Vehicle Systems: Comfort & Security Enhancement of Face/Speech Fusion with Compensational Biometric Modalities

Michael Biermann, Tobias Hoppe, Jana Dittmann
ITI Research Group on Multimedia and Security
Universitätsplatz 2
39106 Magdeburg, Germany
{michael.biermann, tobias.hoppe, jana.dittmann}
@iti.cs.uni-magdeburg.de

Claus Vielhauer
Department of Informatics and Media PSF 2132
14737 Brandenburg an der Havel, Germany

ITI Research Group on Multimedia and Security
Universitätsplatz 2
39106 Magdeburg, Germany

claus.vielhauer@
{fh-brandenburg.de, iti.cs.uni-magdeburg.de}

ABSTRACT
Biometric modalities can be used to improve security, safety and comfort in different applications. For example it is possible to restrict access to computer systems or buildings by biometric authentication. Since the usage of biometric modalities is considered more and more also for vehicles, in this paper we review two existing approaches of fusing speech, face and additional biometric modalities in automotive applications. We also combine them to an extended concept for an improvement of the achievable comfort and security. Especially we include additional soft biometric modalities to compensate failures of the biometric sensors to ensure business continuity through enhanced availability. However, enhancements of the comfort of biometric authentication systems on one side, often lead to a decrease of their security on the other. In a first theoretical simulation we show the overall comfort improvement, compare the new concept with the selected two existing approaches and discuss potential security implications.

Categories and Subject Descriptors
K.6.5 [**Management of Computing and Information Systems**]: security and protection – *Authentication.*

General Terms
Measurement, Performance, Design, Reliability, Experimentation, Security, Human Factors, Theory, Verification.

Keywords
automotive, biometric modalities, characteristics fusion, speech, face

MM&Sec'08, September 22–23, 2008, Oxford, United Kingdom.

1. INTRODUCTION
A lot of different biometric characteristics can be used to identify or verify a person, for example fingerprints, voice or face [1]. Different biometric modalities give degrees of different discriminatory power. Therefore, they are often combined into multi-modal systems to achieve a possibly enhanced authentication performance. Thereby different fusion strategies exist. For instance it is possible to combine the respective features that are extracted from the data like an image or audio signals. Another alternative is the fusion of the resulting matching scores (MS) [2]. Most of the MS-fusion approaches use weights to merge the different scores, e.g. by summing up the weighted scores. The weights might for example be determined by the Equal Error Rates (EER) of each single uni-modal subsystem [3], which is a possible property to characterize a biometric authentication system (defined by the intersection of the False Acceptance Rate/FAR and False Rejection Rate/FRR [4]).

To use the efforts of multi-modal biometric authentication systems also in the car, the special requirements in the automotive domain have to be respected. Unlike in stationary setups, in this environment the quality of the collected biometric data especially depends on factors like the environment conditions (e.g. light or noise levels) or the state of the sensor (e.g. due to mechanical shocks or dirty sensors). To also include such influences we propose to collect and evaluate respective data. Even existing sensor information like light levels, microphone input or window positions could support the adaptive calculation of the fused matching score (also see [5]). To evaluate the current quality of a given sensor, its input could be checked against templates from earlier enrollments (this implies that the templates would probably have to be stored as raw data). We described another method in [6] that uses additional sensors to determine the current environment conditions (see Figure 1.1 and 2.1). For example a camera takes images in different qualities at daytime than at night. By an additional light sensor it is then possible to include partial information to estimate the quality of the captured images, depending on the illumination situation.

In this paper we focus on the application of biometric modalities in the automotive domain. Here, visual and acoustic sensors have been supported more and more [7] for various safety and convenience purposes since the past few years.

Beneath biometric authentication these sensors can be used to assist the driver and higher his/her concentration [8] (e.g. by allowing to

control different systems like the navigation by voice commands, maximizing the driver's attention to road and traffic conditions).

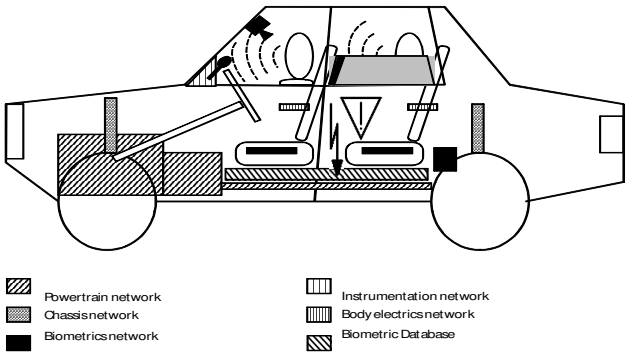

Figure 1.1 Simplified car [6] (translated and enhanced)

Biometric modalities can also be used to address security requirements like access control issues. In one possible scenario, a fingerprint sensor might be included in the door handle. The car will only open, if an authorized person tries to open the door and is recognized by her/his fingerprint [9]. Other applications could also be safety related, e.g. to detect if a person is out of position or if a child's safety seat is mounted in an unsafe way. In these cases it might be possible that an airbag, which is mainly used for protection, could cause additional damage [10 & 11]. Therefore, if such a situation is detected, the system could deactivate the airbag or adjust the inflation parameters [10]. Another example is the use in driver monitoring systems: a camera records the driver and an algorithm analyzes the movement of the eyelids. As a suspicious sequence is detected, e.g. due to fatigue, the system can warn the driver [12].

Although a wide variety of face and speech fusion concepts and applications exists, we can only refer to a minimum number of related work approaching the changing environments and special sensors for the automotive domain. For example, [17] considered the fusion of multi-modal biometric modalities (like audio and video) for the application in vehicles, in particular for speaker identification. Especially in this domain we see a high relevance to respect the frequently changing environment conditions in a car and want to focus on adaptive approaches. Previous work on general environment-adaptive multi-modal biometrics has for example been presented in [18]. With a special focus on a car and its special features including numerous, versatile sensory input, we expect a focus on this to be very effective. Because the existing approaches neglect the usage of additional soft biometric modalities we propose the usage of such biometric modalities, as additional, less distinctive characteristics, that can support the authentication of a person. Within a car, these might be data indicating personal characteristics. To identify person relatable characteristics, e.g. the personal manner of driving, several features or sensor information can be evaluated to enhance existing multi-modal systems. In this paper we exemplarily consider an evaluation of the actuation style of pedals (e.g. based on accelerator or brake pressure) or the steering wheel; also the vehicle speed could be evaluated (also see [15, 21]). Furthermore beside additional soft biometric modalities we use the known concept of Biometric Hashing (see for example [23]) to address the privacy requirements of biometric reference data.

