Bimodal Biometric Person Identification System Under Perturbations

Miguel Carrasco[1], Luis Pizarro[2], and Domingo Mery[1]

[1] Pontificia Universidad Católica de Chile
Av. Vicuña Mackenna 4860(143), Santiago, Chile
mlcarras@puc.cl, dmery@ing.puc.cl
[2] Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science
Saarland University, Bldg. E11, 66041 Saarbrücken, Germany
pizarro@mia.uni-saarland.de

Abstract. Multibiometric person identification systems play a crucial role in environments where security must be ensured. However, building such systems must jointly encompass a good compromise between computational costs and overall performance. These systems must also be robust against inherent or potential noise on the data-acquisition machinery. In this respect, we proposed a bimodal identification system that combines two inexpensive and widely accepted biometric traits, namely face and voice information. We use a probabilistic fusion scheme at the matching score level, which linearly weights the classification probabilities of each person-class from both face and voice classifiers. The system is tested under two scenarios: a database composed of perturbation-free faces and voices (ideal case), and a database perturbed with variable Gaussian noise, salt-and-pepper noise and occlusions. Moreover, we develop a simple rule to automatically determine the weight parameter between the classifiers via the empirical evidence obtained from the learning stage and the noise level. The fused recognition systems exceeds in all cases the performance of the face and voice classifiers alone.

Keywords: Biometrics, multimodal, identificacion, face, voice, probabilistic fusion, Gaussian noise, salt-and-pepper noise, occlusions.

1 Introduction

Human beings possess a highly developed ability for recognising certain physiological or behavioral characteristics of different persons, particularly under high levels of variability and noise. Designing automatic systems with such capabilities comprises a very complex task with several limitations. Fortunately, in the last few years a large amount of research has been conducted in this direction. Particularly, *biometric systems* aim at recognising a person based on a set of intrinsic characteristics that the individual possesses. There exist many attributes that can be utilised to build an identification system depending on the application domain [1,2]. The process of combining information from multiple biometric

D. Mery and L. Rueda (Eds.): PSIVT 2007, LNCS 4872, pp. 114–127, 2007.

traits is known as *biometric fusion* or *multimodal biometrics* [3]. Multibiometric systems are more robust since they rely on different pieces of evidence before taking a decision. Fusion could be carried out at three different levels: (a) fusion at the feature extraction level, (b) fusion at the matching score level, and (c) fusion at the decision level [4].

Over the last fifteen years several multimodal schemes have been proposed for person identification [5,6,7]. It is known that the face and voice biometrics have lower performance compared to other biometric traits [8]. However, these constitute some of the most widely accepted by people, and the low cost of the equipment for face and voice acquisition makes the systems inexpensive to build. We refer to [9] for a relatively recent review on identity verification using face and voice information. We are interested in setting up a bimodal identification system that makes use of these two biometrics.

Traditional recognition systems are built assuming that the biometrics used in the learning (or training) process are noiseless. This ideal condition implies that all variables[1] susceptible to noise must be regulated. However, keeping all these variables under control might be very hard or unmanageable under the system's operation conditions. There are two alternatives to handle this problem. On the one hand, if the nature of the noise is known a suitable filter can be used in a preprocessing step. On the other hand, without any filtering, it is possible to build the recognition system with noisy data and make the biometric classifiers as robust as possible to cope with the perturbations. In this paper we are concerned with the latter alternative.

We propose a probabilistic fusion scheme performed at the matching score level, which linearly combines the classification probabilities of each authenticated person in both the face and the voice matching processes. The identity of a new input is associated with the identity of the authenticated person with the largest combined probability. We assess the robustness of the proposed bimodal biometric system against different perturbations: face images with additive Gaussian and salt-and-pepper noise, as well as with partial occlusions, and voice signals with additive white Gaussian noise. The performance of the fused system is tested under two scenarios: when the database is built on perturbation-free data (ideal case), and when it is built considering variable perturbations. Moreover, we develop a simple rule to automatically determine the weight parameter between the classifiers via empirical evidence obtained from the learning stage and the noise level. We show that combining two lower performance classifiers is still a convenient alternative in terms of computational costs/overall performance.

In Section 2 we describe classical techniques utilised in face and voice recognition. Section 3 details the proposed fused biometric system, which is tested under several perturbation conditions in Section 4. We conclude the paper in Section 5 summarising our contribution and delineating some future work.

[1] In the case of face and voice signals: calibration of audio/video recording devices, analog-digital data conversion, illumination conditions, background noise and interference, among others.

