Multimodal Information Fusion

Norman Poh and Josef Kittler

CVSSP, University of Surrey, Guildford, GU2 7XH, Surrey, UK
normanpoh@ieee.org, j.kittler@surrey.ac.uk

1 Introduction

Humans interact with each other using different modalities of communication. These include speech, gestures, documents, etc. It is only natural that human computer interaction (HCI) should facilitate the same multimodal form of communication. In order to capture this information, one uses different types of sensors, i.e., microphones to capture the audio signal, cameras to capture life video images, 3D sensors to directly capture the surface information, in real time. In each of these cases, Commercial Off-the-shelf (COST) devices are already available, and can be readily deployed for HCI applications. Examples of HCI applications include audio-visual speech recognition, gesture recognition, emotional recognition, and person recognition using biometrics.

The provision of multiple modalities is not only motivated by human factors (i.e., usability, and seamless transition from human-human interaction to HCI) but is also warranted from the engineering point of view. Some of the reasons for not relying on a single mode of communication are given below:

1. Noise in sensed data: One can distinguish three kinds of noise in the sensed data: sensor, channel and modality-specific noise. Sensor noise is the noise that is introduced by the sensor itself. For instance, each pixel in a camera sensor contains one or more light sensitive photo-diodes that convert the incoming light (photons) into an electrical signal. The signal is encoded as the colour value of the pixel of the final image. Even though the same pixel would be exposed several times by the same amount of light, the resulting colour value would not be identical, but have small variation called "noise". Channel noise, on the other hand, is the result of degradation introduced by the transmission channel or medium. For example, under slightly changed lighting conditions, the same HCI modality may change. The best well known example perhaps is in person recognition, where the same face under changing lighting condition appear *more differently* then two different faces captured under the same lighting conditions. Finally, modality-specific is the noise that is due to the difference between the captured data and the *canonical* representation of the modality. For instance, it is common to use frontal (mugshot) face to represent the face of a person. As a result, any pose variation in a captured face image can become a potential source of noise.
2. Non-universality: A HCI system may not be able to acquire meaningful data from a subset of individuals. For example, a speech recognition system is totally useless to recognise sentences conveyed by a dumb person using a sign language; lip reading would have some successes. Similarly, in a biometric application, an iris recognition system may be unable to obtain the iris information of a subject with long

eyelashes, dropping eyelids or certain pathological conditions of the eye. However, a face recognition system would still be able to provide a useful biometric modality. Clearly, while no one modality is perfect, a combination of them should ensure wider user coverage, hence improving accessibility, especially to the disabled person.

3. Upper bound on system performance: The matching performance of a speech-only based recognition system cannot be continuously improved by tuning the feature extraction and classifier. There is an implicit upper bound on the number of distinguishable patterns (i.e., the intrinsic upper and lower limits of speech frequency) that can be represented using features derived from spectral envelope. Automatic lip reading, in this case, may improve the performance further when used in combination with the speech recognition system. In person recognition using fingerprint, for instance, this upper bound is dependent on the biometric feature set, whose capacity is constrained by the variations observed in the feature set of each subject (i.e., *intra*-class variations) and the variations between feature sets of different subjects (i.e., *inter*-class variations). Using multiple biometrics for person recognition has been shown to attain performance beyond that achievable by any single biometric system alone [2] (and references herein).

From the above arguments, it is obvious that a successful HCI system should rely on multiple modalities. Indeed, apart from audio and visual information as mentioned here, human also relies on the haptic modality, smell and taste. From these basic sensory information, higher cues such as 3D and temporal information, as well as emotional (e.g., stress, frustration) and psychological state (e.g., interest), can also be derived. While it may take some time before one can design a general purpose HCI system capable of emulating a human (e.g., a humanoid robot), it is already possible to design a specialised system with some of the functionality needed for effective HCI.

