IET Biometrics

Research Article

Analysis of results of large-scale multimodal biometric identity verification experiment

ISSN 2047-4938
Received on 12th February 2018
Revised 23rd July 2018
Accepted on 20th August 2018
E-First on 31st October 2018
doi: 10.1049/iet-bmt.2018.5030
www.ietdl.org

Andrzej Czyżewski[1], Piotr Hoffmann[1] ✉, Piotr Szczuko[1], Adam Kurowski[1], Michał Lech[1], Maciej Szczodrak[1]

[1]Multimedia Systems Department, Gdańsk University of Technology, Faculty of Electronics, Telecommunication and Informatics, ul. Narutowicza 11/12, Gdańsk, Poland

✉ *E-mail: phoff@multimed.org*

Abstract: An analysis of a large set of biometric data obtained during the enrolment and the verification phase in an experimental biometric system installed in bank branches is presented. Subjective opinions of bank clients and of bank tellers were also surveyed concerning the studied biometric methods in order to discover and to explore relations emerging from the obtained multimodal dataset. First, data acquisition and identity verification methods are described in this study. Then, relationships between ratios of successful and failed verifications between pairs, triplets, and quartets of biometric modalities are studied. An analysis of the sentiment of clients and of banking tellers related to each identity verification attempt was performed based on linguistic methods. The data mining process is described, based on the rough sets methodology, aimed at deriving rules pertaining to consecutive identity verification attempts.

1 Introduction

Multimodality in biometrics can be considered as applying multiple sensors for the same trait, or collecting multiple biometric traits, acquiring multiple units of the same nature (etc. prints of many fingers), taking multiple data snapshots (scans, photos), or using multiple approaches to feature extraction [1]. Moreover, in multimodal biometrics one must consider the most appropriate approach to data fusion, as various levels are possible: (i) pre-classification fusion based on raw data or features, and (ii) post-classification fusion, e.g. combining match scores from different modalities into a single measure or made by processing decisions provided by independent classifiers. Each approach influences the final accuracy of the system [1, 2]. Various multimodal databases and fusion strategies were created and applied in experiments. Many significant databases are not publicly available, but the efforts are documented and can be referred to as a baseline for accuracy [3]. Among those the largest ones are: BIOMET (2003) with faces, voice, fingerprint, hand and signatures of 327 individuals [4], MCYT (2003) with fingerprints and signatures of 330 participants [5], BIOSEC (2007) with face images, speech, fingerprint scans by three different sensors, and iris from 250 subjects [6], BiosecurID (2010) with face, fingerprint, voice, iris, written signature, handwriting, keystroking, palmprint, and hand geometry of 400 individuals [7], MMU-GASPFA database (2013) containing video recordings of gait with depth information acquired by Kinect camera, and single accelerometer located on the hips, speech, and face image of 82 participants [8]. Publicly available multimodal dataset is NIST BSSR1 (2004) containing face and fingerprint of 517 individuals [9], and one yet to be made available database of XJTU (2017) with fingerprints, face, iris scans, and voice recordings from 102 persons [10]. The largest scale effort in multimodal biometry was until now an artificially made combination of separate databases, with an assumption that FERET face images [11] and authors' fingerprint database have the same origin, including records of 972 individuals [12].

No effort was made until now on assessing the user confidence, ease of use, comfort, and acceptance for biometric technologies, and effects of repeated interaction with the verification system, including an improvement in verification rates as the independent user operates the system in uncontrolled conditions. Meanwhile, our research addresses these problems.

In our work, the research was conducted using 100 multimodal biometric stations engineered by the authors [13–15] and distributed in real conditions, equipped with sensors for data acquisition and with processing software enabling registration, and then verification of identity employing four modalities: digital handwritten signature, voice, face image, and hand vein scan. One more modality – face contour – was also registered, hence data were collected during the study, to develop the modality based on the side profile of the face in the future.

First, the pilot study with participation of 126 persons was conducted [13], proving the methodology was appropriate for multimodal biometry identity verification. Later on, the experimental work was extended to 7166 individuals [14, 15], allowing to validate developed methods, extensively. This paper presents the final stage of the work with participation of over 10,000 individuals, allowing for a significant improvement of hitherto made attempt results. As a detailed description of individual biometric methods was published previously [13–15], thus here we focus on the final analysis employing linguistic method for sentiment assessment, and rough set (RS) modelling to predict users' future performance.

In the presented research, a method of sentiment analysis was employed to assess participant comfort of use of tested modalities and to identify main advantages or problems associated with the tested hardware and software. This approach to usability evaluation is a common analysis tool used previously for categorisation of patients' opinions on quality of healthcare in hospitals [16], quality of hotel services assessment [17], or evaluation of political moods during a presidential election [18].

The sentiment analysis is based on statistical analysis of frequency of words and groups of words which is used to identify the most common phrases in the corpus. Such a method is simple to implement and it is efficient in identifying keywords, e.g. most commonly reported problems in the written form. Moreover, the analysis does not require large training dataset as opposed to solutions based on machine learning [16]. Therefore, it was directly applied to data obtained in the experiment presented in this paper. The only prerequisite was an implementation of a language model, taking into account inflection of processed words. Without the model, the sentiment analysis algorithm could not distinguish between two forms of the same word. Meanwhile, our approach is capable of processing all the input provided in the written form in Polish language, without any need for prior collection and processing of training data.

Novelty of our approach lies also in collecting subjective data describing many factors of user experience with the biometric technologies, along with objective performance and accuracy metrics. Linguistic sentiment analysis and RS based modelling allow for assessing new key features of multimodal biometry, as well as for predicting future users' performance.

The reminder of the paper is organised as follows. Section 2 introduces methods for collection of multimodal biometric data, including a description of each modality, database system, user questionnaires, sentiment analysis, and RS modelling. Section 3 documents data collection procedure, and Section 4 presents the results. Section 5 contains conclusions and discussion of the presented approaches.

2 Data collection methods

2.1 Digital handwritten signature

The handwritten signature was acquired using the special wireless biometric pen, developed within the scope of the project. The biometric pen is equipped with a three-axis accelerometer, a three-axis gyroscope, and a surface pressure sensor, a Bluetooth low energy transmitter/receiver, and a rechargeable battery. The signature is being put on a touch screen providing the visual feedback for the user. The similarity measures for verification employ the dynamic time warping (DTW) algorithm applied for three-axis acceleration and angular position signals. The method is based on the assumption that time-domain functions of an acceleration and angular position of two arbitrary authentic signatures entail less warping than for authentic and forged signature. Therefore, the information derived from the DTW method, further used for the verification process, represents the convergence of a diagonal path for the identical signals and an optimal cost path for two allegedly similar signals in the accumulated cost matrix. The cost $\gamma_{i,j}$ in each cell (i, j) of the DTW accumulated cost matrix is being calculated in a standard manner (1), in which the distance $d(f_i,\ g_j)$ between every sample of functions F and G, representing values of a particular parameter, such as acceleration or angular position, of two arbitrary signatures of lengths m and n, is interpreted as an absolute difference (1)–(4). A more detailed description of the method can be found in an earlier publication by the authors [13]

$$\gamma_{i,j} = d(f_i, g_j) + \min\left(\gamma_{i-1,j-1}; \gamma_{i-1}; \gamma_{i,j-1}\right) \tag{1}$$

$$d(f_i, g_j) = |f_i - g_i| \tag{2}$$

$$F = f_1, f_2, \ldots, f_i, \ldots, f_n \tag{3}$$

$$G = g_1, g_2, \ldots, g_j, \ldots, g_m \tag{4}$$

The applied DTW allows for typical variations between individual signatures of the user, still it is required that the user signatures have constantly the same content (e.g. full name, or initial and surname etc.)

