Temporal Integration for Continuous Multimodal Biometrics

Alphan Altinok and Matthew Turk
Computer Science Department
University of California, Santa Barbara
Santa Barbara, California 93106
{*alphan, mturk*}*@cs.ucsb.edu*

Abstract

Typically, biometric systems authenticate the user at a particular moment in time, granting or denying access to resources for the complete session. This model of authentication does not appropriately address environments where a different individual may take over a system from the original user (either willingly or otherwise). We propose a multimodal system that performs authentication continuously *by integrating information temporally as well as across modalities. Such continuous authentication provides ongoing (rather than one-time) verification and can easily be coupled with another system for dynamically adjusting access to privileges accordingly.*

We present an initial approach for temporal integration based on uncertainty propagation over time for estimating channel output distribution from recent history, and classification with uncertainty. Our method operates continuously by computing expected values as a function of time differences. Our preliminary experiments show that temporal information improves authentication accuracy. These empirical results are promising and justify further investigation.

1. Introduction

Biometric user authentication is typically formulated as a "one-shot" process, providing verification of the user when a resource is requested (e.g., logging in to a computer system or accessing an ATM machine). Once the user's identity has been verified, the system resources are available for a fixed period of time or, more typically, until the user logs out or exits the session. While perhaps appropriate for short sessions or low-security environments, this model for authentication is flawed, as it is based on two strong assumptions: (1) a single verification is sufficient, and (2) the identity of the user is constant during the complete session. If the user leaves the work area for a while, or is forcibly removed in a hostile environment, the system continues to provide access to the resources that should be protected. Continuous biometrics attempts to improve on this situation by addressing these assumptions and making user authentication an ongoing process, rather than a one-time, point-of-access occurrence.

One way to approximate continuous biometrics is to require active user authentication on a regular basis, e.g., requesting a password or thumbprint verification every few minutes or so. In most environments, this is not an acceptable requirement. Passive verification, via modalities such as face recognition, can be used to authenticate at a much higher rate, perhaps several times per second, without requiring active user participation. This raises other questions that affect usability: What if, due to a lighting change, noise, or any of several other conditions, the verification fails momentarily? What if the modality in use cannot provide any authentication report for a time?

To be truly useful, continuous biometrics requires temporal integration. In general, a continuous biometric authentication system should be able to provide a meaningful estimate of authentication certainty at any time. This requires analyzing the temporal characteristics of biometric modalities and user behavior to provide a model of user identity that is a continuous function of time (or a discrete function with a reasonably small update rate). Intuitively, the certainty of an authentication result should be relatively high at the moment the score is reported (depending on the characteristics of the modality), and then decrease monotonically over time, until a new report is received.

Temporal integration is particularly relevant and useful in the case of multimodal biometrics. When multiple modalities are used in concert to provide user authentication, there is usually an implicit temporal model — even though the different modalities may report at slightly different times, the results are treated as if they had arrived simultaneously. This is equivalent to assuming a constant user model during this short period.

The most interesting and potentially useful case is when there are multiple modalities in use, where the characteristics of the various modalities may differ significantly.

For example, consider a high-security workstation situation where the biometric modalities are fingerprint, face, voice, and keyboard (keystroke pattern), representing a range of temporal characteristics (frequency and regularity of reports) and accuracies. Keystroke pattern recognition is likely to be the least reliable as an authentication technique, but at times it will give almost continuous output, while the other modalities may have nothing to report. Fingerprint recognition may be quite accurate, but will only be available occasionally. In this situation, we envision a system that monitors all the modalities and makes the best possible decision at any given point in time — even if there has been no information in the recent past. With this model of continuous authentication, a system can constantly communicate the degree of belief in a user's identity, and a monitoring system can implement an appropriate program of action for the particular security environment. A slight decline in authentication certainty may cause certain sensitive areas to be made inaccessible to the user (in many cases not at all disturbing the benign activity of the user), while a large decline may result in the system shutting down access.

Integrating biometric modalities into decision-making has produced successful results in terms of accuracy and robustness [1, 5, 8]. Still, this model of authentication fails to address the temporal nature of the problem. The main goal of this work is to present a temporal integration method to investigate potential benefits of time information for the realization of a continuous authentication system. As such, the system could generate continuous results in terms of confidence in the identity of the user, which would enable adjusting the security level accordingly in real time. In relation with behavioral traits, which are under investigation as admissible biometrics [7], temporal integration would be useful for detecting gradual or abrupt changes or variations in fitness to perform a task.

