Image Compression Effects in Face Recognition Systems

Kresimir Delac, Mislav Grgic and Sonja Grgic
University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb
Croatia

1. Introduction

With the growing number of face recognition applications in everyday life, image- and video-based recognition methods are becoming important research topic (Zhao et al., 2003). Effects of pose, illumination and expression are issues currently most studied in face recognition. So far, very little has been done to investigate the effects of compression on face recognition, even though the images are mainly stored and/or transported in a compressed format. Still-to-still image experimental setups are often researched, but only in uncompressed image formats. Still-to-video research (Zhou et al., 2003) mostly deals with issues of tracking and recognizing faces in a sense that still uncompressed images are used as a gallery and compressed video segments as probes.

In this chapter we analyze the effects that standard image compression methods - JPEG (Wallace, 1991) and JPEG2000 (Skodras et al., 2001) - have on three well known subspace appearance-based face recognition algorithms: Principal Component Analysis - PCA (Turk & Pentland, 1991), Linear Discriminant Analysis - LDA (Belhumeur et al., 1996) and Independent Component Analysis - ICA (Bartlett et al., 2002). We use McNemar's hypothesis test (Beveridge et al., 2001; Delac et al., 2006) when comparing recognition accuracy in order to determine if the observed outcomes of the experiments are statistically important or a matter of chance. Following the idea of a reproducible research, a comprehensive description of our experimental setup is given, along with details on the choice of images used in the training and testing stage, exact preprocessing steps and recognition algorithms parameters setup. Image database chosen for the experiments is the grayscale portion of the FERET database (Phillips et al., 2000) and its accompanying protocol for face identification, including standard image gallery and probe sets. Image compression is performed using standard JPEG and JPEG2000 coder implementations and all experiments are done in pixel domain (i.e. the images are compressed to a certain number of bits per pixel and then uncompressed prior to use in recognition experiments).

The recognition system's overall setup we test is twofold. In the first part, only probe images are compressed and training and gallery images are uncompressed (Delac et al., 2005). This setup mimics the expected first step in implementing compression in real-life face recognition applications: an image captured by a surveillance camera is probed to an existing high-quality gallery image. In the second part, a leap towards justifying fully compressed domain face recognition is taken by using compressed images in both training

Source: Face Recognition, Book edited by: Kresimir Delac and Mislav Grgic, ISBN 978-3-902613-03-5, pp.558, I-Tech, Vienna, Austria, June 2007

and testing stage (Delac, 2006). We will show that, contrary to common opinion, compression does not deteriorate performance but it even improves it slightly in some cases. We will also suggest some prospective lines of further research based on our findings.

2. Image compression basics

First let us briefly explain some basic concepts needed to fully understand the rest of the chapter. Image compression will be introduced with scarce details and an interested reader is referred to cited papers for further exploration.

There are two standard image compression schemes that are of interest here: JPEG (Wallace, 1991) and JPEG2000 (Skodras et al., 2001). These image compression standards are widely used in many applications and are expected to be employed in face recognition as well. Generally, compression seems to be imperative for any reasonable implementation where a large quantity of images need to be stored and used. Both JPEG and JPEG2000 use the general transform coding scheme shown in Figure 1.

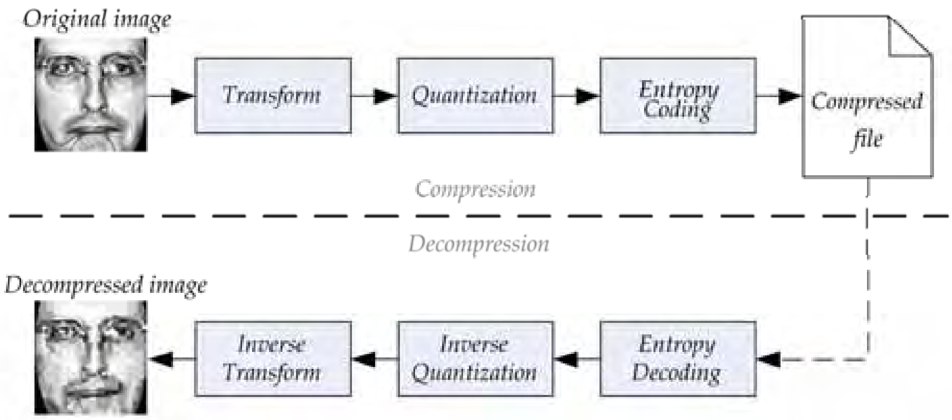


Figure 1. Basic steps of transform coding (compression) of images

The images are first transformed into a form (domain) more suitable for compression. Transforms used are the Discrete Cosine Transform (DCT) in JPEG and Discrete Wavelet Transform (DWT) in JPEG2000. This procedure assigns values to different spatial frequency components present in the image. Since the human visual system is less sensitive to higher frequencies, the coefficients representing such frequencies can be discarded, thus yielding higher compression rates. This is done through quantization and entropy coding, creating the compressed file as an output. Decompression follows the exact inverse procedure. JPEG and JPEG2000 are irreversible, meaning that the original image can not be reconstructed from the compressed file (this is because some coefficients were discarded). The distortions are introduced by coefficients quantization in JPEG and both quantization and entropy coding in JPEG2000. The resulting reconstructed images now have artifacts present, like the checker-board effect in JPEG images or the smear effect in JPEG2000 images. Some examples of these effects in face images can be seen in Figure 2. A closer look at these images and having the former analysis in mind will give us the feel of what actually happens. As the transform coefficients that represent higher frequencies are more and more discarded (or are rounded to lower precision) with higher compression rates, the images become more and more low-pass filtered. This is quite obvious for the JPEG2000 example at 0.2 bpp where we can see that the finer details of the face (like wrinkles) are eliminated in the reconstructed image. It remains to be seen how will this low-pass filtering affect recognition results.