2.2 Voice

A single microphone was used for registering the bank clients' voice at 44 kSa/s sampling rate with 16-bit resolution. It was assumed that the content of speech is not important as only voice timbre is parameterised. The mel-frequency cepstral coefficients are used for speech parametrisation. The features were extracted within 10 ms timebase, the length of the frame corresponded to 25 ms and 13 cepstral channels were employed. The acoustic feature vector was formed by combining zero-order mel-frequency cepstrum coefficients (MFCC) with delta and delta-delta features, which resulted in 39 features in total related to each 10 ms sample. The MFCCs denoted as c_i ($i = 1$–13) were calculated as

$$c_i = \sqrt{\frac{2}{N}\sum_{j=1}^{N} m_j \cos\left(\frac{\pi i}{N}(j - 0.5)\right)} \tag{5}$$

where m_j are the log filterbank amplitudes and N denotes the number of filterbank channels.

For the voice recognition, the Gaussian mixture model (GMM)/ universal background model (UBM) method is employed to verify the speaker's identity [19]. The Alize framework has been used as the speaker recognition back-end [20]. The UBM was trained on the recordings prepared in real bank branches environment, employing 84 participants. Additionally, a speech material from the MOBIO dataset [21] was utilised in order to increase UBM inner variance. Both stages: training of the person-specific voice model and the voice model verification require prior existence of the UBM. The voice verification process follows the rule as

$$\ln\frac{p(s \mid \lambda_{\mathrm{tgt}})}{p(s \mid \lambda_{\mathrm{bkg}})} \gtrless \theta \tag{6}$$

where s represents new voice sample to be verified against known voice model λ_{tgt}, λ_{bkg} is the background voice model, and θ stands for experimentally chosen value of the threshold set to 1.08 in our experiments. The background model remains unchanged during the registration of new users, as it was confirmed that the initial form of UBM assured correct performance even for very large group of individuals [13, 14].

2.3 Face image

Face image was acquired by RGB camera with resolution of 1920 × 1080 pixels at 30 frames/s rate. Extraction of facial image parameters was preceded by face detection in the scene [22], to focus only on the face location. The parameterisation procedure begins with 77 landmarks calculation on the detected face. Based on the landmarks, distinctive face regions are found, i.e. eyes, eyebrows, nose, mouth. Each fragment of the image representing those regions and the whole face image are parameterised using histogram of oriented gradients and local binary pattern features. In the last step, each parameter set calculated from the face image and its subregions are concatenated. Subsequently, the linear discriminant analysis is performed what results in 768-element feature vector [23]. A detailed description of the procedure was published in the pilot study [13]. An approach with time-adaptive, self-learning sliding models was also developed [15].

2.4 Time of flight (ToF) image

The ToF camera image was used as an input for an experimental biometric method exploiting facial depth data to verify the bank clients' identity. The SoftKinetic DS325 ToF sensor was employed, which delivers depth data with 320 × 240 resolution at 60 fps. From the face depth map the contour is retrieved using self-developed algorithm described earlier in the paper of Bratoszewski and Czyżewski [24], and then it can be used in the face verification process. However, at the time of data acquisition made in bank outlets, the contour modality processing was in an initial phase, thus a formal comparison with remaining methods was impossible. Nevertheless, the acquisition of contour biometric samples was performed by the bank clients in order to collect subjective opinions on this modality by appropriate questionnaires, allowing for an assessment of ergonomics, sentiment, and acceptance.

2.5 Hand vein

In contrast to the previous modalities, the registration of samples of the hand blood vessel distribution was conducted using a commercially available solution (Fujitsu PalmSecure [25]). The sensor analysed the spatial distribution of veins in the palm of an individual. As a result of different reflection of the near-infrared light by veins and by the tissue of human body it is possible to verify the identity of a person [25]. The solution used for the study served initially as a reference to other modalities developed for testing purposes, because it should ensure a very low level of false acceptance rate (FAR) and false rejection rate (FRR) errors, as it is declared by its manufacturer. However, it turned out that some external factors such as user unfamiliarity with the technology, imprecise hand positioning, and lowered sensor cleanliness can affect the final effectiveness of the solution by lowering the FAR and FRR measures substantially. In Section 4, the results obtained in the uncontrolled test conditions are presented.

2.6 Database system

The biometric data were registered in the dedicated database. A relational database was built employing the SQL 2012 management system. Each of the individual modality has its own database subsystem containing individual biometric samples. In the database system, a key ‘BioID’ connects all individual databases. The BioID number uniquely identifies a person (a bank client). Therefore, the number of BioIDs generated in the biometric system is interpreted as a number of persons participating in the study.

After registering the biometric samples, the system suggests filling electronic surveys by both the client and the banking teller.

2.7 Subjective data collection

One of this work purposes was data mining and modelling of relations occurring between objective and subjective features. Therefore, the users’ and tellers’ experience and opinions on voice, signature, and face image modalities were characterised by subjective features collected by means of the electronic survey.

Questions for the clients and answer options were defined as follows (question, feature short name, and possible values of answers):

- ‘How fast was the biometric samples registration?’: *uFast* = {1 (very slow), 2 (slow), 3 (average), 4 (fast), 5 (very fast)}
- ‘Was voice/signature/face registration easy and intuitive?’: *uEasyVoi*, *uEasySig*, *uEasyFac* = {1 (definitely not), 2 (no), 3 (yes), 4 (definitely yes)}
- ‘Was the signature/voice/face biometry a reliable and convenient way of verification?’: *uReliVoi, uReliSig, uReliFac* = {1 (too complex), 2 (complex), 3 (average), 4 (rather straightforward), 5 (very convenient), 6 (hard to judge)}
- ‘Are you willing to use your biometric traits during each banking activity?’: *uWill* = {1 (definitely not), 2 (no), 3 (yes), 4 (definitely yes)}
- ‘Was the biometry registration environment private?’: *uPriv* = {1 (definitely not), 2 (no), 3 (yes), 4 (definitely yes)}
- ‘Will biometry increase safety of banking operations?’: *uSecu* = {1 (definitely not), 2 (no), 3 (yes), 4 (definitely yes)}
- ‘Were you afraid about registering your biometric traits in the biometric database?’: *uAfra* = {1 (definitely not), 2 (no), 3 (yes), 4 (definitely yes)}
- Other remarks or opinion in the written form can be provided.

