Fusion of mSSIM and SVM for Reduced-Reference Facial Image Quality Assessment

Pengjun Ji, Yuchun Fang, Zhonghua Zhou, and Jiazhen Zhu

School of Computer Engineering and Science, Shanghai University,
200072 Shanghai, China
ycfang@shu.edu.cn

Abstract. Image Quality Assessment (IQA) is a critical part in face recognition system for helping to pick out the better quality images to assure high accuracy. In this paper, we propose a simple but efficient facial IQA algorithm based on Bayesian fusion of modified Structural Similarity (mSSIM) index and Support Vector Machine (SVM) as a reduced-reference method for facial IQA. The fusion scheme largely improves the facial IQA and consequently promotes the precision of face recognition when comparing to mSSIM or SVM alone. Experimental validation shows that the proposed algorithm works well in multiple feature spaces on many face databases.

Keywords: Image quality assessment, face recognition, biometrics, fusion.

1 Introduction

With the exploding increase in the number of installed video surveillance cameras and the deployment of face recognition software, the demand for high performance face recognition system is obvious, which also impels facial IQA to play a more important role. While factors such as illumination, pose variation, blur and focus change may drastically deteriorate recognition precision [1], thus it is necessary to take these factors into account to automatically evaluate quality of facial images.

For general IQA, there are several most popular full-reference schemes such as the Mean Squared Error (MSE) [2], Peak Signal-to-Noise-Ratio (PSNR) [3], the SSIM index [4] and Tone Rendering Distortion Index (TRDI) [5], but they all need reference images which are hard to obtain in practical application. Thus, no-reference and reduced-reference IQA arouse special attention. Natural Scene Statistics (NSS) model was applied by Sheikh et al. to blindly measure the quality of images [6]. Li et al. deployed a general regression neural network to assess quality of distorted images [7]. Li and Wang extracted a set of distortion detection features from a divisive normalization representation of the image [8].

For concrete application such as biometrics, IQA is designed to specially take the recognition performance into consideration. El-Abed et al. utilized a multi-class SVM-based method combining image quality and pattern-based quality to predict biometric sample quality [9]. Breitenbach and Chawdhry [10] assessed the impact of

W.-S. Zheng et al. (Eds.): CCBR 2012, LNCS 7701, pp. 75–82, 2012.

image quality on multimodal biometric recognition performance, including iris and face. Jain et al. [11] proposed a likelihood ratio-based fusion scheme that takes the quality of multi-biometric samples of iris and fingerprint into consideration.

In this paper, we propose a Bayesian fusion model to specially handle the quality degradation of facial image based on our previous work of facial image quality ranking with mSSIM and SVM [12]. Experimental results show that the fusion algorithm works well in several popular feature spaces on different benchmark databases.

This paper is organized as follows. In Section 2, we introduce the Bayesian fusion model of mSSIM and SVM. Section 3 analyzes the experimental results and the conclusions are drawn in Section 4.

2 Bayesian Fusion for Facial IQA

For biometric images such as face, pose variation and lighting condition are two main factors affecting the quality of image. It is hard for the general no-reference and reduced-reference methods to handle both problems. In our previous work [12], we proposed a Reduced-Reference Automatic Ranking (RRAR) method for facial IQA, which utilizes the mSSIM and SVM to handle the two problems respectively. In this paper, we use Bayesian decision rule to fuse the two modules.

2.1 mSSIM and SVM

Due to its relatively low computational complexity and coherence with human visual system, SSIM is adopted as the full-reference metric to solve the problem of “illumination unevenness” in facial IQA. It is proposed based on the hypothesis that human vision system is highly adaptive to structure [13]. Frontal facial image is a good example with strong structural appearance. However, usually there is no so-called “original image” as reference in face recognition system. By constructing the automatically updated average face as reference, we modify the normal SSIM into a reduced-reference IQA method. The mSSIM is simple yet robust [12].

In order to solve the problem of pose variation, we train a binary SVM classifier in the Uniform Local Binary Pattern (ULBP) feature space to discriminate the frontal and non-frontal facial images, and eliminate facial images with “out-of-plane” rotation [14]. The ULBP is not precise enough in personal identification but proves good enough in discriminating frontal and non-frontal face in our experiments with much lower feature dimension.