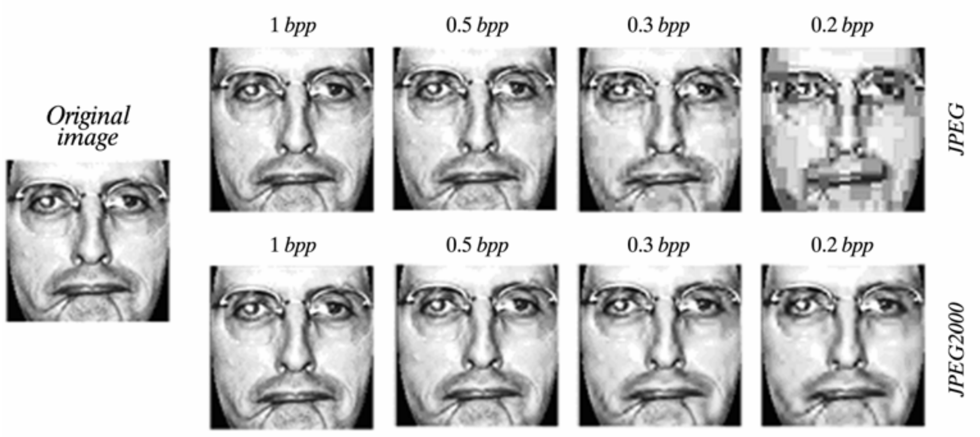


Figure 2. Examples of image distortions introduced by JPEG or JPEG2000 compression

The main tool for measuring the magnitude of compression is *compression ratio*, expressed in the form of *bits per pixel* (bpp). Given that the original (uncompressed) grayscale images that we will consider throughout this chapter are normally 8 bpp, the compression ratio of 1 bpp represents the 8:1 compression. In other words, the compressed file is eight times smaller than the original file (image).

As can be seen in Figure 2, there is practically no difference between the original image and images compressed at 1 bpp, as far as the human visual system is concerned. This comes naturally from the basic idea that the creators of JPEG and JPEG2000 had in mind when creating the standards. Loosely speaking: as little visible distortions as possible. However, the difference can be objectively measured by Peak Signal to Noise Ratio (PSNR), calculated as:

$$\mathrm{PSNR} = 20\,log\left(\frac{2^n - 1}{\mathrm{RMS}}\right)\ [\mathrm{dB}], \tag{1}$$

where n is the number of bits per pixel in the original image and RMS is the Root Mean Square Error defined as:

$$\mathrm{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(I_i - I'_i\right)^2} \tag{2}$$

where I_i is pixel value in the original image, I'_i is corresponding pixel value in the reconstructed image and N is the total number of pixels in the image. PSNR values for images in Figure 2. are shown in Table 1. We can see that JPEG and JPEG2000 behave similarly at moderate compression rates (1 bpp and 0.5 bpp). More apparent differences arise at higher compression rates (0.3 bpp and 0.2 bpp), where JPEG2000 is clearly superior.

	1 bpp	**0.5 bpp**	**0.3 bpp**	**0.2 bpp**
JPEG	34.02	30.00	26.30	19.88
JPEG2000	35.96	30.28	28.12	25.96

Table 1. PSNR values in dB for images in Figure 2

Similar conclusions on JPEG and JPEG2000 efficiency can be found in (Grgic et al., 2001). Through using additional objective image quality measures it was shown that DCT-based and DWT-based compression yield similar results at lower compression rates. At higher compression rates, DWT-based compression retains rather high quality while DCT-based compression quality deteriorates rapidly. In (Ebrahimi et al., 2004) authors showed that there is no significant difference in the quality of JPEG and JPEG2000 compressed images at lower and moderate compression rates. JPEG2000 was determined to be superior at higher compression rates. In (Santa-Cruz et al., 2000) authors concluded that JPEG2000 is both subjectively and objectively superior to JPEG.

In the literature review that follows, we will see how compression effects were tested in face recognition so far and what still remains to be done.

3. Related work

Before proceeding to related work review, one basic term should be clarified. It has to emphasized that all the experiments described in this chapter, including the ones in the following literature review, are conducted in *pixel domain*. This actually means that the images are compressed and then uncompressed prior to being used in the experiments. This way the actual influence that the distortion introduced by compression has on recognition rate is measured.

There has been little investigation of the effects of image compression on face recognition systems do far. As will be seen, mostly JPEG compression is covered and mainly at a single compression ratio.

In (Blackburn et al., 2001) the authors tried to measure the effects of image compression on face recognition systems by simulating the expected real-life setup: images of persons known to the system (gallery) were of high quality (non-compressed) and images of persons unknown to the system (probes) were taken in uncontrolled environment and compressed. Naturally, images were decompressed prior to recognition and thus we can say that experiments were conducted in the pixel domain. JPEG compression was used and face recognition system was tested using the FERET database and its *dup1* (temporal changes) probe set. Images were compressed to 0.8, 0.4, 0.25 and 0.2 bpp. The authors conclude that compression does not affect recognition significantly across wide range of compression rates. Significant performance drop is noted at 0.2 bpp and below. Recognition rate is even slightly higher in some cases when using compressed images (compared to results using original images).

Moon and Phillips (Moon & Phillips, 2001) tested the effects of standard JPEG compression and of a variant of wavelet compression with a PCA+L1 method. Probe images were in both cases compressed to 0.5 bpp, decompressed (so the experiments were conducted in pixel domain) and then geometrically normalized. The training set of images was uncompressed. FERET database was used along with its standard probe sets (only *fb* and *dup1* in this experiment). Results indicate no performance drop for JPEG compression and a slight increase for wavelet compression. Whether this increase in recognition rate is significant or not is unclear.

JPEG2000 compression effects were tested in (McGarry et al., 2004) as part of the development of the ANSI INCITS 385-2004 standard: "Face Recognition Format for Data Interchange" (ANSI, 2004), later to become an ISO/IEC IS 19794-5 standard: "Biometric Data Interchange Formats - Part 5: Face Image Data" (ISO, 2004). The experiment included compression at a compression rate of 10:1, as recommended in (ANSI, 2004; ISO, 2004). A commercial face recognition system was used for testing a vendor database. Again, since there are no details on the exact face recognition method used in the tested system and no details on a database used in experiments, it is difficult to make any comparisons to this work. In a similar setup as in previously described papers, it was determined that there is no significant performance drop when using compressed probe images. Based on their findings, the authors conjecture that compression rates higher than 10:1 could be used.

In (Wat & Srinivasan, 2004) the authors test the effects of JPEG compression on PCA and LDA face recognition methods using the same experimental setup as in (Blackburn et al., 2001). Results are presented as a function of JPEG quality factor. This fact makes any comparison with these results very difficult since the same quality factor will yield different compression rates for different images, dependent upon the statistical properties of a given image. This is why we decided to used bits per pixel as a measure of compression ratio in our experiments. The authors used the FERET database and tested the standard probe sets against a standard gallery. Results indicate a slight increase in performance for the LDA method with the *fc* probe set. For all other probe sets and methods the results were practically the same as with uncompressed images.

