Application of the central weighted structural similarity index for the estimation of the face recognition accuracy

Paweł Forczmański[1*], Krzysztof Okarma[2] [†]

[1] *Chair of Multimedia Systems,*
West Pomeranian University of Technology, Żołnierska 49, 71-210 Szczecin, Poland
[2] *Chair of Signal Processing and Multimedia Engineering,*
West Pomeranian University of Technology, 26. Kwietnia 10, 71-126 Szczecin,
Poland

Abstract

In the paper a novel method for the estimation of the face recognition accuracy based on the modified Structural Similarity is presented. A typical application of the Structural Similarity index is related to the full-reference objective image quality assessment. Growing popularity of this metric is caused not only by the fact of its relatively low computational complexity but also by its sensitivity to three common types of distortions: the loss of contrast, luminance distortions and the loss of correlation.

Taking into account the output of the SSIM metric as the quality map with the resolution nearly the same as that of the input images, it is possible to use any two–dimensional central weighting function to control the level of importance of each image region. The approach proposed in this article is based on the usage of the Central Weighted SSIM index for the prediction of the face recognition accuracy using the images contaminated by several common types of distortions e.g. salt and pepper noise, lossy compression, filtration etc. The described

———————
[*]*E-mail address:* pforczmanski@wi.zut.edu.pl
[†]*E-mail address:* okarma@zut.edu.pl

method is based on the assumption that facial portraits are cropped and centered, which is true for almost all biometric systems. Finally, the results of face recognition by means of PCArc method has been used, as the state–of–the art in this domain. The experiments were conducted on the Olivetti Research Lab database [1].

1. Introduction

Rapid progress in the field of digital image quality assessment, taking place at the beginning of the 21st century, radically changed the methods used for the image and video quality estimation. Instead of well known traditional metrics based on the Mean Squared Error and similar measures, some modern approaches have been proposed. Most of them are also full-reference metrics but their main advantage is much better correlation with the human perception of some typical image distortions like noise, blur, lossy compression artifacts etc. Probably the most popular metric of that kind is the Structural Similarity index analyzed in this paper.

Full–reference approach to the digital image quality assessment is currently the only universal one since the blind assessment methods proposed so far are specialized so their usefulness is strongly limited. The knowledge of the original image is the price we should pay for that universality.

In our paper the modification of the Structural Similarity index is proposed as the method useful for the estimation of the face recognition accuracy assuming the centrally located faces in the previously cropped images which is a typical situation in most face recognition systems.

A reliable image quality information can be used in many image recognition and classification systems in order to forecast the recognition accuracy in the presence of some typical distortions caused e.g. by low bandwidth transmission or lossy compression.

In this paper the usefulness of the modified (Central Weighted) Structural Similarity metric to forecast the face recognition accuracy in the presence of some typical distortions is analyzed. As the face recognition method the two–dimensional Principal Component Analysis algorithm is used. All the experiments are based on standard benchmark database of faces: AT&T Research Lab (known formerly as Olivetti Research Lab) [1]. Sample images used in the experiments are presented below in Fig. 1. There are ten different images of each of 40 distinct subjects. For most subjects, the images were taken in one session, varying the lighting, facial expressions and facial details. All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with a slight tolerance for some side movement). The size of each image is 92×112 pixels, with 256 gray levels per pixel.

Fig. 1. Example images from AT&T database used in experiments

2. PCArc algorithm for face recognition

The problem of automatic image recognition involves gathering, processing and storing large amounts of data. Reducing very high-dimensional feature spaces leads to speeding-up the computations. A method of Principal Component Analysis (PCA, also known as Karhunen–Loeve Transform – KLT) is commonly used to reduce their dimensionality, hence providing much help when it comes to recognition. Classical PCA is applicable only for images with rather low spatial resolution [4]. Moreover, using PCA for tasks such as image recognition or image retrieval can be challenging because it treats the data as one–dimensional, when, in fact, they are two–dimensional. That is why almost all presented algorithms involve some sort of dimensionality pre–reduction discarding in many cases the spatial relations between pixels.