2.2 Bayesian Fusion Model

Depending only on the mSSIM, the influence of pose variation is ignored. While just considering the SVM classifier, the illumination condition is left behind. It is necessary to find a way to combine their merits. We adopt the Bayesian decision rule as a simple parametric fusion model, which aims at finding a classification surface in two-dimensional space constructed by the outputs of mSSIM and SVM.

Denote the sample as $x = (c_x, r_x)$ where c_x , r_x separately denotes the outputs of mSSIM and SVM classifiers. Let $p(F \mid x)$,$p(\overline{F} \mid x)$ be the posterior probability of good quality and poor quality images.

According to Bayesian formula,

$$p(F \mid x) = \frac{p(x \mid F)p(F)}{p(x \mid F)p(F) + p(x \mid \overline{F})p(\overline{F})} \tag{1}$$

$$p(\overline{F} \mid x) = \frac{p(x \mid \overline{F})p(\overline{F})}{p(x \mid F)p(F) + p(x \mid \overline{F})p(\overline{F})} \tag{2}$$

Thus we get the Bayesian decision rules with negative log likelihood ratio format as,

$$-\ln p(x \mid F) + \ln p(x \mid \overline{F}) \begin{matrix} > \\ < \end{matrix} \frac{\ln p(F)}{\ln p(\overline{F})} = Threshold \qquad x \in \begin{cases} F \\ \overline{F} \end{cases} \tag{3}$$

Because of the complexity of facial images, it is acceptable to assume that the outputs of SSIM and SVM are independent, thus

$$p(x \mid F) = p(c_x, r_x \mid F) = p(c_x \mid F)\, p(r_x \mid F) \tag{4}$$

$$p(x \mid \overline{F}) = p(c_x, r_x \mid \overline{F}) = p(c_x \mid \overline{F})\, p(r_x \mid \overline{F}) \tag{5}$$

The final decision rule is,

$$-[\ln p(c_x \mid F) + \ln p(r_x \mid F)] + [\ln p(c_x \mid \overline{F}) + \ln p(r_x \mid \overline{F})] \begin{matrix} > \\ < \end{matrix} \ln \frac{p(F)}{p(\overline{F})} = Threshold,\ x \in \begin{cases} F \\ \overline{F} \end{cases} \tag{6}$$

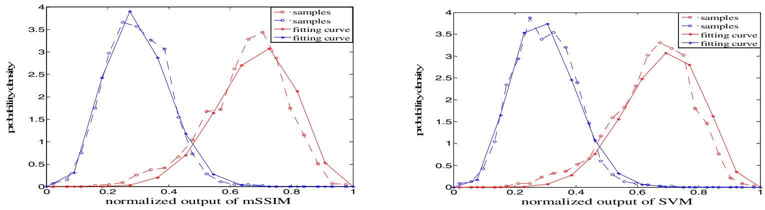
2.3 Fitting of Probability Density Function

To estimate the four posterior probabilities mentioned above, Maximum Likelihood Estimate (MLE) is utilized to fit several distributions, including Normal distribution, Beta distribution, Gamma distribution, Weibull distribution and the non-parametric Parzen Window respectively, then select the most suitable distribution with statistical Kolmogorov-Smirnov test.

Table 1 shows the statistical results for each distribution where the confidence coefficient is 0.05. It can be observed that $p(c_x \mid F)$, $p(c_x \mid \overline{F})$, $p(r_x \mid F)$ and $p(r_x \mid \overline{F})$ separately fit the Parzen Window distribution, the Parzen Window distribution, the Gamma distribution and the Beta distribution. The fitting results on our self-collected database are shown in Figure 1 (a) and (b), where the dotted line denotes histogram of samples, while solid line is the curve of probability density function. As can be observed, the estimation is very close to the real distribution of samples.