Questions and predefined answers of the teller were as follows:

- ‘How long the registration process lasted in minutes?’: *cFast* = {number of minutes}
- ‘Was your assistance required during the registration?’: *cHelp* = {1 (definitely not), 2 (no), 3 (yes), 4 (definitely yes)}
- ‘Was voice/signature/face registration easy for the client?’: *cHardVoi, cHardSig, cHardFac* = {1 (definitely not), 2 (no), 3 (yes), 4 (definitely yes)}
- ‘Was voice/signature/face registration by the client hard and cumbersome for you or requiring assistance?’: *cCumberVoi, cCumberSig, cCumberFac* = {1 (definitely not), 2 (no), 3 (yes), 4 (definitely yes)}
- Values of *sentiment* = *negative*, *positive* obtained from the semantic analysis of optional written comments (Section 2.8).
- Time elapsed from first registration to attempt number *N*: *timeN* = {hours}
- Other remarks or opinion in the written form can be provided.

In the RS methodology, the respective subsets of above lists were used as features allowing for a prediction of validation accuracy.

2.8 Sentiment analysis methods

A dedicated method for the assessment of text input content had to be developed in order to perform an automated analysis of all data gathered by surveys collected from clients and tellers. As it was shown in Section 2.7, data contained in both surveys were twofold:

i. numeric integer values representing the choices made by a client or by a teller while answering a multiple-choice question,
ii. textual, related to descriptive tellers’ remarks related to the process of data registration.

The analysis of the first data type is straightforward, as answers are gathered in the form of an array of numbers. However, the other data type needs a more sophisticated approach to the analysis, in order to facilitate a comparison of results with the rest of the collected feedback.

A semantic analysis algorithm was developed in order to estimate a sentiment connected with each descriptive answer. The approach is based on a detection of key words and phrases which are, in turn, associated with positive or negative sentiments. Such an approach is relatively easy to implement, but it is prone to errors associated with the so-called negation phrase phenomenon [26, 27].

Each text sample can be split into groups of non-whitespace characters which are compared to definitions of words in a dictionary. The dictionary should be formed in accordance with the term frequency-inverse document frequency (TFIDF) score of each word. TFIDF allows for an assessment of significance of each word in the context of its frequency of its occurrence in all analysed text excerpts, as well as with regard to occurrence frequency in the group of analysed surveys [28, 29]. The measure is calculated according to the following formula:

$$\mathrm{TFIDF(word) = TF(word)/log(}N\mathrm{/DF(word))} \quad (7)$$

where DF(word) denotes term frequency of a word – document frequency and N is the total number of analysed surveys. In this case, a single survey is referred to as a document. For the purpose of this paper, 775 definitions of words consisting of name of word, its flexion, assignment to a particular part of speech and associated with one of three possible sentiments (*positive, neutral, negative*) were prepared.

A single word analysis cannot reflect the sentiment being the result of negation. Therefore, a dictionary of phrases consisting of a group of word-representing tokens was also introduced. Moreover, there is a possibility that a whole phrase reflects an extreme sentiment despite the fact that each of component words may have neutral sentiment. A dictionary of 182 phrases was prepared. In order to find out which phrases are occurring in analysed texts, an analysis of most significant groups of words was performed. For the purpose of our research an analysis of sequences of the length of 2 up to 5 words turned out to be sufficient to find repeating sequences (for longer sequences no repetitions were found).

2.9 User progress metric

To perform uniform comparison of collected multimodal biometry data, first the similarity between the registered template and the nth verification sample was expressed in a scale of 0–100, separately for each BioID. The similarity measure was derived as a max–min normalised distance [2] in the set of all samples for the particular BioID and the biometric trait, scaled by a factor 100

$$similarity_n = 100 \times \frac{d(sample_n, template) - \min}{\max - \min} \quad (8)$$

where $d()$ is an Euclidean distance between two n-dimensional descriptors of the sample, e.g. between descriptor of a sample face being verified and previously stored template face, while constants min, max are calculated by determining minimal and maximal distance $d()$ in a set of all samples for the particular BioID and the biometric trait. The same approach is performed for distances between the biometric template and the verification attempt in the domain of descriptors of face, voice, and signature. Consequently, capability for a uniform comparison is achieved.

The absolute value of *similarity* of template and verification sample in authors' opinion depends on many unobservable variables, such as timbre or harshness of voice caused by emotions or sore throat, complexity of signature, lighting conditions for face image, and others. Instead, a relative similarity increases or decreases over time were calculated from verified samples, expressing personal progress in achieving improvement in verification rates. The similarity value in nth verification attempt was always related to the first verification sample $similarity_1$, thus the progress was measured as

$$progress_n = similarity_n\text{-}similarity_1 \quad (9)$$

where $n = 2–6$ is the verification attempt number. A $progress_n > 0$ would indicate a verification improvement comparing to the first verification attempt.

The *progress* is expressed in real values. Meanwhile, the RS method is dedicated to the classification of objects into discrete categories, only. Therefore, continuous domains have to be discretised prior the analysis. The discretisation was performed twofold, employing:

- median model (MM) – discretised into two ranges based on the median value in all values of $progress_n$ registered in the attempt number n in the set of all BioIDs,
- quartile model (QM) – discretised into four quartile ranges in all values of $progress_n$ registered in the attempt number n in the set of all BioIDs.

2.10 RS modelling method

The RS method is proposed in this paper as a tool for modelling dependency between subjective responses collected from surveys, the time elapsed from the first registration to the verification attempt, and the outcome objective measures of verification accuracy. The value of accuracy is divided into sets and approximated by applying the RS methodology. The analysis based on RS allows for finding patterns in data [30, 31], hence the verification accuracy prediction will be made based on the model derived from data.

The RS method consists in determining the set of features, called reduct, required to distinguish objects with different decisions, correctly. Based on the reduct and on training samples, the decision rules are generated.

The authors previously used RS on a pilot dataset, and since it was proven accurate [13], therefore here this method was extended with new decision classes, and was applied to a considerably larger biometric database, as is shown further in this paper.

In the RS method, a decision set is not required to be precise, but instead it is defined by upper and lower approximations: the former including objects that may belong to the set, and the latter including objects that certainly belong to the set, while considering the available knowledge. Objects are characterised by a set of attributes P. In this work P is a set of features, and objects are particular biometric identities. In order to express the quality of P, the positive region POS(P) is defined as a set of objects included within only one decision class. If $\mathrm{POS}(P) = U$ then each object of the universe is correctly classified, whereas the proportion of $|\mathrm{POS}(P)|$ to $|U|$, being the ratio of correctly classified objects to all objects, expresses the quality of RS model.

