OPTIMISATION OF THE PRE-PROCESSING STAGE OF A SMART CARD FACE VERIFICATION SYSTEM

Thirimachos Bourlai and Josef Kittler

University of Surrey, SEPS/CVSSP Group
Guildford, GU2 7XH, Surrey, U.K.
phone: +44 1483 686030, fax: +44 1483 686031, email: {t.bourlai, j.kittler}@surrey.ac.uk
web: www.ee.surrey.ac.uk/CVSSP

ABSTRACT

We investigate the effect of the pre-processing stage parameters on performance of a smart card face verification system (SCFVS). We show that their optimal choice can improve not only the system accuracy but also the speed of verification. Some unexpected behaviour in the system performance is explained by a frequency space analysis of the effect of pre-processing. The analysis further leads to the broadening of the scope of parameter optimisation with beneficial results.

1. INTRODUCTION

The methodology of designing a smart card face verification system has been the focus of research for some time now [2, 1]. In a typical system [4] a biometric (face image, finger print) is acquired and compared with a stored template that has been constructed during enrolment. To alleviate many privacy and security issues raised by conventional architectures of face verification systems an alternative solution was reported in [2]. The key distinguishing feature of the proposed system is that the verification process is carried out on the card. This contrasts with other solutions where the card is used simply as a means of storing the biometric template (reference face model). In this architecture the input image (probe) is first filtered and then geometrically, as well as photometrically normalised in a local host. The registered and normalised probe is then transmitted to the card where the verification step is executed and the decision communicated to the service provider. The proposed system implements the revolutionary client specific linear discriminant analysis technique (CS-LDA) [5], which combines face representation and decision making into a single step that requires a template of the size of the input image. The solution avoids the need to implement the proprietary feature extraction computation in the host. This enhances the portability and security of the system as well as privacy.

The problem is that smart card computing platforms impose many engineering constraints and limitations [3, 7]. Therefore special considerations have to be given to the system design issues in order to improve performance while increasing system speed[1]. An important aspect of the design approach is the optimisation of the parameter settings of the complete system in general and the pre-processing stage in particular, with the criterions of optimisation being accuracy and system speed. We shall investigate the effect of the system optimisation in both the spatial and frequency domain.

The work in this paper starts with the investigation in the spatial domain. After the raw face image is acquired by the camera, a number of pre-processing steps are applied to it. These include different combinations of geometric normalisation, filtering, histogram equalisation, homomorphic filtering and JPEG compression. Different combinations of the above steps are compared in terms of system performance which is measured by using as a criterion of optimality the *half-total error rate* (HTER) on the test set of the benchmarking face database XM2VTS[2].

To gain better understanding of the effect of the various pre-processing steps, we shall study their impact on the probe images in the frequency domain. For this purpose the power spectrum of example probe images is computed. The experimental results show that the system performance is heavily affected by the pre-processing stage. They also reveal why some pre-processing steps produce unexpected behaviour in terms of performance (i.e. high compression resulted in lower error rates) and have motivated us to proceed with the pre-processing optimisation. It transpires that the optimal selection of the parameter settings can result in more than a 30% improvement in system performance.

The rest of the paper is organised as follows. In Section 2 we describe SCFVS. In Section 3 we summarise the previous work. In Section 4 the effect of different parameter settings of the pre-processing stage on system performance is presented. This is followed by the analysis of the effect in the frequency domain in Section 5. Conclusions are drawn in Section 6.

2. SCFVS: SYSTEM DESCRIPTION

Our SCFVS (see Figure 1) has a structure of a typical face-verification system. The face verification method adopted for the implementation on a smart card is the CS-LDA technique, which combines face representation and decision making into a single step, requiring a template of the size of the input image. That means that the *input image* (*live probe image* sent to the card for verification) has the same spatial resolution as the biometric template stored on the card to allow for a valid computation of the metric used in the decision making stage.

The steps involved in the SCFVS are described below:

Face Registration: The aim of the face registration (pre-processing) stage is to normalise the pose and resolution of the face image after face detection. Initially low-pass filtering is applied to the original image so as to remove the high-frequency noise.

Then *geometric normalisation* is performed by an eye position dependent utility. It is a fast, flexible, semi-automatic geometric alignment method based on the positions of the two eyes. This utility is used to crop the face part of the original image (all the image variations that are not directly related to the face with the verification process are removed) and scale

[1] A typical advanced specification smart card (used for the purpose of the experiments of this work) boasts a 13.5MHz processor, 1Mbyte of EEPROM, 8Kb of RAM, and can operate with a data transfer rate of up to 115.2Kbits per second (contact mode).