This paper is structured as follows: In section 2 we give a short overview of the two aforementioned existing approaches for face and speech fusion in automotive domain and discuss their properties. In the third section we develop an improved version that combines the two existing approaches to enhance the comfort and security and compare all these strategies. Additionally, in section 3 we introduce Biometric Hashing for privacy enhancement. In section 4 we summarize this paper and refer to future work. In the later section we also give a guideline how the usage of additional soft biometric modalities can enhance availability issues.

2. Existing Approaches

In this section we present two existing approaches on a fusion of multi-biometric modalities in the automotive domain. We consider a fusion at the matching score level, which means that the different biometric input is processed by separate algorithms. For example, a first algorithm could be used for face analysis, a second one for speech analysis (e.g. [20]) and a third one might calculate a matching score depending on the body weight. After giving some basics of both fusion approaches in the next subsection, we further introduce the *Adaptive Dynamic Fusion* (ADF) from [6] in subsection 2.2 and *Simplified Face Speech Fusion* (SFSF) from [13] in subsection 2.3.

Table 2.1 summarizes the identifiers used in the following subsections.

Table 2.1 Used identifiers and short description

Identifier	Description	Approach
t	discrete point in time	ADF
s_1	camera	
s_2	microphone	
s_3	body weight sensor	
s_4	light sensor	
s_5	window position sensor	
s_6	speed sensor	
$B_{1,t}$	lip movement flag at time t	
$B_{2,t}$	seat usage flag at time t	
$V_{j,t}$	confidence factor for s_j at time t	
W_j	fixed weight for s_j	
N_t	normalization factor at time t	
$MS_{j,t}$	uni-modal matching score for s_j at time t	ADF & SFSF
$MS_{fus,t}$	resulting matching score at time t	
FAR	false acceptance rate	SFSF
f_j	known failure through software	
$d_{j,t}$	disturbance factor at time t	
$g_{j,t}$	dynamically adapted weight at time t	
$w_{j,t}$	normalized weight at time t	

2.1 Basics on Both Chosen Approaches

Since the fusion of different biometric modalities has already been proven to be effective in other applications (see for example [19]), we expect it to be also useful in the automotive domain. The collection of sufficient biometric data in future cars is realistic; some of them (like voice) are even already being used today. The fusion of such biometric signals could allow for many new applications, e.g. to address comfort, safety or security issues. The biometric modalities for such a fusion that have been selected in [6] are face, speech and body weight. Therefore it is necessary that the car

provides the required sensor equipment camera, microphone and body weight sensor, which can partially already be found in today's cars.

In the automotive domain, the environment can change very frequently, e.g. in terms of noise (see [20]) or lighting conditions. It is not possible to assume a quiet and well illuminated environment like it is often being done in static setups. Therefore, in the automotive domain it is important to use a concept that is able to adapt dynamically to different kinds of disturbances. In [6] additional information was respected by including additional sensors (e.g. light sensors). The approach described in [13] uses reference data instead (like a standard image or tone), which is used to estimate the quality of the current signal and the functionality of the corresponding sensor.

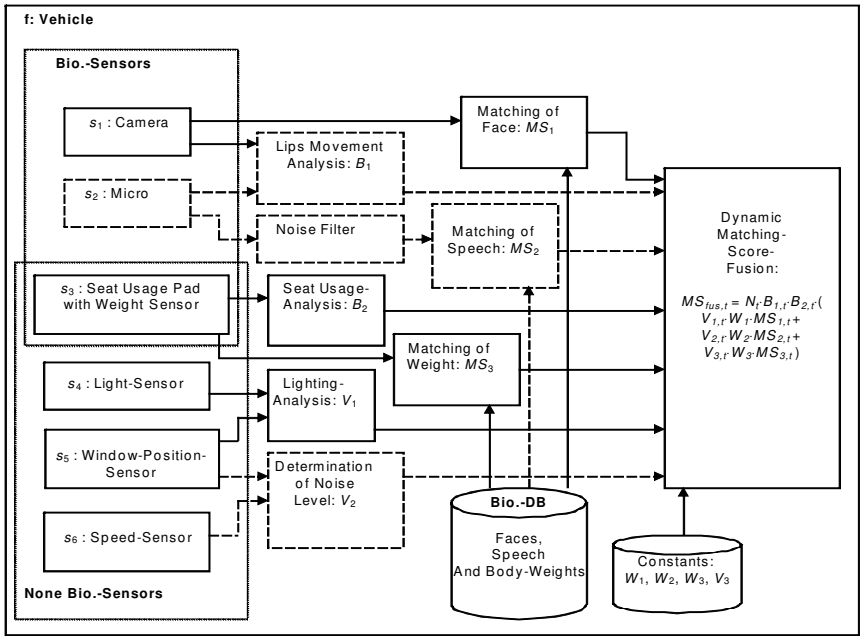

Figure 2.1 Biometric authentication of the driver (translated from [6])

2.2 Adaptive Dynamic Fusion

The original concept for vehicle systems called Adaptive Dynamic Fusion (ADF) [6] considers three types of biometric modalities for the automotive domain: face, speech and body weight. Therefore a camera, microphone and a body weight sensor are used. After the collection of the respective biometric modalities at time *t* the input data is processed by corresponding algorithms like it is common on various biometric systems. Simplified, first the required features are extracted. The second step can be a further processing of the obtained features. Third, a comparison is being done and a matching score is calculated [14]. The original concept performs the fusion of the separate matching scores according to Equation 2.2.1. As the single matching scores are usually defined in the interval [0, 1], a normalization (2nd line of Equation 2.2.1) is used to obtain a fused matching score that also matches this interval.