Practical applications often require a minimal subset RED $\subseteq P$, called a reduct, resulting in the same quality of approximation as P. Attributes with continuous values are discretised, prior to the reduct calculation. Maximum discernibility discretisation algorithm was applied [32, 33]. Numerous algorithms to calculate reducts are available. Herewith, dynamically adjusted approximate reducts heuristic was used [34, 35]. Once the reduct is obtained, all cases in the decision table are analysed, then decision rules are generated. At the classification stage rules are applied for every test object. The final decision is made by 'voting' mechanisms, where all matching rules are taken into account. More information on RS can be found in the literature [31–33, 36] and in our previously published work [13].

3 Data collection procedure

3.1 Experiment environment

It was observed that the hand vein scan and digital signature were not influenced by external conditions. In turn, the effectiveness of voice biometry and face image acquisition depends directly on the conditions in which they are used. The study was conducted in real banking branches, during a routine customer service, therefore the acoustic and lighting conditions could not be fully controlled.

It was observed during the pilot study that the acoustics is dependent on the size of the branch, and on the number of visiting clients, as the major disturbance is the presence of other human voices. The applied correctly trained GMM-UBM model allowed for the operation in the noisy environment [23].

For the purpose of reliable face image acquisition, the general assessment of lighting conditions for each of the stations was carried out based on a direct examination made during the installation of the biometric stations. The following categories and outcomes were identified:

- normal conditions: correct lighting and photo exposition, 75% of cases;
- too bright: overexposed images, 3% of cases;
- too dark: underexposed images, 16% of cases;
- uneven: side lighting, face image partially underexposed, 6% of cases.

Differences are determined mainly by the camera setup and orientation. The daylight cycle includes sunlight, and artificial light sources.

The analysis of customer data who have resubmitted their verification samples at an interval >1 day since the registration date reveals 0.21 probability of the correct verification under the condition that the lighting is different than during the enrolment. On the other hand, the probability of the correct verification at the same lighting conditions was equal to 0.99. That analysis indicates strong dependency of face verification result on lighting conditions, therefore in practical exploitation of the system, the camera orientation and light sources were set up intentionally to reduce the variations of lighting, and increase verification rates. The following results are based on samples collected in corrected environments.

3.2 Data filtration

By dispatching 100 biometric stations located at 60 bank branches 10,078 BioIDs were collected during the conducted study. Completing the survey has been a voluntary process, thus not each enrolment or verification attempt resulted in a completed survey. In total, 9592 surveys were started: 4978 filled in by advisors, and 4614 by clients.

Among all collected biometric samples first a subset of 7166 BioIDs containing data of a proper quality was selected for a close examination: incomplete records, tests, duplicates and were removed. Finally, a strict subset of 3591 records was identified, matching conditions of the current study, i.e. described with correct biometric data and accompanied by fully completed questionnaires by both banking tellers and clients. The rigorous focus on entirely completed surveys is motivated by the overarching goal to study opinions about tested biometric solutions along with objective user performance. Therefore, BioIDs without complete questionnaires were rejected.

Table 1 Number of registered samples corresponding to individual biometric modalities

Modalities	Registration phase	Identity verification phase
handwritten signature	5 (3 templates + 2 control attempts)	2 attempts
hand vein	2 (2 templates + 0 control attempts)	1 attempt
face image	15 (10 templates + 5 control attempts)	5 attempts
face contour	15 (10 templates + 5 control attempts)	5 attempts
voice	4 (3 templates + 1 control attempt)	1 attempt
total	41 biometric samples	14 biometric samples

Table 2 Total number of samples collected for individual modalities

Modalities	All samples	Verifying samples
handwritten signature	32,839	15,357
hand vein	16,777	1121
face image	122,722	26,329
face contour	48,847	24,286
voice	16,534	4756
total	237,719	71,849

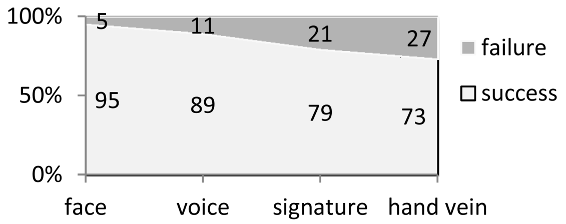

Fig. 1 *Relation between percentage of falsely rejected samples and correctly accepted samples*

4 Data analysis results

4.1 Database

A specified number of enrolment samples is required by each biometric modality to be collected. As is shown in Table 1, the number of registered templates depends on the phase of the performed process (registration or verification). The process of registering biometric samples for individual modalities consists first in collecting and storing a biometric template from a person and then in making a direct verification attempt in order to validate the acquired template. The biometric modalities were designed to register templates together with a verifying sample. However, in case of HandVein, a positive registration of the template is automatically considered as a valid one [25]. Since all other biometric solutions were engineered entirely by the authors of the article, the requirement for storage of verification samples had been made and implemented. Consequently, the total number of all collected samples for each BioID after the template registration amounts to 41, and after the verification it is increased by further 14 samples (Table 1).

The total number of registered biometric samples for the analysed dataset is presented in Table 2. The prepared set of registered samples consists of 237,719 biometric samples, out of which 71,849 served for the verification. The largest number of samples among the registered modalities has been recorded for face image modalities.

Table 3 Co-verification between modalities grouped in pairs, ordered by *p*-value of chi-squared test

	Modalities	*p*
1	face image – voice	<0.0001
2	face image – signature	<0.0001
3	signature – face image	<0.0001
4	signature – voice	<0.0001
5	voice – face image	<0.0001
6	face image – hand vein	<0.0001
7	voice – signature	0.0001
8	signature – hand vein	0.016
9	hand vein – voice	0.052
10	voice – hand vein	0.065
11	hand vein – signature	0.134
12	hand vein – face image	0.553

Table 4 Co-verification between chosen modality and paired modalities, ordered by the *p*-value of chi-squared test

	Modalities	*p*
1	face image – voice, signature	<0.0001
2	signature – voice, face image	<0.0001
3	face image – signature, hand vein	<0.0001
4	voice – signature, face image	<0.0001
5	signature – hand vein, face image	<0.0001
6	face image – voice, hand vein	0.0005
7	voice – hand vein, face image	0.002
8	hand vein – voice, signature	0.007
9	hand vein – voice, face image	0.027
10	voice – signature, hand vein	0.029
11	signature – voice, hand vein	0.173
12	hand vein – signature, face image	0.314

4.2 Co-occurrence of verification success rates

For the purpose of an assessment of co-occurrence of verification success rates that might occur between individual modalities, first, for the subset of 3591 biometric records, collected for all modalities, FRR measures were analysed, which brought results presented in Fig. 1. The above limitation of the number of processed records resulted from the availability of complete biometric samples (all modalities represented in each record).

The co-occurrence between modalities success rates (co-verification) reflects an ability to verify the client positively using a selected modality, under the condition that a positive verification for another modality or modalities occurred as well. It has been assessed using the chi-squared test. Each modality verification result has been compared with another one's result in pairs or grouped with two other modalities or with three other modalities. In each case, the test was performed employing two populations. The first population contained verification results for a particular modality in question, provided positive verification results were obtained for the second paired modality or a group of modalities. The second population contained data associated with the negative verification results of the other modality or modalities. Tests were performed in all possible combinations. The results of the chi-squared tests are shown in Tables 3–5, proving the hypothesis that the result of first modality verification will also indicate significant possibility of successful verification of another modality.