[2] HTER is obtained using the *equal error rate* threshold determined from the Receiver Operation Characteristic (ROC) curve computed on an independent evaluation set

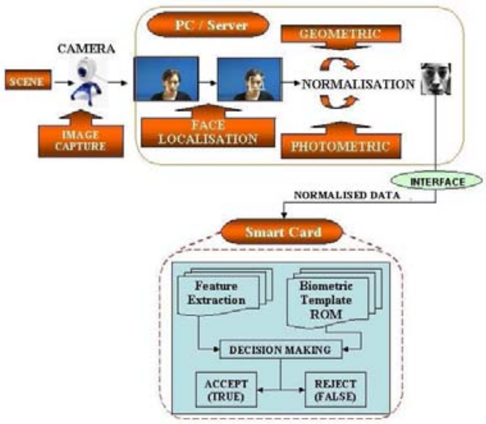

Figure 1: The proposed SCFVS Architecture.

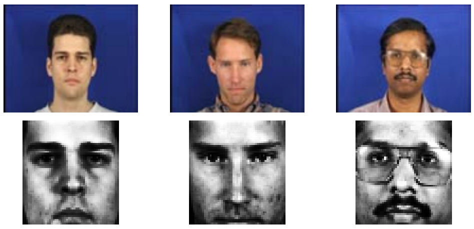

Figure 2: XM2VTS sample of unregistered and registered images using our system.

it to any desired resolution by using bi-linear interpolation. It adjusts the face in a standard position by using rotation, scaling and translation of the centre of the eyes to fixed locations, therefore removing variations in orientation, size, and location of the face in an image. We know that the fine registration of the face depends on the accurate localisation of the eyes. However we did not have such a problem because in the simulated experiments performed on the XM2VTS database the eye localisation was based on the manual registration of the eyes.

Photometric normalisation is carried out to minimise the detrimental effect of illumination changes on the system performance. It is performed after applying the image filtering to the raw (original) images to remove the high-frequency noise. Although various algorithms exist in the literature, Short et al. [8] reported that the use of homomorphic filtering in face verification is critical to achieve good performance. In our system we have followed this recommendation and employed a homomorphic filter (HF) and histogram equalisation (HEQ) [9].

In Figure 2 samples of the original XM2VTS unregistered images of the size 720×576 are shown, as well as the same images after registration, re-sized, converted to grey scale, filtered and finally geometrically and photometrically normalised).

Feature Extraction: The aim of feature extraction is to extract a compact set of interpersonal discriminating geometrical and/or photometrical features of the face. The CS-LDA transformation is determined in a two stage process. Initially a *PCA model* is constructed to achieve a dimensionality reduction and then an *LDA model* is derived to obtain the overall client *i* specific linear discriminant transformation. The transformation defines the client specific Fisher face for testing the claimed identity. The overall theory is described in [5].

Verification: The verification process is performed on the smart card. It involves computing a score quantifying the degree of match between the photometrically normalised image transmitted to the smart card and the user biometric template stored on the card [6].

Another pre-processing step that is considered is image compression that can be applied to the probe image before it is transmitted to the smart card. It usually involves a loss of the high frequency content. Image compression interacts with image filtering, and in turn, with the feature extraction process, as the amount of variance retained by the image data will depend on image smoothing.

We experimented with different spatial resolutions of the geometrically normalised and re-sized images to identify the optimum one in terms of performance (a task that proved to be protocol dependent). However, the *reference parameter set* (REF) uses a relatively low resolution for the face images, namely 55x51, with a grey level resolution of 8bpp. Note that if floats are used for each pixel, the size of the face image stored on the card is 11,22KByte. The selection of the REF parameter set is based on the work of Li [10], where the CS-LDA method was proposed and certain issues were recognised that can affect a smart-based face verification system (storage space of small platforms).