An initial detailed experiment of the effects of compression on face recognition was conducted in (Delac et al., 2005). We tested both JPEG and JPEG2000 compression effects on a wide range of subspace algorithm - metric combinations. Similar to other studies, we also concluded that compression does not affect performance significantly. We supported our conclusions with McNemar's hypothesis test. Some performance improvements were also noted, but none of them were statistically significant.

Wijaya et al. in (Wijaya et al., 2005) performed face verification on images compressed to 0.5 bpp by JPEG2000 and showed that high recognition rates can be achieved using correlation filters. Their conclusion was also that compression does not adversely effect performance.

We can see that the described experiments were mainly done in the same setup: training and gallery images are uncompressed and probe images are compressed to various compression ratios. Most authors conclude that compression does not affect recognition rate significantly, but these conclusions still need to be statistically confirmed. Most of these experiments are limited to a single compression rate and a single recognition method. We will try to address some of these shortcomings in the experiments presented in this chapter.

4. Experimental setups and results

4.1 Database and protocol

We use the standard FERET data set including the data partitions (subsets) for recognition tests, as described in (Phillips et al., 2000). The gallery consists of 1,196 images and there are four sets of probe images that are compared to the gallery images in recognition stage. The *fb* probe set contains 1,195 images of subjects taken at the same time as gallery images with the only difference being that the subjects were told to assume a different facial expression. The *fc* probe set contains 194 images of subjects under different illumination conditions. The *dup1* (duplicate I) set contains 722 images taken anywhere between one minute and 1,031 days after the gallery image was taken, and *dup2* (duplicate II) set is a subset of *dup1* containing 234 images taken at least 18 months after the gallery image was taken. All images in the data set are of size 384 × 256 pixels and grayscale.

4.2 Preprocessing

Original FERET images were first spatially transformed (to get the eyes at the predefined fixed points) based upon a ground truth file of the eye coordinates supplied with the original FERET data. All images were then cropped to 128 × 128 pixels (using the eyes coordinates) and an elliptical mask was used to further eliminate the background. Finally, image pixel values were histogram equalized to the range of values from 0 to 255. These preprocessing steps were carried out on all images prior to preforming the experiments (including compression).

4.3 Algorithms

Three well known appearance-based subspace face recognition algorithms were used to test the effects of compression: Principal Component Analysis - PCA (Turk & Pentland, 1991), Linear Discriminant Analysis - LDA (Belhumeur et al., 1996) and Independent Component Analysis - ICA (Bartlett et al., 2002). It is important to mention that we use ICA *Architecture 2* from (Bartlett et al., 2002) since ICA *Architecture 1* was shown to be suboptimal for face identification tasks (Delac et al., 2005; Delac et al. 2006). For both LDA and ICA, a PCA dimensionality reduction was done as a preprocessing step.

To train the PCA algorithm we used a subset of classes for which there were exactly three images per class. We found 225 such classes (different persons), so our training set consisted of 3 × 225 = 675 images (M = 675, c = 225). The effect that this percentage of overlap has on algorithm performance needs further exploration and will be part of our future work. PCA derived, in accordance with theory, M - 1 = 674 meaningful eigenvectors. We adopted the FERET recommendation and kept the top 40% of those, resulting in 270-dimensional PCA subspace $\mathbf{W}$ (40% of 674 = 270). It was calculated that 97.85% of energy was retained in those 270 eigenvectors. This subspace was used for recognition as PCA face space and as input to ICA and LDA (PCA was the preprocessing dimensionality reduction step). ICA yielded a 270-dimensional subspace, and LDA yielded only 224-dimensional space since it can, by theory, produce a maximum of c - 1 basis vectors. All of those were kept to stay close to the dimensionality of PCA and ICA spaces and thus make comparisons as fair as possible.

Based on our previous findings in (Delac et al., 2005; Delac et al., 2006) we chose the following combinations of algorithms and metrics (one metric for each algorithm) to be used in these experiments: PCA+L1, LDA+COS and ICA+COS. These combinations yielded the highest recognition rates in our previous experiments.

4.4 Measurement methods

Performance of face recognition systems (algorithms, methods) will be presented as rank one recognition rate, as described in (Phillips et al., 2000). Let T represent the training set, G

gallery and P probe set of images. T and G can be the same set but this is not a good testing practice. The actual performance of an algorithm is always rated relative to how well the images in P are matched to images in G. This is the basis of automatic face recognition. Intuitively, it is obvious that P and G should be disjoint; otherwise, the stated problem becomes trivial. We will use the identification scenario in our experiments. To calculate the recognition rate for a given probe set P, for each probe image P_i, we need to sort all the gallery images by decreasing similarity, yielding a list $L = \{L_1, L_2, \ldots , L_K\}$, where K is the total number of subjects in the gallery (assuming that there is one image per subject, K also becomes the number of images and the size of the gallery). Now L_1 is the gallery image most similar to the given probe image (according to the algorithm), L_2 is the next closest match and expanding this to L_k being the kth closest gallery match. Rank one recognition rate answers a simple question: is the top match correct? If L_1 (labeled as the closest gallery match to the given probe image) is really the correct answer, we say that the algorithm correctly recognized the probe image. In other words, the algorithm successfully recognizes a probe image if the probe image and the top ranked gallery image in L are of the same subject. This is called rank one recognition rate (RR) and can be formally defined over the whole set of probe images P as follows: let R_1 denote the number of correctly recognized probe images in L at $k = 1$ and $|P|$ be the probe set size, then:

$$RR = \frac{R_1}{|P|} . \qquad (3)$$

A usual way to report rank one performance is to give it in a form of percentage. That way we actually say that some algorithm has e.g. 86% rank one recognition rate on a given gallery and probe set. Another possible formulation would be that there is 86% chance that the correct answer is the top match (the image L_1).

To measure the significance of the differences in performance at two different compression ratios, we will use McNemar's hypothesis test (Beveridge et al., 2001; Delac et al., 2006). We think that, when comparing recognition algorithms, it is important (yet often neglected) to answer the following question: when is the observed difference in performance statistically significant? Clearly, the difference in performance of 1% or 2% could be due to pure chance. However, we felt the need to investigate these intuitive presumptions using standard statistical hypothesis testing techniques. Generally, there are two ways of looking at the performance difference (Yambor et al., 2002): 1) determine if the difference (as seen over the entire set of probe images) is significant, 2) when the algorithms behave differently, determine if the difference is significant. As argued in (Yambor et al., 2002), the first way to evaluate performance difference fails to take the full advantage of the standard face recognition protocol, so we will focus on the second way. In order to perform this test we recorded which of the four possible outcomes, when comparing two algorithms A1 and A2 (SS – both successful, FF – both failed, FS – first one failed and the second one succeeded, SF – first one succeeded and the second one failed), is true for each probe image. Let N_{SS} represent the number of probe images for which SS outcome is true, N_{SF} the number of probe images for which SF outcome is true, etc. We then formulated our hypotheses as: H0) the probability of observing SF is equal to the probability of observing FS; H1) the probability of observing SF is not equal to the probability of observing FS. H0 is the null hypothesis and H1 the alternative hypothesis.