Combining the HCI information which is inherently multimodal introduces new challenges to machine learning. For example, face and speech modalities are often sampled at different rates (30 frames per second versus 44K samples per second), each often have different representations (2D versus 1D in this case), and as a result, often are processed with different machine-learning algorithms (Ref. Chapter 6). This chapter focuses on issues raised by combining multiple modalities in HCI systems. Combining several systems has been investigated in pattern recognition [3] in general; in applications related to audio-visual speech processing [4] [5, 6]; in speech recognition – examples of methods are multi-band [7], multi-stream [8, 9], front-end multi-feature [10] approaches and the union model [11]; in the form of ensemble [12]; in audio-visual person authentication [13]; and, in multi-biometrics [14–17, 2] (and references herein), among others. In fact, for audio-visual person authentication, one of the earliest work addressing multimodal biometric fusion was reported in 1978 [18]. Therefore, biometric fusion has a history of 30 years.

This chapter is organised as follows: Section 2 presents different levels of information fusion. Section 3 presents a relatively recent type of fusion utilising modality-dependent signal quality. Section 4 discusses issues related to designing a fusion classifier. Finally, Section 5 concludes the paper.

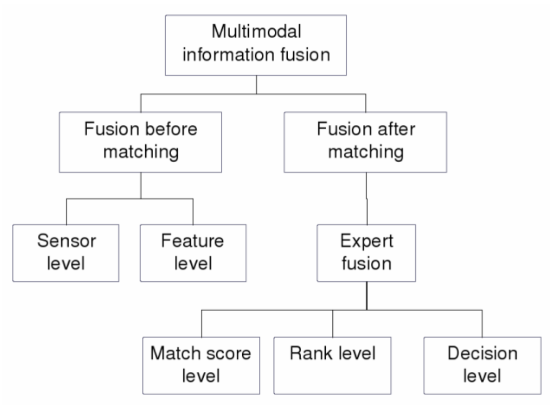


Fig. 1. Fusion can be accomplished at various levels in a biometric system.

2 Levels of fusion

Based on the type of information available in a certain module, different levels of fusion may be defined. Sanderson and Paliwal [19] categorise the various levels of fusion into two broad categories: pre-classification or fusion *before* matching, and post-classification or fusion *after* matching (see Figure 1). The latter has been attracting a lot of attention since the amount of information available for fusion reduces drastically once the matcher has been invoked. Pre-classification fusion schemes typically require the development of new matching techniques (since the matchers/classifiers used by the individual sources may no longer be relevant) thereby introducing additional challenges. Pre-classification schemes include fusion at the sensor (or raw data) and the feature levels while post-classification schemes include fusion at the match score, rank and decision levels (see also [2]).

1. Sensor-level fusion: The raw data (e.g., a face image) acquired from an individual represents the richest source of information although it is expected to be contaminated by noise (e.g., non-uniform illumination, background clutter, etc.). Sensor-level fusion refers to the consolidation of (a) raw data obtained using multiple sensors, or (b) multiple snapshots of a biometric using a single sensor [20, 21].

2. Feature-level fusion: In feature-level fusion, the feature sets originating from multiple feature extraction algorithms are consolidated into a single feature set by the application of appropriate feature normalisation, transformation and reduction schemes. The primary benefit of feature-level fusion is the detection of correlated feature values generated by different feature extraction algorithms and, in the process, identifying a salient set of features that can improve recognition accuracy. Eliciting this feature set typically requires the use of dimensionality reduction/selection methods and, therefore, feature-level fusion assumes the availability of a large number of training data. Also, the feature sets being fused are typically expected to reside in commensurate vector space

in order to permit the application of a suitable matching technique upon consolidating the feature sets [22, 23].

3. Score-level fusion: At this level, the match scores output by multiple experts are combined to generate a new output (a scalar or vector) that can be subsequently used for decision-making. Fusion at this level is the most commonly discussed approach primarily due to the ease of accessing and processing match scores (compared to the raw data or the feature set extracted from the data). Fusion methods at this level can be broadly classified into three categories [24]: density-based schemes [25, 26] (generative approach), classifier-based schemes (discriminative approach) [27] and transformation-based schemes [28] (see also Section 4).