3. PREVIOUS WORK

In our previous work the SCFVS was evaluated on different datasets and system configurations hoping to achieve good and consistent results across all test databases and protocols for the same parameter setting. It transpired that each testing configuration required different parameter setting with the exception of grey level resolution (8bpp since an 8 bit camera is used) and 10-bit fixed point number representation. However, the optimum spatial resolution, JPEG compression quality factor[3] as well as JPEG operational scenario[4] differed from one experimental condition to another. The two most interesting results of this evaluation are the following:

1. In the experiments investigating the effect of spatial resolution, the initial raw face images of the XM2VTS dataset were geometrically and photometrically normalised from their original resolution to a spatial resolution that was varied from 110×102 down to 8×7 in 16 steps (see Figure 3). These steps were deliberately selected in an exponential form in order to emphasise the lower image resolutions, that can be stored in a lower memory volume, and offering a faster transfer of the normalised probe face to the smart card. Grey-scale resolution was kept at 8 bpp. *Figure* 3 shows the sensitivity of the SCFVS performance when spatial resolution is varied.
Generally speaking we expected the verification performance results to be worse at low-resolution than using high-resolution images. This was not always the case.
2. In the experiments where JPEG was used, we found that in order to optimise the smart card design, a JPEG quality factor should be selected, which is scenario and database dependent. Below this quality threshold, the performance can degrade. Above that, there is a surprisingly wide quality range where *compression does not seem adversely to affect*

[3] Image quality is traded off against file size by adjusting different quality settings for the compressor. The range of the quality factor has been modified from 5 to 100 and the optimum compression ratio in terms of performance has been identified.

[4] JPEG compression was applied to probe images of all experimental sets; to probe images of only evaluation and testing set; to templates; to both probes (training and testing) and templates.

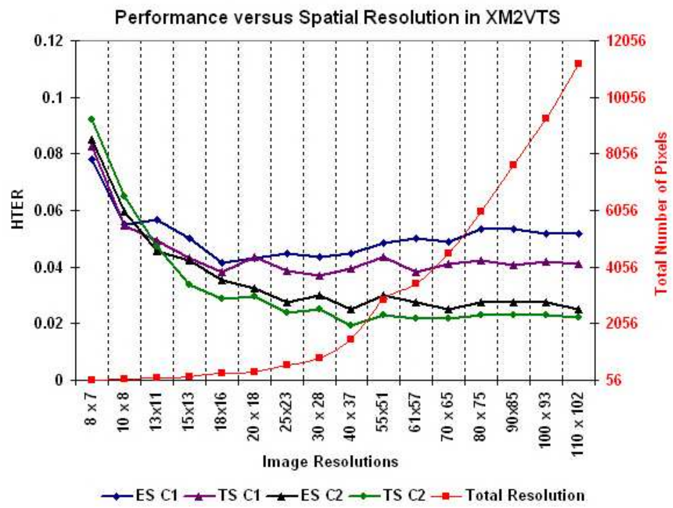

Figure 3: Evaluation and Testing results obtained on the XM2VTS database (C1/C2 configurations).

performance, and for the majority of the testing configurations *it may even improve system performance*. In particular, when JPEG probe images are compressed, the system does not behave as one would expect for large images and low JPEG quality factor settings (i.e. highly compressed images). Higher image resolutions should yield superior performance and, as the compression ratio increases (i.e. the quality factor in the JPEG compression decreases), performance should drop too. Figure 4 shows the sensitivity of the SCFVS performance when JPEG is employed and its quality is varied.

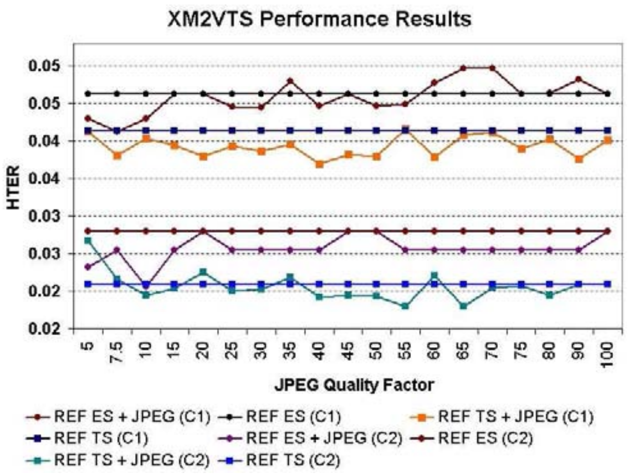

Figure 4: The effect JPEG compression on the images of the training and testing set of the XM2VTS database.ES/TS= Evaluation/Test Set.

These results indicate that the frequency content of the probe images should be further analysed. They suggest that somehow in the pre-processing stage of the verification system, image filtering and dimensionality *m* of the principal component subspace are not properly optimised.

What will follow is a study of the effect of pre-processing on the system performance realised by applying different photometric normalisation techniques, JPEG compression, and the number of PCA components. This study is performed in both the spatial and frequency domain.

