Pattern Recognition Letters 38 (2014) 107–112

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Multimodal biometric system combining ECG and sound signals ☆

Marcin D. Bugdol *, Andrzej W. Mitas

Faculty of Biomedical Engineering, Silesian University of Technology, Charlesa de Gaulle'a 66, 41-800 Zabrze, Poland

ARTICLE INFO

Article history:
Received 4 May 2013
Available online 3 December 2013

Keywords:
Behavioural biometrics
ECG analysis
Sound analysis
Multimodal biometrics

ABSTRACT

In the paper a novel multimodal, behavioural biometric system that combines ECG and sound signal is described. The signal acquisition has been carried out in a small stress condition when the user has been asked to utter sounds at a given pitch. Thanks to the stimulations the ECG signal and especially the HR vary over time and these changes can be used to extract biometric features which represent an individual response to the stimuli. The proposed approach utilised popular statistical coefficients which are computationally effective and simple.

1. Introduction

The most important drawback of the current identity verification methods (based on knowledge or base on possession) is that this process is not certain whether the holder of the specified ID is the entitled person or an imposter. The use of biometrics in the verification system is intended to minimise this problem and reduce the need of carrying the tokens (all what person needs is "with him"). Unfortunately, the biometrics is characterised with a high degree of uncertainty caused by the inaccuracy of the features acquisition. Moreover, the commonly used physiological attributes such as fingerprint or iris, are also vulnerable to theft, which carries significantly greater consequences than the loss of traditional identifiers.

The behavioural biometrics which utilize the human behaviour and reactions seems to be a better solution under the security aspect. The acquisition of these features is more difficult but they are also difficult to steal. This type of biometrics is characterised by smaller distinctiveness, they are used in the verification rather than in identification systems. The combination of two or more biometrics of small uniqueness results in a solution which can be used by a larger group of users. In addition, the level of security is also increased, because all the biometrics have to be stolen in order to enter the system. In this paper a behavioural biometrics system is described which uses two modalities: voice and ECG. The first one was chosen because it is a source of physical features and behavioural ones. Moreover, measuring the voice is completely non-invasive and socially accepted (Bugdol and Bugdol, 2009). The second modality is the ECG, which is another source of biometric features. The main advantages of this modality are that it can be acquired only from a living person and that it is very hard to steal or forge someone ECG. The number of papers that refers to ECG in identity verification suggest that this an emerging biometric modality.

2. Related works

Human voice carries not only information about the physiological features of a person (related to the structure of the vocal cords) but also the behavioural ones: way of speaking, emotion or mood of the speaker (Mitas et al., 2012). The most popular voice analysis methods use coefficients that are based on two main approaches: Linear Predictive Coding (LPC) (Dustor and Szwarc, 2009, 2010) and Fast Fourier Transform (FFT). FFT-based algorithms are the basis for further processing in order to obtain the cepstral coefficients. One of the most commonly used are mel frequency cepstral coefficients (MFCC) which are evaluated in the mel scale which has been empirically estimated and models the human hearing characteristics. These features are used in speech analysis as well as in speaker recognition (Campbell, 1997). The method for human identity recognition (both verification and identification) based on the voice depends on the mode of biometric system: text-dependent or text-independent. In case of text-dependent solutions, techniques based on two types of methods are used (Cole et al., 1997; Furui, 2008): Dynamic Time Warping – DTW and Hidden Markov Model – HMM. HMM-based methods generally yield better results than those based on the DTW. The advantage of the DTW is a much smaller training set required to train the classifier. In text-independent system the most popular classification methods are (Cole et al., 1997; Furui, 2008): Vector Quantization – VQ, Gaussian Mixture Models – GMM and Support Vector Machine – SVM. If a large set of training data is available,

☆ This paper has been recommended for acceptance by M.S. Nixon.

* Corresponding author. Tel.: +48 322371732; fax: +48 322371716.
E-mail address: Marcin.Bugdol@polsl.pl (M.D. Bugdol).

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.11.014

the GMM and SVM-based techniques are characterised by a lower error rate than those based on the VQ. Otherwise, the methods based on VQ should be used.

The ECG signal is a well known diagnostic method which helps in the diagnostics of cardiovascular diseases. Hoekema et al. (2001) have analytically stated that every person has a unique ECG waveform which can be used in the field of biometrics.

The feature extraction methods from the ECG signal which are currently used in biometrics can be divided into two main groups:

- methods based on fiducial points,
- methods not based on the fiducial points.

In the methods of the first group individual parameters are extracted using some characteristic points (called fiducial points) of the ECG waveform (Fig. 1).

Such approach was proposed by Biel et al. (1999) and Biel et al. (2001), where full, twelve-lead ECG monitor was used to register 360 features (30 diagnostic parameters for every lead). Shen et al. (2002) have shown that the number of used leads can be reduced to one.

In several works (Israel et al., 2003, 2005; Irvine et al., 2003; Wiederhold et al., 2006) the examined persons had to do some tasks (i.e. meditation, reading, arithmetic) during which the ECG signal was registered. New characteristic points on the ECG curve have been proposed in addition to the standard ones.

