Development of Face Image Quality Assessment Algorithms for Biometric Identification Tasks

Vladimir Khryashchev, Anatoly Sedov, Andrey Priorov

P.G. Demidov Yaroslavl State University
Yaroslavl, Russia
e-mail: vhr@yandex.ru, agsedov@gmail.com, andcat@yandex.ru

ABSTRACT

Video surveillance has a wide variety of applications for indoor and outdoor scene analysis. When a person is observed by a surveillance camera, usually it's possible to acquire multiple face images of one person. Most of these images are useless due to the problems like motion blur, poor illumination, 3d face rotation, etc. For most biometric applications, getting several best images is sufficient to obtain accurate results. Therefore, there is a task of developing a low complexity algorithm, which can choose the best image from the sequence in terms of quality. Automatic face quality assessment can be used to monitor image quality for different applications. Proposed face quality assessment method has been applied as a quality assessment component in video-based audience analysis system. Using the proposed quality measure to sort the input sequence and taking only high-quality face images, we have successfully demonstrated that it increases the recognition accuracy.

Keywords

Image quality assessment, face recognition, video surveillance.

1. INTRODUCTION

Usually, when a person is in front of a surveillance video camera, several images of his face are saved to storage. Most of them are useless for the biometric identification system for several reasons: human's movement leads to blurring, a person can be in low-light conditions, only a part of the face or significantly turned face may be recorded. Human identification algorithms are computationally complex enough, so the recognition of the entire sequence of images can slow down the work of video surveillance systems [1,2]. Thus, the problem of choosing images of the best quality from all the received images, by which the person identification will be performed, is important.

There are several standards that determine the face image quality - ISO/IEC 19794-5, ICAO 9303 [3,4]. They contain a description of the characteristics that influence the decision on the suitability of image for automatic recognition systems. All the standardized characteristics are grouped into two classes: the texture (sharpness, contrast and light intensity, compression ratio, other distortions), and characteristics directly related to the face features (symmetry, position, rotation, eyes visibility, the presence of glare or shadows on the face).

Unfortunately, in most of practical applications, such as a video surveillance or audience analysis system, it is impossible to satisfy the requirements of ISO/IEC and ICAO standards. Lighting conditions, face posture, compression algorithm and camera type depend on concrete application. Nevertheless, the task of selecting the best quality face image from the video sequence in this case is also important and should be solved in a short period of time.

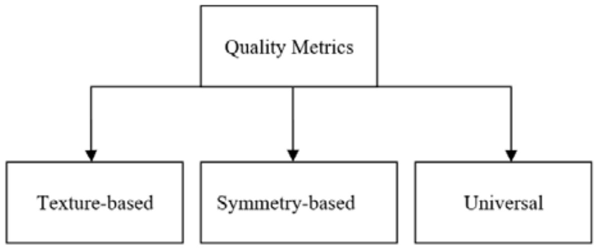

Figure 1. Face quality assessment metrics classification

To solve the problem, various modifications of image quality assessment algorithms are used.

The face quality assessment metrics, that we used for comparison, could be distinguished into three groups as it is shown in Figure 1. The goal of texture-based metrics is to detect if an image is distorted or not. Typical image distortions are compression artifacts, blur, and various types of noise. The goal of symmetry-based metrics is to determine if image is frontal or not. The output of this metrics is a real number showing how the current face pose differs from the frontal one.

The problem is commonly touched in modern scientific literature [5-8]. One of the first approaches to solve this problem is a method based on the application of clustering algorithm according to K-means approach [5]. The experiments have shown that the method has low accuracy when there are many low quality faces in the received sequence. A totally different approach searches the best quality images for face recognition by making quality evaluation of all images [2,6].

The quality of face images is estimated at the pre-processing stage. Depending on the algorithm used in concrete application one (top1) or three (top3) the best quality images are selected. Further they are used for gender classification in audience analysis system or for face verification/identification in video surveillance systems.

Low quality images are removed or archived, the recognition applies only to high-quality images. It is shown in [2] that the use of face image quality evaluation module increase speed of the surveillance system.