For the discussion of Adaptive Dynamic Fusion (ADF), biometric authentication of a driver (Figure 2.1), we use definitions as follows, see also Figure 2.1. *t* is the time when the data from the sensor s_i has been captured. All s_i signify different sensors i=1...6. They are assigned as follows: s_1 camera, s_2 microphone, s_3 body weight sensor, s_4 light sensor, s_5 window position sensor and s_6 speed sensor. Both B_j, j=1..2 are binary factors, which depend on the actual conditions of sensors s_1 … s_6, as outline by the notation $B_{j,t}(s_1 \ldots s_6)$ in Equation 2.2.1. They represent flags used to determine if a) a voice command was spoken ($B_{1,t}$) and b) a person is sitting on a seat ($B_{2,t}$). The $V_{j,t}$ are confidence factors related to the biometric sensors s_1 to s_3 and depend on additional sensors s_4 … s_6. All W_j are constant weights that are based on an estimation of the EER of the related biometric subsystem involved. They are motivated by weight fusion strategies in [3]. All $V_{j,t}$ and W_j have to be in the interval [0, 1]. The $MS_{j,t}$ are the already mentioned uni-modal matching scores. $MS_{fus,t}$ is the resulting score which is used to authenticate the affected person.

The final matching score $MS_{fus,t}$ becomes zero once one of the $B_{j,t}$ operands is set to 0 because of a sensor malfunction. On the one hand the advantage is that the provided access security is high. On the other hand the lower comfort might be a disadvantage. Even if all other sensors are working properly the final matching score will be zero.

$$MS_{fus,t} = N_t * \left[\prod_{j=1}^{2} B_{j,t}(s_1,\ldots,s_6) \right] * \sum_{j=1}^{3} V_{j,t}(s_3,\ldots,s_6) * W_j * MS_{j,t}(s_j)$$

$$where \quad \sum_{j=1}^{3} W_j = 1; \quad N * \sum_{j=1}^{3} V_{j,t} * W_j = 1; \quad V_{j,t} \in [0,1]; \quad W_j \in [0,1]$$

Equation 2.2.1 Adaptive Dynamic Fusion

If all the $V_{j,t}$ are set to 1 it is obvious that the sum of $V_{j,t}*W_j$ is exactly 1. In this case no normalization is needed as the condition is already fulfilled. Assuming that only one of the $V_{j,t}$ is less than 1 (e.g. $V_{1,t}$ due to high illumination), a normalization should be done (see Table 2.2.1).

Table 2.2.1 Exemplary normalization of $V_{j,t}*W_j$

	Modality *j*:		
	1	2	3
$V_{i,t}$ *confidence factor*	0.7	1.0	1.0
W_j *weight*	0.5	0.1	0.4
$V_{i,t}*W_j$	0.35	0.1	0.4
N_t	$1/(0.35 + 0.1 + 0.4) \approx 1.1765$		

To fulfill the condition, the normalization factor N_t is calculated. With the exemplary values of Table 2.2.1, especially the sum of $V_{j,t}*W_j$ (0.85), we get an N_t of about 1.1765. Normalization can be understood as a theoretical alignment of the weights W_j. The ratio between the weights is not affected. The advantage is clearly visible if one of the $V_{j,t}$ is equal to zero while the others remain at a value of one.

Assuming that the $V_{1,t}$ drops to zero (e.g. due to a rapid change of the illumination) it is obvious that the final matching score depends on the remaining $V_{2,t}$ and $V_{3,t}$. In this case the sum of $V_{j,t}*W_j$ is 0.5. Normalization factor N_t is calculated as 1/(1/2) = 2, so all weights are adjusted by multiplying them by 2.

2.3 Simplified Face Speech Fusion

Beside the approach of ADF from [6], Blaschke et. al. formed a method to fuse different biometric modalities with respect to the current performance of the sensors involved. In their work [13] they also focused on the scenario that a sensor could be damaged or has a malfunction. The idea is to equip the system with pre-recorded reference data for all sensors involved (such as microphone or camera) to compare all samples recorded at runtime with this pre-recorded reference data. The result is a factor describing the performance of the corresponding sensor that should be included in the calculation. This concept is called "Simplified Face Speech Fusion" (SFSF) within this paper. It is described in the Equation 2.3.1.

In the SFSF the variables are used as follows: $MS_{1,t} \ldots MS_{n,t}$ are the processed sensor data and represent the matching scores. The failure that is known through the software is represented by the f_i that can be understood as FAR dependent weights. Values $g_{i,t}$ are weights for the different biometric modalities, including the prior known failure f_i and also the disturbance variables $d_{i,t}$ for all involved biometric modalities at time t. Weights $w_{i,t}$ are normalized $g_{i,t}$ using Equation 2.3.1.

$$\begin{pmatrix} MS_{1,t} \\ MS_{2,t} \\ \ldots \\ MS_{n,t} \end{pmatrix}^T * \begin{pmatrix} w_{1,t} \\ w_{2,t} \\ \ldots \\ w_{n,t} \end{pmatrix} = MS_{fus,t}$$

$$and\ g_{i,t} = f_i * d_{i,t};\ f_i = 1 - FAR;\ 0 < i \le n$$

$$where\ w_{i,t} = \frac{g_{i,t}}{\sum_{j=1}^{n} g_{j,t}}$$

Equation 2.3.1 Simplified Face Speech Fusion (SFSF)

2.4 Comparison of the Existing Approaches

On the one hand, the most characteristic difference between both described approaches is that the first approach ADF uses additional sensor information to obtain knowledge about the environment conditions while the second simplified fusion does not use this kind of input. This additional information in the ADF approach is used to adapt the calculation by introduced confidence values $V_{j,t}$. On the other hand, ADF does not consider the current functionality of the according sensor as done in SFSF. For this purpose, the simplified fusion (SFSF) uses a reference quality measure like a defined tone to evaluate the current functionality of the corresponding sensor like the microphone.

Of course the influence of the environment is still included by this concept, but it cannot be determined explicitly. None of the mentioned approaches is able to distinguish if a malfunction and/or unfavorable environment has a bad impact on the results. To address this, in the next section we propose and discuss a new approach that combines the efforts of both existing approaches.

3. Enhanced Fusion Concept and Evaluation

In this section we extend the existing Adaptive Dynamic Fusion approach by the advantage of sensor quality measure of the simplified fusion concept to a new Enhanced Fusion Strategy (EFS). Furthermore we include the option that an application can also modify the provided matching scores and filter relevant information. We also discuss a method to enhance the privacy of the reference data by using a biometric hashing method. Additionally, we give a comparison of achieved results for the four different methods referring to exemplary sets of input values. These are theoretical values chosen to simulate the effects of the fusion strategies and their enhancements. Furthermore we discuss their impact to security, safety and comfort issues.

For the enhancement of the *Adaptive Dynamic Fusion* (ADF) we use the identifiers already introduced in Table 2.1. Additional we use variables shortly described in Table 3.1.