2 Face and Voice Recognition

Face recognition. At present there are three main approaches to the problem of face recognition: i) based on appearance, ii) based on invariant characteristics, and iii) based on models [10,11]. In the first approach the objective is to extract similar characteristics present in all faces. Usually statistical or machine learning techniques are used, and dimensionality reduction tools are very important for improving efficiency. One of the most widely used unsupervised tools in this respect is *principal component analysis* (PCA) [12]. This method linearly projects the high-dimensional input space onto a lower-dimensional subspace containing all the relevant image information. This procedure is applied over all the face images –training set– used for the construction of the identification system. This projection space is known as *eigenfaces space*. To recognize a new face the image is transformed to the projection space, and the differences between that projection and those of the training faces are evaluated. The smallest of these differences, which in turn is smaller than a certain threshold, gives the identity of the required face. The second approach is based on the invariant characteristics of the face, e.g., color, texture, shape, size and combinations of them. The objective consists in detecting those patterns that allow the segmentation of the face or faces contained in an image [13]. The third approach consists in the construction of models in two and three dimensions. Control points that identify specific face positions are determined robustly, and they are joined to form a nonrigid structure. Then this structure is deformed iteratively to make it coincide with some of the structures recognized by the identification system [14]. Unfortunately, this technique is very slow and requires the estimation of precise control points, and therefore the image must have high resolution. Also, because of the iteration process, it can be trapped in local optima, and is therefore dependent on the position of the control points chosen initially.

The different face recognition algorithms depend on the application's domain. There is no system that is completely efficient under all conditions. Our study is limited to developing an identification mechanism considering images captured in controlled environments. The approach chosen is that based on appearance and its implementation through PCA-eigenfaces.

Voice recognition. Voice recognition is the process of recognizing automatically who is speaking by means of the information contained in the sound waves emitted [15,16]. In general, voice recognition systems have two main modules: extraction of characteristics, which consists in obtaining a small but representative amount of data from a voice signal, and comparison of characteristics, which involves the process of identifying a person by comparing the characteristics extracted from its voice with those of the persons recognized by the identification system. Voice is a signal that varies slowly with time. Its characteristics remain almost stationary when examined over a sufficiently short period of time (ca. 5-100 ms). However, over longer time periods (more than 0.2 s) the signal's characteristics change, reflecting the different sounds of voice. Therefore, the most natural way of characterizing a voice signal is by means of the so-called *short-time*

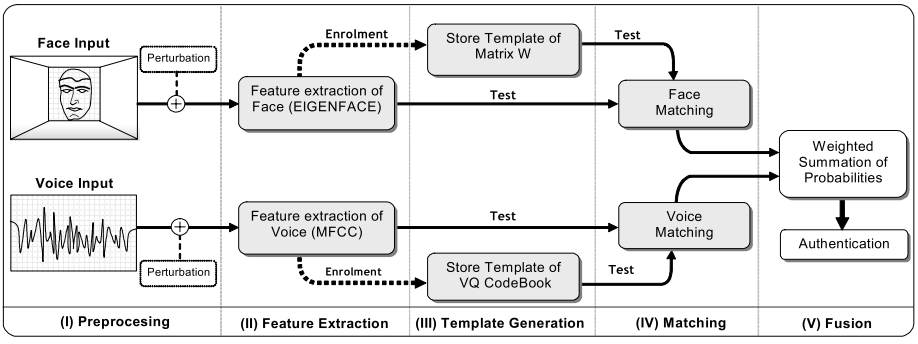


Fig. 1. Proposed framework for person identification

spectral analysis. One of the most widely used techniques in voice recognition is *mel-frequency cepstrum coefficients* (MFCC) [17,18], which we also use in this study. Basically, MFCC imitates the processing by the human ear in relation to frequency and band width. Using filters differentiated linearly at low frequencies (below 1000 Hz) and logarithmically at high frequencies, MFCC allows capturing the main voice's characteristics. This is expressed in the literature as the *mel-frequency scale*. We use this approach for voice characterisation.

3 Fusion of Face and Voice Under Perturbations

As previously mentioned, face and voice biometrics have lower performance compared to other biometric traits [8]. Nevertheless, it is relatively inexpensive to set up systems based on such biometrics. Moreover, PCA-eigenfaces and MFCC techniques require simple computation compared to other more sophisticated techniques [11].

Probabilistic fusion framework. Our proposal consists in fusing these lower performance classifiers by means of a simple probabilistic scheme, with the aim of obtaining an identification system with better performance and robust against different perturbations. The construction of such a system consists of the following five phases outlined in Fig. 1 and described next.