According to the results presented in Tables 3–5, the co-verification between modalities is not symmetric (it does not occur in both directions with the same significance), e.g. a user able to verify himself or herself positively using hand vein will, in general, also be able to verify with signature ($p = 0.016$), however the opposite rule is not statistically significant ($p = 0.134$).

For six pairs of modalities (Table 3) and for five combinations of a given modality or coupled modalities (Table 4), some very low *p*-values have been obtained, represented in tables as values

Table 5 Co-verification between chosen modality and triplet of modalities, ordered by the p-value of chi-squared test

	Modalities	p
1	face image – signature, hand vein, voice	0.001
2	signature – hand vein, face image, voice	0.311
3	voice – signature, hand vein, face image	0.619
4	hand vein – signature, face image, voice	0.822

Table 6 Number of answers containing given amount of words or phrases found to be associated with a negative or with a positive sentiment

Number of answers				
No. negative sentiment	No. positive sentiment			
	0	1	2	3
5	4	0	0	0
4	15	0	0	0
3	2	59	0	0
2	162	4	0	0
1	533	12	2	0
0	2561	217	18	1

Table 7 Top ten pairs of parameters associated with highest absolute value of Pearson correlation coefficient

Parameter A – sentiment	Parameter B – subjective feature	Pearson correlation factor
negative	*cCumberVoi*	−0.243
negative	*cHardVoi*	−0.203
negative	*cHelp*	0.196
negative	*cFast*	0.186
positive	*cHelp*	−0.185
positive	*gender*	−0.176
negative	*cHardCont*	−0.154
positive	*uEasyFac*	−0.149
negative	*uEasyVoi*	−0.147
positive	*advisor age*	−0.144

‘<0.0001’. Nevertheless, the order of the combinations in Tables 3–5 was still possible to be ranked according to chi-squared test results. The face image modality verification success was the one which indicated success in other modalities with the highest statistical significance, i.e. the ability of a user to verify themselves positively with another modality ensures with the high probability that they will be able to verify positively with a face modality. The hand vein modality was the one for which the relation with other modalities was not statistically significant ($p > 0.05$). It can be assumed that the relation between the face image and the handwritten signature modalities occurs in both directions, since p-values are close to each other in the whole range of p-values for all of modalities. This finding seems advantageous, because it can be utilised in biometric systems due to two aspects. First, in financial operations demanding a high level of security, both modalities can be combined together providing a solution which offers a stable FRR level, since a user's ability to verify himself or herself with any of these two modalities is similar. Employing the fusion of modalities instead of using just a single one may increase the rejection rate of forgeries. Moreover, the hand-written signature is the most acceptable form of authorisation in banking systems, as it is already being used commonly in various confirmation-requiring procedures.

4.3 Semantic analysis of survey results

Following the process of semantic analysis introduced in Section 2.8, each answer collected in the survey was split into a sequence of words and phrases with one of three possible sentiments: positive, negative, and neutral. Then, for each answer the occurrences of each word or phrase associated with each sentiment were counted. Therefore, each answer was described by two integer numbers: the number of words or phrases associated with a positive sentiment and the number of words or phrases associated with a negative sentiment. Results of this stage of analysis could be treated as a numerical interpretation of the content of textual answers in surveys and it can be analysed with the use of machine learning algorithms. Frequencies of answers containing various numbers of sentiment-associated words or phrases are summarised in Table 6.

An additional step of filtering was applied before the analysis of surveys. Only surveys containing more than one word were taken into consideration. This allowed for the processing surveys only which were likely to contain feedback about the system.

For 2561 of analysed surveys there were no words or phrases found to be associated with a positive or with a negative sentiment. A group of 77 text excerpts contained both positive and negative statements. For 236 answers it was found 1 or more positive statement and for the negative sentiment, 716 such answers were identified (Table 6). A number of answers containing given amount of words or phrases were found to be associated with a negative or with a positive sentiment.

According to the data collected during the research, a conclusion that bank tellers tend to write more negative-biased comments may be drawn. The Pearson correlation factor reflecting sentiment assessment and length of the comment were also calculated, and it was found that for the positive sentiment evaluation case the coefficient was equal to 0.04, whereas for the negative sentiment it was equal to 0.73. Therefore, if a comment was associated with a negative sentiment, then it also was more likely to be written as a longer one. In case of shorter comments, bank tellers usually provided a feedback informing that the system works correctly. In the longer ones, they described the problem.

The Pearson correlation factor was also calculated for each numerical answer gathered from client and teller surveys. Examples of pairs of parameters extracted from the survey associated with the highest absolute values of the factor are given in Table 7. There is a weak correlation between the negative sentiment estimate and the difficulty estimate of the voice modality for the client and for the teller. Such a correlation also exists for the registration process length and for the amount of help needed during the data acquisition. Correlations between estimates of sentiments and the rest of answers are significantly smaller than the correlation between the negative sentiment estimate and the length of the comment.

4.4 Prediction of identity verification improvement

The main purpose of the RS model was to validate a methodology of predicting verification accuracy based on objective and on subjective factors observed during the enrolment. Subjective responses collected from surveys reflect user opinions on: comfort, ergonomics, intuitiveness, and other aspects of the enrolment. In turn, objective features reflect accuracy of the first verification performed directly after the registration of a new identity, and the time elapsed from the first registration to the verification attempt.

The RS was used to explore relations between objective and subjective features of the whole verification process and result improvement measures. The experiment was conducted in six scenarios aimed at modelling measures of $similarity_1$ and $progress_{2,3,4,5,6}$.

First, Pearson correlations between performance metrics in consecutive attempts were calculated (Table 8), revealing a significant correlation existence.

The similarity and progress values were discretised with MM and QM methods, and then RS lower and upper approximation procedure was performed, resulting in:

- selection of features (calculation of reduct),
- rules induction based on all cases,
- prediction of decision classes for $similarity_1$ and for $progress_n$, $n = 2–6$, with values located within ranges defined by MM and QM models.