2. Multimodal Biometrics

There has been a good deal of research in recent years on integrating multiple modalities to identify or authenticate a user. In such a multimodal biometric system, the method of integration is very important, as the accuracy of a strong biometric could suffer when integrated with a weaker biometric [3, 6]. To our knowledge, there has been no published research in the biometrics community to date that focuses on temporal integration as formulated here.

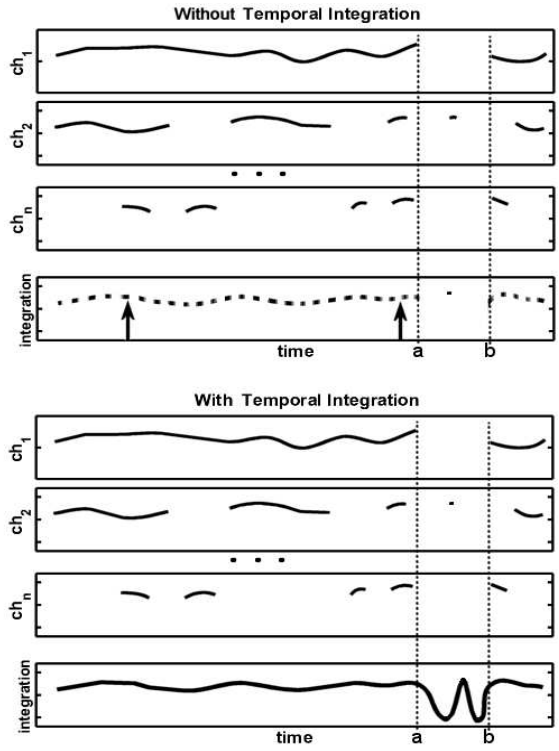

Figure 1: A static multimodal system (top) vs. one with temporal integration ($bottom$). Normalized scores from three channels are shown, with the integrated authentication score below. The multimodal system at top can not integrate information from all channels. For most of the time from a to b, the static multimodal system cannot perform authentication.

Figure 1 shows a qualitative comparison between a multimodal system that performs integration across modalities (without integration over time) and one which does temporal integration as well. The first system would be ineffective when there is no channel reporting — e.g., for most of the time between a and b. Through the entire sequence, the system would have to make decisions based on only partial observations, except where all channels are reporting an opinion (as indicated by arrows in Figure 1). In reality, due to the nature of biometric modalities involving lengthy computations or sample collection times, this should not be expected to happen frequently.

Interestingly, most accurate biometrics (iris scan, fingerprint, DNA matching and the like) are either lengthy procedures in collection or verification, or they are intrusive and cannot be performed frequently. A static multimodal system can only use such accurate indicators once they are observed.

2.1. Channel Integration

A multimodal biometric system can integrate modality information ("vertical" integration) at *feature*, *score*, and *figurethree* levels [1, 11, 5, 9]. In general, the most information is available at the feature level; integrating at this level is considered to be "early" integration. However, training at this level can be very complex and require an inordinate amount of data; later (higher) levels of integration are easier to build and often yield higher degrees of robustness. For decision level integration, it can be shown analytically that a strong biometric can achieve better accuracy alone than combined with a weaker biometric if both are operating at their cross-over points [6]. Unless the cross-over point of the weaker biometric is shifted, integration at the decision level would not be more accurate. Incorporating temporal information could change this limitation by shifting the cross-over point of weaker biometrics.

Since modality integration can be handled independent of temporal integration, it is possible to use various channel integration methods to improve overall accuracy of the system. In this work, channel integration is not our primary goal, so we chose a simple naive Bayes classifier to handle channel integration as a binary classification problem incorporating uncertainty measures. Similarity scores from individual biometric channels are normalized to the interval $[0, 1] \in \Re$ and integrated using the Bayes classifier. Our temporal integration method generates an expected score distribution and an estimated related uncertainty about this distribution. We weight class priors by the associated uncertainty before classification. It should be noted that weighting class priors would not scale well with larger data sets [4] presenting a potential limitation, especially since we are concerned with real-time operation.

2.2. Temporal Integration

There are several challenges for temporal ("horizontal") integration of a multimodal authentication system. First, as mentioned in the introduction, individual biometric channels cannot always provide simultaneous observations. One channel might provide information at a much higher frequency than another channel. Second, some channels might only provide sporadic observations over time. For example, we could not expect the user to provide a fingerprint at certain times. Third, for sporadic channels alone, temporal integration could be useless or statistically meaningless, if not impossible, to formulate, since there might be unexpectedly long intervals between observations. Fourth, the system should provide a way of making decisions during time intervals even if none of the individual channels provide any observations in that instant. For example, if we made observations δ milliseconds ago, then the system should be able to make decisions based on recent observations as we would not expect the user to be away in such a short interval. Our method addresses all of these challenges.