In case where one algorithm performs better than another algorithm, H0 can be rejected if the observed difference in performance of the compared algorithms is statistically significant. Therefore, H0 is tested by applying a one-tailed test. Suppose that *Pr*(SF) and *Pr*(FS) are the probabilities of observing SF and FS outcomes under H0. For example, if it appears that *Pr*(SF) > *Pr*(FS), i.e. A1 performs better than A2, then we calculate:

$$Pr\ (\text{A1 better than A2 at least as many times as observed}) = \sum_{i=N_{SF}}^{n} \frac{n!}{i!\,(n-i)!} \cdot \left(\frac{1}{2}\right)^{n} \tag{4}$$

where $n = N_{SF} + N_{FS}$ is the number of probe images for which only one algorithm incorrectly classify them. This probability is usually called *p*-value for rejecting H0 in favor of H1. H0 is rejected when the *p*-value is lower than some predefined threshold α (usually $\alpha = 0.05$, i.e. 5%), and in this case we can conclude that *the observed difference in performance of the compared algorithms is statistically significant*.

We will report the outcomes of McNemar's test in our results as "○" when there is no statistically significant difference when using images at a given compression ratio compared to using original images, "×" the recognition ratio is significantly worse than with original images and "✓" when the recognition ratio using compressed images is significantly higher than with original images.

Another handy tool that can be used here is the Normalized Recognition Rate (*NRR*), defined as the ratio between recognition rate (*RR*) for compressed images and recognition rate for original images (Delac, 2006):

$$NRR = \frac{RR_{\text{compressed}}}{RR_{\text{original}}}\ . \tag{5}$$

So, at a given bitrate (number of bits per pixel), if *NRR* = 1, the performance is the same as with original images, if *NRR* < 1, performance is worse, and if *NRR* > 1, performance is better then with original images. We will present *NRR* curves (*NRR* as a function of compression ratio) for some interesting results just as an example of their possible usage. Full analysis of the results with *NRR* is out of scope of this chapter.

4.5 Experiments

As stated before, most of the experiments presented in the literature so far use the scenario where only probe images are compressed. We will here try to perform another experiment where all the images are compressed to a given compression ratio. This will be a good foundation for possible new area in face recognition research - *face recognition in compressed domain*. Compressed domain means that instead of decompressing the compressed images and then using (distorted) pixel values as input to face recognition methods, transformed coefficients are used as inputs. The decoding process should be interrupted after the entropy decoding and the obtained coefficients (DCT or DWT) used as inputs to classification methods. This way it is possible to achieve large computational time saves by avoiding the inverse DCT or DWT.

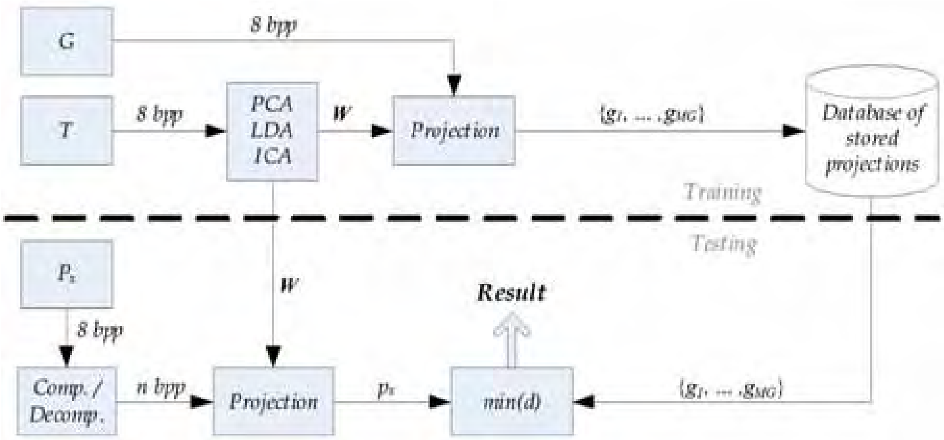


Figure 3. Experimental setup 1 (EXP1)

Scenario that was used in studies so far (only probe images are compressed) will be addressed as EXP1 in further text and a block-scheme of this approach can be seen in Figure 3. The setup where all images (training, gallery and probe) are compressed to the same compression ratio will be addressed as EXP2 and a block-scheme can be seen in Figure 4. The illustrations in Figure 3 and Figure 4 represent the training and recognition stage of a PCA, LDA or ICA-based system for a single probe image P_x. T and G represent training and gallery sets of images, respectively. Original (uncompressed) images have 8 bpp and compressed images have a hypothetical value of n bpp. In the module $min(d)$ the distance between the projected probe image p_x and the list of gallery images $\{g_1, g_2, \ldots, g_{MG}\}$ is calculated and a minimal distance is determined (MG is the number of images in the gallery). The identity of the person on a gallery image determined to be the closest to P_x in the subspace is the identity of the unknown person returned by the system. This is a standard rank one identification scenario.

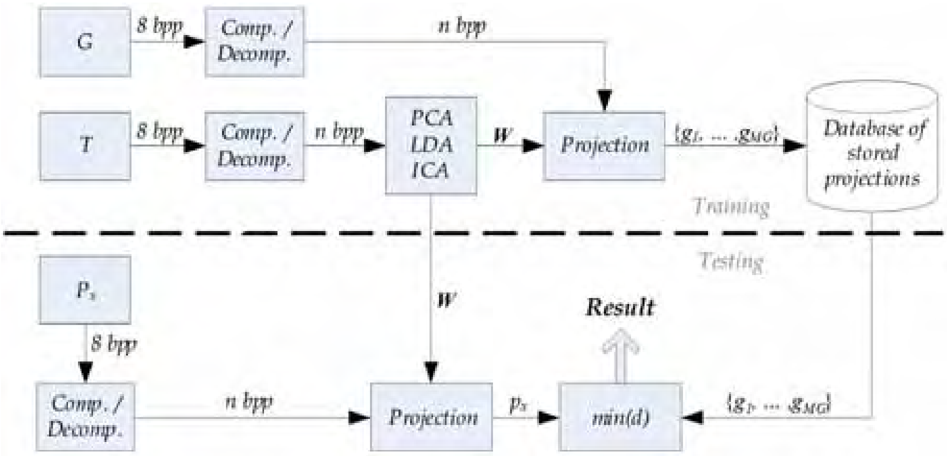


Figure 4. Experimental setup 2 (EXP2)

