to achieve independence of sharpness score from image background.

Fig. 3. (a) Face image with detected local facial features and mask contour. (b) Masked image. (c) Response of modified Laplace operator on masked image

2.3. Illumination quality

Variations caused by changes in illumination constitute yet another significant challenge encountered by automated biometric systems. In order to compensate for different lighting conditions some implementations may perform histogram equalization or similar histogram dependent techniques in order to normalize an image before its analysis. For this reason, it is highly important to begin with images which make the best (maximum) use of the available dynamic range.

We estimate quality of illumination by determining the length R of available range of gray intensities excluding 5% of the darkest and brightest pixels (see Fig. 4). The illumination quality score S_3 is simply the percentage of the total dynamic range represented in R:

$$S_3 = Illumination(I) = \frac{R}{256}. \quad (4)$$

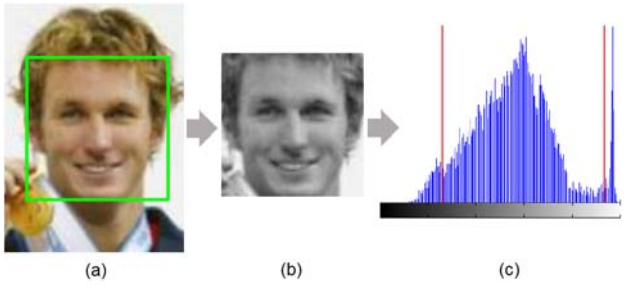

Fig. 4. (a) Face image with marked bounding box. (b) Cropped grayscale image. (c) Pixel intensity histogram; used range R marked with red vertical lines

2.4. Face size

The face image resolution score is perhaps the easiest of the aforementioned quality features to measure. The face size quality score S_4 is defined to be the linear function of the size of its bounding box. In general, high resolution images are preferred over low resolution images. We define lower threshold for face size as 50 pixels and upper threshold as 150 pixels. If bounding box size is below the lower threshold, face size score takes the minimum value, and in the case of bounding box size is beyond the upper threshold, face size score takes the maximum value, because it is no longer useful to achieve higher resolution:

$$S_4 = Size(x) = \begin{cases} 0, & x < 50 \\ 0.01x - 0.5, & 50 \le x \le 150, \\ 1, & x \ge 150 \end{cases} \quad (5)$$

where x is the bounding box size.

2.5. Overall quality score

Each of the four facial quality features discussed in the previous sections can score in the range $[0,1]$, but these features should not contribute equally to an overall quality score $q(I)$ of image I. For this reason, they are combined according to the weighted sum

$$q(I) = w_0 + \sum_{i=1}^{4} w_i S_i, \quad (6)$$

where w_0 is the bias term and the coefficients $w_i, i = 1,2,3,4$ determine the impact, which the quality features have on the overall score. For automatic weights tuning Ozay's et al.[9] technique has been adapted.

We consider the problem of weights tuning as the problem of linear regression learning, where facial quality features S_i act as predictor variables and overall image quality score $q(I)$ acts as dependent variable. The main problem that arises with such approach is the difficulty of training set preparation. Indeed, it is difficult for human to quantify the image quality, especially if it is necessary to consider several factors. For this reason, we decided to obtain overall quality scores for objects of training set using the information about their mutual similarity (in terms of face recognition match scores). Let us define it more formally.

A matching algorithm $\mathcal{A}$ produces a score for a given pair of images:

$$s_{t_k t_l} = \mathcal{A}(t_k, t_l), \quad (7)$$

where t_k denotes the k^{th} image of the i^{th} individual in training set. Considering the match score as a similarity measure, a quality measurement algorithm should satisfy the following property: face image of a subject should be assigned a high quality score if it is similar to other images of the same subject while it is different from the image of other subjects. This rule makes it necessary to define a measure of the match score quality. In [11], the normalized match score (NMS) was proposed for this purpose. The NMS between the k^{th} and l^{th} images of i^{th} individual is defined as:

$$NMS(t_k, t_l) = \frac{s_{t_k t_l} - \mu_{t_k}(s_{non-match})}{\sigma_{t_k}(s_{non-match})}, \quad (8)$$

where $\mu_{t_k}(s_{non-match})$ and $\sigma_{t_k}(s_{non-match})$ are respectively the mean and standard deviation of the match scores between the image t_k and the images from other individuals $f \neq i$.