Table 3.1 Additional identifiers for EFS

Identifier	Description
$A_{j,t}$	weight adjustment operand for application at time *t*
$F_{j,t}$	functionality of the sensor s_j at time *t*
$B_{j,t}$	binary operand at time *t*; to select/unselect a biometric modality *j* in general
CB_j	compensational biometric modalities (*j*=1: steering properties; *j*=2: acceleration behavior; *j*=3: brake pedal pressure)
$MS\_CB_{j,t}$	matching score of CB_j at time *t*
MB_j	main biometric modalities (*j*=1: face; *j*=2: voice; *j*=3: body weight)
$MBU_{j,t}$	usage of main biometric modality *j* at time *t*
$F\_CB_{j,t}$	functionality of CB_j at time *t*
$V\_CB_{j,t}$	confidence factor of CB_j at time *t*
W_CB_j	weight of CB_j
WN_t	weight normalization at time *t*

3.1 Enhancement of the Fusion Strategy

Beneath the main biometric modalities which are used in ADF (face, speech, body weight) we suggest the use of additional hard (strong discriminatory power) and/or soft biometric modalities (weak discriminatory power) to increase the comfort and security. To combine both concepts of ADF and SFSF into a new Enhanced Fusion Strategy (EFS), new factors $F_{j,t}$, similar to the confidence factors $V_{j,t}$ are introduced for the functionality of the sensors involved. To evaluate the suitability of the measured data, the $V_{j,t}$ address the environment conditions that are measured through additional sensors. The functionality factors $F_{j,t}$ are based on data of the respective sensor to determine the current quality (state of the sensor). If, for example, the confidence factor for the microphone is at its maximum value of 1, this can be understood as the fact that the system should work properly and the environment is not expected to have bad influence on the result. But if in such a situation there is a loud noise in the recorded audio signal even if the environment is known to be silent, the microphone probably is defective. At this point we do not take care of the exact way how disturbing signals from the environment are detected. One solution might be to use two or more microphones and let control them each other.

Using functionality factors it is possible to give a graduation on how good a sensor is currently working. Assumed that there is some reference data for typical functionality it should be possible to detect abnormal behavior. Different levels of reference samples might be stored classifying ideal and regular functionality, different kinds of disturbed functionality as well as absolute failure. This would allow multiple levels of potential quality classifications.

The second aspect that has been added in this context are additional compensational biometric modalities (CB). This can also be soft biometric modalities that have some relation to the driver. For example, the individual manner of driving can in certain extends be derived from the input data like temporal behavior in pedal pressures (e.g. acceleration and braking), steering angles or global data like vehicle speed (see for example [15; 21]). We now include such information into our fusion by defining matching scores of such additional biometric modalities: For every main biometric matching score $MS_{j,t}$ we define an additional matching score $MS\_CB_{j,t}$. that compensates the respective main biometric in case it fails (e.g. due to component failures or difficult environment conditions). If the sensors s_1 … s_3 are working at their maximum functionality and in an optimal environment, the compensational biometric modalities are not included, but if the functionality of the main sensors (respectively the confidence regarding the environment) is decreasing, the CB will be included more and more in the calculation to compensate for the failures. Therefore Equation 3.1.1 introduces a new factor that describes the main biometric usage ($MBU_{j,t}$). We use the product of confidence $V_{j,t}$ and functionality $F_{j,t}$ of the corresponding sensor s_j at time t to estimate the quality of the respective main biometric modality MB_j as $MBU_{j,t}$. Individually for a given MB_j, the amount of the usage of the respective CB_j at time t is defined as $1\text{-}MBU_{j,t}$.

$$MBU_{j,t} = V_{j,t} * F_{j,t}$$

$$V_{j,t} \in [0,1]; F_{j,t} \in [0,1]$$

Equation 3.1.1 new Form of the Normalization

For all compensational biometric modalities we also introduce values for the functionality of the sensor $F\_CB_{j,t}$, the confidence factor $V\_CB_{j,t}$ and the weight $W\_CB_{j,t}$. To avoid false weighting that leads to an incorrect fused matching score we introduce a different kind of normalization than described in Equation 2.2.1. The weightings are still calculated depending on the EER, but an adjustment is carried out. First the weightings are calculated as in the ADF strategy separately for the main biometric modalities and the CB. This way, the weights are already adjusted within the class of the main and within the compensational biometric modalities. To also keep the relation between a mixed selection of main and compensational biometrics, the respective weights are additionally adapted to be in an appropriate relation to each other. In general, the CBs have a less discriminatory power and their influence should be less than those of the main biometric modalities, if they are included.

Therefore at the calculation of the weights during the adjustment of the system, the weights of the compensational biometric modalities are additionally adapted to be in relation to the main biometric modalities, instead of simple normalization to 1 as performed in ADF approach. Therefore the weights of the currently used biometric modalities are summed up and normalized to 1.0. Equation 3.1.2 shows the calculation of the corresponding factor. This factor is used in an individual multiplication with the used weights.

$$WN_t = \frac{1}{\sum_{j=1}^{3} MBU_{j,t} * W_j + (1 - MBU_{j,t}) * W\_CB_j}$$

$$where \sum_{j=1}^{3} W_j = 1; W_j \in [0,1]$$

Equation 3.1.2 Used Weights normalization

Another enhancement of the existing approaches from section 2 is the additional provision for application related weighting. In the automotive domain it is useful to group applications and commands into three different classes. Usually, these classes are not disjoint so it might be possible that an application could belong to more than one class. We use three classes: safety, security and comfort applications. Safety applications are relevant for safety tasks, e.g. to protect the humans inside of the car. Security applications protect the car and/or its usage against unauthorized access whereas comfort applications are related to comfort issues like the calibration of the air conditioner. In order to give an application the option to adjust the standard weighting of the different matching scores, new application dependent operands $A_{j,t}$ have been introduced that can be set to any value in [0, 1]. They depend on the respective class of the command and the additional information about environment and sensor properties (that is also included in $V_{j,t}$ and $F_{j,t}$). The $A_{j,t}$ are dynamically calculated by the application itself to achieve any application-dependent purposes. They have been added to the inner part of the sum so that the weighting of a specific biometric modality can additionally be decreased or increased by the application. Furthermore the binary operands of the ADF are now included in the inner part of the sum and are also available for all possible biometric modalities. This way it is possible that an application can request the usage of only a subset of biometric modalities available. The new final fusion concept is shown in Equation 3.1.3.