All methods listed above are based on features extracted from the fiducial points of the ECG, because they are widely used in medical diagnosis and an approach to use them in the field of biometrics is natural. However, for biometric purposes such fuzzy localisation of the mentioned fiducial points causes that the verification can be incorrect. Moreover, there is no universal rule clearly defining the boundaries of waves (Martinez et al., 2004) and the number of publications that concern methods of automatic fiducial points detection (compared for example in Karpagachelvi et al. (2010)) suggests, that this matter is still a subject of research.

A different approach has been presented by Wübbeler et al. (2007). No features from the cardiac function were extracted but only their waveforms were compared.

A transformation of the ECG signal, derived from one and three lead ECGs, into a three-dimensional space was proposed by Fang and Chan (2009, 2013). On the basis of five subsequent heart beats one averaged value was computed and it was reconstructed in the phase space. The signals were averaged using the R peak, which was the only characteristic feature extracted directly from the ECG waveforms.

Wang et al. (2006, 2008) extracted for every, 800 ms long, frame two set of features: based on fiducial points and those representing the potential difference between the selected leads. In case of the second set a single heart beat was transformed to a M-dimensional space and the features were obtained using the PCA method.

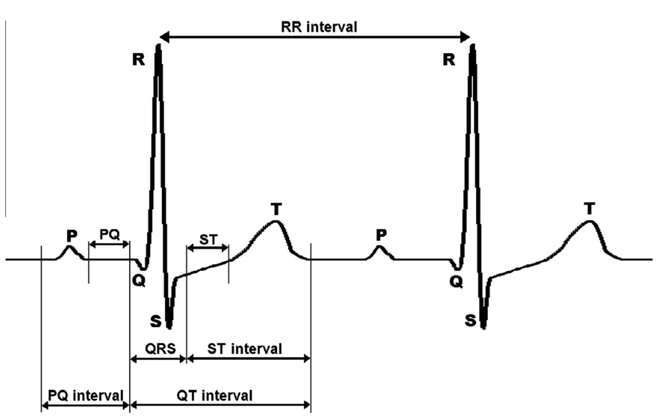

Fig. 1. ECG wave with fiducial points.

In Agrafioti and Hatzinakos (2008), Plataniotis et al. (2006) and Wang et al. (2008) the autocorrelation function have been used to compute the coefficients and the DCT or LDA methods have been used to reduce features number. This method has been also used to verify people with cardiac arrhythmia Agrafioti and Hatzinakos (2009).

Different approach has been presented in Li and Narayanan (2010) where the combination of the temporal and cepstral coefficients of the ECG was proposed.

The ECG signal has been also combine with popular physiological biometrics: face (Israel et al., 2003) and fingerprint (Singh et al., 2012).

3. Multimodal behavioural biometric system

The human reactions in response to the given visual and sound stimulation were used to extract the biometric features. The influence of sound on humans is the subject of research in the field of music therapy. In the papers Mitas (2008, 2009) the individual responses to a specific type of music have been discussed. The thesis, which have been there presented, have been the basis for constructing a stimulus based biometric system. Such systems can be useful for people whose disability prevents them from using traditional biometric systems. They can use also a single modality biometric systems based on biomedical signals, but systems enhanced with behavioural elements better secure the access to resources. The detailed information about the stimulations used, the acquisition process and measurement device construction is presented below.

3.1. Stimulation

The sound stimulation is a single sound from a real instrument and namely an electric piano. This intrument has been chosen because a noiseless, equal-loud sound could be recorded. Moreover, keyboard instruments such as piano are popular, well known and commonly used in music education. The sound frequency is randomly generated and the sounds from the human speech range are three times more frequent than others. The second type of stimulation is a visual one. It is a square of uniform colour. The colour depends on the match degree between the sound emitted by user and the exemplary sound generated by the stimulator. The colours are green if two sounds match almost perfectly, then dark green, yellow, orange, dark red and red when these sounds are completely different. The colour schemes have been chosen based on natural perception of colours. Red has been selected as a warning colour, which in this case represents the worst possible adjustment of the person to the given sound. The green colour is its opposite; people associate it with peace and relaxation. Intermediate colours have been chosen so that the changes between green and red are smooth and the sequence is the same as in the rainbow. The colour ranges have been defined based on the difference between the fundamental frequencies of emitted and exemplary sound ($f_{exemp} - f_{sign}$) and they are given by the following formula:

$$colour = \begin{cases} green & \text{for} \quad |f_{exemp} - f_{sing}| \in [0, 1/4\ \text{tone}] \\ dark\ green & \text{for} \quad |f_{exemp} - f_{sing}| \in (1/4\ \text{tone}, 3/4\ \text{tone}] \\ yellow & \text{for} \quad |f_{exemp} - f_{sing}| \in (3/4\ \text{tone}, 5/4\ \text{tone}] \\ orange & \text{for} \quad |f_{exemp} - f_{sing}| \in (5/4\ \text{tone}, 7/4\ \text{tone}] \\ dark\ red & \text{for} \quad |f_{exemp} - f_{sing}| \in (7/4\ \text{tone}, 11/4\ \text{tone}] \\ red & \text{for} \quad |f_{exemp} - f_{sing}| > 11/4\ \text{tone} \end{cases} \tag{1}$$