I. **Preprocessing.** In this phase k face images and k voice signals are considered for each one of the t persons in the system. With the purpose of examining the behaviour of the classifiers constructed with altered data, both signals are intentionally contaminated with different kinds of perturbations. The face images are contaminated with Gaussian, salt-and-pepper noise, or partial occlusions, while the voice signals are perturbed with additive white Gaussian noise. This also allows us to verify the performance of the algorithms used in our study for the extraction of characteristics. All signals belonging to a person j, perturbed or not, are associated with the person-class $C(j)$, for all $j = 1, \ldots, t$.

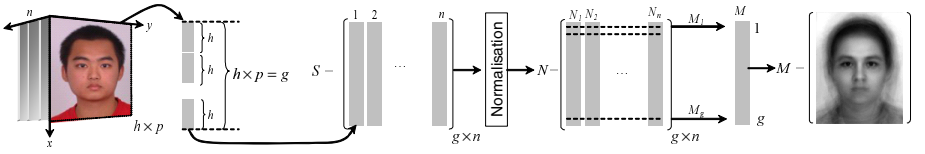


Fig. 2. Vector transformation of each image of the training set and later normalization and calculation of the mean image of the set of training images

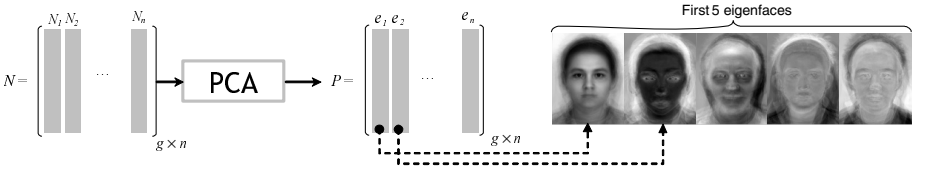


Fig. 3. Generation of the eigenfaces by means of PCA using the normalized data

II. **Face feature extractor.** To extract the face features we use the method known as *eigenfaces* [19]; see figures 2 and 3. All the images of the training set are transformed into column vectors and are concatenated in a matrix S. This matrix is normalized (N) by subtracting the mean and dividing by the standard deviation of each column. This improves contrast and decreases the effect of changes in illumination. Then, by averaging its rows, matrix N is reduced to a column vector M which represents the mean image of the training set. Then, applying PCA to normalization matrix N we obtain eigenfaces matrix P. Column vectors $e_1, \ldots, e_n$ represent the eigenfaces, and they are ordered from more to less information content. Finally, matrix W is obtained which contains the characteristics of the training set. This is calculated as the cross product between corresponding columns in the normalization and projection matrices, i.e. $W_i = N_i \cdot P_i$, for all columns $i = 1, \ldots, n$.

Voice feature extractor. The process of generation of the MFCC coefficients requires a set of steps that transform a voice signal into a matrix that contains its main characteristics. Initially, the audio signal is divided into a set of adjacent frames. Then each frame is filtered through a Hamming window, allowing the spectral distortion to be minimized both at the beginning and at the end of each frame. Then a transformation is made in each frame to the spectral domain with the Fourier transform; the result of this transformation is known as a *spectrum*. The next step transforms each spectrum into a signal that simulates the human ear, known as a *mel scale*. Finally, all the mel-spectra are transformed into the time domain by means of the discrete cosine transform. The latter step generates as a result the *mel frequency cepstrum coeficients* (MFCC) of the voice signal. For details we refer to [20].

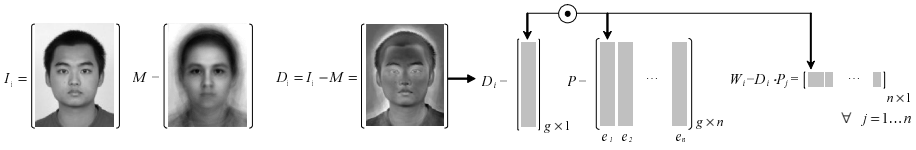


Fig. 4. Determination of the difference image using the general mean of the training set and the calculation of the characteristics vector W_i of face i

III. **Template generation.** The process of storing the biometric characteristics extracted before is called *enrolment*. In the case of the face, the characteristics matrix W is stored. For the voice, a compressed version of the signals of each person is stored. For that purpose, we make use of the LBG clustering algorithm [21], which generates a set of vectors called *VQ-Codebook* [22]. The registered features are considered as templates with which the features of an unknown person must be compared in the identification process.

IV. **Face matching.** To determine probabilistically the identity of an unknown person i, first the difference D_i between its normalized image I_i and the mean image M of the training set is calculated. Then the characteristics vector W_i is generated as the dot product between D_i and each column of the projection matrix P; see Fig. 4. Later, the Euclidian distances between the vector W_i and all the columns of the characteristics matrix W are computed. The k shortest distances are used to find the most likely person-class $C(j)$ to which the unknown person i is associated with.

