Multi-Modal Biometrics: An Overview

Kevin W. Bowyer,[1] K. I. Chang,[1] P. Yan,[1] P. J. Flynn,[1] E. Hansley,[2] S. Sarkar[2]
1. Computer Science and Engineering / University of Notre Dame / Notre Dame, IN 46556 USA
2. Computer Science and Engineering / University of South Florida / Tampa, FL 33620 USA

kwb@cse.nd.edu

Abstract

The topic of multi-modal biometrics has attracted strong interest in recent years. This paper categorizes approaches to multi-modal biometrics based on the biometric source, the type of sensing used, and the depth of collaborative interaction in the processing. This paper also attempts to identify some of the challenges and issues that confront research in multi-modal biometrics.

1. Introduction

Biometrics has become a "hot" area. Governments are funding research programs focused on biometrics. Conferences in the area readily attract attendees from government and industry. Research advances find their way into commercial products quite quickly. The area of multi-modal biometrics seems especially interesting and exciting, because it is in some sense at the frontier of biometrics, and presents complex design problems.

This paper is part overview of multi-modal biometrics, part review of some results from the Notre Dame research group, and part speculation about challenges to progress in multi-modal biometrics. The coverage and ideas are inevitably influenced by our own research experiences.

2. Categories of Biometric Algorithms

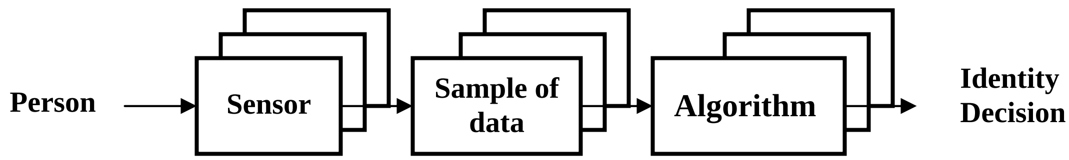
The "simple" approach to biometrics is to sense a single sample (image) of a biometric source (body part) from a person and then process that to obtain a recognition result. The vast majority of face recognition research has implicitly assumed this framework. The same is true of iris recognition. How one understands fingerprint on this issue is perhaps ambiguous, since law enforcement has historically used a "ten print" (five fingers of each hand) acquisition when possible. But commercial application of fingerprint for identity verification typically uses one impression of one finger [15].

The term "multi-modal" is used in the literature with various meanings, as illustrated in Figure 1. Perhaps the least ambiguous example of multi-modal would be two different biometric sources on a person, say face and fingerprint, sensed by different sensors. Two different properties, say infra-red and reflected light, of the same biometric source, say the face, would be another unambiguous example of multi-modal. An ambiguous example would be two different biometric sources, say face and ear, imaged by the same sensor. Another ambiguous example would be two different properties, say 3D shape and reflected light, of the same source, say face, sensed by the same sensor. We will take an expansive view in this paper, and consider all of these variations as "multi-modal," and consider "multi-biometric" as an equivalent term. We are basically in agreement here with the terminology suggested by Wayman [20].

2.1. Multi-algorithm

One step beyond a "simple" biometric is what we might call a multi-algorithm approach. This approach still employs a single sensor, and acquires a single biometric sample. Two or more different algorithms process the single sample, and the individual results are fused to obtain an overall recognition result.

The multi-algorithm approach would seem to be attractive, both from an application point of view and



Is it a "multi-modal" biometric when …

- **Two different types of sensors are used?**
  Such as infra-red and visible face? face and fingerprint?
- **One sensor senses two different properties of one source?**
  Such as 3D and 2D face or ear acquired simultaneously?
- **Two different biometric sources are sensed?**
  Such as 2D images of face and ear, or face and iris?
- **Multiple samples are sensed by one sensor of the same source?**
  Such as multiple 2D face images as in FRGC program?
- **Multiple algorithms are applied to one sample of one source?**
  Such as multi-vendor combination of 2D face as in FRVT 2002?