Table 1. Kolmogorov-smirnov test results

	Nomal	Beta	Gamma	Weibull	Parzen	critical value
$p(c_x \mid F)$	0.1383	0.0952	0.1325	0.4632	**0.0496**	0.0837
$p(c_x \mid \overline{F})$	0.0473	0.1373	0.0864	0.0415	**0.0247**	0.0289
$p(r_x \mid F)$	0.0636	0.1453	**0.0620**	0.6118	0.0932	0.0657
$p(r_x \mid \overline{F})$	0.0435	**0.0271**	0.0382	0.0329	0.0577	0.0315



(a) (b)

Fig. 1. Fitting results of probability density function on self-collected database: (a) mSSIM ($p(c_x \mid F)$ and $p(c_x \mid \overline{F})$ fit with the Parzen Window distribution); (b) SVM ($p(r_x \mid F)$ and $p(r_x \mid \overline{F})$ separately fits the Gamma and the Beta distribution)

3 Experimental Results and Analysis

Normally, IQA measures quality based on reference images. In the case of face recognition, the IQA serves to improve the recognition accuracy. Hence, we use the accuracy of face recognition to evaluate the effect of the IQA method. To validate the results of IQA, we perform face recognition tests in several popular feature spaces on several face databases.

3.1 Face Databases and Feature Spaces

The experiments are performed on four face databases listed in Table 2. D1 is collected in our lab without lighting constraint. D2, D3 and D4 are subsets randomly selected respectively from expression database of CAS-PEAL [15], FERET [16] and FRGC. D2 contains expression variations. D3 contains lighting, aging, pose, expression and decoration variations. D4 contains images collected in uncontrolled environments with complex lighting variations.

Table 2. Face databases

	Origin	#Subjects	#Images per Subject	Complexity
D1	Self-collected	50	10	Uncontrolled
D2	CAS-PEAL	100	5	Controlled
D3	FERET	256	4	Controlled
D4	FRGC	459	6	Uncontrolled

For each database, we classify facial images of each subject as better and poorer quality subsets in the following way: the facial images of each subject are sorted in descending order according their quality scores obtained from the fusion model, the better half are regarded as the Better Quality Set (BQS), while the other half are denoted as the Poorer Quality Set (PQS). The union of BQS and PQS is the set of all images (ALLS). On BQS, PQS and ALLS of all 4 databases, we respectively extract the ULBP, Principal Component Analysis (PCA) feature and Linear Discriminant

Analysis (LDA) features to evaluate the proposed facial IQA method with the Nearest Neighbor classifier.

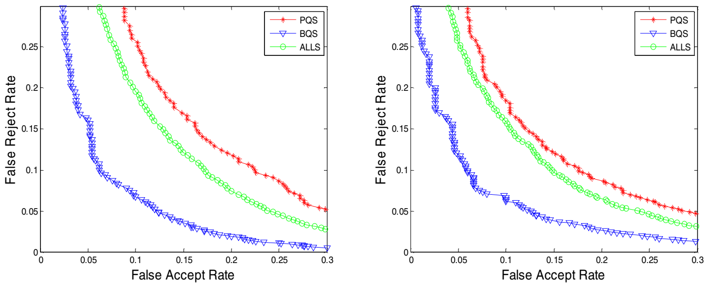
3.2 Effectiveness of the Proposed IQA Method

To verify the effect of the proposed IQA model, we perform face verification experiments and adopt the Equal Error Rate (EER) point of False Rejection Rate (FRR) and False Acceptance Rate (FAR) to measure accuracy as in Table 3.

Table 3. ERRs of the BQS, ALLS and PQS in 3 feature space on 4 face databases

Set	Method	D1	D2	D3	D4
PCA	BQS	**0.0949**	**0.0650**	**0.1500**	**0.3428**
	ALLS	0.1435	0.0654	0.2056	0.3435
	PQS	0.1453	0.0908	0.2241	0.3698
LDA	BQS	**0.0573**	0.0412	**0.1590**	**0.2800**
	ALLS	0.0643	**0.0407**	0.2047	0.2928
	PQS	0.0692	0.0516	0.2252	0.3241
ULBP	BQS	**0.0974**	**0.0820**	**0.0860**	**0.0746**
	ALLS	0.1294	0.1364	0.1318	0.1272
	PQS	0.1594	0.1596	0.1581	0.1356