Voice matching. In the case of voice, the process consists in extracting the ceptrum coefficients of the unknown speaker i by means of the calculation of the MFCCs, and calculating their quantized vector qv_i. Then the Euclidian distances between qv_i and all the vectors contained in the VQ-codebook are determined. The same as with the face, the k shortest distances are used to find the most likely person-class $C(j)$ to which the unknown speaker i is associated with.

V. **Fusion.** Finally, the response of the fused recognition system is given as a linear combination of the probabilistic responses of both the face classifier and the voice classifier. Since each person in the database has k signals of face and voice, the nearest k person-classes associated to an unknown person i represent those that are more similar to it. Thus, if the classification were perfect, these k classes should be associated with the same person, such that the classification probability would be $k/k = 1$. The procedure consists of two steps: Firstly, we determine the classification probability of each person j for face matching $P_f(j)$, as well as for voice matching $P_v(j)$:

$$P_f(j) = \frac{V_f(j)}{k}, \qquad P_v(j) = \frac{V_v(j)}{k}, \qquad \text{for all } j = 1, \ldots, t; \tag{1}$$

where $V_f(j)$ and $V_v(j)$ is the number of representatives of the person-class $C(j)$ out of the k previously selected candidates in the face matching and

in the voice matching stages, respectively. Secondly, we infer the identity of an unknown person i with the person-class $C(j)$ associated with the largest value of the combined probability

$$P(j) = \alpha \cdot P_f(j) + (1 - \alpha) \cdot P_v(j), \quad \text{for all } j = 1, \ldots, t. \tag{2}$$

The parameter $\alpha \in [0, 1]$ weights the relative importance associated with each classifier. In the next section we present a simple rule to estimate this parameter.

Estimation of the weight parameter α. The weight α is the only free parameter of our probabilistic fusion model and it is in connection with the reliability that the recognition system assigns to each classifier. Therefore, its estimation must intrinsically capture the relative performance between the face classifier and the voice classifier in the application scenario. In general, as it will be shown in the experimental section, estimating this parameter depends on the input data.

Heuristically, the feature learning process provides empirical evidence about the performance of the face and voice classifiers. Once the learning is done, the identification capabilities of the system are tested on faces and voices belonging to the set of t recognisable persons, though these data have not been previously used for learning. In this way, we have quantitative measurements of the classifiers' performance at our disposal. Thus, a simple linear rule for estimating α based on these measurements is given by

$$\hat{\alpha} = \frac{1 + (q_f - q_v)}{2}, \tag{3}$$

where q_f, $q_v \in [0, 1]$ are the empirical performance of the face and voice classifiers, respectively. This formula assigns more importance to the classifier that performs better under certain testing scenario. When both classifiers obtain nearly the same performance, their responses are equally considered in equation (2). This scheme agrees with the work by Sanderson and Paliwal [9], since assigning a greater weight to the classifier with better performance clearly increases the performance of the fused recognition.

4 Experimental Results

The data base used consists of 18 persons, with eight different face and voice versions for each one. The faces used are those provided by the Olivetti Research Laboratory (ORL) [23]. The faces of a given person vary in their facial expressions (open/closed eyes, smiling/serious), facial details (glasses/no-glasses), and posture changes. The voices were generated using an electronic reproducer in MP3 format at 128 kbps. A total of 144 recordings (8 per person) were made.

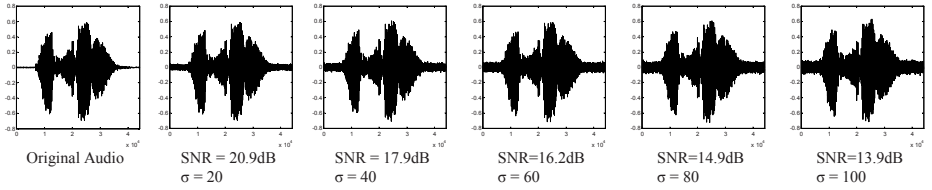


Fig. 5. One of the voice signals used in the experiments and some of its noisy versions