4.6 Results

The results for both experiments can be seen in Tables 2 through 9. The figures presented on tables represent rank one recognition rates. "McN" presents the result of McNemar's hypothesis test (result at a given compression ratio compared to the result using original uncompressed images). By looking at the results of McNemar's test, we can immediately conclude that compression to 1 bpp and 0.5 bpp does not significantly influence the results in any method and/or experiment. This is consistent with previous studies and it additionally gives strong statistical basis for such a conclusion. In the following text we will give an analysis for each probe set in both experiments and present two possible real life applications of the conclusions drawn from this study.

fb	JPEG		Orig.	1 bpp	0.5 bpp	0.3 bpp	0.2 bpp
PCA+L1	EXP1	RR	79,4	79,4	79,4	78,9	77,2
		McN	-	○	○	○	×
	EXP2	RR	79.4	78.9	79.4	79.0	75.4
		McN	-	○	○	○	×
LDA+COS	EXP1	RR	75.4	75.4	75.2	75.3	73.6
		McN	-	○	○	○	×
	EXP2	RR	75.4	75.5	75.5	74.5	72.6
		McN	-	○	○	○	×
ICA+COS	EXP1	RR	83.0	82.8	83.0	82.0	80.0
		McN	-	○	○	○	×
	EXP2	RR	83.0	83.1	83.0	82.2	75.6
		McN	-	○	○	○	×

Table 2. The results for JPEG compression, *fb* probe set ("○" - no statistically significant difference compared to using original images; "×" - RR significantly worse than with original images; "✓" - RR significantly higher than with original images)

fc	JPEG		Orig.	1 bpp	0.5 bpp	0.3 bpp	0.2 bpp
PCA+L1	EXP1	RR	47.9	46.4	45.9	47.9	44.3
		McN	-	○	○	○	×
	EXP2	RR	47.9	50.0	49.5	51.0	42.3
		McN	-	○	○	✓	×
LDA+COS	EXP1	RR	11.3	11.3	11.3	11.3	10.8
		McN	-	○	○	○	○
	EXP2	RR	11.3	11.3	11.3	11.9	11.3
		McN	-	○	○	○	○
ICA+COS	EXP1	RR	68.6	68.0	67.5	69.6	66.5
		McN	-	○	○	○	○
	EXP2	RR	68.6	67.5	68.6	66.5	57.7
		McN	-	○	○	○	×

Table 3. The results for JPEG compression, *fc* probe set

*dup*1	JPEG		Orig.	1 bpp	0.5 bpp	0.3 bpp	0.2 bpp
PCA+L1	EXP1	RR	38.5	38.6	38.5	38.2	35.1
		McN	-	○	○	○	×
	EXP2	RR	38.5	39.2	39.2	38.8	35.7
		McN	-	○	○	○	×
LDA+COS	EXP1	RR	35.6	35.6	35.3	35.8	33.8
		McN	-	○	○	○	×
	EXP2	RR	35.6	35.6	35.3	35.7	33.4
		McN	-	○	○	○	×
ICA+COS	EXP1	RR	44.3	44.9	44.5	42.9	41.1
		McN	-	○	○	×	×
	EXP2	RR	44.3	45.3	44.5	43.6	36.4
		McN	-	✓	○	○	×

Table 4. The results for JPEG compression, *dup1* probe set

*dup*2	JPEG		Orig.	1 bpp	0.5 bpp	0.3 bpp	0.2 bpp
PCA+L1	EXP1	RR	19.7	20.1	20.1	19.2	15.8
		McN	-	○	○	○	×
	EXP2	RR	19.7	20.5	21.4	19.2	17.2
		McN	-	○	○	○	○
LDA+COS	EXP1	RR	12.8	12.8	12.8	13.6	12.4
		McN	-	○	○	○	○
	EXP2	RR	12.8	13.2	13.2	12.4	13.2
		McN	-	○	○	○	○
ICA+COS	EXP1	RR	30.8	32.0	30.7	29.9	27.3
		McN	-	○	○	○	×
	EXP2	RR	30.8	31.2	30.3	31.2	24.8
		McN	-	○	○	○	×

Table 5. The results for JPEG compression, *dup2* probe set

fb	JPEG2000		Orig.	1 bpp	0.5 bpp	0.3 bpp	0.2 bpp
PCA+L1	EXP1	RR	79.4	79.4	79.6	79.1	78.6
		McN	-	○	○	○	○
	EXP2	RR	79.4	79.2	79.2	79.7	75.4
		McN	-	○	○	○	×
LDA+COS	EXP1	RR	75.4	75.4	75.3	75.2	75.0
		McN	-	○	○	○	○
	EXP2	RR	75.4	75.5	75.2	75.1	72.6
		McN	-	○	○	○	×
ICA+COS	EXP1	RR	83.0	83.1	83.1	83.0	83.4
		McN	-	○	○	○	○
	EXP2	RR	83.0	83.4	83.5	83.8	76.7
		McN	-	○	○	○	×

Table 6. The results for JPEG2000 compression, *fb* probe set

fc	JPEG2000		Orig.	1 bpp	0.5 bpp	0.3 bpp	0.2 bpp
PCA+L1	EXP1	RR	47.9	46.4	46.4	45.9	45.8
		McN	-	○	○	○	○
	EXP2	RR	47.9	51.0	51.5	52.6	42.3
		McN	-	✓	✓	✓	×
LDA+COS	EXP1	RR	11.3	11.3	11.3	10.8	11.3
		McN	-	○	○	○	○
	EXP2	RR	11.3	11.3	11.3	10.8	11.3
		McN	-	○	○	○	○
ICA+COS	EXP1	RR	68.6	69.0	68.5	68.5	68.6
		McN	-	○	○	○	○
	EXP2	RR	68.6	67.0	67.0	64.4	56.2
		McN	-	○	○	×	×