It can be observed that for PCA, LDA and ULBP, the BQSs selected with the Fusion model are all of prominently better performance compared with PQS for the 4 face databases. On D2, the accuracy of BQS in LDA feature space is slightly worse than that in ALLS. This is because that the LDA feature works well on the relatively good quality image dataset D2. However, in all the other cases, the accuracy on BQS outperforms that on ALLS. Such results demonstrate that our IQA model adapts to different kinds of features and works well on different databases.



(a) (b)

Fig. 2. Comparison among BQS, PQS and ALLS in ULBP feature space on (a) D2 and (b)D4

We also compare the performance with the Receive Operating Characteristic (ROC) Curve of FRR with respect to FAR. An example with the ULBP feature is shown in Figure 2(a) for D2 with relatively high quality images from CAS-PEAL and Figure 2(b) for D4 with images obtained in uncontrolled environment from FRGC. The promotion of accuracy with the proposed model is about the same for both image sets.

3.3 Comparison of the Fusion Algorithm with the Other Methods

In the four experimental databases, we also compare the performance of the proposed IQA method with the non-fusion IQA methods through face verification test in the ULBP feature space. The EERs of FAR and FRR are shown in Table 4.

Table 4. ERRs of three algorithms on ALLS

Method	D1	D2	D3	D4
SVM	0.1789	0.1707	0.2339	0.2645
mSSIM	0.1520	**0.1359**	0.1817	0.1966
Proposed method	**0.1294**	0.1364	**0.1318**	**0.1272**

It can be observed that the fusion algorithm performs much better than both mSSIM and SVM in all the other cases except for the results on D2. In D2, the pose variation is so small that SVM does not contribute much for the fusion algorithm. Hence, the fusion algorithm has very close performance with that of mSSIM. From Table 4, it can also be observed that mSSIM outperforms SVM due to its flexibility to structural information in IQA.

3.4 Examples of Quality Ranking

To further check the effectiveness of the proposed facial IQA method, visual comparisons are performed on the four databases. For each subject, his or her images are ranked according to the quality scores obtained with the proposed fusion model in descending order. Some of the results are shown in Figure 3, quality scores of three methods are listed under each subject.

It can be observed that in the first two columns, the illumination around facial region is more even and the pose is more upright than the other two columns. The facial IQA results obtained with the proposed model reflect the quality of image when taking illumination and pose variations into account together. It's obvious that our fusion method shows reasonable consistency with human perception.

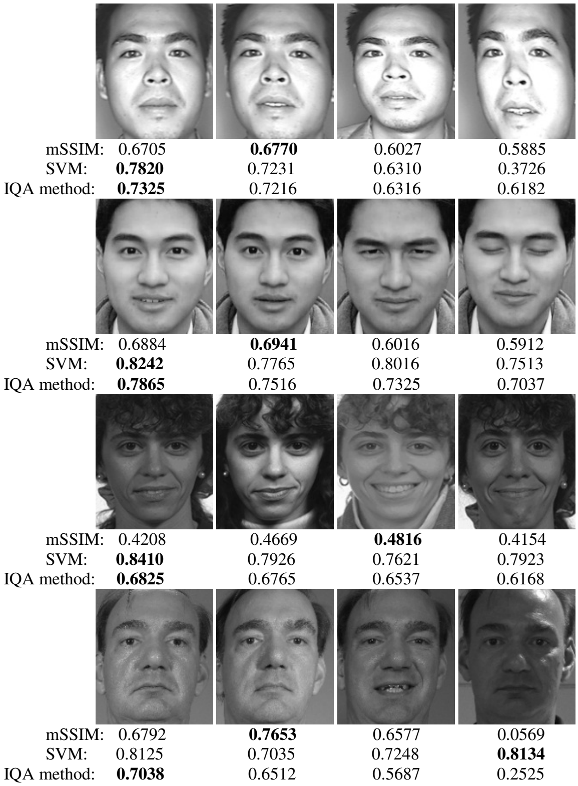


Fig. 3. Examples of quality ranking results of the fusion model in descending order from left to right, each row one subject from one database (Row 1: Pose changes from upright to slanting; Row 2: expression changes from slight to severe; Row 3: combined changes of expression and lighting; Row 4: illumination becomes more uneven)