One of the possible solutions of this problem is using two-dimensional transformation based on PCA. The first algorithm from this group was presented in [3], where a novel, two–dimensional version of PCA for face recognition task was developed. An extension of this method (as PCArc – for row and column representations) was presented in [4].

The algorithm of recognition discussed in this paper is composed of three basic elements. It is presented in Fig. 2. First, two–dimensional Fast Fourier Transform (FFT) is performed on each image giving a matrix containing spectral components selected from the lower part of the amplitude spectrum [5]. Then it is reduced by means of PCArc producing final matrix of output features [6]. The last stage is the classification performed using the Euclidean distance calculation.

The detailed algorithm of recognition is presented below. Let us assume that the whole dataset contains K classes of images. Each class represent a single
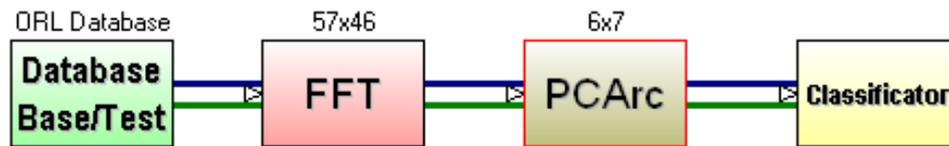
Fig. 2. Recognition scheme

person and consists of L images. Each image $X^{(k,l)}$ contains $M \times N$ gray–scale pixels.

In the first step of learning stage, for all images in the dataset we calculate global mean image $\overline{X}$ and mean image for each class $\overline{X}^{(k)}$ (used for distance calculation in the recognition stage). Then we remove the global mean image from each single image:

$$\hat{X}^{(k,l)} = X^{(k,l)} - \overline{X}, \qquad \forall k = 1, \ldots, K, \qquad \forall l = 1, \ldots, L.$$

In the next step we calculate two covariance matrices for both row representations of images:

$$R = \sum_{k=1}^{K} \sum_{l=1}^{L} \hat{X}^{(k,l)} \left(\hat{X}^{(k,l)} \right)^{T};$$

$$C = \sum_{k=1}^{K} \sum_{l=1}^{L} \left(\hat{X}^{(k,l)} \right)^{T} \hat{X}^{(k,l)}.$$

For both R and C we calculate eigenvalue matrices $\{\Box^{(R)}, \Box^{(C)}\}$ (diagonal matrices of eigenvalues of $M \times M$ and $N \times N$ elements, respectively), and eigenvector matrices $\{V^{(R)}, V^{(C)}\}$(orthogonal matrices of $M \times M$ and $N \times N$ elements, respectively, whose columns contain eigenvectors) [7]:

$$\Lambda^{(R)} = \left(V^{(R)} \right)^{T} R V^{(R)};$$

$$\Lambda^{(C)} = \left(V^{(C)} \right)^{T} C V^{(C)};$$

From the diagonals of $\Box^{(R)}$ and $\Box^{(C)}$ we select the largest elements and memorize their positions. From $\left(V^{(R)} \right)^{T}$ we select r rows corresponding to the r largest eigenvalues and from $V^{(C)}$ we select c columns, respectively. Then we build two new matrices: $F^{(R)}$ of $r \times M$ elements and $F^{(C)}$ of $N \times c$ elements, which will be used as the transformation matrices for PCArc. The transformation is performed according to the following equation [6, 8]:

$$Y^{(k,l)} = F^{(R)} \left(X^{(k,l)} - \overline{X} \right) F^{(C)},$$

giving a feature matrix Y for each image X.