Figure 1 – What Does the "Mode" In Multi-Modal Refer To?

from a research point of view. From an application point of view, it appears to minimize sensor and sensing cost. There is just one sensor and just one sample sensed in order to obtain a recognition result. However, relatively little work has been done in this area. As one example, the 2002 Face Recognition Vendor Test documented increased performance in 2D face recognition by combining the results of different commercial recognition systems [18]. More recently, Gokberk et al have looked at combining multiple algorithms for 3D face recognition [9]. Xu et al [22] have also combined different algorithmic approaches for 3D face recognition.

A variation of the multi-algorithm approach builds an ensemble of the same basic type of algorithm, with intentional variation between instances. For example, Chawla et al used the random subspaces concept to create an ensemble and obtain improved recognition rates from an eigen-face algorithm [6,7].

2.1. Multi-sample

Another approach might be called "multi-sample" or "multi-instance." Multiple samples of the same biometric are sensed, the same algorithm processes each of the samples, and the individual results are fused to obtain an overall recognition result. For example, Chang et al [3] used a multi-sample approach with 2D face images as a baseline against which to compare the performance of multi-sample 2D + 3D face. Also, a multi-sample style of experiment was part of the Face Recognition Grand Challenge [16,17].

An overview of the results in Chang et al [3] is depicted in Figure 2. The multi-modal 2D + 3D performance is greater than that of either 3D or 2D alone. However, the multi-sample performance with two 2D images also performs better than simple 2D or 3D, though not as good as multi-modal 2D + 3D. However, multi-sample with four 2D images outperforms multi-modal 2D+3D. This suggests that with "enough" samples multi-sample has the potential to outperform multi-modal. This result is potentially important for applications because plain 2D sensors are generally cheaper than multi-modal sensors. We first noticed this "multi-sample compared to multi-modal" effect in the context of eigen-face / eigen-ear biometrics, depicted in Figure 3.

In comparison to the multi-algorithm approach, multi-sample has advantages and disadvantages. One advantage is that using multiple samples may overcome poor performance due to one sample that has unfortunate properties. For example, a person might be blinking in one face image, and this might present problems for the recognition algorithm, but if multiple samples in time are used, it is unlikely that the person is blinking in all of them. One disadvantage is that acquiring multiple samples requires either multiple copies of the sensor, or that the user be available for sensing over a longer period of time. Thus, comparing to multi-algorithm, multi-sample seems to require