Logically, we have the choice of first integrating temporally or over channels (horizontally or vertically). If we first integrate over channels, then the problem is equivalent to temporal integration using a single biometric channel. On the other hand, integrating temporally first enables us to work with asynchronous biometric channels, since within some neighborhood in time of an observation we will have very good estimates from that observation. For making decisions in the absence of observations at a given point in time, we use expected values of observations from channels with varying degree of uncertainty. Perhaps the best approach, but also the most complex to formulate, is to integrate in both directions (across channels and across time) simultaneously, rather than sequentially.

3. Method

Just as in integrating channels, for temporal integration we can choose to integrate information at level of features, scores, or decisions. Our method works in continuous time by computing expected values of scores as a function of time difference between the last observation and current time. The main idea is based on the assumption that an authentication score is still valid for some amount of time, δt. As time passes, we should be less and less certain about this value. To formulate this idea as a function of time we estimate an uncertainty measure of scores per channel from the recent past, until a new observation is recorded. The joint posterior distribution of a score is approximated and then propagated over time until we obtain a new score from that channel. Due to the propagation of the score distribution over time, we use a degeneracy model for the uncertainty measure of each score.

The most important reason in favor of working with scores, rather than at the feature or decision level, is the way of modeling uncertainty of channel opinions. In lower levels, uncertainty has a related physical meaning. For example, at the physical measurement level, uncertainty is related to signal noise, which might not necessarily map well into an uncertainty about the decision. Treating scores as random variables is in fact this mapping, statistically backed by the Central Limit Theorem. Another reason to work with scores, aside from the underlying mathematical difficulty of using many features, is the fact that feature selection is still as much

art as it is science. Naturally, we would prefer our integration method to be as general as possible. On the other hand, the later the integration, the more information is discarded, so early integration may achieve better results, using an appropriate set of features. After establishing promising results with scores, we plan to continue investigating such directions in the future.

Each channel is assumed to provide a normalized similarity score s, and an expected variance σ_{ch} as a characteristic parameter of the channel. If σ_{ch} is not provided, it is computed for each channel offline. This measure is equivalent to inherent uncertainty in a channel's decisions. This variance is only used as the default variance of the channel if computing the channel variance is not possible from recent past. For example, σ_{ch} is needed for initial few scores or for channels which provide scores at longer intervals. One might ask that if the uncertainty is known, why compute it from the past again? The reason is that the σ_{ch} measure itself varies over time. For example, if lighting conditions were the underlying reason for the face recognition channel to report highly variable scores over the past 5 seconds, this variability should be corrected in par with the lighting conditions.

We normalize channel scores to $[0, 1] \in \Re$, where 1 indicates perfect similarity to the user model and 0 indicates an unknown person. For channels with higher frequency, we compute the uncertainty σ_p from past scores within a τ_{ch} time period. Note that this duration is the crucial part of our method and it has a different value for each channel.

We model each channel with a Gaussian $\tilde{N}(\mu, \sigma_{ch})$ or $\tilde{N}(\mu, \sigma_p)$, where μ is the reported score for the channel, as discussed above. (We will refer to σ_{ch} and σ_p as σ from now on.) Consequently, scores are random variables with $s \sim \tilde{N}(\mu, \sigma)$. This distribution is propagated over time with increasing uncertainty in the score value as a function of time.

Figure 2 shows conceptually how a score s is treated. The darker lines over the Gaussian show the change in shape of Gaussian over time.

When a score is recorded, a timestamp t is generated and the uncertainty σ is computed over the past $t - \tau$, if applicable, otherwise $\sigma = \sigma_{ch}$. The idea is that we will be less and less certain about this score and probabilities of all possible scores will increase as time passes by.

The increase of uncertainty over time is computed as a function of time from the last score. We used an exponential degeneracy function $\phi(\tau)$ to estimate the mode ($\frac{1}{\sigma\sqrt{2\pi}}$) of the $\tilde{N}(\mu, \sigma)$ at $t + \tau$. The degeneracy function $\phi(\tau) = k \exp^{\alpha\tau}$ depends only on α which we take as the mean variability over the last τ_{ch} time period.