The NMS provides some information about the quality of the images t_k and t_l, but it is not symmetric in its arguments as the non-match score distribution will vary when NMS arguments are interchanged. This variation could be especially strong when the images are of different quality. Hence, it is not a good measure of the quality of the match score, which is symmetric by its definition.

In order to avoid mentioned problem, in [9] symmetric normalized match score ($SNMS$) was defined:

$$SNMS(t_k, t_l) = \frac{1}{2}\left(NMS(t_k, t_l) + NMS(t_l, t_k)\right). \quad (9)$$

$SNMS$ satisfies desired property: it will be high for high-quality image pairs and will be low for low-quality image pairs. Therefore, $SNMS$ can be used as a measure of the quality of a match.

Once we have a way to measure the quality of the match, the next step is to separate it into the qualities $q(t_k)$ and $q(t_l)$ of matched images. We model the quality of the match as the average quality of compared images:

$$\frac{q(t_k) + q(t_l)}{2} \approx SNMS(t_k, t_l). \quad (10)$$

If we have at least three images for each individual in the training dataset, the separation problem can be solved. In particular, by combining all the equations for an individual i, we obtain the following least squares problem:

$$\mathbf{A}\mathbf{q}_\mathbf{i} + \boldsymbol{\varepsilon} = \mathbf{y}_\mathbf{i}, \quad (11)$$

where

$$\mathbf{A} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & \dots & 0 \\ 0.5 & 0 & 0.5 & 0 & \dots & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & \dots & 0 & 0 & 0.5 & 0.5 \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{bmatrix},$$

$$\mathbf{q}_\mathbf{i} = \begin{bmatrix} q(t_1) \\ \vdots \\ q(t_N) \end{bmatrix}, \quad \mathbf{y}_\mathbf{i} = \begin{bmatrix} SNMS(t_1, t_2) \\ SNMS(t_1, t_3) \\ \vdots \\ SNMS(t_{N-1}, t_N) \end{bmatrix}.$$

Here, $\mathbf{A}$ is a $T \times N$ matrix with two non-zero elements in each row where N is the number of samples from individual i, and T is the number of possible pairs of $\mathbf{i}^{\mathbf{th}}$ individual images. When $N \ge 3$, $\mathbf{A}$ is a full column rank; hence the solution with minimum squared error is given by $\mathbf{q}_\mathbf{i} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}_\mathbf{i}$.

The sample qualities for all the images in the training set can be obtained using this separation scheme. Once we have a quality value assigned to each image in the training dataset, we can find coefficients $\mathbf{W} = (w_0, w_1, w_2, w_3, w_4)^T$ of the linear regression that would predict face image quality $q(I)$ based on values of facial quality scores $\mathbf{f}_\mathbf{I} = (1, S_1, S_2, S_3, S_4)^T$:

$$q(I) = W^T f_I. \quad (12)$$

3. EXPERIMENTS

First of all we have performed visual analysis of the proposed FQA method by estimating quality of video frames and sorting them in descending order. Fig. 5 illustrates the example of such ranking. In most cases, the relative position of sorted frames corresponds to the intuitive idea.

Further, we evaluated our face quality assessment method for the video-based face verification task on real-world YouTube Faces [12] dataset.

Fig. 5. Example of image ranking based on the proposed face quality assessment method. Numbers below images indicate corresponding quality score

3.1. Experimental setup

YouTube Faces (YTF) is a video dataset, which contains 3,425 videos of 1,595 different subjects downloaded from YouTube. It is collected in unconstrained conditions and contains large variations in pose, expressions, illumination, etc. All YTF video clips are provided with labels indicating the identity of a person appearing in each video. It also contains meta-data defining benchmark protocols for video-based face verification task.

YouTube Faces dataset follows a ten-fold, cross validation, pair-matching ('same'/'not-same') test. Specifically, 5,000 video pairs were randomly selected and divided into 10 subject-mutually exclusive splits. Each split contains 250 'same' and 250 'not-same' pairs. The goal is to determine for each split, which are the same and which are the not-same pairs, by training on nine remaining splits.