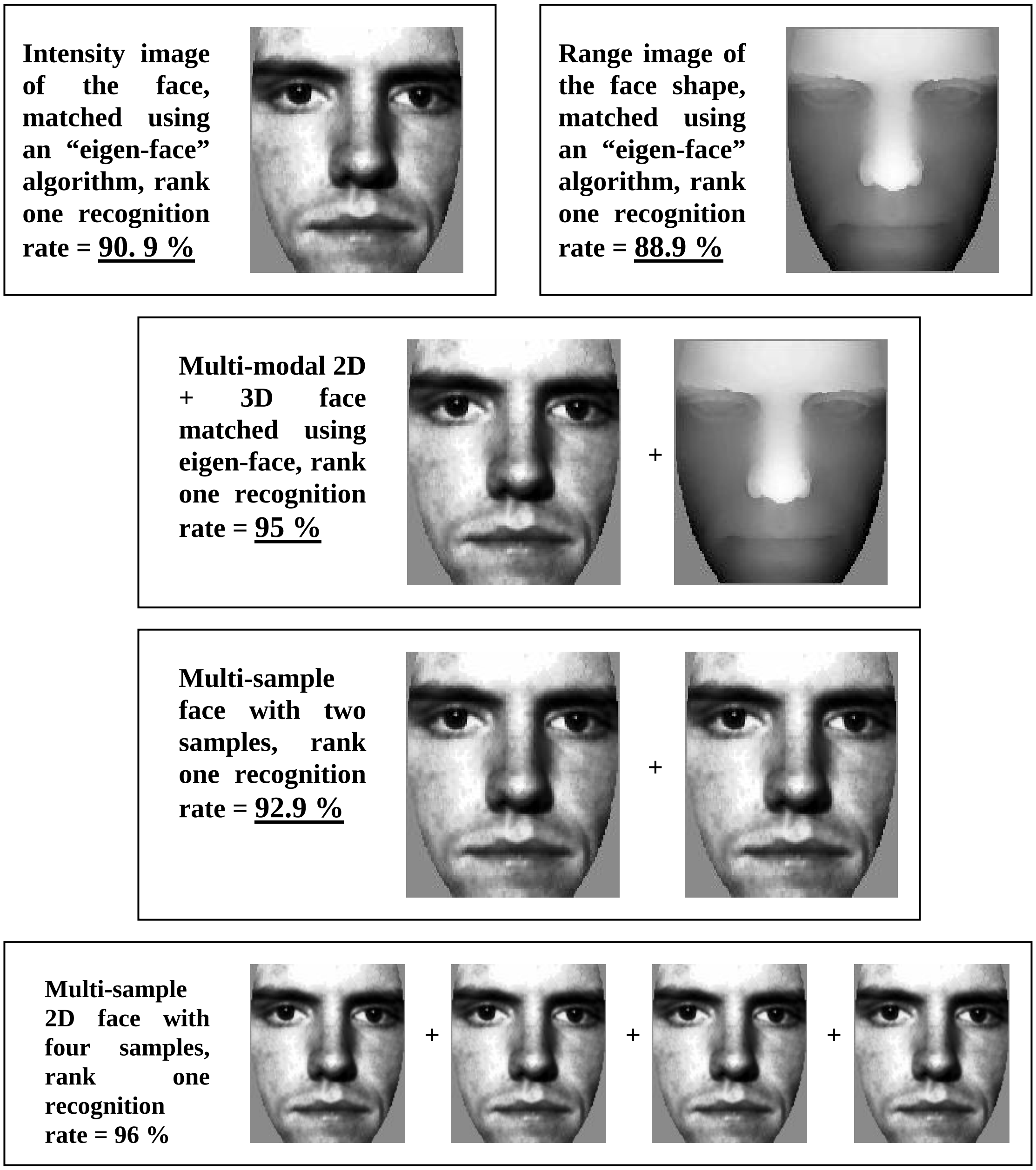


Figure 2 – Comparing Multi-Modal Against a Default Multi-Sample.

either greater expense for sensors, greater cooperation from the user, or a combination of both.

2.3. Multi-modal: "orthogonal"

One common category of multi-modal biometrics might be called "orthogonal." By "orthogonal" we mean to indicate that the biometric sources are different; that is, different parts of the body are involved. An example would be face recognition and fingerprint used together. In fact, the most publicly visible use of multi-modal biometrics is likely the use of face and fingerprint planned in the "US VISIT" program. In a speech about this program in 2003, an