4 Conclusion

In this paper, we propose a fusion facial IQA algorithm based on the Bayesian rule, it fuses mSSIM and SVM to rank the quality of facial images with a simple decision function. The proposed model can prominently improve the precision of face recognition compared to mSSIM or SVM alone. Experimental results show that the proposed algorithm works well in multiple feature spaces on multiple face databases. The ranking results of facial image quality are coherent with the human visual system considering illumination and pose variations.

Acknowledgments. The work is funded by the National Natural Science Foundation of China (No.61170155) and the Shanghai Leading Academic Discipline Project (No.J50103).

References

1. Zamani, A.N., Awang, M.K., Omar, N., Nazeer, S.A.: Image Quality Assessments and Restoration for Face Detection and Recognition System Images. In: Second Asia International Conference on Modeling & Simulation, AICMS 2008, pp. 505–510 (2009)
2. Zhou, W., Bovik, A.C.: A universal image quality index. IEEE Signal Processing Letters 9, 81–84 (2002)
3. Zhou, W., Qiang, L.: Information Content Weighting for Perceptual Image Quality Assessment. IEEE Transactions on Image Processing 20, 1185–1198 (2011)
4. Zhou, W., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13, 600–612 (2004)
5. Jang-Kun, S., Seung Beom, P.: Assessment of Image Quality Degraded by Tone Rendering Distortion. Journal of Display Technology 7, 365–372 (2011)
6. Sheikh, H.R., Bovik, A.C., Cormack, L.: No-reference quality assessment using natural scene statistics: JPEG2000. IEEE Transactions on Image Processing 14, 1918–1927 (2005)
7. Chaofeng, L., Bovik, A.C., Xiaojun, W.: Blind Image Quality Assessment Using a General Regression Neural Network. IEEE Transactions on Neural Networks 22, 793–799 (2011)
8. Qiang, L., Zhou, W.: General-purpose reduced-reference image quality assessment based on perceptually and statistically motivated image representation. In: 15th IEEE International Conference on Image Processing, ICIP 2008, pp. 1192–1195 (2008)
9. El-Abed, M., Giot, R., Hemery, B., Charrier, C., Rosenberger, C.: A SVM-based model for the evaluation of biometric sample quality. In: 2011 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM), pp. 115–122 (2011)
10. Breitenbach, L., Chawdhry, P.: Image quality assessment and performance evaluation for multimodal biometric recognition using face and iris. In: Proceedings of 6th International Symposium on Image and Signal Processing and Analysis, ISPA 2009, pp. 550–555 (2009)
11. Nandakumar, K., Yi, C., Jain, A.K., Dass, S.C.: Quality-based Score Level Fusion in Multibiometric Systems. In: 18th International Conference on Pattern Recognition, ICPR 2006, pp. 473–476 (2006)
12. Jiazhen, Z., Yuchun, F., Pengjun, J., Abdl, M.E., Wang, D.: RRAR: A novel reduced-reference IQA algorithm for facial images. In: 2011 18th IEEE International Conference on Image Processing (ICIP), pp. 3313–3316 (2011)
13. Zhou, W., Guixing, W., Sheikh, H.R., Simoncelli, E.P., En-Hui, Y., Bovik, A.C.: Quality-aware images. IEEE Transactions on Image Processing 15, 1680–1689 (2006)
14. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 971–987 (2002)
15. Wen, G., Bo, C., Shiguang, S., Xilin, C., Delong, Z., Xiaohua, Z., Debin, Z.: The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 38, 149–161 (2008)
16. Phillips, P.J., Hyeonjoon, M., Rauss, P., Rizvi, S.A.: The FERET evaluation methodology for face-recognition algorithms. In: 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Proceedings, pp. 137–143 (1997)