Once an estimate of score distribution $\tilde{N}(\mu, \sigma)$ at $t+$

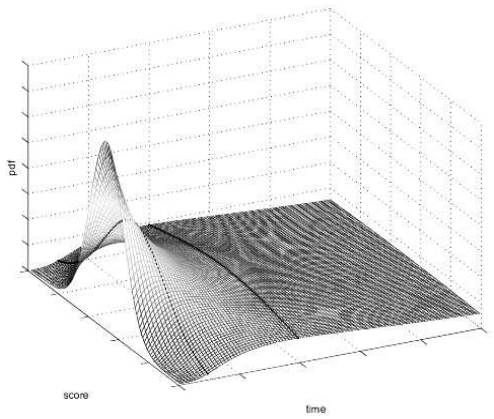

Figure 2: Propagation of scores and associated uncertainties over time. As time passes, σ increases from a recently computed σ_p.

τ is obtained, we compute the expected value of a score at $t + \tau$ from this distribution by evaluating

$$E_{\tilde{N}_{past}}\{N_{now}(s)\} = \int_{-\infty}^{\infty} \tilde{N}_{now}(s)\tilde{N}_{past}(s)ds$$

Note that the limits of the integral we are interested in are not $-\infty$ and ∞, but 0 and 1. Hence the distribution at $t + \tau$ is not a proper Gaussian anymore. However, the error resulting from ignoring the tails of this distribution is insignificant. Although we could opt for a proper distribution, such as a triangular distribution, this would introduce a larger modeling error. Alternatively, this Gaussian can easily be scaled to cover unit area, which would not change the expected value of the score. To evaluate the expected value we use the following approximation.

Suppose $X = \{X_1, X_2, ..., X_n\}$ is the set of random variables that characterize the model, with values $x_1, x_2, ..., x_n$. The expectation, $E(a)$, of a function $a(X_1, X_2, ..., X_n)$ can be approximated by

$$\sum_{x_1} ... \sum_{x_n} a(x_1, ..., x_n)P(X_1 = x_1, ..., X_n = x_n)$$

$$\approx \frac{1}{N} \sum_{k=0}^{N-1} a(x_1^k, ..., x_n^k)$$

where x_i^k are the values for point k in a sample of size N.

It should be noted that we want to minimize the filtering effect of our method, where occasional false positives and false negatives are *corrected* by subsequent scores. Therefore a predictor-corrector style modeling, such as a Kalman Filter, is not a model of choice. Also,

the choice of the exponential function was based on lifetime modeling studies, which could be better modeled with $(1 - \tanh(x))$ or a similar function. The crucial heuristic of our method is the length of considered past, and how many correct scores it includes. Clearly, the degeneracy model leaves room for refinement. Incorporating contextual information successfully into the model and learning appropriate parameters from data are possible refinements.

4. Experiments

We chose face, voice, and fingerprint as individual biometric modes for simulating channels with different temporal characteristics. The lack of a suitable multimodal corpus with face recognition, voice verification, and fingerprints of individuals forced us to simulate individuals by matching independently collected data into virtual identities for 24 individuals. Scores from each channel are obtained as detailed below. Our goal is to achieve continuous multimodal authentication which is more accurate than the component channels and gives meaningful results at any point in time. A second set of experiments was run with different lengths of past scores in consideration.

4.1. Face Recognition

This is the channel with the highest reporting frequency. Face scores are obtained from a face recognizer based on Eigenfaces [12]. Images are obtained using a face detector built on [13] from 20fps video. For each individual, there is a 2 min video containing $\sim$80 frames at (near) frontal pose. 20 images from frontal images were used for training. The data does not have frontal pose throughout the entire video sequence, hence the recognition does not provide good scores every $50ms$.

4.2. Voice Verification

A subset of the TIMIT database [10] was used. The subset contains LPC cepstrum feature vectors. The energy in all recordings was normalized to compensate for possible differences in loudness. After preemphasis, $16th$-order LPC-cepstra were calculated for $32ms$ frames centered at $16ms$ intervals. The feature vectors are the rows of the resultant matrix. Each frame is used as an independent sample drawn from the distribution of that speaker. Each speaker is modeled as a Gaussian. In total just under $15s$ of training data per speaker are available. Log-likelihoods are the scores for voice verification.

Table 1: Recognition rates of individual channels vs temporal multimodal integration.

Integrated	304	47.50%
Face	210	32.81%
Integrated	173	97.74%
Voice	171	96.61%

Table 2: Correct recognition at variable history lengths.