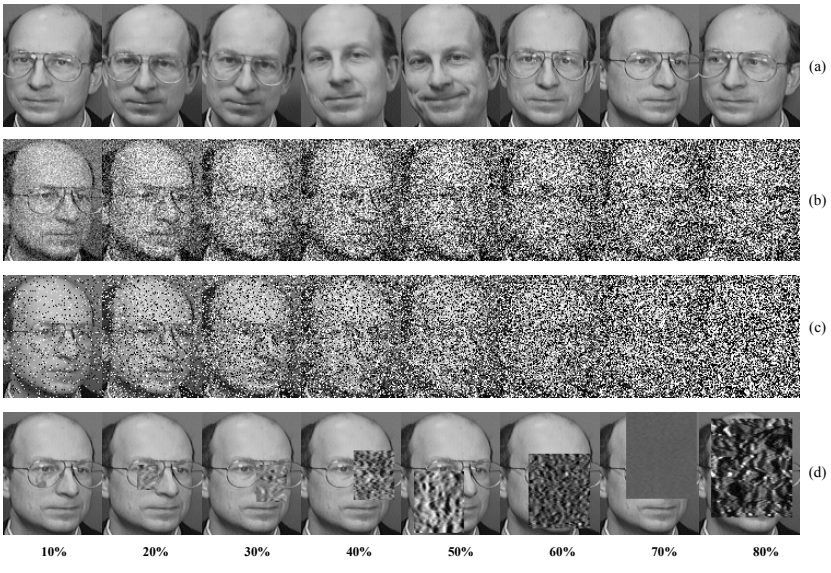


Fig. 6. (a) Original face sequence of an individual with eight different expressions. (b) Sample with variable Gaussian noise. (c) Sample with variable salt-and-pepper noise. (d) Sample with variable textured occlusions.

Face and voice classifiers alone. We performed two types of experiments to analyse the effect of noisy data on the performance of the face and voice classifiers without fusion. In the first experiment (*Exp.1*), the recognition system is constructed with perturbation-free data, but later it is tested on noisy data. In the second experiment (*Exp.2*), the recognition system is constructed considering various perturbations of the face and the voice signals, and tested then on perturbation-free data. Different perturbation levels were considered. The voice signals contain additive white Gaussian noise with zero mean and variable standard deviation $\sigma = \{0, 10, \ldots, 100\}$ of their mean power weighted by a factor of 0.025. The faces contain additive Gaussian noise with zero mean and standard deviation $\sigma = \{0, 10, \ldots, 100\}$ of the maximal grey value, additive salt-and-pepper noise that varies between 0% and 100% of the number of pixels, or randomly located textured occlusions whose size varies between 0% and 100%

of the image area [24]. Figures 5 and 6 show examples of the data utilised in testing.

The experiment *Exp.1* in Fig. 7(a) shows, on the one side, that the MFCC method has a low capability of recognising noisy data when only clean samples have been used for training. On the other side, we observe that PCA-eigenfaces[2] deals quite well with all types of noise till 70% of perturbation, and it is specially robust against Gaussian noise. Surprisingly, the experiment *Exp.2* in Fig. 7(b) reveals an improvement on the voice recognition when this classifier is constructed considering noisy samples. However, the face recognition is now able to satisfactorily manage up to 30% of perturbations. Notice that when no perturbations at all are considered (ideal case), the performance of the classifiers is around 90%.

Fused voice and face recognition. In this section we aim at combining the responses of both the face classifier and the voice classifier using the relation (2). A crucial aspect of this objective is the proper estimation of the weight parameter α. In the experiments of the previous section we varied the noise level over a large range, and the results logically depended on the amount of noise. We would like to use the formula (3) to adjust the computation of the parameter α to the noise level. This assumes that we should have quantitative measurements of the noise level on the voice and face samples, but in a real application the amount of noise is not known in advance. The estimation of these quantities for the different signals used here is out of the scope of this paper. However, we cite several strategies appear in the literature for noise estimation in audio [25,26,27] and image [28,29,30,31,32,33] signals.

If we assume that we have reliable estimations of the noise level in voice and face signals, and since the empirical performances of the classifiers are known from the learning stage under different testing scenarios, it is possible to compute the parameter α using the relation (3). For example, considering voice signals with variable white Gaussian noise and face images with salt-and-pepper noise, the figures 8(a) and 8(b) show the estimated $\hat{\alpha}$ curves for the experiments *Exp.1* and *Exp.2* of the figures 7(a) and 7(b), respectively. Evidently, the weight α increases as the noise in the voice signal increases, because voice recognition is more sensitive to noise than face recognition.

Again, we measure the performance of the fused recognition under two experimental scenarios: *Exp.3*: system built with perturbation-free data and tested then on noisy samples; and *Exp.4*: system built with noisy data and tested then on noiseless samples. Figures 9 and 10 show the recognition performance for these two operation settings, respectively. The missing $\hat{\alpha}$ curves have been omitted for the sake of space and readability. Notice that the performance of the ideal case now reaches 100%. Similarly, under the same experimental settings, the fused recognition outperforms the voice and face classifiers alone. The performance

[2] Although PCA may require a precise localisation of the head, the set of faces used in the experiments were not perfectly aligned, as shown in Fig. 6. However, satisfactory results are still achievable.