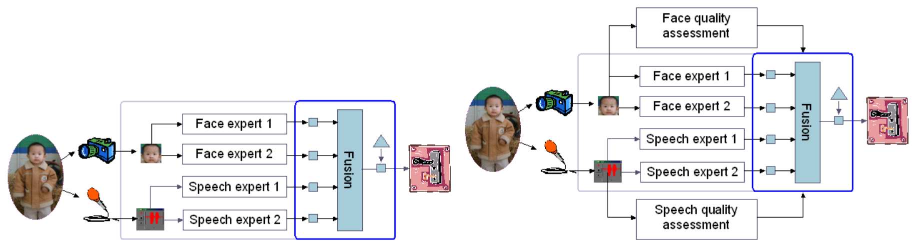
4. Rank-level fusion: In HCI applications, the output of the system can be viewed as a ranking of plausible hypotheses. In other words, the output indicates the set of possible hypotheses sorted in decreasing order of confidence. The goal of rank level fusion schemes is to consolidate the ranks output by the individual expert systems in order to derive a consensus rank for each hypothesis. Ranks provide more insight into the decision-making process of the matching compared to just the best hypothesis, but they reveal less information than match scores. However, unlike match scores, the rankings output by multimodal experts are comparable. As a result, no normalisation is needed and this makes rank level fusion schemes simpler to implement compared to the score level fusion techniques [29].

5. Decision-level fusion: Many COTS HCI systems provide access only to the final recognition decision. When such COTS devices are used to build a multimodal system, only decision level fusion is feasible. Methods proposed in the literature for decision level fusion include “AND” and “OR” rules [30], majority voting [31], weighted majority voting [32], Bayesian decision fusion [33], the Dempster-Shafer theory of evidence [33] and behaviour knowledge space [34].

3 Adaptive versus Non-adaptive Fusion

One can also distinguish fusion classifiers by whether they are adaptive or non-adaptive. Adaptive, or quality-based fusion attempts to change the weight associated with a modality as a function of the signal quality measured on the modality. The idea is to give higher weights to the modality with higher quality. For instance, in biometric person recognition, if the facial image is corrupted by bad illumination, the output of the speech system may be weighed more, and vice-versa when speech system is corrupted by noise.

The quality of an incoming modality is measured by quality measures. *Quality measures* are a set of criteria designed to assess the quality of incoming signal of a modality. Examples of quality measures for face images are face detection reliability, presence of glasses, brightness, contrast, etc [35]. For speech, this would be signal-to-noise ratio and speech-likeness (versus noise). An ideal quality measure should correlate, to some extent, with the performance of the classifier processing the modality [36]. For instance, if a face recognition system can degrade in performance due to change of head pose, than head pose is an ideal quality measure candidate. In practice, more than one quality measures are often needed as each measure can only quantify one particular aspect of the signal quality; and a pool of measurements will, in principal, gauge as many de-



(a) Non-adaptive fusion (b) adaptive fusion with quality

Fig. 2. Adaptive vs. non-adaptive fusion

grading factors as possible. Figure 2 shows the difference between an adaptive and a non-adaptive fusion scheme.

One can show that such a fusion strategy requires a non-linear solution in the underlying system output space. Let $y_{com} \in \mathbb{R}$ be the combined score and $y_i \in \mathbb{R}$ be the output of each modality (after matching; hence producing, for instance, posterior probability of a class label). A non-adaptive linear fusion classifier will take the form of:

$$y_{com} = \sum_{i=1}^{N} w_i y_i + w_0$$

where $w_i \in \mathbb{R}$ is the weight associated to the output y_i and w_0 is a bias term[1]. In contrast, the adaptive fusion classifier would compute

$$y_{com} = \sum_{i=1}^{N} w_i(q) y_i + w_0(q) \tag{1}$$

where $w_i(q)$ changes with the quality signal q. For the sake of simplicity, we shall assume for now that $q = \{q_1, \ldots, q_N\}$, where q_i is the quality measure of the i-th modality. In general $w_i(q)$ could be of any functional form. However, we shall assume that weights vary linearly as a function of quality, i.e.,

$$w_i(q) = \sum_i w_i^{(2)} q_i + w_i^{(1)} \tag{2}$$

and

$$w_0(q) = \sum_j w_i^{(0)} q_i + w_0^{(0)} \tag{3}$$

[1] Essentially, y_{com} is a the output of a *discriminant function*. In order to consider y_{com} as probability, one can use a logistic or sigmoid output instead, i.e., $y_{com} = \frac{1}{1+\exp(-a)}$, where $a = \sum_{i=1}^{N} w_i y_i + w_0$. Generalisation to multiple-class hypotheses is straightforward, as discussed in [37, 38].