We perform video-based face recognition with FQA component in the following way:

1. Find the highest quality frame of a video sequence.
2. Use it as input for the image-based face recognition algorithm.

3.2. Experimental comparisons

The proposed FQA method was compared against two other methods: Fourney's et al.[3] and Wong's et al.[13] patch-based method. Our and patch-based methods were trained on PubFig[7] dataset. For selected frames matching, face recognition module from the FaceSDK library[2] was used. All experiments were conducted on the machine equipped with a 2.3GHz Intel Core i7-3610QM processor and 8GB of RAM.

Results for MATLAB implementations are presented in Table.

Method	Accuracy	FQA speed
Fourney et al.[3]	69.82%	142.5 fps
Wong et al.[13]	79.92%	1.6 fps
Proposed	74.46%	28.5 fps

The comparison on the YouTube Faces dataset

The results indicate that patch-based FQA outperforms our method when considering verification accuracy. However, it works much slower. So to handle a pair of videos with a frame rate of 24 fps and a duration of 3 seconds each, patch-based method will take approximately 90 seconds for frames quality estimation, while the proposed method can do it in 5 seconds. In this way, our method provides a trade-off between speed and accuracy.

4. CONCLUSION

In this paper, we presented a new face image quality assessment algorithm, which is based on evaluating a set of facial features. The proposed approach is capable of handling issues such as pose and illumination variations, motion blurring and insufficient face image resolution. We developed a new face symmetry estimation method, which is based on analysis of symmetrically located face features. To get the overall face image quality weighted averaging of facial features is perform. For automated weights learning we have adapted Ozay's et al [9] technique.

Our FQA method has been applied as a quality assessment component in video-based face verification system. Comparison with other methods has demonstrated that our method provides a trade-off between accuracy and efficiency of the verification system.

5. ACKNOWLEDGMENTS

This paper was supported by RFBR, research project No. 14-01-00849 a.

6. REFERENCES

[1] Dalal N. and Triggs B. Histograms of Oriented Gradients for Human Detection. In Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 2, 2005. – P. 886 – 893.

[2] FaceSDK – face analysis library. http://www.tevian.ru/ru/products/facesdk

[3] Fourney A. and Laganiere R. Constructing Face Image Logs that are Both Complete and Concise. In 4th Canadian Conference on Computer Vision and Robot Vision, 2007. – P. 488 - 494

[4] Gao X., Li S.Z., Liu R. and Zhang P. Standardization of face image sample quality. In Proc. Int. Conf. Biometrics, 2007. – P. 242 - 251.

[5] Kazemi V. and Josephine S. One Millisecond Face Alignment with an Ensemble of Regression Trees. In Computer Vision and Pattern Recognition (CVPR), 2014.

[6] Konushin V., Lukina T., Kuharenko A., Konushin A. Simile classifiers for face classification. In Proc. GraphiCon, 2012. – P. 108 - 112.

[7] Kumar N., Berg A., Belhumeur P. and Nayar S. Attribute and Simile classifiers for face verification. In Proc. ICCV, 2009. – P. 365 - 372.

[8] Nasrollahi K. and Moeslund T.B. Face quality assessment system in video sequences. In BIOID, Lecture Notes in Computer Science (LNCS), vol. 5372, 2008. – P. 10 - 18.

[9] Ozay N., Tong Y., Frederick W. and Liu X. Improving face recognition with a quality-based probabilistic framework. In Computer Vision and Pattern Recognition (CVPR) Biometrics Workshop, 2009. – P. 134 - 141.

[10] Ren S., Cao X., Wei Y. and Sun J. Face Alignment at 3000 FPS via Regressing Local Binary Features. In Computer Vision and Pattern Recognition (CVPR), 2014.

[11] Tabassi E., Wilson C.L. and Watson C.I. Fingerprint image quality. Technical report, NIST, 2004.

[12] Wolf L., Hassner T. and Maoz I. Face Recognition in Unconstrained Videos with Matched Background Similarity. In IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2011. – P. 529 - 534.

[13] Wong Y., Chen S., Mau S., Sanderson C. and Lovell B.C. Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2011. – P. 74 – 81.