The classification employs feature space created for all face images in the database. A query face is projected into it and Euclidean distance is being calculated. Two alternative distances are taken into consideration: the first one is a distance to all elements, whereas the second one is the distance to the centers of all classes only $\left(\overline{X}^{(k)}\right)$. The latter approach is easier to perform and requires much fewer computations, however, its efficiency is lower in the case of irregular clusters. One of the recognition experiments is presented in Fig. 3. It was performed in "FaReS–Modeller" Environment [**9**]. As it can be seen from the figure below, it gives 5 nearest images from the database for a query image ("source image" in the picture), sorted in the decreasing similarity order.
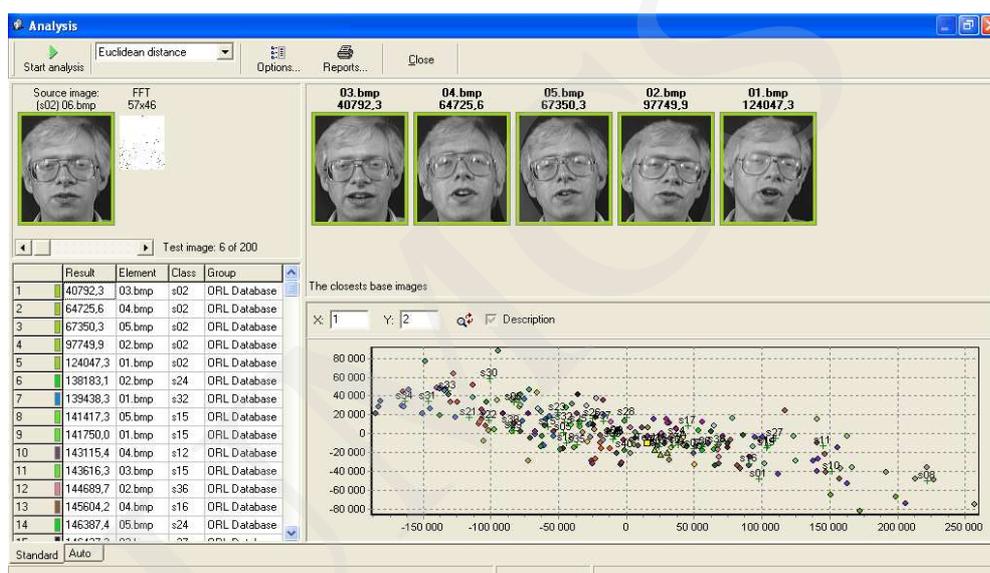

Fig. 3. Sample results of recognition in the FaReS environment (overall recognition rate equal to 95%)

3. Image quality assessment using the Structural Similarity index

Taking into account the poor correlation of some classical image quality metrics, like the Mean Squared Error (MSE), Mean Absolute Error (MAE), Signal to Noise Ratio (SNR) etc. [**10**] with the subjective evaluations, there is a need of using some modern metric which should be much more useful in that aspect. Regardless of the correlation with the subjective quality scores, a good correlation with the pattern (e.g. face) recognition accuracy can be expected assuming the presence of some typical image distortions.

A great majority of the image quality metrics used in computer vision and image analysis applications can be classified as the full–reference metrics. It can be treated as one of the disadvantages but, on the other hand, currently available blind (no-reference) metrics [11] are not universal so the number of distortion types influencing on their values is limited. For instance, there are many metrics sensitive on JPEG compression artifacts [12, 13] and Gaussian blur [14] but their applicability for some other types of distortions (e.g. various types of noise) is strongly limited and may lead to unpredictable and even wrong results. Similar situation takes place in the field of the reduced-reference image quality assessment methods utilizing the partial information related to the original image. Another approach based on the modelling of the Human Visual System is computationally demanding so the complexity of such algorithms causes the similarly limited applicability. For the relatively low performance and embedded systems with strongly limited amount of memory, the statistical quality estimation approach [15] can be also used.