History length (secs)	0.5	1.0	2.0	5.0
Correct recognition	304	310	318	301
Recognition rate (%)	47.5	48.4	49.7	47.0

4.3. Fingerprint

A subset of fingerprint data was obtained from the FVC2002 fingerprint verification competition. A demo version of fingerprint identification/verification software [14] was used to obtain similarity scores between fingerprints. The software extracts minutiae-based features. It handles rotation and intensity variations. For successful operation it requires a minimum of 10 features for each fingerprint.

4.4. Results

We expect that temporal integration would be useful by enabling continuous authentication and by improving accuracy of a multimodal biometric system. Figure 3 shows decisions made by our method over a period of 32 seconds (each tick = 1 frame). The simulated user is the authentic (virtual) identity over the entire period, so that a 1 indicates a correct authentication, and a 0 marks where the system fails to authenticate the identity correctly. The varying face recognition scores are due to face motion, where it becomes frontal 6 times during the 32 second period. Better recognition scores are obtained when the face became full frontal in view.

The top three graphs show individual channel scores. The bottom graph shows the decisions obtained by our method with a history length of 0.5 second for all channels. The first few points are not affected by temporal integration due to insufficient history. In the case of a non-temporal multimodal system, all (if any) decisions would have to be based on what is observed at that point in time, regardless of what happened in the instant before. We can poll our system at any time for an authentication.

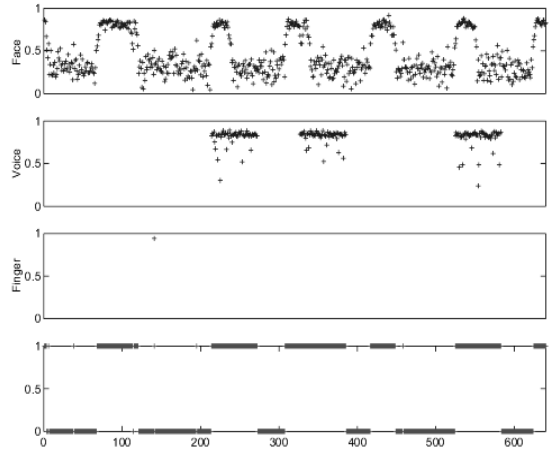

Figure 3: Temporal integration over a period of 32 seconds. Individual channels report scores in real time as they become available; note the single fingerprint score in frame 141. The bottom graph shows binary verification decisions made at every frame, a 1 being valid authentication.

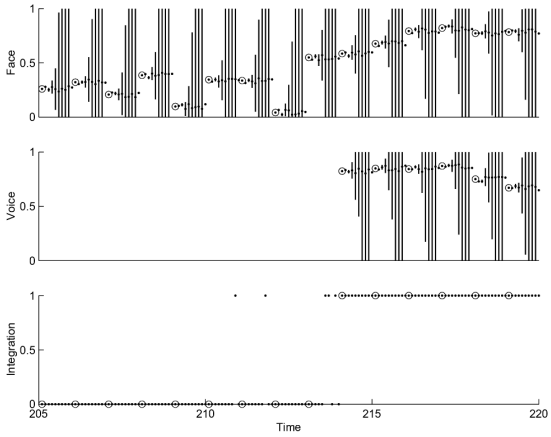

Figure 4: An enlarged version of Figure 3 between frames 205 and 220. Each frame is polled 10 times within the frame. Vertical lines show the variances of propagated distributions around the means since the last score. Fingerprint channel is not shown since the only score is beyond the history window of 0.5 seconds. Circles show actual scores from Figure 3.

To verify that integrated results are actually comparable to individual channel rates, we compared the correct recognition counts of integrated and individual channels. Table 1 shows this comparison over the periods when each individual channel is active.

Table 2 shows the effect of history length on recognition. The history length is applied to all channels. Our results suggest that there is a cross-over point for the length of relevant history, although more extensive study is necessary.

Figure 4 shows an enlarged sequence between frames 205 and 220 (0.75 seconds). Vertical lines show the variances of propagated distributions around the means since the last score. Fingerprint channel is omitted from both Figure 4 and Figure 5 since the only score lies beyond the relevant history of 0.5 seconds. An authentication result was requested 10 times within each frame. Our method is only limited by the underlying hardware in terms of temporal resolution, and an authentication score can be obtained given any point in time.

Figure 5 shows the same enlargement for a system that only integrates channels. Authentication is only possible when at least of the channels report an opinion. Note that in Figure 4 and Figure 5 the authentication is based only on face recognition scores for the first half of the sequence as no previous data was recorded from other channels within the last 0.5 seconds. Depending on the length of relevant history, our system can evaluate what has been seen within the last n seconds even if there were no scores reported from any channel, which would be impossible without temporal integration.

