ENVIRONMENT-ADAPTIVE MULTI-CHANNEL BIOMETRICS

Stephen M. Chu[1], Minerva Yeung[2], Luhong Liang[1] and Xiaoxing Liu[1]

[1]Beijing, China, [2]Santa Clara, USA
Microprocessor Research Labs, Intel Corporation.

ABSTRACT

This paper looks into multi-channel/multimodal biometric systems that are adaptive to environmental variations. In this work, we introduce a general formulation that addresses the environmental robustness of multi-channel fusion in biometric systems. Based on the formulation, two audio-visual biometric systems are developed. The first relies on confidence measures derived from the environmental conditions to dynamically weight the contributions of the biometric channels; whereas the second considers the multiple channels jointly to optimally adjust the fusion parameters according to the current environmental conditions. Experimental evaluations with varying testing conditions show that both systems achieve lower recognition error rate comparing with a baseline non-environment-adaptive audio-visual system. It is further shown that incorporating joint-optimization of multi-channel fusion parameters to cater to environmental changes as in the second system consistently leads to improved recognition accuracy over other systems, and at the same time guarantees to perform no worse than any of the individual biometric channels under all environmental conditions.

1. INTRODUCTION

Biometrics is an emerging technological field that potentially has a wide range of applications including access control, surveillance, and intelligent interfaces. While human biological traits are ubiquitous and can offer much higher security, the state-of-the-art biometric systems still lag behind the simple non-biological based methods like using the badge or password with respect to reliability.

One promising approach to improve the reliability of biometric systems is to consider multiple biometric channels. Different biometric channels may carry complementary information about the person being identified. Therefore, the conjunction of the multiple information sources could lead to more robust cues about the subject. Furthermore, consider the situation where a biometric signal is unavailable, for example: the voice characteristics of a person using speaker recognition system may be altered because of illness; or consider the situation when single-channel biometrics become less reliable due to environmental changes, such as variation in lighting conditions for a face recognition system. In these situations, the multi-channel approach could achieve higher accuracy than any of the individual biometric channels alone. In fact, encouraging results of systems using multiple modalities for speaker verification have been reported in the literature [1-4].

However, existing works on multimodal/multi-channel biometrics usually assume and often optimize for fixed environmental conditions. It is well-known that variations in the operating environments pose serious challenges to biometric systems. In fact, improving person recognition performance in adverse environment for a single modality has been extensively studied in the respective fields: for example, in the field of audio-based speaker identification, much effort has been make to compensate the acoustic noise; in the field of face recognition, research seeks to compensation for the variation in the lighting conditions. In other words, existing works on environmental robustness mostly concern a single biometric channel.

In this work, we aim to develop a general framework that addresses the robustness issues in biometrics through the multi-channel approach. Inherently, the following two problems are implied: 1) how to carryout fusion of multiple biometric channels; and 2) how to incorporate environmental factors in the fusion architecture. The proposed framework considers multi-channel fusion and environmental adaptation in a unified fashion. Based on the framework, we developed two environment-adaptive multi-channel biometric systems. The first use a confidence measure derived from real-time environmental measurements and an empirical function to dynamically weight the contribution of a given biometric channel. The second system considers the multiple channels and their relevant environmental factors jointly and is able to optimally adjust the fusion parameters to track the environmental changes.

In the next section we will describe the problem formulation and introduce the multi-channel environmental adaptation framework. In Section 3, we describe the proposed approaches towards the biometric systems. We will discuss the fusion/adaptation experiments and show the results in Section 4, and conclude the paper in Section 5.

2. PROBLEM FORMULATION

Among the potential biometric modalities, some are closely coupled, e.g. audio and visual speech; while others are loosely coupled or uncorrelated, e.g. one's fingerprints and face. In general, the fusion of the multiple information channels can take place at either lower level in the feature domain or at higher level in the decision domain. The former usually assumes the existence of a close coupling between the modalities. In order to develop a multi-modal/multi-channel biometrics system that is robust and adaptive to environ-

mental variations, we need to design fusion architectures that can handle an arbitrary set of information sources; hence we shall not make rigid assumptions on the type of coupling among the modalities.

We adopt a decision fusion approach to combine the multiple biometric channels as shown in Figure 1. In the

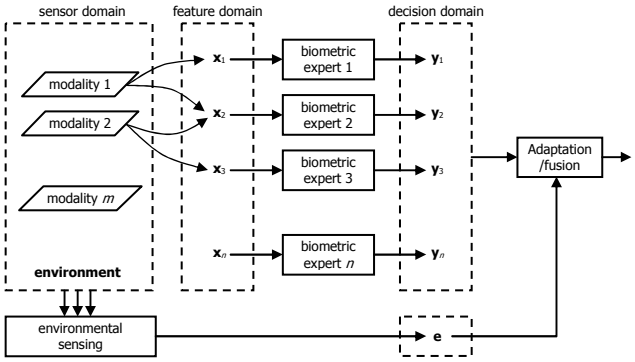

Figure 1. Framework for environment-adaptive multi-channel biometrics

decision fusion setting, the information carried in a single channel i is processed by a dedicated biometric expert, which gives its opinion $\mathbf{y}_i$, which is usually in the form of the likelihood of the observed data $\mathbf{x}_i$, in the given channel. The goal of the decision fusion is to make the recognition decision based on the opinions of the multiple experts.