4. SPATIAL DOMAIN

The purpose of the study is to evaluate the effect of four different configurations of the pre-processing stage on the SCFVS system performance. The first case simply involves geometrically normalised probe images. Then three other cases are considered that represent different combinations of photometric normalisation procedures. The relative sensitivity of the REF and BEST case when JPEG is employed is also investigated. These approaches are labelled as follows:

1. **INIT**: geometrically normalised image.
2. **noHOM**: Binomial filtering (BF) with kernel size 1×11, geometric normalisation (GEO), Histogram Equalisation (HistEq) and no Homomorphic Filtering (HomF).
3. **REF** or **REF+JPEG**: The reference parameter set *without/with* applying JPEG on the probe images (training and testing sets) - BF(1×11), GEO , HistEq and HomF.
4. **BEST** or **BEST+JPEG**: The optimum parameter sets for XM2VTS C1/C2 protocols *without/with* applying JPEG on the probe images (training and testing sets). Optimum parameter set of **XM2VTS C1** - Gaussian Filtering with kernel size 1×21 and $\sigma = 6.75$, optimum number of PCA components = 106, GEO, HistEq and HomF. For the **XM2VTS C2** - Gaussian Filtering with kernel size 1×25 and $\sigma = 1.5$, optimum number of PCA components = 330, GEO, HistEq and HomF.

The experimental results are shown in Table 1 and Figure 5 where we can see the system performance for the four approaches as well as the sensitivity of the REF vs. BEST+JPEG on the XM2VTS C1 protocol (C2 yields similar results). Since time and memory constraints are very important in smart card processing the user access time **UAT**[5] is also measured before and after using *JPEG compression* on the REF and BEST cases.

The experimental results reveal that by optimising the pre-processing stage for all compression ratios the system performance improves. Moreover the choice of filter, the filtering parameters and the amount of compression (that can be applied to achieve maximum performance) affect UAT and the memory usage of the card.

5. FREQUENCY DOMAIN

The purpose of this investigation is to evaluate the effect of different parameter settings in the frequency domain for each of the two protocols of the XM2VTS dataset. This is realised by applying the Fourier transform, which positions *Low Frequency Components near* the origin, while *High Frequency Components* are *further away*. The lowest frequency component (for zero frequency) is the d.c. component, which represents the average value of samples.

Four different ways are used to visualise the spectrum of the probe images. *Colour Spectrum*: obtained by pseudo colour coding of the spectrum. *3-dimensional plot*: demonstrates a *3D representation* and therefore a more direct way to visualise the Fourier transform. Z axis represents the magnitude of the

[5] It is the Total CPU Time in msec that the process spends in user and kernel mode that can be measured in face detection, normalisation, compression, transfer of the probe image to/from the card ($\frac{ImageSize}{Compression\ Ratio \times 14.4KBytes\ /s}$), decompression and matching. Since face detection and matching times are very small (a few msec), they are excluded from the measurements.

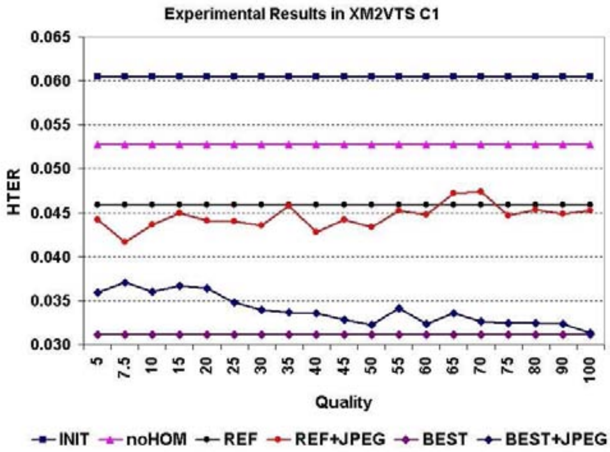

Figure 5: XM2VTS C1: Performance of the four different configurations of the pre-processing procedures and the sensitivity of the REF vs. BEST+JPEG.

Table 1: Performance vs different pre-processing approaches in XM2VTS. (A/M=Automatic/Manual selection of the number of PCA components, PROT=protocol, PRE=pre-processing approach, UAT=user access time(msec), CR=Compression ratio. Note that the size of the images before compression is 2805 bytes.)