Table 8 Correlations between similarity and verification progress metrics for given modality

Correla-tions	$similar_1$	$progre_2$	$progre_3$	$progre_4$	$progre_5$	$progre_6$
			Voice modality			
$similar_1$	—	**−0.748**	**−0.717**	**−0.782**	−0.269	**−0.637**
$progre_2$	**−0.748**	—	0.454	0.588	0.561	**0.685**
$progre_3$	**−0.717**	0.454	—	**0.636**	**0.674**	**0.797**
$progre_4$	**−0.782**	0.588	**0.636**	—	**0.742**	**0.649**
$progre_5$	−0.269	0.561	**0.674**	**0.742**	—	**0.726**
$progre_6$	**−0.637**	**0.685**	**0.797**	**0.649**	**0.726**	—
			Face image modality			
$similar_1$	—	−0.580	**−0.788**	−0.632	**−0.691**	**−0.809**
$progre_2$	−0.580	—	0.660	0.637	0.638	0.601
$progre_3$	**−0.788**	0.660	—	0.588	**0.710**	0.509
$progre_4$	−0.632	0.637	0.588	—	**0.789**	0.644
$progre_5$	**−0.691**	0.638	**0.710**	**0.789**	—	0.660
$progre_6$	**−0.809**	0.601	0.509	0.644	0.660	—
			Signature modality			
$similar_1$	—	−0.629	**−0.725**	**−0.738**	**−0.656**	**−0.637**
$progre_2$	−0.629	—	0.606	0.599	0.306	0.427
$progre_3$	**−0.725**	0.606	—	0.515	0.587	0.511
$progre_4$	**−0.738**	0.599	0.515	—	0.501	**0.630**
$progre_5$	**−0.656**	0.306	0.587	0.501	—	0.562
$progre_6$	**−0.637**	0.427	0.511	**0.630**	0.562	—

Voice modality (for bold numbers $p < 0.003$, for all other $p < 0.04$)
Face image modality (for bold numbers $p < 0.001$, for all other $p < 0.06$)
Signature modality (for bold numbers $p < 3.2e\text{-}8$, for all other $p < 6e\text{-}4$)

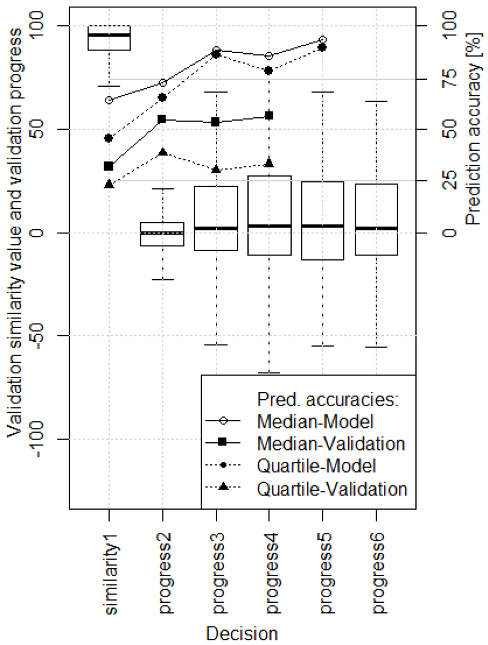

Fig. 2 *Signature modality modelling and prediction results: validation samples and progress characteristics (boxplots), accuracy of prediction (lines)*

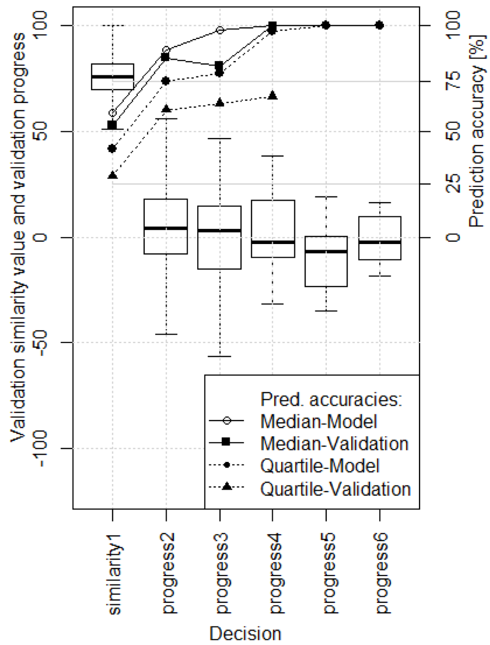

Fig. 3 *Voice modality modelling and prediction results: validation samples and progress characteristics (boxplots), accuracy of prediction (lines)*

Accuracy of prediction of modelled *progress* classes was interpreted as the RS model accuracy. Also, an averaged accuracy over ten cross-validation runs was determined. The cross-validation involved splitting the dataset into ten equinumerous subsets, then taking one to act as a validation set and use other nine to calculate rules for classification. Rule induction and validation was repeated ten times and results were averaged.

Results of prediction of *progress* measures are characterised in Figs. 2–4. Box-whiskers plots depict median and quartile ranges of each decision attribute obtained by the calculated improvement in similarity (9). Prediction accuracies obtained with all models are plotted as lines.

It was assumed that the minimal size of a single decision class to be a subject of RS modelling was 30 biometric records. Therefore, the MM with two classes required at least 60 records, and QM with four classes required at least 120 records. Only in case of a few first visits to the bank this requirement was fulfilled. In the time span of the project <60 clients were inclined to visit the bank in order to attend the biometric verification five or more times, and <120 clients paid visits more than four times. Therefore, relevant RS models were not validated for some $progress_n$ values.

The *similarity* characteristic, denoted as box-whiskers plots in Figs. 2–4, revealed that the signature had the highest similarities among validation and registration samples, reflecting a high stability of the signature modality. Voice and face sample had lower values, with medians below 78.

The *progress* value for biometric verification usually was small. Particularly for signature and for voice the median is close to 0 – meaning that no noticeable progress was made. This is justified by the verifications samples from attempts no. 2, 3, … being very similar to the samples of the first verification. The face modality had a negative value of *progress*, interpreted as a significant change in the face appearance, when a verification is made after a large number of days, in other lighting conditions, or in case of changed makeup or facial hair.

The results of modelling and predicting the $similarity_1$ and *progress* by RS approach are shown in Figs. 2–4 as lines. Accuracy of $similarity_1$ prediction was low, as features used for the modelling and classification were not sufficient. Observed values of $similarity_1$ cannot be explained by survey answers and RS methodology. MM and QM models performed poorly, resulting in almost random classification with the MM of accuracy close to 50% and QM close to 25%.

However, accuracies for *progress* prediction were considerably higher, namely:

- signature *progress* prediction accuracy with MM: 64 –94%, and with QM: 45.6–90%,
- voice *progress* prediction accuracy with MM: 60–100%, and with QM: 42–100%,
- face *progress* prediction accuracy with MM: 60–100%, and with QM: 39–100%.

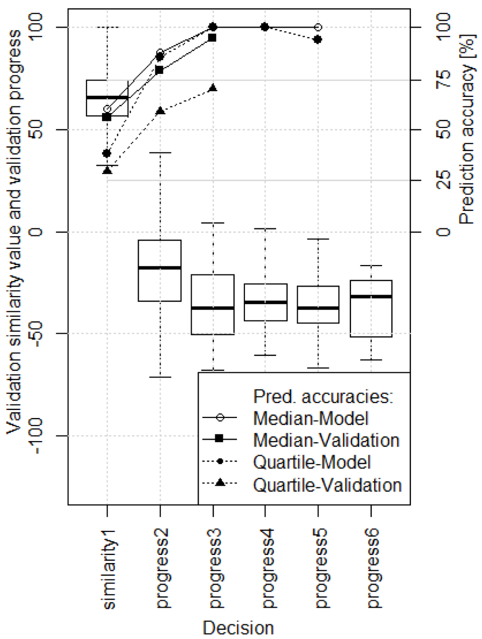

Fig. 4 *Face modality modelling and prediction results: validation samples and progress characteristics (boxplots), accuracy of prediction (lines)*

Generally, the classification among four classes based on quartiles was less accurate than the one made among two classes based on median.