Here, we make a generalization on the concept of multimodal decision fusion by relaxing the one-to-one mapping between the modalities and the learning channels. In particular, we shall allow each channel to consider more than one modality, to permit fusion for intimately coupled modalities at multiple levels. It is also viable to have multiple channels consider a single modality, each using a different learning algorithm and possibly different feature set. Hence, each learning channel can be viewed as an *expert* specialized in one or more modalities.

It is worthwhile to point out that the focus of our approach is to improve overall recognition performance through multi-channel fusion, rather than to perfect each of the biometric channels. Therefore, we shall treat the biometric experts as black boxes, and concentrate on the environmental conditions they are operating in and the partial decisions they make.

Implied by the formulation is that the environment relevant to the biometric system is quantifiable through the measurement of a set of *environmental factors* **e** (Figure 1), e.g. the acoustic noise level and the illumination intensity. The environmental factors effectively add a set of new dimensions in the decision domain. Under this formulation, the recognition decision is then made jointly on the expert opinions and the measurements of the environmental conditions. In practice, the environmental factors are constantly evolving over time. Therefore a successful algorithm should also be able to track the environmental changes and quickly adapt to them.

3. ENVIRONMENT-ADAPTIVE FUSION

3.1. Naïve Bayes Fusion Approach

For the baseline fusion system, we adopt the naïve Bayes approach. Specifically, the information carried in each biometric channel is processed by a dedicated learner/classifier, which gives a quantity f_i that approximates the likelihood of the observed data in the given channel:

$$y_i = f_i(\mathbf{x}_i \mid C_j, \theta_i) \propto p(\mathbf{x}_i \mid C_j) \tag{1}$$

where $\mathbf{x}_i$ is the observation for channel i, C_j is the class label, and θ_i denotes the particular parameterization scheme used to model the target distribution in this channel. In fusion, it is assumed that the channels are conditionally independent given the class label, thus the joint likelihood of the observations form all channels can be factorized as

$$p(\mathbf{x}_1, \mathbf{x}_1, \ldots \mathbf{x}_m \mid C_j) = \prod_{i=1}^{m} p(\mathbf{x}_i \mid C_j) \propto \prod_{i=1}^{m} f_i(\mathbf{x}_i \mid C_j, \theta_i) \tag{2}$$

Classification is then carried out using the maximum likelihood decision rule.

3.2. Environmental Confidence Weighting Approach

The baseline fusion scheme does not take the environmental factors into account. The channels are essentially weighted equally under all conditions. Empirically, this orthodox Bayesian approach usually yields sub-optimal classification results. One plausible approach to improve over the baseline fusion is to weight the contribution of an individual channel to the overall decision by a confidence measure of the channel given the environmental measurements. We seek a mapping $h_i(\cdot)$ that projects the environmental factors relevant to the channel i at time t, $\mathbf{e}_i(t)$, to a scalar $w_i(t)$ which quantify the confidence for the given channel at time t. We shall refer $w_i(t)$ as the *environmental confidence*.

$$w_i(t) = h_i\left(\mathbf{e}_i(t)\right) \tag{3}$$

This confidence measure is then used to weight the biometric channels dynamically to track environmental changes.

$$g_i = \prod_{i=1}^{n} [f_i(\mathbf{x}_i \mid C_j, \theta_i)]^{w_i(t)} = \prod_{i=1}^{n} y_i^{\,w_i(t)} \tag{4}$$

The possible formulations of $h(\cdot)$ are myriad. One reliable and straightforward choice of such a mapping is the recognition rate versus environmental factors curve for the particular biometric channel, which can be obtained experimentally.

3.3. Optimal Channel Weighting Approach

The environmental confidence weighting approach essentially biases the channels based on their local confidence, thus does not promise global optimality. Particularly, the multi-channel recognition rate is not guaranteed to be higher than that of the individual channels.

For a given combination of environmental factors, the channel weights that give the global optimal recognition performance on a dataset can be found through empirical search.

In the two-channel case, if we introduce the following constraint,

$$\sum_{i=1}^{2} w_i^{optimal}(t) \equiv 1,\ w_i^{optimal} \geq 0 \text{ for all } i. \quad (5)$$

then the search is reduced to only one dimensional and confined in the real interval [0, 1].