$$MS_{fus,t} = \sum_{j=1}^{3} B_{j,t} * A_{j,t} * \left[\left\{MBU_{j,t} * WN_t * W_j * MS_{j,t}\right\} + \left\{(1 - MBU_{j,t}) * V\_CB_{j,t} * F\_CB_{j,t} * WN_t * W\_CB_j * MS\_CB_{j,t}\right\}\right]$$

$$\sum_{j=1}^{3} W_j = 1; W_j \in [0,1]; V\_CB_{j,t} \in [0,1]; F\_CB_{j,t} \in [0,1]$$

Equation 3.1.3 Enhanced Strategy

3.2 Privacy Enhanced Fusion Strategy (PEFS)

With reference to Figure 1.1 a potential vulnerability of the system can be identified in the biometric database. If the needed templates are stored as original data like images or speech samples, a thief could steal them and misuse them for fake identifications and verifications, respectively. Additionally, the privacy of the driver and other enrolled occupants should be protected, e.g. in cases of re-selling or rental scenarios. Therefore the stored data should not contain any privacy related information. One possible way to achieve a privacy enhancement is to store information in a more compact form. For example, only features might be stored that have initially been extracted from the chosen biometric modalities. Another possibility is to store information in form of biometric hash values. For example, for handwriting there already exists an approach that makes use of such a biometric hash value [23]. Also for features obtained from a facial image [24] it is possible to generate a similar hash value. In the following we use the terms voice hash, face hash and body weight hash regarding the biometric modalities voice, face, and body weight that are used to calculate the biometric hash.

For different persons, the hash values of their feature vectors should be different in almost every element (intra class sensitivity), even if their feature vectors only vary in a small number of elements. But additionally, for the same person (approximately) the same hash vector should be generated each time (intra class stability). With such a hash function it is possible to enhance the privacy of the biometric system. Two different modes could be chosen for the decision: In one mode (distance mode) the distance between a stored hash and the hash of the current data is calculated using different distance measures [3]. The other mode (verification mode) either retrieves 1 for identical hash codes or 0 for different hash values and thus requires stable intra-class hash values. As we recommend the use of a hash method in verification mode, the result of a comparison is no longer a matching score like we used in Equation 3.1.3. Therefore, instead of the $MS_{j,t}$ the result of the verification mode ($HV_{j,t}$) is used and interpreted as a matching score. This means, if the hash verification yields a result of zero, the priority of the partial result, due to the environment and sensor conditions, has no effect on the final result (now $HV_{fus,t}$ instead of $MS_{fus,t}$). For example if the environment is optimal, but the person is not in an optimal position, the verification might lead to a result of 0.

3.3 Comparison of EFS with the existing approaches

To compare different biometric authentication systems, several tests and fusion strategies like [16] have been developed and performed. In this section we perform a comparison of the ADF concept, the SFSF strategy and our new approaches EFS and PEFS. Since (unlike [16]) we do not yet have real data for our tests, we perform the comparison within a simulation using theoretical values. This way we show how the approaches react on several input combinations in different situations. We show how the comfort and security are affected. This is not a complete list of possible situations but we use some extreme choices that illustrate the general differences.

First we discuss three different scenarios that affect the mentioned aspects safety, security and comfort. They also show, that the inclusion of additional information about environment and sensor properties could lead to an improvement of some of these aspects.

Scenario 1 – Comfort and Safety related

In this scenario a speech command is used to adjust the seat for general comfort purposes by ensuring the safety of the car and its occupants at the same time. For example if the driver has to use the brake pedal, a seat movement could hinder him to do it with an appropriate pressure which might lead to an accident that could have been prevented otherwise. Therefore the safety is affected since people can get injured and material can get damaged by a seat movement in an inappropriate situation.

Scenario 2 – Comfort and Security related

The second scenario is the voice-operated multimedia system. Different classes of movies exists e.g. child (level 0), teen (level 1) and adult (level 2). All occupants registered in the biometric system of the car also belong to one class and are only able to start movies of their own or a lower class level. By using voice commands it is comfortable to control the system, but also the security is affected. If a person cannot be recognized well enough, only movies of the lowest level are unlocked.

Scenario 3 – Security related

Third, the following scenario concerns security. Imagine a thief that forges different sensors in a car to get access to the internal systems in order to unlock the car and be able to drive. When the owner recognizes the loss of his car, he could send additional information to it, e.g. by using future GSM communication infrastructure. With this additional knowledge the system is able to lock all comfort systems so it might only grant a minimum amount of functionality in a restricted state. This might be necessary because even the current driver (who is probably the thief), should be able to guide and stop the car safely without endangering other occupants or bystanders. For instance it might be possible to limit the driving speed to 6 mph like it is considered in [22].

In the mentioned scenarios both ADF and SFSF do not take care of the situation. They neither take care of the possibly dangerous situation of scenario 1 nor of the additional information provided in scenario 2 and 3. The EFS now is able to include the mentioned parameters and adaptively adjust the impact of the matching score of a biometric characteristic on the overall decision.

Among other analyses, during the following comparison we show how the new approach solves the problem mentioned in section 2. Our simulation will start in a basic scenario assuming that everything is fine and all sensors are working normally. The environment is in a state where it does not have negative effects on the sensors. Table 3.3.1 shows these basic sensor data assumed and the initial additional factors for a discrete time t=0. As mentioned before, the comparison is based on a simulated evaluation and does not rely on real data. We define a matching score *MS* of 1.0 as a full match, i.e. the best value that the employed biometric subsystem or the fusion can provide. Further we imply that the body weight is a biometric modality that has a less discriminatory power. Thus, the *MS* for camera and microphone can be expected to be more precise than the one for the body weight sensor. Also the selected *MS*

weights will be different for the different biometric modalities involved. For a better comparison, the application operands $A_{j,t}$ are set to 1 because these operands do not have an influence on the ADF and SFSF concepts. Additionally, for SFSF we assume a FAR of 0.1 for the biometric subsystems depending on camera and microphone (i.e. 10% of the attackers are accepted falsely while 90% are rejected correctly). The subsystem depending on the weight is assumed to have a FAR of 0.7 (i.e. 30% of the attackers will be rejected correctly).