The usefulness of some universal image quality assessment methods, including the full–reference ones, which require the knowledge of the original image without any distortions, is evident, especially in such areas as the development of some new image filtration algorithms, lossy compression methods or video transmission techniques.

The most popular modern approach to the image quality assessment is based on the Universal Image Quality Index [16], further extended into Structural Similarity [17] defined as:

$$SSIM = \frac{\left(2 \cdot \overline{X} \cdot \overline{X'} + C_1\right) \cdot \left(2 \cdot \sigma_{XX'} + C_2\right)}{\left(\overline{X}^2 + \overline{X'}^2 + C_1\right) \cdot \left(\sigma_x^2 + \sigma_{X'}^2 + C_2\right)} \tag{1}$$

where: X and X' denote the original and distorted images respectively. The stability constants C_1 and C_2 are introduced in order to prevent dividing by zero for large flat and dark regions of the image but they should not introduce any significant changes of the results for the rest of the image. The authors of paper [17] suggest using the following values: $C_1 = (0.01 \cdot 255)^2$ and $C_2 = (0.03 \cdot 255)^2$. The other elements in formula (1) are defined as:

$$\overline{X} = \frac{1}{N \cdot M} \sum_{n=1}^{N} \sum_{m=1}^{M} X(n,m)$$

$$\overline{X'} = \frac{1}{N \cdot M} \sum_{n=1}^{N} \sum_{m=1}^{M} X'(n,m)$$

$$\sigma_x^2 = \frac{1}{N \cdot M - 1} \sum_{n=1}^{N} \sum_{m=1}^{M} \left(X(n,m) - \overline{X} \right)^2$$

$$\sigma_{X'}^2 = \frac{1}{N \cdot M - 1} \sum_{n=1}^{N} \sum_{m=1}^{M} \left(X'(n,m) - \overline{X'} \right)^2$$

$$\sigma_{XX'} = \frac{1}{N \cdot M - 1} \sum_{n=1}^{N} \sum_{m=1}^{M} \left(X(n,m) - \overline{X} \right) \left(X'(n,m) - \overline{X'} \right)$$

Applying such a method using the 11×11 pixels Gaussian sliding window ($N = M = 11$) the quality map of the image can be obtained with the resolution nearly the same as the analyzed image (reduced by $N - 1$ rows and $M - 1$ columns). The overall image quality metric is obtained as the mean value of the whole quality map. One of the most relevant properties of the SSIM index is its sensitivity to three common types of distortions introduced by many image processing algorithms: the loss of correlation (structural distortions), luminance distortion and the loss of contrast.

4. Proposed method of image quality estimation

Taking into account the fact of sensitivity of face recognition methods to many typical distortions, including the presence of noise, blurring, lossy compression etc., the influence of the image quality on the recognition accuracy is obvious. The general relation is straightforward: the worse image quality means the weaker recognition. Nevertheless, more accurate estimation of the face recognition accuracy on the base of the type and amount of distortions present in the input images is not an easy task. For that purpose the usage of the previously described SSIM index is proposed, also with the modification related to the central weighting using the 2-D Gaussian function (Fig. 4). The idea of using the Central Weighting SSIM is based on the assumption of central location of the most important elements of the input images, so the quality of the central parts of the image should be the most crucial. For the typical face recognition systems the input images are usually cropped in order to eliminate any redundant elements except the central part of face.

In our experiments the calculations of the linear correlation coefficients between the recognition accuracy by the PCArc method using the nearest element and the centers of classes as the classification criteria and the values of the SSIM index, as well as proposed Central Weighted SSIM index, have been performed.

The recognized images have been distorted using several typical methods such as contamination by the salt and pepper noise, lossy JPEG compression,

low–pass filtration, median filtering, contrast reduction and intensity change. The examplary facial portrait with various distortions is presented in Fig. 5.