Derived rules were based on reducts composed of the most important features related strongly to the decision. Voice MM for $progress_2$ had a reduct $\{similarity_1, uWill\ cHelp\}$, with the largest correlation of $similarity_1$: −0.748, and marginally low correlation of *uWill*, *cHelp* equal to 0.127, 0.066, respectively. Other reducts were structured in the same manner: the first feature being $similarity_1$ with the high correlation (Table 8), and two or three values from questionnaires with correlation close to 0. Negative and positive sentiments were not included in reducts, because of low correlation with decisions.

Relative sizes of positive regions POS were calculated (Table 9), as introduced in Section 2.10. Small values indicate low number of cases with strictly defined deterministic decision. This was a result of the presence of validation samples matching many conflicting rules. The final decision was made by ‘voting’, where contributions of all rules were taken into account. In consequence, the finally achieved accuracy of predicting user *progress* proven to be satisfactorily high.

Cases with the positive region size close to 1.0 indicate a capability of given RS model to properly describe with deterministic rules the relations between features collected at the user registration phase and the features of consecutive validation attempts.

It can be seen that $similarity_1$ has an extremely low POS, and with $progress_2$ it improves, e.g. for face MM and QM models reaching 0.7. It can be hypothesised that the $similarity_1$ of the first verification is not related to the initial user experience reported in questionnaires and it is impossible to be predicted by RS modelling. Next attempts are based on the user subjective experience and the objective similarity, and RS models are capable to extract and to model this relation in a more deterministic manner, up to a very high degree (1.0 for face and voice), with lowest results in case of signature. The signature models low POS reflect indeterministic nature of relations between subjective user experience and verification rate improvements, as well as changes in samples proximity to the template.

5 Conclusions

Several analysis techniques were used for determining the relationship between various types of biometrics in the presented study. Moreover, the sentiment analysis was performed over subjective outcomes from surveys, and the RSs methodology was employed to predict improvements in identity verification accuracy. To the best of the authors’ knowledge, this was the first research attempt comprising acquisition and analysis of a multimodal biometrics database of this volume in uncontrolled real conditions (it involved 100 multimodal biometric stations operated by over 10,000 individuals).

The outcomes of this project led us to draw some generalised conclusions related to biometry applications in the banking environment. However, the research performed on the collected data has also some limitations, for example: subjective opinion surveys were voluntary, what significantly limited the amount of data appropriate for these experiments. The number of participants willing to visit the bank and use biometric verification more than 5

Table 9 Relative sizes of positive regions POS for given modality and model (MM – median model, QM – quartile model)

Pos.reg. for class and model	Voice models		Signature models		Face models	
Class	MM	QM	MM	QM	MM	QM
$similarity_1$	0.04	0.07	0.03	0.02	0.04	0.03
$progress_2$	0.41	0.37	0.18	0.21	0.70	0.68
$progress_3$	0.96	0.37	0.57	0.64	1.0	1.0
$progress_4$	1.0	0.92	0.13	0.39	1.0	1.0
$progress_5$	1.0	1.0	0.73	0.77	1.0	0.88
$progress_6$	1.0	1.0	—	—	—	0.04

times was less than 120, what reduced the model prediction accuracy.

Along with biometric data also subjective opinions on user confidence, ease of use, comfort, and acceptance for biometric technologies were collected. This research allowed to assess the effects of repeated interaction with the verification system, by measuring and by predicting improvement in user verification rates.

An important finding is that according to the FRR measure and co-validation between modalities, the face image can be combined with the signature, providing a reliable biometric solution. It can work independently on the ability of a user to verify positively with any of these modalities. As the co-validation between both modalities is bi-directional, thus such a fusion of modalities can have a great potential in increasing true rejection rate of forgeries.

It was possible to analyse sentiment of textual surveys performed after the acquisition of biometric data and to select reviews which are associated with a positive or with a negative sentiment. Also, it was found that texts contained in the surveys associated to the negative sentiments are likely to be longer than other fragments of surveys. Such surveys contain a useful information for creators of this kind of systems. A large collection of textual feedback can be automatically processed by the presented algorithm. The sentiment estimate may be used as a feature for the machine learning applied to an automatic assessment of the performance of large-scale biometric systems.

By applying RS analysis it was observed that progress metrics derived from the biometric records were strongly correlated with initial sample similarity values. Users capable of using biometric sensors were improving their successful verification rates after every consecutive interaction with the system. Numerous subjective features obtained in surveys were also included in decision rules, and many modelled relations achieved a high degree of accuracy. It was made possible to predict future performance of biometric identity validation expressed in terms of similarity between collected samples and the registered biometric template.

In the future, a greater focus should be put on mandatory collection of, possibly shorter, surveys, as the subjective data proven to be a valuable information for assessing usability of the system and for predicting identity verification performance.

6 Acknowledgments

This work was supported by the grant no. PBS3/B3/26/2015 entitled 'Multimodal biometric system for bank client identity verification' co-funded by the Polish National Center for Research and Development. The bank and the project steering committee approved this study for publication.

7 References

[1] Ross, A., Jain, A.K.: 'Multimodal biometrics: an overview'. 12th European Signal Processing Conf. (EUSIPCO), Vienna, Austria, 2004, pp. 1221–1224

[2] Jain, A., Nandakumar, K., Ross, A.: 'Score normalization in multimodal biometric systems', *Pattern Recognit.*, 2005, **38**, (12), pp. 2270–2285, doi: 10.1016/j.patcog.2005.01.012

[3] Faundez-Zanuy, M., Fierrez-Aguilar, J., Ortega-Garcia, J., *et al.*: 'Multimodal biometric databases: an overview', *IEEE Aerosp. Electron. Syst. Mag.*, 2006, **21**, (8), pp. 29–37, doi: 10.1109/MAES.2006.1703234

[4] Garcia-Salicetti, S., Beumier, C., Chollet, G., *et al.*: 'BIOMET: a multimodal person authentication database including face, voice, fingerprint, hand and signature modalities'. Audio- and Video-Based Biometric Person Authentication (AVBPA 2003), Berlin, (LNCS, 2688), 2003, doi: 10.1007/3-540-44887-X_98

[5] Ortega-Garcia, J., Fierrez-Aguilar, J., Simon, D., *et al.*: 'MCYT baseline corpus: a bimodal biometric database', *IEE Proc., Vis. Image Signal Process.*, 2003, **150**, (6), pp. 395–401, doi: 10.1049/ip-vis:20031078

[6] Fierrez, J., Ortega-Garcia, J., Toledano, D.T., *et al.*: 'Biosec baseline corpus: a multimodal biometric database', *Pattern Recognit.*, 2007, **40**, (4), pp. 1389–1392, doi: 10.1016/j.patcog.2006.10.014