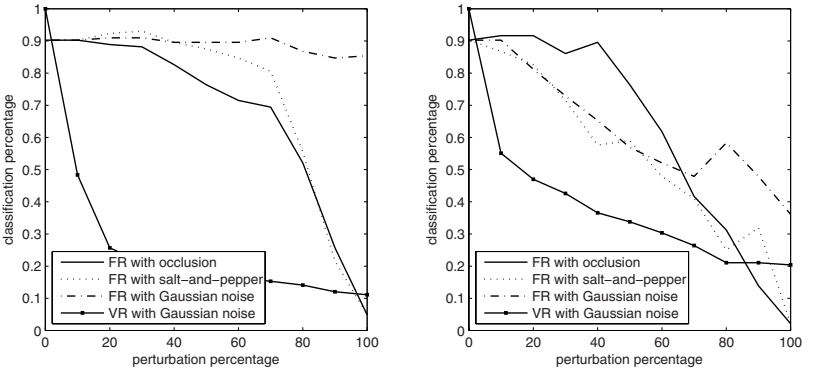


Fig. 7. Independent performance of voice recognition (VR) and face recognition (FR) systems. (a) *Exp.1*: Recognition systems built with perturbation-free data and tested on samples with variable noise. (b) *Exp.2*: Recognition systems built with variable noisy data and tested on noiseless samples.

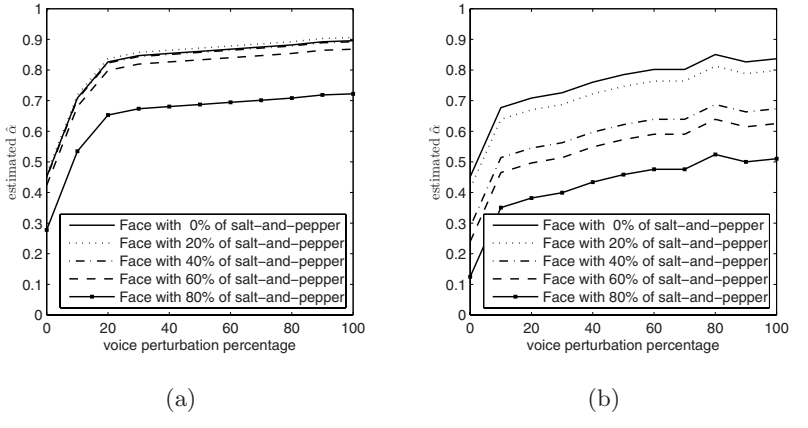
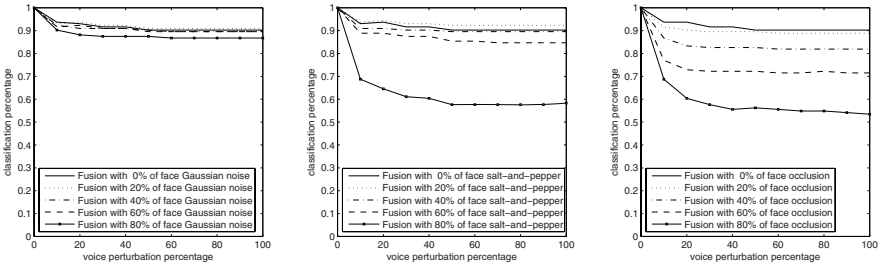


Fig. 8. Estimated $\hat{\alpha}$ curves when voice signals with variable white Gaussian noise and face images with salt-and-pepper noise are considered in (a) *Exp.1*, and (b) *Exp.2*

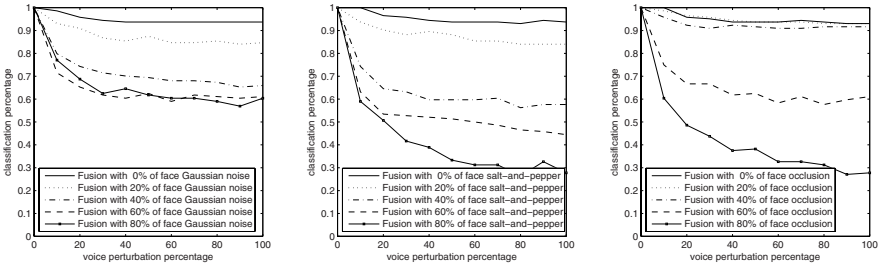
stability of the experiment *Exp.3* is in accordance with the much larger influence of the face classifier as outlined the Fig. 7(a). Although such an influence is not so large in Fig. 7(b), the experiment *Exp.4* also enjoys certain stability.