For each of 40 classes from the AT&T database 5 images have been used for training and the remaining 5 images have been used for tests, also in their distorted versions. The average recognition accuracy for the images without any distortions is 95.0% (considering the distance to the nearest element) and 90.5% if the distance to the centers of classes has been used. The average results of recognition for different distortions are presented in Table 1.
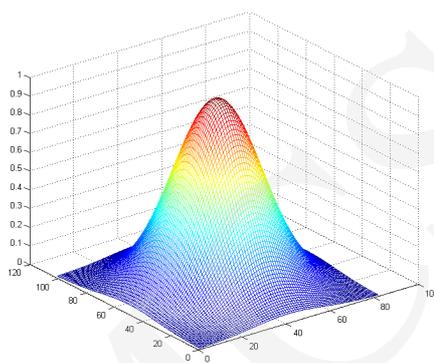

Fig. 4. Central Weighting Gaussian function used in our experiments (scaled according to the resolution of images in the dataset)
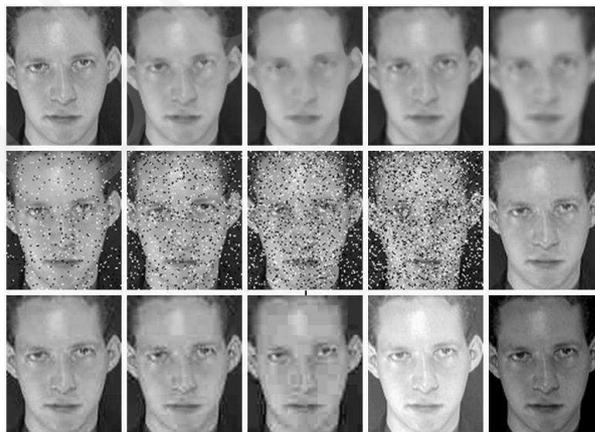

Fig. 5. Example original image and its distorted variants: median filtered (mask 3×3 and 5×5), low–pass filtered (mask 3×3 and 5×5), contaminated by the salt and pepper noise (5%, 10% and 20% of pixels), JPEG compressed (quality: 60%, 40%, 20% and 10%) and intensity offset change (+20,-20), respectively

Table 1. Average recognition accuracy for the images with various distortions

Distortion	Recognition accuracy [%]	
	Nearest element	Centers of classes
Median filtering – mask 3x3	93.0	82.0
Median filtering – mask 5x5	92.5	80.5
Low-pass filtering – mask 3x3	93.0	81.5
Low-pass filtering – mask 5x5	88.5	77.5
5% salt and pepper noise	95.5	83.5
10% salt and pepper noise	91.5	75.0
15% salt and pepper noise	78.5	69.5
20% salt and pepper noise	57.5	52.0
JPEG – quality 60%	94.5	82.5
JPEG – quality 40%	94.5	82.0
JPEG – quality 20%	94.5	81.5
JPEG – quality 10%	94.5	82.0
contrast reduction (70%)	92.5	81.5
contrast reduction (50%)	93.5	81.5
intensity offset change (+20)	93.5	82.5
intensity offset change (-20)	90.5	78.5
No distortions	95.0	90.5

Table 2. Pearson correlation coefficient between the recognition accuracy and the image quality metrics for all distortions used in experiments

Distortions	Classification approach	Correlation coefficient	
		SSIM	Central Weighted SSIM
salt and pepper noise	Nearest element	0.8340	0.8400
	Centers of classes	0.8808	0.8854
low-pass and median filtering	Nearest element	0.8501	0.8501
	Centers of classes	0.9404	0.9427
both noise and filtering	Nearest element	0.6470	0.6561
	Centers of classes	0.7186	0.7273
JPEG compression, noise and filtering	Nearest element	0.7024	0.7096
	Centers of classes	0.7607	0.7669
JPEG compression and filtering	Nearest element	0.5837	0.6829
	Centers of classes	0.7171	0.7903
All distortions except the intensity offset	Nearest element	0.7042	0.7107
	Centers of classes	0.7662	0.7718
All distortions	Nearest element	0.7071	0.7092
	Centers of classes	0.7712	0.7714

5. Conclusions

Summarizing the experiments performed, it can be noticed that the usage of the proposed method does not lead to worsening of the results in any case. The increase of the correlation coefficients is especially evident for the set of filtered (by low–pass and median filters) and JPEG compressed images. Considering the relatively low computational cost of the additional weighting, the proposed method can be successfully applied in many face recognition systems.