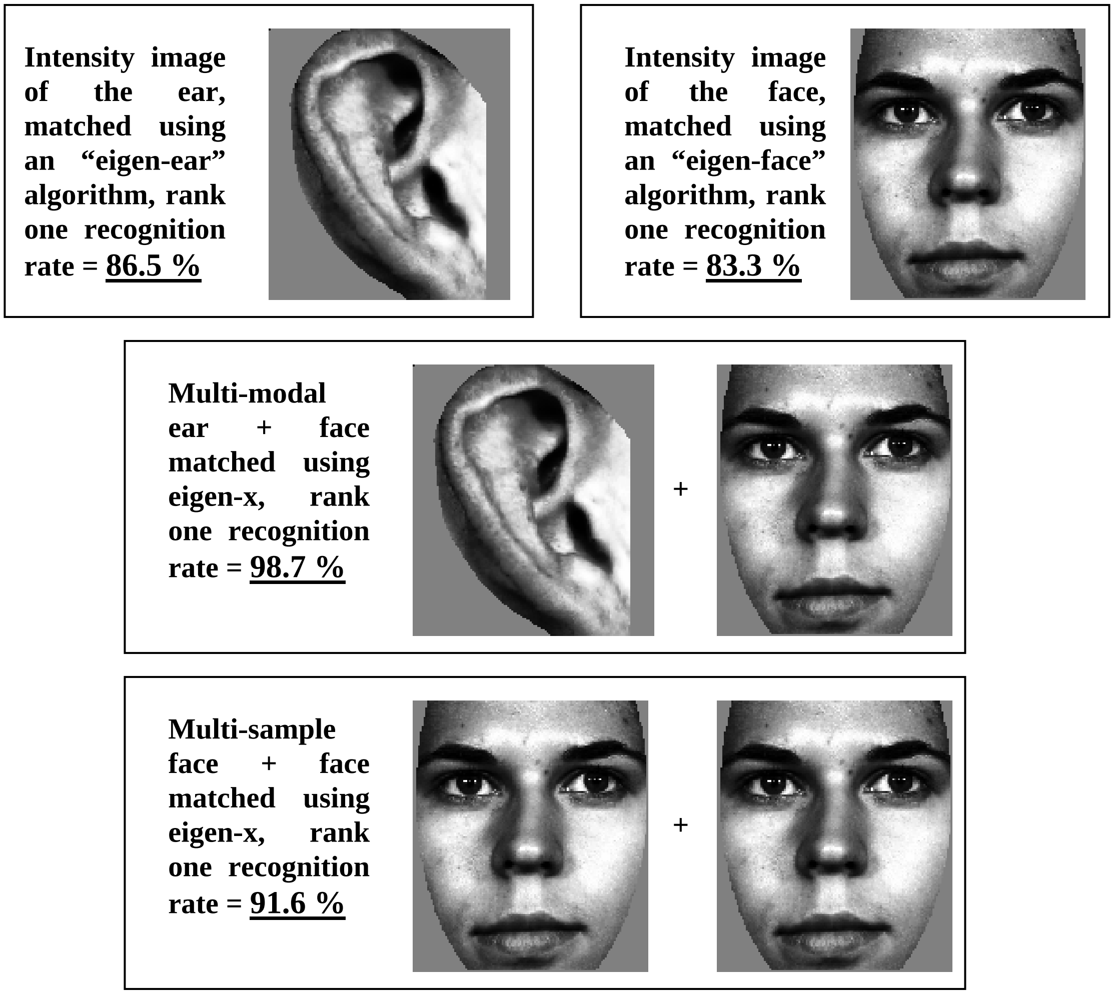


Figure 3 – Comparing Multi-Modal Against a Default Multi-Sample.

official actually mentioned face, fingerprint and iris – "We'll do so through a minimum of two biometric identifiers - initially, fingerprints and photographs; later, as the technology is perfected, additional forms such as facial recognition or iris scans may be used as well" [11].

In this category, there appears to be little or no opportunity for interaction between the individual biometrics. For instance, it is difficult to see how the processing of either face or fingerprint could be used to help the other. However, other multi-modal biometrics may present opportunities for collaboration.

2.4. Multi-modal: "independent"

Another category of approach to multi-modal biometrics might be called "independent." By "independent" we mean to indicate that the individual biometrics are processed independently of each other. It would seem that orthogonal biometrics are processed independently by necessity. But when the biometric source is the same and different properties are sensed, then the processing may be independent, but there is at least the potential for gains in performance through collaborative processing.