Substituting (2) and (3) into (1), and re-arranging, we find:

$$
\begin{aligned}
y_{com} &= \sum_{i=1}^{N} y_i \left(w_i^{(2)} q_i + w_i^{(1)} \right) + \left(\sum_{i=1}^{N} w_i^{(0)} q_i + w_0^{(0)} \right) \\
&= \sum_{i=1}^{N} w_i^{(2)} y_i q_i + \sum_{i=1}^{N} w_i^{(1)} y_i + \sum_{i=1}^{N} w_i^{(0)} q_i + w_0^{(0)} \qquad (4)
\end{aligned}
$$

where we note that $w_i^{(2)}$ is the weight associated to the pairwise element $y_i \cdot q_i$, $w_i^{(1)}$ is the weight associated to y_i, and $w_i^{(0)}$ is the weight associated to q_i.

From (4), it is obvious that one way to realise (1) is to consider a classifier taking $\{y_i q_i, y_i, q_i\}$ as features. The adaptive fusion can be seen as a function $\{y_i, q_i | i = 1, \ldots, N\} \rightarrow y_{com}$ taking the product of y_i and q_i into consideration. This shows that adaptive fusion, in the simplest case, is a linear function of quality, and such a function is non-linear in the space of $\{y_i, q_i | i = 1, \ldots, N\}$.

There are two ways quality measures can be incorporated into a fusion classifier, depending on their role, i.e., either as a control parameter or as evidence. In their primary role, quality measures are used to modify the way a fusion classifier is trained or tested, as suggested in the Bayesian-based classifier called "expert conciliation" [39], reduced polynomial classifier [40], quality-controlled support vector machines [41], and quality-based fixed rule fusion [42]. In their secondary role, quality measures are often concatenated with the expert outputs to be fed to a fusion classifier, as found in logistic regression [43] and the mixture of Gaussians Bayesian classifier [44].

Other notable work includes the use of Bayesian networks to gauge the complex relationship between expert outputs and quality measures, e.g., Maurer and Baker's Bayesian network [45] and Poh *et al.*'s quality state-dependent fusion [35]. The work in [35] takes into account an array of quality measures rather than representing quality as a scalar. By means of grouping the multi faceted quality measures, a fusion strategy can then be devised for each cluster of quality values.

Other suggestions include the use of quality measures to improve HCI device interoperability [46, 47]. Such an approach is commonly used in speaker verification [48] where different strategies are used for different microphone types.

Last but not least, another promising direction in fusion is to consider the reliability estimate of each biometric modality. In [49], the estimated reliability for each biometric modality was used for combining decision-level decisions, whereas in [50–53], score-level fusion was considered. However, in [50–52], the term "failure prediction" was used instead. Such information, derived solely from the expert outputs (instead of quality measures), has been demonstrated to be effective for single modalities [50], fusion across sensors for a single modality [51], and across different machine learning techniques [52]. In [53], the notion of reliability was captured by margin, a concept used in large-margin classifiers [54]. Exactly how the reliability is defined and estimated for each modality, and how it can be effectively used in fusion, are still open research issues.