In the future work the further extension of the proposed approach using the color images, as well and some other weighting functions, is planned, also in the case of other pattern recognition methods.

References

[1] AT&T Research Laboratories, *The Database of Faces*, http://www.cl.cam.ac.uk/research/dtg/attarchive/ facedatabase.html (1994)

[2] Swets, D. L., Weng, J., *Using Discriminant Eigenfeatures for Image Retrieval*, IEEE Transactions on Pattern Analysis and Machine Intelligence 18(8) (1996) 831.

[3] Tsapatsoulis, N., Alexopoulos, V., Kollias, S., *A Vector Based Approximation of KLT and Its Application to Face Recognition*, Proceedings of the IX European Signal Processing Conference EUSIPCO-98, Island of Rhodes, Greece, September, 3 (1998) 1581.

[4] Swets, D. L., Weng, J., *Hierarchical Discriminant Analysis for Image Retrieval*, IEEE Transactions on Pattern Analysis and Machine Intelligence 21(5) (1999) 386.

[5] Forczmański, P, Kukharev, G., *Comparative Analysis of Simple Facial Features Extractors*, Journal of Real-Time Image Processing 1(4) (2007) 239.

[6] Kukharev, G., Forczmański, P., *Facial Images Dimensionality Reduction and Recognition by Means of 2DKLT*, Machine Graphics and Vision 16(3/4) (2007) 401.

[7] Fukunaga K., *Introduction to Statistical Pattern Recognition*, 2nd Ed., New York, Academic Press (1990).

[8] Kukharev, G., Forczmański, P., *Face Recognition by Means of Two-Dimensional Direct Linear Discriminant Analysis*, Proceedings of 8th International Conference on Pattern Recognition and Information Processing, Minsk, Belarus, May 18-20, (2005) 280.

[9] Kukharev G, Kuźmiński A., *Biometric techniques. Part I. The Methods of Face Recognition* (in Polish). Szczecin (Poland), Pracownia Poligraficzna WI PS (2003) 310.

[10] Eskicioglu, A., Fisher, P., Chen, S., *Image Quality Measures and Their Performance*, IEEE Transactions on Communications 43(12) (1995) 2959.

[11] Li, X., *Blind Image Quality Assessment*, Proceedings of IEEE International Conference on Image Processing (2002) 449.

[12] Wang, Z., Sheikh, H., Bovik, A., *No-Reference Perceptual Quality Assessment of JPEG Compressed Images*, IEEE International Conference on Image Processing (2002) 477.

[13] Meesters, L., Martens, J.–B., *A Single-Ended Blockiness Measure for JPEG-Coded Images*, Signal Processing 82(3) (2002) 369.

[14] Marziliano, P., Dufaux, F., Winkler, S., Ebrahimi, T., *A No-Reference Perceptual Blur Metric*, Proceedings of IEEE International Conference on Image Processing, Rochester, USA (2002) 57.

[15] Okarma, K., Lech, P., *Monte Carlo Based Algorithm for Fast Preliminary Video Analysis*, Lecture Notes in Computer Science 5101, Springer-Verlag, Heidelberg (2008) 790.

[16] Wang, Z., Bovik, A., *A Universal Image Quality Index*, IEEE Signal Processing Letters 9(3) (2002) 81.

[17] Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E., *Image Quality Assessment: From Error Measurement to Structural Similarity*, IEEE Transactions on Image Processing 13(4) (2004) 600.