In practice, the environmental condition the system will encounter could be any point in the span of the environmental factors. It is therefore infeasible to find optimal weights for all possible points by directly applying the above method. However, if we view the optimal channel weight for a given channel as a function of the *complete* set of environmental factors,

$$w_i^{optimal}(t) = H_i(\mathbf{e}(t)) \quad (6)$$

then it is possible to get an empirical approximation of H by sampling in the environmental factors domain. Furthermore, experiments indicate that H is usually slow varying and monotonic, thus a rather sparse sampling grid can be used in constructing the function. This observation gives rise to a fast and effective method to perform multi-channel fusion and environmental adaptation. Specifically, through sampling and interpolation, H can be specified *a priori* using a set of training data. During recognition, given a set of measured environmental factors, the optimal channel weights can be quickly looked up to carryout fusion. Note that the method is computationally very efficient as it does not introduce any additional calculations in the recognition stage besides the weight lookup.

4. EXPERIMENTAL EVALUATIONS

4.1. Experimental Setup

Two biometric channels are considered in the experiments: the acoustic speaker recognition expert and the face recognition expert. The XM2VTSDB database [5] is used to evaluate the experimental systems on a closed-set person recognition task. The database provides four sets of data for each subject; each set consists of two audio-visual sequences. Among the four sets, we use the first two as the training set for the individual experts, the third as a held-out set for the fusion and adaptation training, and the fourth set to perform evaluations.

Because the emphases of the work are not in the individual biometric channels, we adopt existing algorithms in their implementations. The speaker recognition expert is based on utterance-independent Gaussian mixture models (GMM) trained using the universal background model plus MAP adaptation paradigm [6]. For the face recognition expert, a system using the embedded hidden Markov models (EHMM) similar to [7] is developed.

4.2. Environmental Characterization

To facilitate the experiments on environmental adaptation, environmental conditions are simulated so that a meaningful range of variations can be covered. For the acoustic environment, we simulated noisy ambient conditions by adding background babble noise at varying levels to achieve a signal-to-noise (SNR) range between 5dB and 30dB. The baseline performance of the speaker recognition expert is shown in Figure 2.

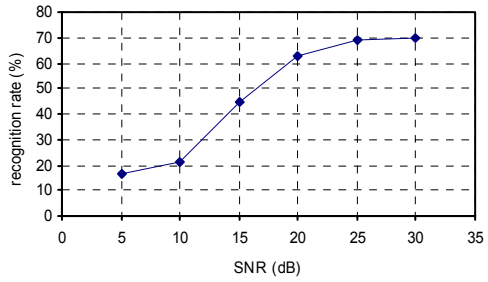

Figure 2. Performance versus acoustic noise curve for the speaker recognition channel.

The most common environmental factor that is relevant to the face recognition expert is the lighting condition. However, illumination intensity changes can be easily compensated by normalization procedures built in the face recognition system; whereas the effect of lighting direction change on face images is difficult to simulate. In the following experiments, we consider the visual sensor noise as the environmental factor for the face recognition channel. In reality, the sensor noise arises naturally in video and still-image sensors and becomes more prominent as the ambient light decreases. We model the sensor noise as white Gaussian and synthesize the noisy visual data. The recognition performance of the face recognition expert under the varying environmental conditions is shown in Figure 3.

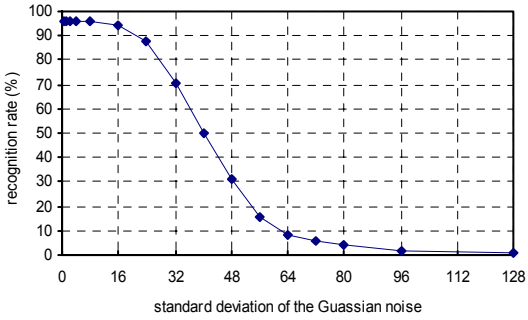

Figure 3 Performance versus visual noise curve for the face recognition channel.

4.3. Environmental Adaptation Experiments

Audio-visual biometric systems are implemented following the three approaches described in Section 3. For the environmental confidence weighting system, the recognition-rate-versus-environmental-factor curves are used as the mapping functions. For the optimal channel weighing system, 64 points (8 levels in each dimension) in the combined audio-visual environmental factors' domain are sampled to construct the empirical optimal weight function. Figure 4 shows the function obtained for the speaker identification channel.

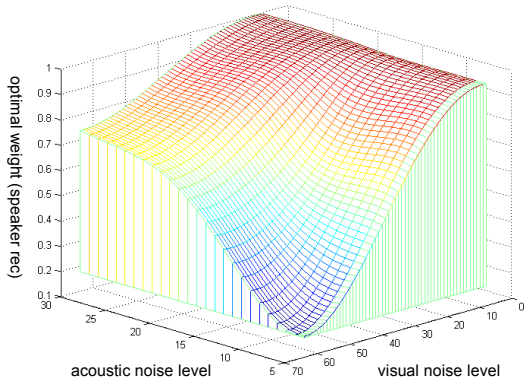

Figure 4. Optimal channel weights for the speaker recognition channel as function of the environmental factors.