Table 3.3.1 Basic Assumption: Simulation Data

Sensor	**cam**	**micro**	**body weight**	
$MS_{i,0}$ *matching score*	0.9	0.7	0.4	
$HV_{i,0}$ *verification hash code*	1	1	0	
$V_{j,0}$ *confidence factor*	1	1	1	sum: 3
W_j *weight*	0.5	0.36	0.14	sum: 1
$B_{(ADF)j,0}$ *binary operator*	1	n/a	1	
$B_{(EFS)j,0}$ *binary operator*	1	1	1	
$A_{j,0}$ *application operand*	1	1	1	
$F_{j,0}$ *functionality factor*	1	1	1	
FAR_i *selected false acceptance rate*	0.1	0.1	0.7	
Compensational Sensor	**steering properties sensors**	**acceleration behavior sensors**	**brake pedal**	
$MS\_CB_{j,0}$ *matching score compensational biometric modalities*	0.3	0.1	0.2	
$HV\_CB_{j,0}$ *verification hash code compensational biometric modalities*	1	0	0	
$V\_CB_{j,0}$ *confidence factor compensational biometric modalities*	1	1	1	sum: 3
W_CB_j *adjusted weight compensational biometric modalities*	0.11	0.11	0.11	sum: 0.33
$F\_CB_{j,0}$ *functionality factor compensational biometric modalities*	1	1	1	

As Table 3.3.2 shows, no normalization has to be done for the Adaptive Dynamic Fusion (ADF), because either the sum of all $V_{j,0}*W_j$ is 1 or, respectively, the sum of the $V_{j,0}$ is 3, which depends on the respective normalization function. In this case, the calculated fused matching score is simply the sum of the products of the initial matching scores and the corresponding weights for each individual modality involved.

Table 3.3.2 Basic Assumption: ADF Results

Sensor	**cam**	**micro**	**body weight**	
$V_{j,0}*W_j$	0.5	0.36	0.14	sum: 1
$N_{(ADF)j,0}$ *normalization factor*				1
$MS_{fus,0}$ *fused matching score*	0.45	0.252	0.056	sum: 0.758

The additional data and results for the Simplified Face Speech Fusion (SFSF) are shown in Table 3.3.3.

Table 3.3.3 Basic Assumption: SFSF Results

Sensor	**cam**	**micro**	**body weight**	
f_i *through software known failure*	0.9	0.9	0.3	
$d_{j,0}=F_{j,0}$ *current disturbance*	1	1	1	
$g_{j,0}$ *unnormalized weights*	0.9	0.9	0.3	sum: 2.1
$w_{j,0}$ *normalized weights*	0.429	0.429	0.143	sum: 1.001
$MS_{fus,0}$ *fused matching score*	0.3861	0.3003	0.0527	sum: 0.7391

As it can be seen, the fused matching score is nearly the same than the one from ADF concept.

Table 3.3.4 shows the resulting matching score of the Enhanced Fusion Strategy (EFS) and the Privacy Enhanced Fusion Strategy (PEFS).

Table 3.3.4 Basic Assumption: EFS and PEFS Results

Sensor	**cam**	**micro**	**body weight**	
MBU_0 *main biometric usage*	1			
WN_0 *used weights normalization*	1			
$MS_{fus,0}$ *fused matching score*	0.45	0.252	0.056	sum: 0.758
$HV_{fus,0}$ *fused verification hash code*	0.5	0.36	0	sum: 0.86

ADF and EFS give the same fused matching score because the $B_{j,0}$ in Adaptive Dynamic Fusion, the corresponding $F_{j,0}$ in the Enhanced Fusion Strategy as well as the corresponding $V_{j,0}$ are 1. Simplified, the resulting $MS_{fus,0}$ therefore equals the sum of the weighted input matching scores. The different calculated score of the Simplified Face Speech Fusion is caused by the additional use of the FAR and the different kind of the calculation of the weights. The results are still similar. The best result is given by the PEFS but in our theoretical example it is heavily dependent on the biometric modalities that are used and their respective weights. If the *HV* is 0 for the cam and 1 for the body weight, the result is the sum of the lowest two weights, which is 0.5. The results of the other strategies are not affected, so PEFS give the lowest result in that case.

Now we show what happens if one of the binary operators for ADF, $B_{j,0}$ is set to zero (e.g. $B_{1,0}$ because of a camera malfunction). Accordingly, we also set the corresponding values for the concepts SFSF ($d_{1,0}$) and EFS/PEFS ($F_{1,0}$) to this value. All other parameters are the same as in Table 3.3.1.

Obviously, as mentioned in section 2, the fused matching score for ADF will be zero because it is a product of factors one of which is zero. Table 3.3.5 and 3.3.6 show how this affects the Simplified Face Speech Fusion and the enhanced concepts.

Table 3.3.5 Modified Assumption (Failure of First Main Biometric): SFSF Results

Sensor	**cam**	**micro**	**body weight**	
$d_{j,0}=F_{j,0}$ *current disturbance*	0	1	1	
$g_{j,0}$ unnormalized weights	0	0.9	0.3	sum: 1.2
$w_{j,0}$ *normalized weights*	0	0.75	0.25	sum: 1.0
$MS_{fus,0}$ *fused matching scores*	0	0.525	0.1	sum: 0.625

The matching score $MS_{1,0}$ of the malfunctioning sensor becomes zero and only the two other matching scores are included in the fusion and the resulting fused matching score $MS_{fus,0}$. The weights corresponding to the two properly working sensors are adjusted. They are increased but their ratio is not changed. Therefore, a lower result can be expected compared to Table 3.3.3.

Table 3.3.6 Modified Assumption (Failure of First Main Biometric): EFS and PEFS Results

Sensor	**cam**	**micro**	**body weight**	
$\boldsymbol{MBU_{j,0}}$ *main biometric usage*	0	1	1	
1-$\boldsymbol{MBU_{j,0}}$ *compensational biometric usage*	1	0	0	
$\boldsymbol{WN_0}$ *used weights normalization*	1.639			
$\boldsymbol{MS_{fus,0}}$ *fused matching score*	0.054	0.413	0.092	sum: 0.559
$\boldsymbol{HV_{fus,0}}$ *fused verification hash code*	0.18	0.59	0	sum: 0.77

The fused matching score for EFS and the fused verification hash code for PEFS are also decreasing as it is shown in Table 3.3.6. However, this is also obvious since the discriminatory power of the *CB* is not expected to be as good as the discriminatory power of the corresponding main biometric modalities. The matching score of the used *CB* is nearly one third of the main matching score, as shown in Table 3.3.1. Additionally, the weights of the *CB* are adjusted in a way that they are in relation to the main biometric modalities. This way the SFSF, EFS and PEFS concepts still provide a more or less good result allowing a decision to be done on. The ADF concept would have rejected any person. The best result is still given by the PEFS.