Numerous examples of independent processing of different properties of the same biometric source can be found in face recognition. Much of what might be called the "first generation" of multi-modal 2D+3D face recognition took this approach. For example, Chang et al [3,5] applied an eigen-face approach independently to the 2D intensity image of the face and the 3D shape of the face rendered as a range image. Chen et al [8] applied a similar approach to infra-red images and 2D intensity images of the face. Also, Chang et al compared the combination of ear + face to the use of eigen-face or eigen-ear alone [4].

We should acknowledge that some of the initial enthusiasm for some multi-modal biometrics has dampened. For example, initial enthusiasm for infra-red imaging for face recognition was based in part on the infra-red image being independent of the illumination in the scene. Experimental results showed that this was true. But results also showed that face appearance in infra-red varies based on physiologic and emotional factors that have less effect on typical 2D images [8]. Similarly, part of the enthusiasm for 3D shape as a biometric is the idea that shape is independent of illumination. It is true that shape is defined independent of illumination, but it is not necessarily sensed independent of illumination [1].

2.5. Multi-modal: "collaborative"

A less common approach to multi-modal biometrics might be called "collaborative." By "collaboration" we mean the degree to which the processing of one biometric is influenced by the results of processing another biometric.

Considering multi-modal face recognition again, there are some approaches which do not fit neatly into this independent / collaborative categorization. For instance, Papatheodorou and Reuckert [14] approached multi-modal 2D + 3D face recognition by treating the data as points in a 4D space of (x, y, z, intensity). They were then able to use a 4D iterative closest point (ICP) algorithm for the matching stage. The two properties of the face are treated in an integrated manner in the matching, so that it is not quite independent, but also certainly not collaborative in the sense that we want to suggest here.

There are now a number of examples in the literature of what might be called weakly collaborative approaches to multi-modal biometrics. Husken et al [10] describe a Viisage approach to multi-modal 2D+3D face recognition that locates the feature points (e.g., eyes) on the face in the 2D image, and then transfers these locations over to the (registered) 3D data to process the features there. Socolinsky et al [19] follow a similar approach in their multi-modal infra-red + visible-light face recognition. Their sensor is able to obtain registered images from the two modes, and they find the eye location in the visible-light image and transfer the locations over too the infra-red image. These approaches are collaborative in the sense that intermediate results of processing in one modality are used to help the processing in the other modality. But the degree of collaboration is not extensive, and is only in the direction of 2D to the other modality.

We can imagine that much more extensive collaborative processing might be possible. Again, take the example of 2D+3D face recognition. There are artifacts in both types of images, and it may be possible to exploit the ease of finding a certain type of artifact in one mode to improve the reliability of processing the other mode. For example, if specular highlights are found in the 2D face image, this might inform the processing of the 3D shape of the face, since specular highlights in 2D often result in artifacts in the 3D sensing. Also, once something of the general shape of the face is known, it may be possible to use this to consistently interpret regions of the 2D image as affected by shadows. In this way, the intermediate results of processing each modality might be used to improve the reliability and accuracy in processing the other. It seems that the area of "collaborative" processing among multi-modal biometrics, although relatively less explored currently, should hold great potential for important gains.

3. Challenges to Multi-Modal Biometrics

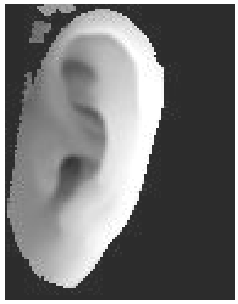
A number of challenges and issues confront multi-modal biometrics research. The items mentioned here are not presented as an exhaustive list, but as suggestions for discussion. The items listed range from technical to social, and are meant to suggest discussion topics rather than present conclusions.

3.1. Are appropriate datasets available?

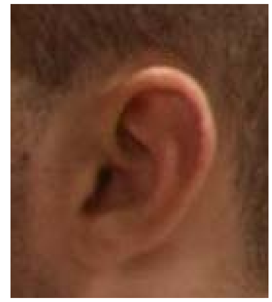
One mundane but important practical issue is whether or not appropriate datasets exist to support research in multi-modal biometrics. A dataset is not a research result in and of itself, but a well-designed dataset can greatly facilitate research and a poorly designed dataset can be a hindrance. The importance of good datasets has been recognized by some major journals modifying their editorial policy to consider papers that primarily describe a dataset available to the research community.