Table 7. The results for JPEG2000 compression, *fc* probe set

*dup*1	JPEG2000		Orig.	1 bpp	0.5 bpp	0.3 bpp	0.2 bpp
PCA+L1	EXP1	RR	38.5	38.3	38.5	38.2	38.5
		McN	-	○	○	○	○
	EXP2	RR	38.5	38.8	38.9	38.0	35.7
		McN	-	○	○	○	×
LDA+COS	EXP1	RR	35.6	35.6	35.5	35.4	35.1
		McN	-	○	○	○	○
	EXP2	RR	35.6	35.5	35.5	35.3	33.4
		McN	-	○	○	○	×
ICA+COS	EXP1	RR	44.3	44.7	44.5	44.5	44.3
		McN	-	○	○	○	○
	EXP2	RR	44.3	45.0	43.8	42.4	35.5
		McN	-	○	○	×	×

Table 8. The results for JPEG2000 compression, *dup1* probe set

*dup*2	JPEG2000		Orig.	1 bpp	0.5 bpp	0.3 bpp	0.2 bpp
PCA+L1	EXP1	RR	19.7	19.7	20.1	19.7	19.6
		McN	-	○	○	○	○
	EXP2	RR	19.7	20.5	19.7	18.8	17.9
		McN	-	○	○	○	○
LDA+COS	EXP1	RR	12.8	13.3	13.7	13.6	13.2
		McN	-	○	○	○	○
	EXP2	RR	12.8	13.2	13.7	13.7	13.2
		McN	-	○	○	○	○
ICA+COS	EXP1	RR	30.8	32.5	32.0	29.5	30.0
		McN	-	○	○	○	○
	EXP2	RR	30.8	32.5	30.8	29.1	22.7
		McN	-	○	○	○	×

Table 9. The results for JPEG2000 compression, *dup2* probe set

5. Analysis

5.1 Different expressions (*fb*)

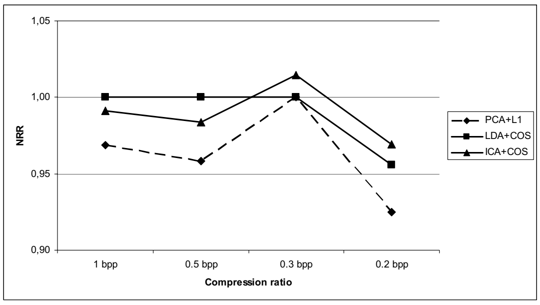
All methods exhibit great stability for both JPEG and JPEG2000 compression and in both EXP1 and EXP2 setups (Table 2 and Table 6). Even though there are a few recognition rate increases when the images are mildly compressed, none of those increases are statistically significant. If we take a look at the example of visual deformations introduced by compression (Figure 2), this level of stability is quite surprising. In spite of the fact that an image compressed to 0.3 bpp using JPEG is virtually unrecognizable and, on average, has PSNR = 25 dB, there seems to be no effect on face recognition performance. If we have a closer look at the results in Table 2 and Table 6, we can see that both JPEG and JPEG2000 do not significantly deteriorate performance until 0.2 bpp. At 0.2 bpp all recognition methods experience significant performance drop. We can conclude that, for the different expressions task, all compression ratios above 0.2 bpp are acceptable and can be used in a face recognition system. Unfortunately, rarely are such easy tasks (ideal imaging conditions and face images varying only in facial expressions) put before the systems designers and this is why we have to consider other possible influences on recognition accuracy as well (different illuminations and temporal changes).

JPEG2000 seems to be more efficient (in terms of image quality) if an image is to be presented to a human operator that has to make a final decision about someone's identity. This is an expected scenario in high confidence applications, like law enforcement applications. In such an application, a list of the most likely matches are presented to the user which now has to make the final choice. JPEG2000 images seem to be visually less distorted at higher compression rates and thus more appropriate for such uses. JPEG images can also be used, but at moderate or low compression rates (0.5 bpp and above).

The overall rank one recognition rates for the *fb* probe set are above 75%, which was expected and is consistent with previous studies of the same face recognition algorithms in pixel domain (Delac et al., 2006; Bartlett et al., 2002; Yambor et al., 2002; Beveridge et al., 2001; Belhumeur et al., 1996). ICA+COS yielded highest recognition rates in both experiments. For JPEG - 83% at 0.5 bpp in EXP1 and 83.1% at 1 bpp in EXP2 and for JPEG2000 – 83.1% at 0.5 bpp in EXP1 and 83.8% at 0.3 bpp in EXP2. It is interesting to notice that overall best results was achieved at a surprisingly high compression of 0.3 bpp (≈ 26:1).

5.2 Different illumination (*fc*)

The results for the *fc* probe set in both experiments can be seen in Table 3 and 7 and Figure 5 and 6. If we take a look at the results of both experiments for JPEG compression (Table 3 and Figure 5), we can see that compression again does not deteriorate performance up to 0.3 bpp. Only at 0.2 bpp the differences become statistically significant. These results are mainly quite similar to the *fb* probe set results. However, there are some differences, namely, the statistically significant recognition rate improvement for PCA+L1 with JPEG compression at 0.3 bpp in EXP2, and consistent significant improvement for JPEG2000 compression at 1, 0.5 and 0.3 bpp in EXP2. Both mentioned differences are clearly visible in Figure 5 and 6. In those figures the *NRR* curves are shown as a function of compression rate (in bpp) for all experiments with the *fc* probe set (Figure 5 for JPEG and Figure 6 for JPEG2000 compression). As already mentioned, PCA+L1 exhibits some statistically significant improvements in these experiments and this is clearly visible as the curves in Figure 5 and 6 exceed the value of one in those cases. This is a good example of the advantages of presenting results of similar experiments using the *NRR* curve.



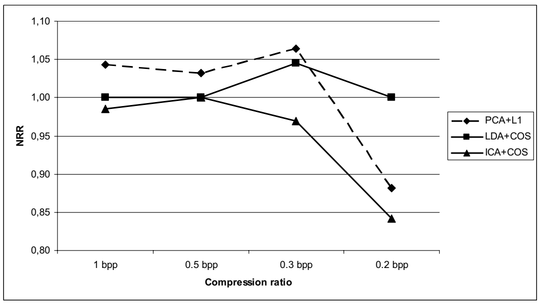
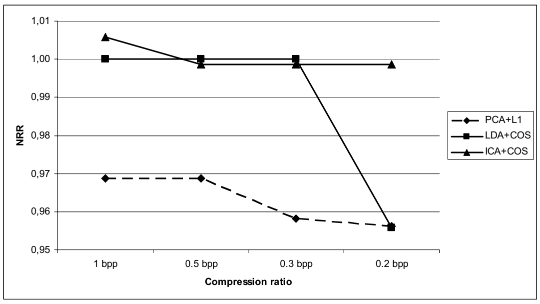