[7] Fierrez, J., Galbally, J., Ortega-Garcia, J., *et al.*: 'BiosecurID: a multimodal biometric database', *Pattern Anal. Appl.*, 2010, **13**, p. 235, doi: org/10.1007/s10044-009-0151-4

[8] Ho, C.C., Ng, H., Tan, W.W., *et al.*: 'MMU GASPFA: a COTS multimodal biometric database', *Pattern Recognit. Lett.*, 2013, **34**, (15), pp. 2043–2050, doi: 10.1016/j.patrec.2013.01.027

[9] NIST BSSR1 biometric score set, Available at: www.nist.gov/itl/iad/image-group/nist-biometric-scores-set-bssr1

[10] Shang, D., Zhang, X., Han, J., *et al.*: 'MultiModal-database-XJTU: an available database for biometrics recognition with its performance testing'. IEEE 3rd Information Technology and Mechatronics Engineering Conf. (ITOEC), Chongqing, China, 2017, pp. 521–526, doi: 10.1109/ITOEC.2017.8122351

[11] NIST FERET Face Recognition Technology, www.nist.gov/programs-projects/face-recognition-technology-feret

[12] Snelick, R., Uludag, U., Mink, A., *et al.*: 'Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005, **27**, (3), pp. 450–455

[13] Szczuko, P., Czyżewski, A., Hoffmann, P., *et al.*: 'Validating data acquired with experimental multimodal biometric system installed in bank branches', *J. Intell. Inf. Syst.*, 2017, pp. 1–31, doi: 10.1007/s10844-017-0491-2

[14] Czyżewski, A., Bratoszewski, P., Hoffmann, P., *et al.*: 'Performance analysis of developed multimodal biometric identity verification system', *Elektron. Konstrukcje, Technologie, Zastosowania*, 2018, **59**, (4), doi: 10.15199/13.2018.4.7

[15] Szczuko, P., Czyżewski, A., Szczodrak, M.: 'Variable length sliding models for banking clients face biometry', in '*Multimedia tools and Applications*' (Springer, 2018), pp. 1–18, doi: 10.1007/s11042-018-6432-4, in print

[16] Greaves, F, Ramirez-Cano, D, Millett, C, *et al.*: 'Use of sentiment analysis for capturing patient experience from free-text comments posted online', *J. Med. Internet Res.*, 2013, **15**, (11), p. 239. doi: 10.2196/jmir.2721

[17] Duan, W., Cao, Q., Yu, Y., *et al.*: 'Mining online user-generated content: using sentiment analysis technique to study hotel service quality'. 46th Hawaii Int. Conf. System Sciences, Wailea, Maui, HI, USA, 2013, pp. 3119–3128. doi: 10.1109/HICSS.2013.400

[18] Wang, H., Can, D., Kazemzadeh, A., *et al.*: 'A system for real-time Twitter sentiment analysis of 2012 U.S. Presidential election cycle'. Proc. ACL System Demonstrations, Jeju, Republic of Korea, 2012, pp. 115–120

[19] Chen, W., Hong, Q., Li, X.: 'GMM-UBM for text-dependent speaker recognition'. Int. Conf. Audio, Language and Image Processing, Shanghai, 2012, pp. 432–435

[20] ALIZE, Open source recognition, University of Avignon. Available at http://mistral.univ-avignon.fr, accessed 21 January 2018

[21] McCool, C., Marcel, S., Hadid, A., *et al.*: 'Bi-Modal person recognition on a mobile phone: using mobile phone data'. IEEE ICME Workshop on Hot Topics in Mobile Mutlimedia, Melbourne, VIC, Australia, 2012

[22] Szczodrak, M., Czyżewski, A.: 'Evaluation of face detection algorithms for the bank client identity verification', *Found. Comput. Decis. Sci.*, 2017, **42**, (2), pp. 137–148

[23] Bratoszewski, P., Czyżewski, A., Hoffmann, P., *et al.*: 'Pilot testing of developed multimodal biometric identity verification system'. Signal Processing, Algorithms, Architectures, Arrangements, and Applications, Poznań, Poland, 2017, pp. 184–189. 20.9.2017-22.9.2017

[24] Bratoszewski, P., Czyżewski, A.: 'Face profile view retrieval using time of flight camera image analysis'. Pattern Recognition and Machine Intelligence, PReMI 2015, Warsaw, Poland, 2015 (LNCS, 9124)

[25] Fujitsu Identity Management and PalmSecure. Available at https://www.fujitsu.com/au/Images/PalmSecure_Global_Solution_Catalogue.pdf, accessed 20 January 2018

[26] Na, J., Sui, H., Khoo, C., *et al.*: 'Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews'. Conf. Int. Society for Knowledge Organization (ISKO), Wurzburg, Germany, November 2004, pp. 49–54

[27] Na, J.-C., Khoo, C., Horng Jyh Wu, P.: 'Use of negation phrases in automatic sentiment classification of product reviews', *Libr. Collect., Acquis., Tech. Serv.*, 2005, **29**, (2), pp. 180–191. https://doi.org/10.1016/j.lcats.2005.04.007

[28] Trstenjak, B., Mikac, S., Donko, D.: 'KNN with TF-IDF based framework for text categorization', *Procedia Eng.*, 2014, **69**, pp. 1356–1364. https://doi.org/10.1016/j.proeng.2014.03.129

[29] Ma, Z., Feng, J., Chen, L., *et al.*: 'An improved approach to terms weighting in text classification'. Int. Conf. Computer and Management (CAMAN), Wuhan, China, 19–21 May 2011

[30] Pawlak, Z.: '*Rough sets theoretical aspects of reasoning about data*' (Kluwer, Dordrecht, 1991)

[31] Pawlak, Z.: 'Rough sets', *Int. J. Comput. Inf. Sci.*, 1982, **11**, p. 341, https://doi.org/10.1007/BF01001956

[32] Bazan, J.G., Peters, J.F., Skowron, A.: 'Behavioral pattern identification through rough set modelling'. Rough Sets, Fuzzy Sets, Data Mining, and Granular Gomputing. RSFDGrC, Springer, Berlin, Heidelberg, 2005 (LNCS, 3642), pp. 688–697

[33] Nguyen, S.H.: 'On efficient handling of continuous attributes in large data bases', *Fundam. Inform.*, 2001, **48**, (1), pp. 61–81

[34] Janusz, A., Stawicki, S.: 'Applications of approximate reducts to the feature selection problem'. Proc. Int. Conf. Rough Sets and Knowledge Technology (RSKT), Banff, Canada, vol. 6954, 2011, pp. 45–50

[35] Janusz, A., Ślęzak, D.: 'Random probes in computation and assessment of approximate reducts'. Proc. of RSEISP, Granada and Madrid, Spain, 2014 (LNCS, 8537), pp. 53–64

[36] Zhong, N., Dong, J., Ohsuga, S.: 'Using rough sets with heuristics for feature selection', *J. Intell. Inf. Syst.*, 2001, **16**, p. 199, doi: 10.1023/A:1011219601502