Multi-modal 2D+3D face may be an area where the answer to the question about adequate datasets is yes. The Face Recognition Grand Challenge program released a version 1 and a version 2 dataset. Together these datasets contain nearly 5,000 3D face scans from approximately 500 different individuals, with time lapse between images of a year or more in some cases [16,17]. Over 100 groups have obtained the version 1 dataset and over 50 of these have obtained the version

3D ear shape, displayed here as a range image, is matched using an ICP algorithm.



Intensity image of the ear, matched using an “eigen-ear” algorithm.



Number of Gallery, Probe Samples	Modalities / Algorithms Compared			
	2D PCA		3D ICP	
	Min	Sum	Min	Sum
1G,1P	73.4 % (no fusion)		81.7% (no fusion)	
1G, 3P	82.2 %	83.4 %	95.3 %	81.1 %
2G, 2P	84.0 %	87.5 %	97.0 %	81.9 %
3G, 1P	81.7 %	80.5 %	91.1 %	81.7 %

Figure 4 – Multi-Sample Biometric In Which “Min” Fusion Performed Best.

2 dataset [21]. This size of this dataset may seem large, but as individual biometrics become more accurate, the size and complexity of dataset required to demonstrate a significant difference between algorithms increases.

A related issue is the use of “chimera” datasets. Say that a researcher wants to experiment with multi-modal face + fingerprint. It is easy to find a face image dataset and also to find a separate fingerprint image dataset. A common approach is to create a dataset in which the images of person 1 in the face image dataset are associated with the images of person 1 in the fingerprint dataset, and so on. Since it seems plausible to expect that there is no correlation between the appearance of a persons’ face and fingerprint, there is not an obvious problem. But it could be that such chimera datasets hide some important property of real biometric data. Wayman has taken essentially this position in saying, “All multi-biometric measures (multi-modal or not) from a single person are by necessity correlated” [20].

3.2. Is sum really the best fusion method?

Almost every research report in multi-modal biometrics considers several possible ways of fusing the results. However, it seems that no fusion approach has emerged that generally achieves a statistically significant improvement over a simple sum of scores.

Figure 4 depicts one case in which we found that a “min” fusion performed significantly better than “sum” [23,24]. The experiment uses ear biometrics data from 169 persons, four intensity images and four 3D scans for each person. For 2D matched with the eigen-ear algorithm, all multi-sample variations clearly improve over single-sample. For 3D matched with the ICP algorithm, improvement from multi-sample depends on the fusion method. Min performs significantly better than sum. It is possible that the presence of outliers in the 3D data leads to outliers in the 3D shape matching, and that the min does a better job of handling this. Some method of fusion that automatically adapts to exploit the properties of the data might be a useful advance.

3.3. Can multi-sample always out-perform multi-modal?

We mentioned earlier some results that suggest that a multi-sample approach using “enough” samples can out-perform a multi-modal approach. However, there are so far only a few studies that have looked at this, and the generality of this effect is not yet clear. Is it really the case that using additional samples of 2D face images can increase performance of 2D face recognition without limit, or is there some number of images at which performance reaches a plateau? And if performance does eventually reach a plateau with multiple samples of one modality, can increased performance then be obtained by using samples of a different modality?

3.4. Design for multi-sample approach?

The improvements seen by using a multi-sample approach suggest a whole set of important design questions. How do we choose a best set of N samples for a particular biometric? As a specific example, for face recognition, what is the relative importance of sampling variation in pose, illumination and expression? What is the smallest size sample set that captures a given level of variation?