Figure 5. *NRR* curves for JPEG compression on the *fc* probe set (EXP1 top; EXP2 bottom)



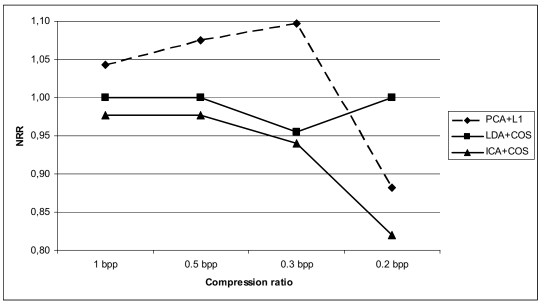


Figure 6. *NRR* curves for JPEG2000 compression on the *fc* probe set (EXP1 top; EXP2 bottom)

Compression drastically improves the results for PCA+L1 algorithm in some cases. For LDA+COS and ICA+COS this effect is not that emphasized. One might actually expect even worse results for compression of images taken in different illumination conditions. The different illumination influences large portions of an image and sometimes even the whole image. This being so, it appears that illumination changes are represented by low frequencies in an image, thus low-pass filtering (such as JPEG or JPEG2000 compression) should not eliminate the differences between various images taken in different illumination conditions. However, in spite of this, all algorithms seem to be very stable across a wide range of compression rates and in both experimental setups. Nastar et al. (Nastar et al., 1997) showed that only the high-frequency spectrum is affected by changes in facial expression. They also conjecture that illumination changes mostly affect the whole image, thus being in the low-frequency part of the spectrum. It is interesting to notice that PCA+L1 yielded the highest recognition rates for both JPEG and JPEG2000 compression at a very high compression rate of 0.3 bpp. The effect that compression has on PCA+L1 results could be further explored by reconstructing the compressed images after projection to PCA subspace and comparing the reconstructed images to original images to capture the differences induced by compression. The overall best rank one recognition rates for the *fc* probe set are achieved by ICA+COS in both experiments. For JPEG - 69.6% at 0.3 bpp in EXP1 and 68.6% at 0.5 bpp in EXP2 and for JPEG2000 – 69% at 1 bpp in EXP1 and 67% at 1 and 0.5 bpp in EXP2.

5.3 Temporal changes (*dup1* & *dup2*)

The results for probe sets that test the effect that aging of the subjects has on face recognition (*dup1* and *dup2*) are shown in Tables 4, 5, 8 and 9. The trend of very stable results across a wide range of compression rates is still noticeable. Additionally, for these probe sets all three algorithms have statistically insignificant performance differences, even at 0.2 bpp. Slight (statistically insignificant) improvements are noticeable at almost all compression rates and for all algorithms. It appears that the low-pass filtering by compression contributes more to the overall stability of the results than to significant improvements.

The overall best rank one recognition rates for the *dup1* probe set are achieved by ICA+COS in both experiments. For JPEG - 44.9% at 1 bpp in EXP1 and 45.3% at 1 bpp in EXP2 and for JPEG2000 – 44.7% at 1 bpp in EXP1 and 45% at 1 bpp in EXP2.

The overall best rank one recognition rates for the *dup2* probe set are achieved by ICA+COS in both experiments. For JPEG - 32% at 1 bpp in EXP1 and 31.2% at 1 and 0.3 bpp in EXP2 and for JPEG2000 – 32.5% at 1 bpp in EXP1 and 32.5% at 1 bpp in EXP2.

Mild compression of 8:1 (1 bpp) seems to be very effective at improving face recognition from images taken at different points in time. The removal of fine details, such as wrinkles and even facial hair, obviously makes images of the same person more similar.

5.4 Possible applications

We will now try to answer a question of where could the results and conclusions presented here be used in real life. We will describe two very basic applications. Firstly, as was previously hinted, the obvious use is in law enforcement applications. An image of an unknown subject is presented to the system, that image is compared to all the images known to the system. There can be hundreds of thousands of such images and any storage requirements save in such application is of crucial importance.

Secondly, there has recently been an increased interest in using face recognition systems in mobile and handheld devices (Wijaya et al., 2005). In such applications the face of the subject is recorded using a camera mounted on a device and transaction/login is approved or rejected based on that image. Recognition is mostly done at the remote server side and images (or some extracted image features) are sent over a telecommunication network. If a device in question is a mobile phone, higher level image processing is usually computationally expensive so the whole image is sent. Cameras usually deliver images in an already compressed format and being able to use this feature and send a compressed file across the network would be a big advantage.

6. Conclusion

We can group the conclusions based on a level of compression and the probe sets into two parts: i) higher compression rates (0.5, 0.3 and in some cases even 0.2 bpp) seem to be suitable for recognizing faces with different expressions (*fb* probe set) and images taken in different illumination conditions (*fc* probe set); ii) lower compression rates (1 bpp) seem to be suitable for recognizing images taken at different points in time (*dup1* and *dup2* probe set). Taking this analysis into account, it seems that the current practice of deciding on the level of compression based on visual distortion of images is wrong. While the images compressed to 0.3 bpp are visually significantly distorted, the recognition results are in almost all experiments statistically indistinguishable from the results achieved by using uncompressed images. In many cases these results are slightly better and in some cases even significantly better than the ones achieved with uncompressed images. The correct criteria for selecting the optimal compression ratio would therefore be: the optimal compression rate is the one yielding the highest recognition rate at given circumstances (classification algorithm, task given etc.). It certainly seems reasonable to allow image compression up to 0.5 bpp (a 16:1 compression) for face recognition purposes.

JPEG2000 compression seems to have less effect on recognition results than JPEG. Significant performance improvements are not as often as with JPEG, but all methods exhibit remarkable stability when JPEG2000 was used. This conclusion is similar to the one presented in (Schaefer, 2004), where the first comprehensive study of the influence of JPEG and JPEG2000 compression on content-based image retrieval was conducted. Schaefer concludes that JPEG2000 gives better results at higher compression rates than JPEG.

From the experiments presented in this chapter in can be concluded that *compression does not significantly influence face recognition performance up to 0.3 bpp*. In other words, there seems to be no reason not to store images in the compressed format. 0.3 bpp corresponds to compression ratio of about 26:1. Even using a more moderate compression of 1 bpp or 0.5 bpp would be a great save in storage requirements while retaining high visual quality of the reconstructed images. As far as the usage scenario (only probe images are compressed or the whole systems works with compressed images) is concerned, no conclusion can be drawn as to which is more suitable. However, since the transition to fully compressed domain recognition seems plausible, in order to be able to directly compare the results in both domains, the second scenario (the whole systems works with compressed images at a given compression rate) should be used when experimenting.