5. Conclusion

We have introduced a new model for temporal integration in biometric user authentication and developed an initial method for a continuous authentication system. Our temporal integration method depends on the availability of past observations, which makes the length of relevant history an important heuristic. Another important design choice is the degeneracy function. The existence of a cross-over point in the history suggests further investigation of the degeneracy.

We have shown on simulated data that our preliminary system can provide continuous authentication results which are consistently better than individual components of the system. Clearly, gathering a true multimodal database is very important for continued work in this field.

When the history length is set to 0, the system ignores temporal integration and degenerates into a multimodal system. Although our approach attempts to minimize the filtering effect of false positives and false negatives, our temporal integration method would suffer from this smoothing behavior to some degree as it stands. The net effect of this behavior is integration of positive decisions, as well as negative ones, as expected.

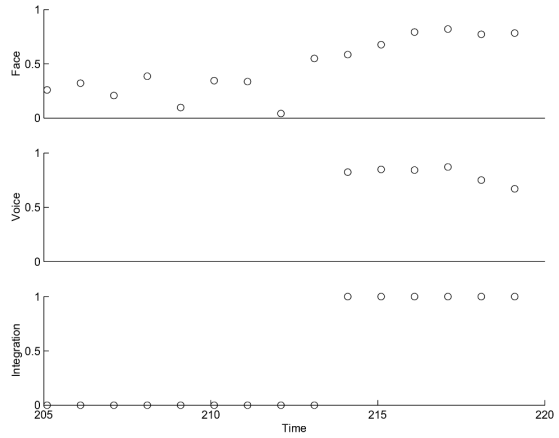

Figure 5: An enlarged version of Figure 3 between frames 205 and 220. Channel integration only, no temporal integration was performed. The system can perform authentication only when a score was reported by at least one channel.

References

[1] R. Brunelli and D. Falavigna, Person Identification using Multiple Cues, *IEEE Transactions on PAMI, Vol 12*, pp. 955-966, Oct. 1995.

[2] T. Choudhury, B. Clarkson, T. Jebara, and A. Pentland, Multimodal Person Recognition using Unconstrained Audio and Video, *Second International Conference on AVBPA*, pp. 176-181, Washington D. C., USA, Mar. 1999.

[3] John Daugman, Biometric Decision Landscapes, *Technical Report, University of Cambridge, UK*, 1999.

[4] R. Duda, P. Hart, and D. Stork, Pattern Classification, *John Wiley & Sons, NY. 2nd Ed.*, 2001.

[5] L. Hong and A. Jain, Integrating Faces and Fingerprints for Personal Identification, *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 12*, pp. 1295-1307, Dec. 1998.

[6] L. Hong, A. K. Jain, and S. Pankanti, Can Multibiometrics Improve Performance?, *In Proceedings AutoID'99, NJ, USA*, pp. 59-64, Oct. 1999.

[7] A. K. Jain, R. Bolle, and S. Pankanti, Multimodal Biometrics: Personal Identification in a Networked Society, *Kluwer Academic Publishers*, pp. 1-38, 1999.

[8] N. Poh and J. Korczak, Hybrid Biometric Authentication System Using Face and Voice Features, *Third International Conference on AVBPA*, pp. 348-353, 2001.

[9] N. Poh, S. Bengio, and J. Korczak, A Multi-sample Multi-source Model for Biometric Authentication, *Proceedings, IEEE 12th Workshop on Neural Networks for Signal Processing*, pp. 375384, 2002.

[10] S. Seneff and V. Zue, Transcription and Alignment of the TIMIT Database, *In Proceedings of the Second Symposium on Advanced Man-Machine Interface through Spoken Language, Oahu, Hawaii*, Nov, 1988.

[11] G. Shakhnarovich, L. Lee, and T. Darrell, Integrated Face and Gait Recognition From Multiple Views, *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition, Lihue, HI*, Dec. 01.

[12] M. Turk and A. Pentland, Eigenfaces for Recognition, *Journal of Cognitive Neuroscience, Vol. 3, No. 1*, pp. 71-86, 1991.

[13] P. A. Viola and M. J. Jones, Robust Real-Time Object Detection, *Technical Report, COMPAQ Cambridge Research Laboratory, Cambridge, MA*, Feb. 2001.

[14] http://www.neurotechnologija.com/verifinger.html.