3.5. Is the goal accuracy or coverage?

We typically think of the motivation as being increased recognition accuracy. But could it be more important in some applications to increase the fraction of the population for which at least one biometric can be acquired? As a specific example, is the purpose of acquiring face and fingerprint to achieve very high accuracy, or to be able to acquire some usable biometric even if a person has a bandage on their face or their hand in a cast? To the extent that the motivation is broader coverage, then multi-sample and multi-algorithm approaches become less interesting.

3.6. Does biometrics deliver on promises?

This is perhaps not a question that is specific to multi-modal biometrics, but applicable to biometrics generally. Certainly there were some in our field who, just after the terrorist attacks of 9-11, made statements that seemed to promise more than the field was ready to deliver. For example, an executive of one face recognition company said– "We know that three out of the nineteen were already known on watch lists and knowing how many checkpoints these people had to go through, we had a high probability to alert, intercept, these individuals maybe August 21 or 23 when they crossed the Canadian border and we would have perhaps foiled the whole plot" [2]. The subjective qualifiers in this statement, such as "high probability, and "perhaps," make it hard to disprove. But it certainly seems that the effect of such statements is to mislead the public about the capabilities of biometrics.

It is an element of professionalism and ethics that researchers should make rational claims that are supported by experiments and data. If we look to something like the Software Engineering Code of Ethics [12], we find statements such as element 6.7, "Be accurate in stating the characteristics of software on which they work, avoiding not only false claims, but also claims that might reasonably be supposed to be speculative, vacuous, deceptive, misleading, or doubtful."

3.7. Should we think about privacy issues?

Researchers in this field are, almost by definition, trying to make it possible and practical to do things that were not possible or practical in the past. The intent of the field is to put new technical capabilities in the hands of government and industry. The "big brother" concern is that governments will use the technology to intrude too far into personal privacy. As we strive to make the technology work, it is perhaps hard to find the additional mental energy to be concerned with possible negative uses of the technology. But it could be useful if we spend a little effort thinking in this direction. Consider a related example that is in the news at the time that this is written. It is likely that the people who run Google and the people who use Google never anticipated that the U.S. government would ask for large amounts of search records, but this is now happening. Biometric systems collect personal data, and if the data is archived, then it is likely that in the future some originally unanticipated use of the data will emerge, and a government may ask for archived data. Should concerns about such possible future events affect how we design biometric systems today?

4. Summary and Discussion

Multi-modal biometrics is an exciting and interesting research area that cuts across traditional to bring together researchers from somewhat different areas. Although work in multi-modal biometrics dates to the 1970s [20], it is a "new" enough area that the terminology is perhaps not always consistent. Even what qualifies as "multi-modal" may be the subject of some disagreement.

Wayman argues that true multi-modal biometrics (using different types of sensors) is not currently a practical solution – "At the current state of biometric system development, information content can be most easily increased within the practical constraints by collecting more images from more body parts or by viewing body parts from multiple angles, and existing information can be better exploited by looking at it in more ways" [20]. We are in general agreement, but see this as a short-term practical emphasis versus a long-term research emphasis. It seems reasonable that there are eventual limits to the performance than can be achieved with one sensor type. If this is so, and if still greater performance is desired, then multi-modal biometrics of this stronger type will be necessary. It also seems likely that true multi-modal biometrics can make "spoofing" more difficult.

5. Acknowledgments

Biometrics research at the University of Notre Dame is supported by the National Science Foundation under grant CNS-013089, by the Central Intelligence Agency, and by the US Department of Justice under grants 2005-DD-BX-1224 and 2005-DD-CX-K078.

The 2D + 3D face results shown come from K. Chang's dissertation; the 2D + 3D ear results from P. Yan's dissertation, and the 2D face + ear results from E. Hansley's dissertation work.

10. References

[1] Bowyer, K. W., K. I. Chang, and P. J. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *CVIU*, 101 (1), 1-15, January 2006.

[2] Bowyer, K. W., "Face recognition technology and the security versus privacy tradeoff," *IEEE Technology and Society*, 9-20, Spring 2004.