There is a group of face image quality assessment algorithms that uses objective methods to determine standardized facial quality characteristics. Overall facial image quality is obtained by combining the results of these methods. This group of algorithms is called metric fusion algorithms. The metric fusion can be done by thresholding each of characteristic value. In this case, the residual quality would be a number of characteristics above threshold. Another approach of metric fusion assigns a weight to every measured standardized characteristic [7]. Machine learning methods are widely used to determine metrics weights. It should be noted that metric fusion algorithms are tied to a specific database of training images, as well as to a specific recognition system. To solve this problem, a fundamentally different approach for measuring facial image quality, without using standardized facial quality characteristics (for

example, a statistical method based on the face model [6] or the method based on the learning to rank [8]) was created.
In none of the available papers the expert evaluation of face image quality was used. At the same time, expert opinions are widely applied in the analysis of images and video sequences quality [9]. It should be noted that a person can easily identify most of the standardized facial image quality characteristics.
Our main contribution is the proposed image quality assessment algorithm based on the method of learning to rank. We evaluate the performance of the proposed algorithm in practical situation of facial recognition.
In Section 2 we describe the universal quality assessment metric based on machine learning theory. In section 3 face recognition architecture is described. Section 4 contains the experimental results of applying this metric to select the best quality faces (top1 and top3) at the face verification/identification task (video surveillance systems).

2. LEARNING TO RANK METRIC

The disadvantage of the methods described in [2,6,10] is the fact that they do not take into account possible differences in recognition algorithms. For example, a recognition algorithm can accurately recognize faces, even if a part of the face is covered by another object, for example, by hand. For such an algorithm, faces with occlusion should not have a poor quality, whereas an algorithm which does not work accurately for faces with occlusion, it should be the opposite. Considering the drawbacks of the existing solutions, an algorithm based on the method of learning to rank is proposed here. This algorithm consists of two stages: normalization and quality control.

2.1. Quality control

Assume that the face recognition algorithm is tested on the databases *A* and *B*, and the algorithm based on the database *A* has a higher accuracy than the one based on the database *B*. In other words, for current recognition algorithm the images from the database *A* are of better quality than the images from the database *B*. Let's write it in the form: $A > B$. Two images I_i and I_j are selected from *A* and *B*, respectively. The function $f(\ )$, which input is an image and the output – a feature vector. Let's define a linear function of image quality $S(I) = w^T f(I)$. The goal is to find a vector w that would meet conditions (1-3) as much as possible, and we should consider that images from one database have the same quality image.

$$w^T f(I_i) > w^T f(I_j);\ \forall I_i \in A, \forall I_j \in B \tag{1}$$

$$w^T f(I_i) = w^T f(I_j);\ \forall I_i \in A, \forall I_j \in A \tag{2}$$

$$w^T f(I_i) = w^T f(I_j);\ \forall I_i \in B, \forall I_j \in B \tag{3}$$

The description above matches with the formulas from the paper [11], and, respectively, may be represented in the following terms:

$$\begin{gathered} \text{minimize}\left(\left\|w^T\right\|_2^2 + \lambda_1 \sum \xi_{i,j}^2 + \lambda_2 \sum \eta_{i,j}^2 + \lambda_3 \sum \gamma_{i,j}^2\right) \\ w^T\left(f(I_i) - f(I_j)\right) \geq 1 - \xi_{i,j};\ \forall I_i \in A, \forall I_j \in B \\ w^T\left(f(I_i) - f(I_j)\right) \leq \eta_{i,j};\ \forall I_i \in A, \forall I_j \in A \\ w^T\left(f(I_i) - f(I_j)\right) \leq \gamma_{i,j};\ \forall I_i \in B, \forall I_j \in B \\ \xi_{i,j} \geq 0, \quad \eta_{i,j} > 0, \quad \gamma_{i,j} \geq 0 \end{gathered} \tag{4}$$

This approach can be extended to a larger number of databases and features. If a mixture of signs is used, a two tier strategy should be used. Assume that m different feature vectors could be extracted from the image *I*. For *i*-th vector the quality will be calculated according to the formula $S_i(I) = w_i^T f_i(I), (i = 1,2,\dots,m)$. In the first phase of learning, vector weights are calculated according to the formula (4) for all of the various features. $\vec{S} = [S_1(I), S_2(I) \dots S_m(I)]^T$ is column vector containing various quality ratings for each attribute, respectively. In the author's implementation $m = 5$ is used. Let's define face image quality at step two as $S_k(I) = w_k f_\phi(\vec{S})$, where f_ϕ – polynomial function, which is represented by the expression (5).