The following theoretical evaluation we performed considers the worst case scenario, which is a potential simultaneous failure of all primary sensors. In this case none of the main biometric modalities could be used. ADF and SFSF would result in a fused matching score of 0. The EFS and PEFS would use the additional soft biometric modalities (*CB*). The resulting score is visualized in Table 3.3.7.

The fused matching score should be the average (0.2) of the different $MS\_CB_{j,0}$ as all are included with the same weight W_CB_j. Compared to ADF and SFSF this still has advantages: even if the main biometric modalities fail, subsystems that use authentications can still perform. As the discriminatory power is probably massively decreased, systems which are mainly related to the security domain will eventually do not grant access to their functionality, but comfort related applications still might provide the requested services. Also in this scenario the PEFS give the best result for comfort related applications.

Table 3.3.7 Third Assumption (Failure of all Main Biometrics): EFS and PEFS Results

Sensor	**cam**	**micro**	**body weight**	
$\boldsymbol{MBU_{j,0}}$ *main biometric usage*	0	0	0	
1-$\boldsymbol{MBU_{j,0}}$ *compensational biometric usage*	1	1	1	
$\boldsymbol{WN_0}$ *used weights normalization*	3.03			
$\boldsymbol{MS_{fus,0}}$ *fused matching score*	0.1	0.03	0.07	sum: 0.2
$\boldsymbol{HV_{fus,0}}$ *fused verification hash codes*	0.33	0	0	sum: 0.33

Table 3.3.8 shows the impact of the application operands $A_{j,0}$ regarding scenario 2 in section 3.3 using the basic data from Table 3.3.1 and a modified $V_{1,0}$ of 0.5. As the classification (level) of the desired movie increases, the application lowers the influence of the respective modality (here: face), also including its compensational biometric.

Table 3.3.8 Impact of Application Operands (EFS & PEFS)

Sensor	**cam**	**micro**	**body weight**	
$\boldsymbol{MBU_{j,0}}$ *main biometric usage*	0.5	1	1	
1-$\boldsymbol{MBU_{j,0}}$ *compensational biometric usage*	0.5	1	1	
$\boldsymbol{WN_0}$ *used weights normalization*	1.242			
$\boldsymbol{A_{j,0}}$ *application operand (level 0)*	1	1	1	
$\boldsymbol{MS_{fus,0}}$ *fused matching score*	0.3	0.313	0.07	sum: 0.683
$\boldsymbol{HV_{fus,0}}$ *fused verification hash codes*	0.379	0.447	0	sum: 0.826
$\boldsymbol{A_{j,0}}$ *application operand (level 1)*	0.5	1	1	
$\boldsymbol{MS_{fus,0}}$ *fused matching score*	0.15	0.313	0.07	sum: 0.533
$\boldsymbol{HV_{fus,0}}$ *fused verification hash codes*	0.19	0.447	0	sum: 0.637
$\boldsymbol{A_{j,0}}$ *application operand (level 2)*	0	1	1	
$\boldsymbol{MS_{fus,0}}$ *fused matching score*	0	0.313	0.07	sum: 0.383
$\boldsymbol{HV_{fus,0}}$ *fused verification hash codes*	0	0.447	0	sum: 0.447

In this scenario the application operand $A_{j,0}$ is useful for purposes like improving the access protection: Because of disturbing lighting conditions, a child trying to start a movie of level 2 might accidentally be recognized as a person registered for this level even while the $V_{j,0}$ includes information regarding the environment. The $A_{j,0}$ additionally lowers the matching score even more, so that the child finally is rejected, anyway. However, this can also decrease the comfort because persons of level 2 could also be rejected this way.

3.4 Security and Comfort Aspects

In this section we give some comments what effects the discussed approaches can have on security and comfort. As it was shown in Tables 3.3.2 to 3.3.4, the different approaches (ADF, SFSF, EFS and PEFS) give nearly the same fused matching score if the system is working at its optimal state with a confidence and functionality factors of 1 for all sensors. In that state neither an improvement nor degradation can be recognized. The original concept was intended to enforce the security by neglecting the comfort. There is no fall-back solution to use the system in any failure case. This has some advantage in case of attacks: An invader might try to fool the system by deactivating a certain sensor. If a binary operator corresponding to this sensor forces the final matching score to zero, the system will not provide any functionality to him.

The Simplified Face Speech Fusion offers a good chance to compensate the failure of sensors. But it does not include explicit information about environmental changes through additional sensors as the functionality of the sensors involved is included.

With the Enhanced Fusion Strategy we can achieve a better comfort compared to the ADF by lowering security. In relation to the SFSF

the EFS includes additional explicit information about the environment. Furthermore it is possible to give a substitution of malfunctioning sensors. In the case that also these sensors fail, the system is not able to authenticate a user; however, the ADF and SFSF approaches also fail if main biometric sensors fail.

4. Summary and Future Work

The proposed modification of the Adaptive Dynamic Fusion includes the possibility of the Simplified Face Speech Fusion to also use information about the current functionality of each sensor. By the inclusion of additional soft biometric modalities from additional sensors, the calculation has been improved. By the use of such compensational biometrics, the system is now able to still provide a (decreased) matching score if main biometric modalities fail. This way, systems which are mainly related to the comfort domain could still provide user dependent functions.

For this paper, we used an exemplary assignment of three chosen compensational biometrics to the main biometrics. As future work, the influence of the choice of different compensational biometrics as well as their assignment to the main biometrics will have to be investigated closer.

Beyond our suggested new fusion model on matching score level, the selection of an appropriate threshold remains an important question. Future work will study adaptive thresholding techniques under consideration of actual confidence factors and availability of sensors in specific security and comfort requirements.