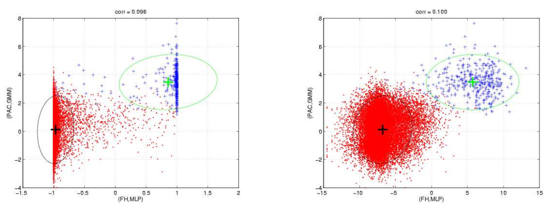
4 Other Design Issues

In the previous section, we highlighted the need of designing an effective fusion mechanism and a set of quality measures *dependent* on the modality and its underlying classifier. Apart from these two, there are also the following issues:

- **Fusion Strategies** An important consideration when adopting a fusion strategy is to consider the statistical dependency among the expert outputs. For instance, in intramodal fusion, several experts may rely on the same biometric sample and so higher dependency is expected among the expert outputs. On the other hand, in a multimodal setting, the pool of experts is likely to be statistically independent (see Figure 4). In [43], three types of frameworks are proposed in order to solve a multimodal fusion problem involving intramodal experts. The first framework simply assumes independence, in which case the fusion classifier reduces to a Naive Bayes one. In this case, the output of each expert can be transformed to the same domain (e.g., probability) and the transformed outputs can be combined using simple rules such as sum or product (see Figure 5(a)). The second framework considers dependency of experts in an intramodal setting (all observing the same information) whereas ignores the dependency at the multimodal setting, hence realising a two-stage fusion process (see Figure 5(c)). Note that the second stage of fusion in this case can be based on the Naive Bayes principal (using simple product rule, but can also be approximated using the sum rule). Finally, the third framework makes no assumption about the expert outputs (see Figure 5(b)). An comparison of these fusion strategies for biometric person authentication can be found in [43].
- **Expert output transformation** Very often, the outputs of different experts are not comparable. Transforming the outputs into a common domain may help improve the ease of design of a fusion classifier. One has the choice of transforming into probability or log-likelihood ratio domain [15]. In the latter case, the output distribution tend to be normal, hence making multivariate analysis such as correlation more reliable. Figure 3 illustrates the transformation of the output of two multimodal experts into log-likelihood ratio space.
- **Choice of architecture** There is a huge space of different fusion architectures that has not been explored. The range of possible configurations encompassing serial, parallel and hybrid structures is immense. While the parallel fusion strategy is most commonly used in multimodal HCI information fusion, there are additional advantages in exploring serial fusion, where the experts are considered one at a time. It offers the possibility of making reliable decisions with only a few experts, leaving only difficult samples to be handled by the remaining experts.

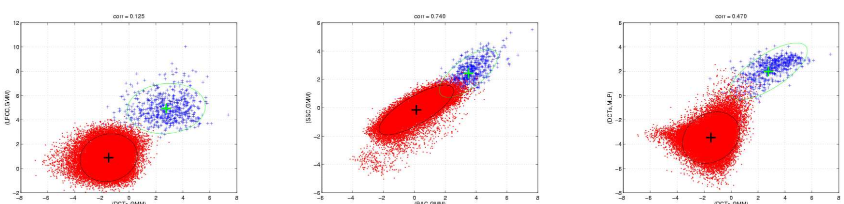
5 Conclusions

This chapter gives an overview of multimodal information fusion from the machine-learning perspective. Although information can be combined at different levels (data, feature, score, rank and decision), score-level fusion offers the best trade-off between information complexity and the flexibility in modelling the dependency among different



(a) Before transformation (b) After transformation

Fig. 3. Transformation of heterogeneous expert output to log-likelihood ratio.

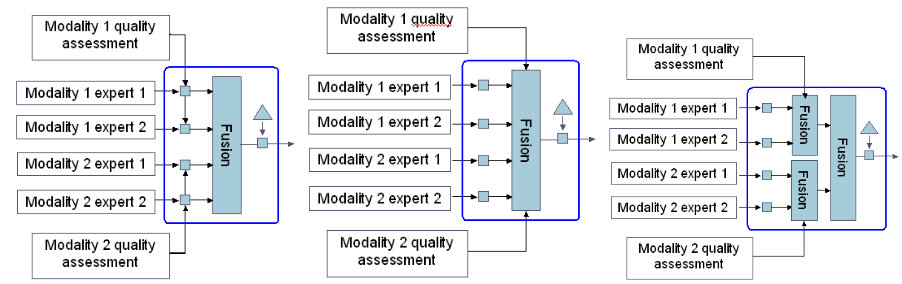


(a) Multimodal fusion (b) Intramodal speech fusion (c) Intramodal face fusion

Fig. 4. Intramodal versus multimodal fusion

information sources. In particular, we presented a relatively recent fusion approach to multimodal fusion, namely, adaptive fusion, where the idea is to measure the signal quality of each modality and then use this information at the fusion level. In this way, the modality expert with better signal quality impacts on the final decision more heavily. Research in adaptive multimodal fusion is gaining momentum. There are still open problems related to its applications. Among others are:

– methodology for designing an effective set of modality-specific and/or classifier-specific quality measures
– methodology for designing an optimal adaptive (quality-based) fusion mechanism
– mechanisms for combining expert opinions regarding the meaning of the communicated information



(a) Transformation-based fusion (b) Classifier-based fusion (c) Hybrid fusion strategy

Fig. 5. Different fusion strategies

6 Acknowledgement

This work was supported partially by the advanced researcher fellowship PA0022_121477 of the Swiss National Science Foundation, and by the EU-funded Mobio project grant IST-214324 (www.mobioproject.org).

References

1. A. K. Jain, P. Flynn, and A. Ross, *Handbook of Biometrics*, Springer Verlag, 2008.
2. A. Ross, K. Nandakumar, and A.K. Jain, *Handbook of Multibiometrics*, Springer Verlag, 2006.
3. S.N. Srihari T.K. Ho, J.J. Hull, "Decision Combination in Multiple Classifier Systems," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 1, pp. 66–75, January 1994.
4. S. Lucey, *Audio Visual Speech Processing*, Ph.D. thesis, Queensland University of Technology, 2002.
5. J. Luettin, *Visual Speech and Speaker Recognition*, Ph.D. thesis, Department of Computer Science, University of Sheffield, 1997.
6. T. Chen and R. Rao, "Audio-Visual Integration in Multimodal Communications," *Proc. IEEE*, vol. 86, no. 5, pp. 837–852, 1998.
7. C. Cerisara, *Contribution de l'Approache Multi-Bande à la Reconnaissance Automatic de la Parole*, Ph.D. thesis, Institute Nationale Polytéchnique de Lorraine, Nancy, France, 1999.
8. S. Dupont, *Étude et Dévelopment de Nouveaux Paradigmes pour la Reconnaissance Robuste de la Parole*, Ph.D. thesis, Laboratoire TCTS, Universite de Mons, Belgium, 2000.
9. Astrid Hagen, *Robust Speech Recognition Based on Multi-Stream Processing*, Ph.D. thesis, Ecole Polytechnique Federale de Lausanne, Switzerland, 2001.
10. M. L. Shire, *Discriminant Training of Front-End and Acoustic Modeling Stages to Heterogeneous Acoustic Environments for Multi-Stream Automatic Speech Recognition*, Ph.D. thesis, University of California, Berkeley, USA, 2001.
11. Ji Ming and F. Jack Smith, "Speech Recognition with Unknown Partial Feature Corruption - a Review of the Union Model," *Computer Speech and Language*, vol. 17, pp. 287–305, 2003.