$$\begin{aligned} f_\phi(\vec{S}) = \Bigg[& \left(\frac{1}{\sqrt{2}}\right) S_1 S_1^2 \left(\sqrt{2}\right) S_2 \left(\sqrt{2}\right) S_1 S_2 S_2^2 \left(\sqrt{2}\right) S_3 \left(\sqrt{2}\right) S_1 S_3 \\ & \left(\sqrt{2}\right) S_2 S_3 S_3^2 \left(\sqrt{2}\right) S_4 \left(\sqrt{2}\right) S_1 S_4 \left(\sqrt{2}\right) S_2 S_4 \left(\sqrt{2}\right) S_3 S_4 S_4^2 \\ & \left(\sqrt{2}\right) S_5 \left(\sqrt{2}\right) S_1 S_5 \left(\sqrt{2}\right) S_2 S_5 \left(\sqrt{2}\right) S_3 S_5 \left(\sqrt{2}\right) S_4 S_5 S_5^2 \Bigg]^T \end{aligned} \tag{5}$$

3. FACE RECOGNITION

Face recognition is a hot topic in computer science and an area of active research [2,6]. It can be used in a variety of applications: surveillance system, human-computer interfaces and audience analysis. A face recognition task can be divided into two separate tasks: identification and verification. During the first task we try to answer the question "who is this person?". During a verification task we validate the claimed identity based on the facial image (one-one matching).
There are many huge companies who are investing a lot of efforts into the face recognition systems: Facebook, Microsoft, NTechLab. But today such systems are available for everyone due to Openface software library [12]. Openface contains all parts of typical facial recognition system:

1. Detection
2. Normalization
3. Representation
4. Classification

Openface facial detector is made using the histogram of oriented gradients features combined with a linear classifier. An image pyramid and sliding window detection scheme is used. This functionality is implemented in dlib library which is used by Openface. Face detector output is the set of bounding boxes around a face. A bounding box with the biggest square is returned as a face detection result.
The variance between the detected face and faces from a database, causes a recognition system accuracy decline. One way of solving this problem is face normalization. Face normalization is a process of transformation of the input image in such a way that all facial landmarks are placed in predefined locations. One computational effective method of doing it is based on affine transform. It is used in Openface and requires facial landmarks as an input. Method [13] is used for facial landmarks detection. Coordinates of eyes and nose on the normalized image are calculated as a mean from appropriate coordinates from faces in a database. In addition the image is cropped to the size 96x96 pixels.
Normalized facial images are transformed to the compact vector representations with 128 dimensions by a deep neural network. The architecture of this network is described in detail in [14].

The neural network was trained on two large publicly available datasets Face-Scrub [15] and CASIA-WebFace[16]. Stochastic gradient decent was used as an optimization algorithm. Loss function was chosen as follows:

$$L = \sum_{i}^{N} f\left(x_i^a\right) - f\left(x_i^p\right)_2^2 - f\left(x_i^a\right) - f\left(x_i^n\right)_2^2 + \alpha, \qquad (6)$$

where x_i^a - the input image of a person i, x_i^p - the image of the same person from training database, x_i^n - the image of another person. The α value was selected during the training process and was equal to 0.2.

Person identity from training database is found at classification step. Open-face does not provide any classification functionality but a simple classifier can be built without any external libraries by measuring Euclidian distance between the facial image representations obtained at the previous step.

4. EXPERIMENTAL RESULTS

Test database containing facial images of 60 persons (60 Person Face Comparison Database – 60PFCD) was collected for experiments. Images were obtained in real-life situation at low lighting conditions (<100lx). There are 10 images for each person (Fig. 2).

Figure 2. Facial images from 60PFCD dataset

Standard face detector [17] was used to detect faces. The following measures were calculated for every detected face: image resolution, sharpness, symmetry, measure of symmetry of landmarks points S, quality measure K (based on learning to rank [8]) and two no-reference image quality metrics NRQ LBP and BRISQUE. In addition, the expert quality assessment had been conducted for each image with values ranging from 1 (worst quality) to 10 (best quality). A group of 10 experts defined the best quality image (top1) and the best 3 quality images for each 60 persons. To obtain a mean, we weighted the experts results – every image from the top-1 got the weight of 3, other images from the top-3 got weight of 2 (there was no difference between the second and the third quality images). We computed the investigated facial quality metrics scores for each image in the dataset.