PROT	PRE	PCA	HTER	CR	UAT
C1	**INIT**	235 A	0.0605	-	-
C1	**noHOM**	135 A	0.0527	-	-
C1	**REF**	211 A	0.0459	-	240
C1	**REF+JPEG**	193 A	0.0417	5.1:1	242
C1	**BEST**	106 M	0.0312	-	370
C1	**BEST+JPEG**	106 M	0.0322	2.76:1	372
C2	**INIT**	277 A	0.0394	-	-
C2	**noHOM**	148 A	0.0420	-	-
C2	**REF**	247 A	0.0264	-	250
C2	**REF+JPEG**	217 A	0.0220	4.76:1	252
C2	**BEST**	330 M	0.0176	-	395
C2	**BEST+JPEG**	330 M	0.0183	2.54:1	397

complex-valued function. This facilitates a more direct observation of the magnitude change of the power spectrum. *Contour*: it gives a clear view of the boundary of the frequency domain where the spectral content is above a given threshold. Finally, *a slice cut in the X direction* is used to display the spectrum distribution along the X-axis at the centre of the spectrum. The power spectrum is plotted for the four different cases discussed in the second paragraph of Section 4.

The study had explained why in the REF case presented in Figure 4 there is a performance improvement when JPEG is employed. Comparing the third (REF) and forth (REF+JPEG) row of Figure 6 we can see that even a low compression reduces the HTER of the system. This motivated us to perform further studies on pre-processing and to finally achieve the optimum (BEST) parameter sets presented in Section 4 and in the last two rows of Figure 6.

6. CONCLUSIONS

The experimental results demonstrate that the optimisation of the parameters of the pre-processing stage of a biometric system results in a considerable improvement in performance. Interestingly, the use of JPEG compression has proved to be beneficial in filtering out those high frequency components injected by photometric normalisation. As the amount of compression increases (increasing the compression quality parameter Q), more high frequency components are filtered out. However, even when a relatively low compression (Q=75) is applied (row four of *Figure 6*), enough high frequency components are filtered out to improve the system performance. This behaviour flags the problem of interaction of the various design parameters and raises pertinent questions about the optimality of the PCA subspace.

In this context we showed that to optimise the performance, the right selection of the filtering parameters and the number of PCA components is required. The spectrum obtained in the optimised cases *BEST C1/C2* can be viewed in the last two rows of *Figure 6* when the PCA space is optimised. Once the PCA space is optimised the use of JPEG compression would result in a loss of important high frequency components and consequently in performance degradation. Note that the required high frequency components for each protocol are different. Thus the optimum number of the PCA components will be data dependent.

Acknowledgements

The authors would like to acknowledge the support received from the Biosecure projects, EPSRC under grant GR/S46543/01(P) and the OmniPerception Ltd.

REFERENCES

[1] B.A.M. Schouten and J.W.H. Tangelder, “Non-intrusive Face Verification by a Virtual Mirror Interface Using Fractal Codes” in *In: CD-ROM Proceedings Biometrics on the Internet - Third COST 275 WORKSHOP*, 2005.

[2] T. Bourlai and J. Kittler and K. Messer, “On Optimisation of Smart Card Face Verification Systems” in *Proc. AVSS*, Sydney, Australia, November 22-24, 2006.

[3] J. Czyz and L. Vandendorpe, “Evaluation of LDA-based face verification with respect to available computational resources” in *Int’l Workshop on Pattern Recognition in Information Systems*, 2002.

[4] J. L. Dugelay and J. C. Junqua and C. Kotropoulos and R. Kuhn and F. Perronnin and I. Pitas “, Recent Advances in Biometric Person Authentication,” in *Proc. ICASSP*, Orlando, U.S.A., May, 2002.

[5] Y.P. Li and J. Kittler and J. Matas, “Face verification using client specific fisher faces” in *Proc. The Statistics of Directions, Shapes and Images*, September 3-5, 2000.

[6] M. Osborne and N. K. Ratha, “A JC-BioAPI Compliant Smart Card with Biometrics for Secure Access Control” in *Proc. AVBPA 2003*, June, 2003, pp. 903–910.

[7] W. Rankl and W. Effing, *Smart Card Handbook*. Address: John Wiley & Sons, 2000.

[8] J.Short and J.Kittler and K.Messer, “A Comparison of Photometric Normalisation Algorithms for Face Verification” in *Proc. Automatic Face and Gesture Recognition*, 2004, pp. 254–259.