With respect to the robustness of the arbitrarily chosen feature extraction tools, it was shown that occlusions cause greater impact than Gaussian or salt-and-pepper noise on the eigenfaces analysis, and the analysis of the voice signals via MFCC is much more sensitive to white noise. However, even when the face and voice classifiers might reach a low performance independently, it is possible



(a) (b) (c)

Fig. 9. *Exp.3*: Performance of a bimodal person identification system by fusing voice and face classifiers. The system is built with perturbation-free data and tested then on noisy samples. Voice signals with white Gaussian noise and image faces with (a) Gaussian noise, (b) salt-and-pepper noise, and (c) textured occlusions, are considered.



(a) (b) (c)

Fig. 10. *Exp.4*: Performance of a bimodal person identification system by fusing voice and face classifiers. The system is built with noisy data and tested then on noiseless samples. Voice signals with white Gaussian noise and image faces with (a) Gaussian noise, (b) salt-and-pepper noise, and (c) textured occlusions, are considered.

to obtain a much better recognition system when the responses of both classifiers are fused in a probabilistic manner. Similarly, by improving the performance of the independent classifiers the overall performance increases too.

It has been shown that, depending on the learning and operation conditions of the identification system, it might be worthwhile to consider not only ideal noiseless samples when building the classifiers, but also inherent or potential sources of noise, which may improve the whole identification process.

For a particular application, the impact of every source of noise in the learning step as well as in the operation step should be evaluated before the identification system is set up. In the light of that study, the decision of building the system under noise samples or not should be taken.

5 Conclusions and Future Work

This work presents a biometric person identification system based on fusing two common biometric traits: face and voice. The fusion is carried out by a simple probabilistic scheme that combines the independent responses from both face and voice classifiers. The performance of the recognition system is assessed under different types of perturbations: Gaussian noise, salt-and-pepper noise and textured occlusions. These perturbations might affect the samples used to build the classifiers, and/or the test samples the system must identify. It is shown that the proposed probabilistic fusion framework provides a viable identification system under different contamination conditions, even when the independent classifiers have low single performance. We present a simple formula to automatically determine the weight parameter that combines the independent classifiers' responses. This formula considers the empirical evidence derived from the learning and testing stages, and it depends in general on the noise level. As future work, we will investigate more robust feature extraction tools that provide better results under this probabilistic scheme. We also seek for alternative ways to estimate the weight parameter.

Acknowledgments. This work was partially funded by CONICYT project ACT-32 and partially supported by a grant from the School of Engineering at Pontificia Universidad Católica de Chile. The authors would like to thank the G'97-USACH Group for their voices utilized in this research.

References

1. Prabhakar, S., Pankati, S., Jain, A.K.: Biometric recognition: Security and privacy concerns. IEEE Security and Privacy 01(2), 33–42 (2003)
2. Jain, A.K.: Biometric recognition: How do i know who you are? In: Roli, F., Vitulano, S. (eds.) ICIAP 2005. LNCS, vol. 3617, pp. 19–26. Springer, Heidelberg (2005)
3. Ross, A., Jain, A.: Multimodal biometrics: An overview. In: Proc. 12th European Signal Processing Conference, EUSIPCO 2004, Vienna, Austria, pp. 1221–1224 (September 2005)
4. Ross, A., Jain, A.K.: Information fusion in biometrics. Pattern Recognition Letters 24(13), 2115–2125 (2003)
5. Brunelli, R., Falavigna, D.: Person identification using multiple cues. IEEE Trans Pattern Anal Mach Intell 17(10), 955–966 (1995)
6. Bigün, E., Bigün, J., Duc, B., Fischer, S.: Expert conciliation for multi modal person authentication systems by bayesian statistics. In: Bigün, J., Borgefors, G., Chollet, G. (eds.) AVBPA 1997. LNCS, vol. 1206, pp. 291–300. Springer, Heidelberg (1997)
7. Snelick, R., Uludag, U., Mink, A., Indovina, M., Jain, A.: Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems. IEEE Trans Pattern Anal Mach Intell 27(3), 450–455 (2005)
8. Jain, A.K., Ross, A.: Multibiometric systems. 47(1), 34–40 (2004)
9. Sanderson, C., Paliwal, K.K.: Identity verification using speech and face information. Digit Signal Process 14(5), 449–480 (2004)