12. G. Brown, *Diversity in Neural Network Ensembles*, Ph.D. thesis, School of Computer Science, Uni. of Birmingham, 2003.
13. C. Sanderson, *Automatic Person Verification Using Speech and Face Information*, Ph.D. thesis, Griffith University, Queensland, Australia, 2002.
14. K. Nandakumar, "Integration of M]ultiple Cues in Biometric Systems," M.S. thesis, Michigan State University, 2005.
15. N. Poh, *Multi-system Biometric Authentication: Optimal Fusion and User-Specific Information*, Ph.D. thesis, Swiss Federal Institute of Technology in Lausanne (Ecole Polytechnique Fédérale de Lausanne), 2006.
16. K. Kryszczuk, *Classification with Class-independent Quality Information for Biometric Verification*, Ph.D. thesis, Swiss Federal Institute of Technology in Lausanne (Ecole Polytechnique Fédérale de Lausanne), 2007.
17. J. Richiardi, *Probabilistic Models for Multi-Classifier Biometric Authentication Using Quality Measures*, Ph.D. thesis, Swiss Federal Institute of Technology in Lausanne (Ecole Polytechnique Fédérale de Lausanne), 2007.
18. A. Fejfar, "Combining Techniques to Improve Security in Automated Entry Control," in *Carnahan Conf. On Crime Countermeasures*, 1978, Mitre Corp. MTP-191.
19. C. Sanderson and K. K. Paliwal, "Information Fusion and Person Verification Using Speech and Face Information," Research Paper IDIAP-RR 02-33, IDIAP, September 2002.
20. R. Singh, M. Vatsa, A. Ross, and A. Noore, "Performance Enhancement of 2D Face Recognition via Mosaicing," in *Proceedings of the 4th IEEE Workshop on Automatic Identification Advanced Technologies (AuotID)*, Buffalo, USA, October 2005, pp. 63–68.
21. A. Ross, S. Shah, and J. Shah, "Image Versus Feature Mosaicing: A Case Study in Fingerprints," in *Proceedings of SPIE Conference on Biometric Technology for Human Identification III*, Orlando, USA, April 2006, pp. 620208–1 – 620208–12.
22. A. Ross and R. Govindarajan, "Feature Level Fusion Using Hand and Face Biometrics," in *Proceedings of SPIE Conference on Biometric Technology for Human Identification II*, Orlando, USA, March 2005, vol. 5779, pp. 196–204.
23. B. Son and Y. Lee, "Biometric Authentication System Using Reduced Joint Feature Vector of Iris and Face," in *Proceedings of Fifth International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, Rye Brook, USA, July 2005, pp. 513–522.
24. A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics*, Springer, New York, USA, 1st edition, 2006.
25. S. C. Dass, K. Nandakumar, and A. K. Jain, "A Principled Approach to Score Level Fusion in Multimodal Biometric Systems," in *Proceedings of Fifth International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA)*, Rye Brook, USA, July 2005, pp. 1049–1058.
26. B. Ulery, A. Hicklin, C. Watson, W. Fellner, and P. Hallinan, "Studies of Biometric Fusion," Technical Report NISTIR 7346, NIST, September 2006.
27. P. Verlinde and G. Cholet, "Comparing Decision Fusion Paradigms using k-NN based Classifiers, Decision Trees and Logistic Regression in a Multi-modal Identity Verification Application," in *Proceedings of Second International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, Washington D.C., USA, March 1999, pp. 188–193.
28. A. K. Jain, K. Nandakumar, and A. Ross, "Score Normalization in Multimodal Biometric Systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, December 2005.
29. T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision Combination in Multiple Classifier Systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 1, pp. 66–75, January 1994.
30. J. Daugman, "Combining Multiple Biometrics," Available at http://www.cl.cam.ac.uk/users/jgd1000, 2000.