With respect to the features of the fusion strategies, further options could be considered in future. One potential enhancement could be to only respect a reduced number of input modalities when applicable. For example, in certain scenarios a reduced set of biometric modalities might be chosen that leads to reduction of power consumption at the cost of a less dependable classification. Such a setup might be sufficient for scenarios without important security requirements. Biometric modalities giving a strong classification but also lead to high power consumption could especially be used for high secure scenarios. For such purposes, a classification of the biometric modalities useable for different scenarios in the automotive domain would be helpful. We also consider to allow an application to additionally decide whether or not the compensational biometric modalities may be used. In case the application does not permit the use of CB, the system will only use the main biometric modalities for the affected calculation, which might then be based on input data of bad quality or from failing sensors.

Based on our first theoretical evaluation presented in this paper we are also planning to calculate fused matching scores for all four concepts on real data. Test data will be collected in a laboratory environment within optimal conditions and under different real-life conditions in vehicles (such as noise or different acoustic and light situations).

Acknowledgements

The work on biometric hashes with regard to verification and reproducibility is partly supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project WritingPrint). The automotive-related work has been supported in part by the European Commission through the EFRE Programme "Competence in Mobility" (COMO) under Contract No. C(2007)5254.

European Commission
European Regional Development Fund
INVESTING IN YOUR FUTURE

5. REFERENCES

[1] 2006. Biometrics Frequently Asked Questions, NSTC, Committee on Technology, Committee on Homeland and National Security, Subcommittee on Biometrics, http://www.biometrics.gov/Documents/FAQ.pdf

[2] Ross, A.A., Nandakumar K., Jain, A.K. 2006. Handbook of Multibiometrics, Springer, ISBN-13: 978-0-387-22296-7

[3] Scheidat, T., Vielhauer, C. und Dittmann, J. 2005. Distance-Level Fusion Strategies for Online Signature Verification. Otto-von-Guericke University Magdeburg, Germany.

[4] 2006. Biometrics Glossary, NSTC, Committee on Technology, Committee on Homeland and National Security, Subcommittee on Biometrics, http://www.biometricscatalog.org/biometrics/GlossaryDec2005.pdf

[5] Nandakumar, K., Chen, Y., Jain, A.K. and Dass, S.C. 2006, Quality-based Score Level Fusion in Multibiometric Systems, http://www.stt.msu.edu/~sdass/papers/NandakumaretaI_QualityFusion_ICPR2006.pdf

[6] Makrushin, A., Dittmann, J., Kiltz, S. and Hoppe, T. 2008. Exemplarische Mensch-Maschine-Interaktionsszenarien und deren Komfort-, Safety- und Security-Implikationen am Beispiel von Gesicht und Sprache. Alkassar, Siekmann (Hrsg.): Sicherheit 2008; "Sicherheit - Schutz und Zuverlässigkeit"; Beiträge der 4. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI), 2.-4. April 2008 in Saarbrücken; pp. 315 - 327, ISBN 978-3-88579-222-2, 2008

[7] Vargyas, C. and Crouch R. 2004, European Security, An analysis of biometric technology for automotive security applications http://www.sbd.co.uk/assets/1325_Biometrics_report_sample.pdf

[8] Sterbak R. 2005, My Car Understands Me, Pictures of the Future | Fall 2005, https://www.ct.siemens.com/en/business/speech/anlagen/siemens_my_car_understands_me.pdf

[9] Giordano, P. Fingerprint based bio-starter and bio-access, Gerardo lovane, D.I.I.M.A. - University of Slerno, http://arxiv.org/pdf/cs.CV/0308034.pdf

[10] Marklund, P.-O. and Nilsson, L. 2003. Optimization of airbag inflation parameters for the minimization of out of position occupant injury, Computational Mechanics 31, Springer-Verlag, DOI 10.1007/800466-003-0457-9

[11] Minister of Public Works and Government Services 2004. Air Bag Deactivation What You Need to Know to Make an Informed Decision, http://www.tc.gc.ca/roadsafety/tp/tp13178/pdf/tp13178e.pdf

[12] Johns, M., Tucker, A. and Chapman R. 2006. Monitoring the Drowsiness of Drivers: A New Method Based on the Velocity of Eyelid Movements, Sleep Diagnostics Pty Ltd, Richmond, Melbourne, Australia http://www.optalert.com/images/research_46ca27b4efb34.pdf

[13] Blaschke, T., Lipaczewski, M., Stengel, M. and Winter, M. 2008. Multi-Modal Data Analysis Project: Biometrics, Team –

Automotive, Technical Report, Otto-von-Guericke University of Magdeburg, FIN, ITI , winter semester 07/08

[14] McDowall R.D. 2000 Biometrics: The Password You'll Never Forget, Pharmaceutical File, http://www.21cfrpart11.com/files/library/compliance/lcgc10_00.pdf

[15] Erdoğan, H., Özyağcı, A.N., Eskil, T. et. al. 2005. Experiments on Decision Fusion for Driver Recognition, Faculty of Engeneering and Natural Sciences, Sabanci University İstanbul

[16] BioSecure Multimodal Evaluation Campaign 2007, http://biometrics.it-sudparis.eu/BMEC2007/index.php

[17] Erzin, E., Yücel Yemez, Y. and Tekalp, A.M. 2005. Joint Audio-Video Processing For Robust Biometric Speaker Identification In Car; In: DSP for In-Vehicle and Mobile Systems, pp. 237-256, Springer, ISBN 0-387-22978-7

[18] Chu, S.M., Yeung, M., Liang, L. and Liu X. 2003. Environment-adaptive multi-channel Biometrics, in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003.

[19] Chetty, G. and Wagner, M., 2006. Video to the Rescue. School of Information Sciences and Engineering, University of Canberra, Australia. http://portal.acm.org/citation.cfm?id=1273403

[20] Iwano, K., Seki, T. and Furui, S. 2005. Noise Robust Speech Recognition Using Prosodic Information; In: DSP for In-Vehicle and Mobile Systems, pp. 139-152, Springer, ISBN 0-387-22978-7

[21] Igarashi, K., Takeda, K., Itakura, F. and Abut, H. 2005. Is Our Driving Behavior Unique?; In: DSP for In-Vehicle and Mobile Systems, pp. 257-274, Springer, ISBN 0-387-22978-7

[22] Borchers, D. 2008. Mit PKI gegen den Autoklau http://www.heise.de/security/Mit-PKI-gegen-den-Autoklau--/news/meldung/104593

[23] Vielhauer, C. 2006. Biometric User Authentication For IT Security: From Fundamentals to Handwriting, Springer

[24] Gu, H., Su, G. and Du, C. 2003. Feature Points Extraction from Faces, Image and Vision Computing NZ, Palmerston North, November 2003