[3] Chang, K. I., K. W. Bowyer, and P. J. Flynn, "An evaluation of multi-modal 2D+3D face biometrics," *IEEE Trans. on PAMI* 27 (4), 619-624, April 2005.

[4] Chang, K. I., K. W. Bowyer, S. Sarkar, and B. Victor. Comparison and combination of ear and face Images for appearance-based biometrics, *IEEE Trans. on PAMI* 25 (9), 1160-1165, September 2003.

[5] Chang, K.I., K. W. Bowyer, and P. J. Flynn, Face Recognition Using 2D and 3D Facial Data, *Multi-Modal User Authentication Workshop*, December 2003, Santa Barbara, 25-32.

[6] Chawla N. V., and K. W. Bowyer, "Random subspaces and subsampling for 2-D face recognition**,"** *CVPR*, San Diego, June 2005, II: 582-589.

[7] Chawla N. V., and K. W. Bowyer, "Designing multiple classifier systems for face recognition," the *Springer-Verlag LNCS 3541: Multiple Classifier Systems* 2005, 407-416.

[8] Chen, X., P. J. Flynn, and K. W. Bowyer, "Infra-red and visible-light face recognition," *CVIU*, 99 (3), 332-358, Sept. 2005.

[9] Gokberk, B., A.A. Salah. and L. Akarun, "Rank-Based Decision Fusion for 3D Shape-Based Face Recognition," *LNCS 3546: AVBPA*, 1019-1028, July 2005.

[10] Husken, M., M. Brauckmann, S. Gehlen and C. von der Malsburg, "Strategies and benefits of fusion of 2D and 3D face recognition," *IEEE Workshop on Face Recognition Grand Challenge Experiments*, June 2005.

[11] Hutchinson, A., "Remarks by Under Secretary Hutchinson on the launch of US VISIT," www.dhs.gov/dhspublic/interapp/speech/speech_0114.xml

[12] IEEE Computer Society, "Software Engineering Code of Ethics and Professional Practice," www.computer.org/computer/code-of-ethics.pdf

[13] Liu, X., K. W. Bowyer, P. J. Flynn, Iris recognition and verification experiments with improved segmentation method, *AutoID 2005*, New York, 118-123.

[14] Papatheodorou, T. and D. Reuckert, "Evaluation of automatic 4D face recognition using surface and texture registration," *Aut. Face and Gesture Recog.*, 321-326, 2004.

[15] Pay By Touch ™, www.paybytouch.com.

[16] Phillips, P. J., P. J. Flynn, T. Scruggs, K. W. Bowyer and W. Worek, "Preliminary Face Recognition Grand Challenge Results," *Aut. Face and Gesture Recog.*, 2006.

[17] Phillips, P. J., P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," *CVPR*, San Diego, June 2005, I:947-954.

[18] Phillips, P.J., P. Grother R.J. Michaels, D.M. Blackburn and E. Tabassi and J.M. Bone, "FRVT 2002: overview and summary", March 2003, available at www.frvt.org.

[19] Socolinksy, D., and A. Selinger, "Thermal face recognition in an operational scenario," *CVPR*, June 2004, II:1012-1019.

[20] Wayman, J. L., "A path forward for multi-biometrics," *ICASSP '06*, to appear.

[21] www.bee-biometrics.org

[22] Xu, C., Y. Wang, T. Tan and L. Quan, Automatic 3D face recognition combining global geometric features with local shape variation information," *Aut. Face and Gesture Recog.*, 308-313, 2004.

[23] Yan P., and K. W. Bowyer, Multi-biometrics 2D and 3D ear recognition, *Springer-Verlag LNCS 3546: AVBPA*, July 2005, New York, 503-512.

[24] Yan P., and K. W. Bowyer, Empirical evaluation of advanced ear biometrics, *Workshop on Empirical Evaluation Methods on Computer Vision*, San Diego, June 2005, 41.