7. Acknowledgement

The work described in this chapter was conducted under the research project: "Intelligent Image Features Extraction in Knowledge Discovery Systems" (036-0982560-1643), supported by the Ministry of Science, Education and Sports of the Republic of Croatia.

Portions of the research in this chapter use the FERET database of facial images collected under the FERET program. The authors would like to thank the FERET Technical Agent, the U.S. National Institute of Standards and Technology (NIST) for providing the FERET database.

8. References

ANSI INCITS 385-2004 (2004). *Face Recognition Format for Data Interchange*

Bartlett, M.S.; Movellan, J.R. & Sejnowski, T.J. (2002). Face Recognition by Independent Component Analysis, *IEEE Trans. on Neural Networks*, Vol. 13, No. 6, November 2002, pp. 1450-1464

Belhumeur, P.N.; Hespanha, J.P. & Kriegman, D.J. (1996). Eigenfaces vs. Fisherfaces: Recognition using Class Specific Linear Projection, *Proc. of the 4th European Conference on Computer Vision, ECCV'96*, 15-18 April 1996, Cambridge, UK, pp. 45-58

Beveridge, J.R.; She, K.; Draper, B.A. & Givens, G.H. (2001). A Nonparametric Statistical Comparison of Principal Component and Linear Discriminant Subspaces for Face Recognition, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR'01*, Kaui, HI, USA, December 2001, pp. 535-542

Blackburn, D.M.; Bone, J.M. & Phillips, P.J. (2001). *FRVT 2000 Evaluation Report*, 2001, available at: http://www.frvt.org/FRVT2000/documents.htm

Delac, K.; Grgic, M. & Grgic, S. (2005). Effects of JPEG and JPEG2000 Compression on Face Recognition, *Lecture Notes in Computer Science, Pattern Recognition and Image Analysis*, Vol. 3687, 2005, pp. 136-145

Delac, K.; Grgic, M. & Grgic, S. (2006). Independent Comparative Study of PCA, ICA, and LDA on the FERET Data Set, *International Journal of Imaging Systems and Technology*, Vol. 15, No. 5, 2006, pp. 252-260

Delac, K. (2006). Face Recognition in Compressed Domain, *PhD Thesis*, University of Zagreb, Faculty of Electrical Engineering and Computing, 2006 (in Croatian)

Ebrahimi, F.; Chamik, M. & Winkler, S. (2004). JPEG vs. JPEG2000: An Objective Comparison of Image Encoding Quality, *Proc. of SPIE*, Vol. 5558, 2004, pp. 300-308

Grgic, S.; Grgic, M. & Zovko-Cihlar, B. (2001). Performance Analysis of Image Compression Using Wavelets, *IEEE Trans. on Industrial Electronics*, Vol. 28, Issue 3, 2001, pp. 682-695

ISO/IEC IS 19794-5 (2004). *Biometric Data Interchange Formats* - Part 5: Face Image Data

McGarry, D.P.; Arndt, C.M.; McCabe, S.A. & D'Amato, D.P. (2004). Effects of compression and individual variability on face recognition performance, *Proc. of SPIE*, Vol. 5404, 2004, pp. 362-372

Moon, H. & Phillips, P.J. (2001). Computational and Performance aspects of PCA-based Face Recognition Algorithms, *Perception*, Vol. 30, 2001, pp. 303-321

Nastar, C.; Moghaddam, B. & Pentland A. (1997). Flexible Images: Matching and Recognition using Learned Deformations, *Computer Vision and Image Understanding*, Vol. 65, No. 2, 1997, pp. 179-191

Phillips, P.J.; Moon, H.; Rizvi, S.A. & Rauss P.J. (2000). The FERET Evaluation Methodology for Face-Recognition Algorithms, *IEEE Transactions on Pattern Recognition and Machine Intelligence*, Vol. 22, No. 10, October 2000, pp. 1090-1104

Santa-Cruz, D.; Ebrahimi, T.; Askelof, J. & Larsson, M. & Christopoulos, C.A. (2000). JPEG 2000 Still Image Coding versus Other Standards, *Proc. of SPIE*, Vol. 4115, 2000, pp. 446-454

Schaefer, G. (2004). JPEG2000 vs. JPEG from an Image Retrieval Point of View, *Proc. International Conference on Image Processing*, *ICIP'04*, Vol. 1, 24-27 October 2004, pp. 437-440

Skodras, A.; Christopoulos, C. & Ebrahimi T. (2001). The JPEG 2000 Still Image Compression Standard, *IEEE Signal Processing Magazine*, Vol. 18, No. 5, September 2001, pp. 36-58

Turk, M. & Pentland, A. (1991). Eigenfaces for Recognition, *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, 1991, pp. 71-86

Wallace, G.K. (1991). The JPEG Still Picture Compression Standard, *Communications of the ACM*, Vol. 34, Issue 4, April 1991, pp. 30-44

Wat, K. & Srinivasan, S.H. (2004). Effect of Compression on Face Recognition, *Proc. of the 5th International Workshop on Image Analysis for Multimedia Interactive Services*, *WIAMIS 2004*, , Lisboa, Portugal, 21-23 April 2004, pp. 1071-1074

Wijaya, S.L.; Savvides, M. & Vijaya Kumar B.V.K. (2005). Illumination-Tolerant Face Verification of Low-Bit-Rate JPEG2000 Wavelet Images with Advanced Correlation Filters for Handheld Devices. *Journal of Applied Optics*, Vol. 44, 2005, pp. 655-665

Yambor, W.S.; Draper, B. & Beveridge, J.R. (2002). Analyzing PCA-Based Face Recognition Algorithms: Eigenvector Selection and Distance Measures, *Empirical Evaluation Methods in Computer Vision, Christensen H., Phillips J. (eds.)*, 2002, World Scientific Press, Singapore.

Zhao, W.; Chellastra, R.; Rosenfeld, A. & Phillips, P.J. (2003). Face Recognition: A Literature Survey, *ACM Computing Surveys*, Vol. 35, Issue 4, December 2003, pp. 399-458

Zhou, S.; Krueger, V. & Chellapa, R. (2003). Probabilistic Recognition of Human Faces from Video, *Computer Vision and Image Understanding*, Vol. 91, 2003, pp. 214-245