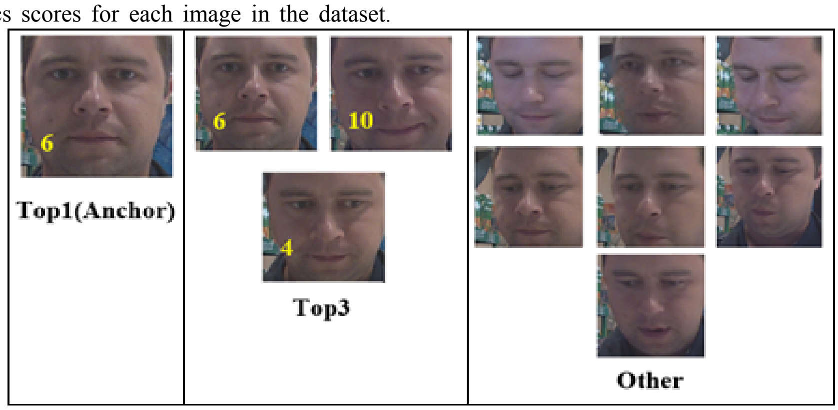

Figure 3. Anchor and top3 illustration

Tables 1-2 contain top-1 accuracy and top-3 accuracy results for objective face quality metrics. It is clear that metric K is more accurate than the other investigated metrics.

Table 1. Top-1 accuracy of facial image quality metrics (60PFCD dataset)

Res.	K	Symmetry	S	Sharpness	BRISQUE	NRQ LBP
6	20	10	7	7	7	4

Table 2. Top-3 accuracy of facial image quality metrics (60PFCD dataset)

Res.	K	Symmetry	S	Sharpness	BRISQUE	NRQ LBP
65	99	68	73	60	52	49

We used 60PFCD database to measure the accuracy of face recognition system described above. By accuracy we mean the number of correctly classified images divided by a total number of classified images. At the classification step we used a classifier based on so-called “anchor” images. An anchor image is a single image which was chosen from a person image sequence (Fig. 3). There were 3 anchor selection schemes:

1. Anchor image is chosen as the highest quality image by MOS metric value.
2. Anchor image is chosen as the first image in the sequence.
3. Anchor image is chosen as the highest quality image by learning to rank metric value.

For each anchor image we saved its vector representation to classifier memory. To classify some test image we measured the Euclidian distance between its vector representation and the anchors vector representation. We chose the anchor representation which had the smallest distance and assign the appropriate person id to the test image. To collect test images we selected a pair of images from each person images (anchor image cannot be in this pair). This pair was selected randomly or from the top3 images chosen by MOS metric.

Fig. 4a plots the accuracy results for 2 classifiers based on “anchor” images. For the first classifier, anchors are of the highest quality images by the MOS metric. For the second classifier, anchors are of the highest quality images by the learning to rank metric. Test images were chosen from the top3 quality image by MOS metric. It can be seen from the results that the classifier with anchors based on MOS metric is a little more accurate than the classifier with anchors based on learning to rank metric. These results are shown in Fig. 4b (test images were chosen randomly).

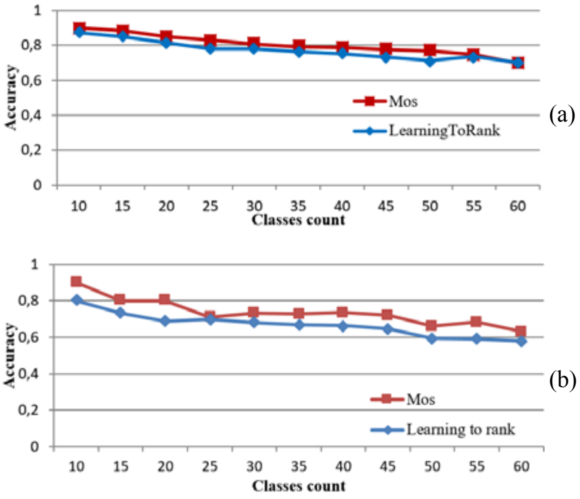

Figure 4. Accuracy of a face recognition system with different classifiers. First classifier uses the highest quality image by MOS metric as anchors. The second one uses the highest quality images by learning to rank metric. a) Test images are chosen from top3 images by MOS metric; b) Test images are chosen randomly.