31. L. Lam and C. Y. Suen, "Application of Majority Voting to Pattern Recognition: An Analysis of its Behavior and Performance," *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, vol. 27, no. 5, pp. 553–568, 1997.
32. L. I. Kuncheva, *Combining Pattern Classifiers - Methods and Algorithms*, Wiley, 2004.
33. L. Xu, A. Krzyzak, and C. Y. Suen, "Methods for Combining Multiple Classifiers and their Applications to Handwriting Recognition," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 3, pp. 418–435, 1992.
34. Y. S. Huang and C. Y. Suen, "Method of Combining Multiple Experts for the Recognition of Unconstrained Handwritten Numerals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 1, pp. 90–94, January 1995.
35. N. Poh, G. Heusch, and J. Kittler, "On Combination of Face Authentication Experts by a Mixture of Quality Dependent Fusion Classifiers," in *LNCS 4472, Multiple Classifiers System (MCS)*, Prague, 2007, pp. 344–356.
36. K. Kryszczuk and A. Drygajlo, "Improving classification with class-independent quality measures: Q-stack in face verification," in *LNCS 4472, Multiple Classifiers System (MCS)*, 2007, pp. 1124–1133.
37. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification and Scene Analysis*, John Wiley and Sons, New York, 2001.
38. C. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1999.
39. J. Bigun, J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, "Multimodal Biometric Authentication using Quality Signals in Mobile Communnications," in *12th Int'l Conf. on Image Analysis and Processing*, Mantova, 2003, pp. 2–13.
40. K-A. Toh, W-Y. Yau, E. Lim, L. Chen, and C-H. Ng., "Fusion of Auxiliary Information for Multimodal Biometric Authentication," in *LNCS 3072, Int'l Conf. on Biometric Authentication (ICBA)*, Hong Kong, 2004, pp. 678–685.
41. J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun, "Kernel-Based Multimodal Biometric Verification Using Quality Signals," in *Defense and Security Symposium, Workshop on Biometric Technology for Human Identification, Proc. of SPIE*, 2004, vol. 5404, pp. 544–554.
42. O. Fatukasi, J. Kittler, and N. Poh, "Quality Controlled Multimodal Fusion of Biometric Experts," in *12th Iberoamerican Congress on Pattern Recognition CIARP*, Via del Mar-Valparaiso, Chile, 2007, pp. 881–890.
43. J. Kittler, N. Poh, O. Fatukasi, K. Messer, K. Kryszczuk, J. Richiardi, and A. Drygajlo, "Quality Dependent Fusion of Intramodal and Multimodal Biometric Experts," in *Proc. of SPIE Defense and Security Symposium, Workshop on Biometric Technology for Human Identification*, 2007, vol. 6539.
44. K. Nandakumar, Y. Chen, S. C. Dass, and A. K. Jain, "Likelihood ratio based biometric score fusion," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 342–347, 2008.
45. D. E. Maurer and J. P. Baker, "Fusing multimodal biometrics with quality estimates via a bayesian belief network," *Pattern Recognition*, vol. 41, no. 3, pp. 821–832, 2007.
46. F Alonso-Fernandez, J. Fierrez, D. Ramos, and J. Ortega-Garcia, "Dealing with sensor interoperability in multi-biometrics: The upm experience at the biosecure multimodal evaluation 2007," in *Proc. of SPIE Defense and Security Symposium, Workshop on Biometric Technology for Human Identification*, 2008.
47. N. Poh, T. Bourlai, and J. Kittler, "Improving Biometric Device Interoperability by Likelihood Ratio-based Quality Dependent Score Normalization," in *accepted for publication in IEEE Conference on Biometrics: Theory, Applications and Systems*, Washington, D.C., 2007, pp. 1–5.

48. R. Auckenthaler, M. Carey, and H. Lloyd-Thomas, "Score Normalization for Text-Independant Speaker Verification Systems," *Digital Signal Processing (DSP) Journal*, vol. 10, pp. 42–54, 2000.
49. Krzysztof Kryszczuk, Jonas Richiardi, Plamen Prodanov, and Andrzej Drygajlo, "Reliability-based decision fusion in multimodal biometric verification systems," *EURASIP Journal of Advances in Signal Processing*, vol. 2007, 2007.
50. W. Li, X. Gao, and T.E. Boult, "Predicting biometric system failure," *Computational Intelligence for Homeland Security and Personal Safety, 2005. CIHSPS 2005. Proceedings of the 2005 IEEE International Conference on*, pp. 57–64, 31 2005-April 1 2005.
51. B. Xie, T. Boult, V. Ramesh, and Y. Zhu, "Multi-camera face recognition by reliability-based selection," *Computational Intelligence for Homeland Security and Personal Safety, Proceedings of the 2006 IEEE International Conference on*, pp. 18–23, Oct. 2006.
52. T. P. Riopka and T. E. Boult, "Classification enhancement via biometric pattern perturbation," in *AVBPA*, 2005, pp. 850–859.
53. N. Poh and S. Bengio, "Improving Fusion with Margin-Derived Confidence in Biometric Authentication Tasks," in *LNCS 3546, 5th Int'l. Conf. Audio- and Video-Based Biometric Person Authentication (AVBPA 2005)*, New York, 2005, pp. 474–483.
54. A. J. Smola and P. J. Bartlett, Eds., *Advances in Large Margin Classifiesr*, MIT Press, Cambridge, MA, 2000.