The experimental results are summarized in Table 1 and compared in Figure 5. Shown are the results for the visual noise level $\sigma = 32$ and acoustic SNR varying from 5dB to 30dB. Comparing with the two single channel systems, the baseline multi-channel system with naïve Bayes fusion gives higher performance at the higher SNR levels. For instance, at 30dB, a 17.19% improvement in recognition rate is achieved by the multi-channel system comparing with the best performance achievable by an individual biometric expert. However, in low SNR conditions, the naïve Bayes fusion yields results that are in-between the recognition rates logged by the two single-channel recognizers. At 10dB, the baseline fusion system is 27.02% behind the face recognition expert in term of recognition rate. The environmental confidence weighting method shows clear improvements over the baseline fusion, for example, the performance gap is narrowed to 13.69% in the 10dB case. Notice that the method does not guarantee to perform no worse than the individual experts at all environmental conditions. Indeed, the results indicate that the performance gain obtained by the locally derived confidence weights is not optimal. This is especially prominent in conditions when the reliabilities of the two biometric channels differ greatly, as suggested in the figure 5.

Finally, the results confirm the superiority of the optimal channel weighting method, of which the recognition rate is consistently at the top among the five test systems at all the given ambient conditions. At 30dB SNR, the optimal channel weighting method attains an additional 2.11% gain in recognition rate comparing with the environmental confidence weighting method, which translates to an 18.8% reduction in recognition error rate. At 10dB, when the speaker

Table 1. Summary of person recognition results of the experimental biometric systems.

SNR	speaker rec.	face rec.	naïve Bayes	env. conf. weighting	opt. ch. weighting
5	0.1719	0.7123	0.4772	0.6632	0.7123
10	0.2351	0.7123	0.4421	0.5754	0.7123
15	0.4632	0.7123	0.6772	0.6947	0.7263
20	0.6351	0.7123	0.8211	0.8316	0.8632
25	0.7018	0.7123	0.8421	0.8421	0.8982
30	0.7158	0.7123	0.8877	0.8877	0.9088

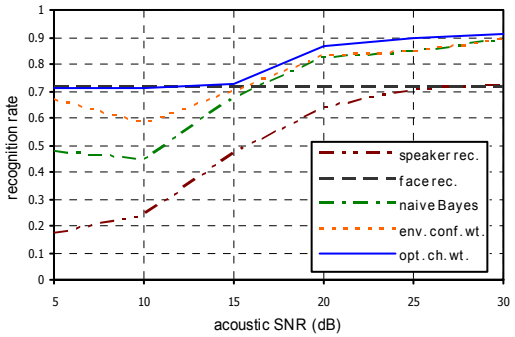

Figure 5. Comparing person recognition results of the experimental biometric systems.

recognition channel is unreliable, the optimal channel weighting system essentially achieves the same performance as the face recognition expert.

5. CONCLUSIONS

In this work we considered multi-channel fusion in the context of environmental adaptation for biometric systems. We introduce a general formulation to the problem and propose several effective approaches to carryout environment-adaptive decision fusion. Experiments show environmental confidence weighting and optimal channel weighting both achieve higher recognition rate comparing with the straightforward naïve Bayes fusion. Moreover, the optimal channel weighing method consistently shows improved recognition accuracy over other systems, while guarantees to perform no worse than any of the individual biometric channels under all environmental conditions. The work also validates that environment-adaptive multi-channel biometric systems can offer significant gains in robustness and reliability, and it is a promising vector for further research.

Acknowledgements: The authors thank Rainer Lienhart and Xiaobo Pi for their insights and valuable discussions.

REFERENCES

[1] R. Brunelli and D. Falavigna, "Person identification using multiple cues," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, pp. 955-966, Oct. 1995.

[2] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, pp. 226-239, March 1998.

[3] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, "Fusion of face and speech data for person identity verification," *IEEE Trans. Neural Networks*, vol. 10, pp. 1065-1074, Sept. 1999.

[4] C. Sanderson and K. K. Paliwal, "Noise compensation in a multi-modal verification system," in *Proceedings of ICASSP 2001*, vol. 1, pp. 157-160, 2001.

[5] The extended m2vts database. http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/

[6] J. P. Campbell, "Speaker recognition: A tutorial," *Proceedings of the IEEE*, vol. 85, pp. 1437-1462, Sept. 1997.

[7] A. V. Nefian, M. H. Hayes, "An Embedded HMM-Based Approach for Face Detection and Recognition" in *Proceedings of ICASSP 1999*, vol. 6, pp. 3553-3556, 1999.