Fig. 5 shows the experimental results for a classifier that uses the first image as an anchor and a classifier which uses the highest quality image based on learning to rank metric. Test images are selected randomly in this case. It can be seen from the results that the second classifier is more accurate. It means that facial image quality assessment can be used in a face recognition systems with a dynamic database (a new person is added during system operation) to select person images for storage.

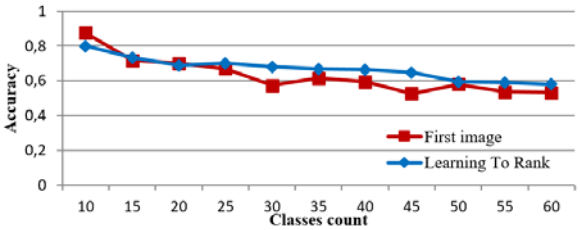

Figure 5. Accuracy of a face recognition system with different classifiers. First classifier uses the first person's image as anchors. The second one uses the highest quality images by learning to rank metric. Test images are chosen randomly.

5. CONCLUSIONS

A set of face image quality metrics were investigated in relation to the problem of selecting the best facial images for biometric identification. The usage of only the highest quality faces in a recognition process leads to accuracy improving and savings in computational resources. We obtained 15-18% accuracy increase in face recognition while using only three top quality images. In the experiment on a choice of three best pictures the measure based on learning to rank shows the best result. The results will be useful to engineers in building video surveillance and biometric identification using a facial image.

6. ACKNOWLEDGEMENT

This work was supported by the Russian Foundation for Basic Research grants (№ 15-07-08674).

REFERENCES

[1] W. Zhao, R. Chellappa, P. Phillips and A. Rosenfeld, "Face recognition: A literature survey", *ACM Computing Surveys (CSUR), vol. 35, № 4*, pp. 399–458, 2003.

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

[3] ISO/IEC 19794-5 (published version). *Information technology – Biometric Data Interchange Formats*, 2005.

[4] Machine readable travel documents. *International Civil Aviation Organization*, 2006.

[5] S. Berrani and C. Garcia, "Enhancing face recognition from video sequences using robust statistics", *IEEE International Conference on Video and Signal Based Surveillance (AVSS),* pp. 324–329, 2005.

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

[7] M. Nikitin, A. Konushin and V. Konushin, "Face quality assessment for face verification in video", Proceedings of GraphiCon'2014, pp. 111–114, 2014.

[8] J. Chen, Y. Deng, G. Bai and G. Su, "Face Image quality assessment based on learning to rank", *IEEE Signal Processing Letters, vol. 22, № 1*, pp. 90–94, 2015

[9] I. Nenakhov, V. Khryashchev and A. Priorov, "No-Reference Image Quality Assessment based on Local Binary Patterns", *Proceedings of the 14th IEEE EAST-WEST DESIGN & TEST SYMPOSIUM*, pp. 529–532, 2016.

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

[11] G. Zhang and Y. Wang, "A symmetry-based quality assessment of face images", *ISVC, Lecture Notes in Computer Science (LNCS), vol. 5876*, pp. 499–508, 2009.

[12] B. Amos, B. Ludwiczuk, and S. Mahadev, "OpenFace: A general-purpose face recognition library with mobile applications", *CMU-CS-16-118, CMU School of Computer Science, Tech. Rep.*, 2016.

[13] D. King, "Dlib-ml: A Machine Learning Toolkit", *Journal of Machine Learning Research, v. 10*, pp. 1755-1758, 2009

[14] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering"*arXiv preprint arXiv:1503.03832*, 2015.

[15] H. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets", *IEEE International Conference on Image Processing*, 2014.

[16] D. Yi, Z. Lei, S. Liao and S. Li. "Learning face representation from scratch", *arXiv preprint ar-Xiv:1411.7923*, 2014.

[17] J. Howse. OpenCV Computer vision with Python. *Birmingem: Packt Publishing Ltd.*, 2013.