10. Yang, M.-H., Kriegman, D.J., Ahuja, N.: Detecting faces in images: A survey. IEEE Trans Pattern Anal Mach Intell 24(1), 34–58 (2002)
11. Lu, X.: Image analysis for face recognition: A brief survey. Personal Notes (May 2003)
12. Ruiz-del-Solar, J., Navarrete, P.: Eigenspace-based face recognition: a comparative study of different approaches. IEEE Trans Syst Man Cybern C Appl Rev 35(3), 315–325 (2005)
13. Guerfi, S., Gambotto, J.P., Lelandais, S.: Implementation of the watershed method in the hsi color space for the face extraction. In: IEEE Conference on Advanced Video and Signal Based Surveillance, 2005. AVSS 2005, pp. 282–286. IEEE Computer Society Press, Los Alamitos (2005)
14. Lu, X., Jain, A.: Deformation analysis for 3d face matching. In: Proc. Seventh IEEE Workshops on Application of Computer Vision, WACV/MOTION 2005, pp. 99–104. IEEE Computer Society Press, Los Alamitos (2005)
15. Doddington, G.R.: Speaker recognition identifying people by their voices. Proc. IEEE 73(11), 1651–1664 (1985)
16. Furui, S.: Cepstral analysis technique for automatic speaker verification. IEEE Trans Acoust Speech Signal Process 29(2), 254–272 (1981)
17. Murty, K.S.R., Yegnanarayana, B.: Combining evidence from residual phase and mfcc features for speaker recognition. IEEE Signal Process Lett 13(1), 52–55 (2006)
18. Picone, J.W.: Signal modeling techniques in speech recognition. Proc. IEEE 81(9), 1215–1247 (1993)
19. Kirby, M., Sirovich, L.: Application of the karhunen-loeve procedure for the characterization of human faces. IEEE Trans Pattern Anal Mach Intell 12(1), 103–108 (1990)
20. Wei, H., Cheong-Fat, C., Chiu-Sing, C., Kong-Pang, P.: An efficient mfcc extraction method in speech recognition. In: Proc. 2006 IEEE International Symposium on Circuits and Systems, ISCAS 2006, may 2006, pp. 145–148. IEEE Computer Society Press, Los Alamitos (2006)
21. Linde, Y., Buzo, A., Gray, R.M.: An algorithm for vector quantizer design. IEEE Trans Comm 28(1), 84–95 (1980)
22. Kinnunen, I., Kärkkäinen, T.: Class-discriminative weighted distortion measure for vq-based speaker identification. In: Caelli, T.M., Amin, A., Duin, R.P.W., Kamel, M.S., de Ridder, D. (eds.) SPR 2002 and SSPR 2002. LNCS, vol. 2396, pp. 681–688. Springer, Heidelberg (2002)
23. Samaria, F., Harter, A.: Parameterisation of a stochastic model for human face identification. In: Proc. 2nd IEEE Workshop on Applications of Computer Vision, pp. 138–142. IEEE Computer Society Press, Los Alamitos (1994)
24. Dana, K.J., Van-Ginneken, B., Nayar, S.K., Koenderink, J.J.: Reflectance and texture of real world surfaces. ACM Transactions on Graphics (TOG) 18(1), 1–34 (1999)
25. Yamauchi, J., Shimamura, T.: Noise estimation using high frequency regions for speech enhancement in low snr environments. In: Proc. of the 2002 IEEE Workshop on Speech Coding, pp. 59–61. IEEE Computer Society Press, Los Alamitos (2002)
26. Reju, V.G., Tong, Y.C.: A computationally efficient noise estimation algorithm for speech enhancement. In: Proc. of the 2004 IEEE Asia-Pacific Conference on Circuits and Systems, vol. 1, pp. 193–196. IEEE Computer Society Press, Los Alamitos (2004)
27. Wu, G.D.: A novel background noise estimation in adverse environments. In: Proc. of the 2005 IEEE International Conference on Systems, Man and Cybernetics, vol. 2, pp. 1843–1847. IEEE Computer Society Press, Los Alamitos (2005)

28. Starck, J.L., Murtagh, F.: Automatic noise estimation from the multiresolution support. Publications of the Astronomical Society of the Pacific 110(744), 193–199 (1998)
29. Salmeri, M., Mencattini, A., Ricci, E., Salsano, A.: Noise estimation in digital images using fuzzy processing. In: Proc. of the 2001 International Conference on Image Processing, vol. 1, pp. 517–520 (2001)
30. Shin, D.H., Park, R.H., Yang, S., Jung, J.H.: Block-based noise estimation using adaptive gaussian filtering. IEEE Transactions on Consumer Electronics 51(1), 218–226 (2005)
31. Liu, C., Freeman, W.T., Szeliski, R., Kang, S.B.: Noise estimation from a single image. In: Proc. of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 901–908. IEEE Computer Society Press, Los Alamitos (2006)
32. Grammalidis, N., Strintzis, M.: Disparity and occlusion estimation in multiocular systems and theircoding for the communication of multiview image sequences. IEEE Transactions on Circuits and Systems for Video Technology 8(3), 328–344 (1998)
33. Ince, S., Konrad, J.: Geometry-based estimation of occlusions from video frame pairs. In: Proc. of the 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, pp. 933–936. IEEE Computer Society Press, Los Alamitos (2005)