[9] Y. P. Li and J. Kittler and J. Matas, “Analysis of the LDA-based matching schemes for Face Verification” in *Proc. BMVC 2000*, Bristol, U.K., September 11-14. 2000.

[10] Yongping Li, “Linear Discriminant Analysis and its application to Face Identification,” in *PhD Thesis, University of Surrey*, September, 2000.

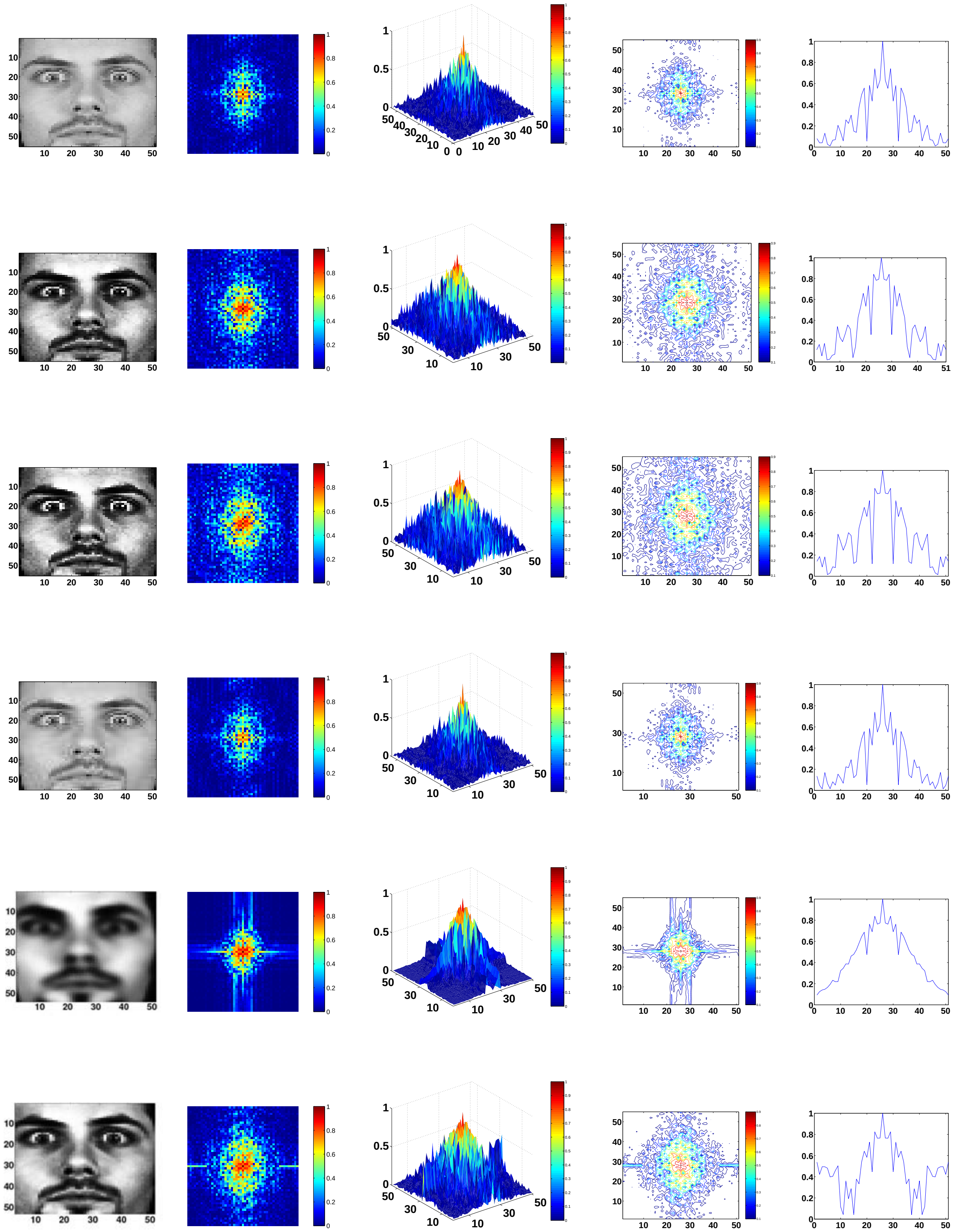

Figure 6: The power spectrum of subject 070_1_2 (XM2VTS) under different pre-processing conditions. From row one to six, the methods presented are, the INIT case (no pre-processing), noHOM, REF, REF+JPEG(Quality factor=75), BEST C1 and BEST C2.