Pattern Recognition Letters 34 (2013) 2043–2050

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

MMU GASPFA: A COTS multimodal biometric database

Chiung Ching Ho *, Hu Ng, Wooi-Haw Tan, Kok-Why Ng, Hau-Lee Tong, Timothy Tzen-Vun Yap, Pei-Fen Chong, C. Eswaran, Junaidi Abdullah

Faculty of Computing and Informatics, Multimedia University, Jalan Multimedia, 63100 Cyberjaya, Malaysia

A R T I C L E I N F O

Article history:
Available online 15 February 2013

Keywords:
GASPFA
COTS
Multimodal biometric
Gait action recognition
Face
Speech

A B S T R A C T

This paper describes the baseline corpus of a new multimodal biometric database, the MMU GASPFA (Gait–Speech–Face) database. The corpus in GASPFA is acquired using commercial off the shelf (COTS) equipment including digital video cameras, digital voice recorder, digital camera, Kinect camera and accelerometer equipped smart phones. The corpus consists of frontal face images from the digital camera, speech utterances recorded using the digital voice recorder, gait videos with their associated data recorded using both the digital video cameras and Kinect camera simultaneously as well as accelerometer readings from the smart phones. A total of 82 participants had their biometric data recorded. MMU GASPFA is able to support both multimodal biometric authentication as well as gait action recognition. This paper describes the acquisition setup and protocols used in MMU GASPFA, as well as the content of the corpus. Baseline results from a subset of the participants are presented for validation purposes.

1. Introduction

Multimodal biometric databases have increased in numbers in recent years due to continuous research efforts being done in this field. The origins of multimodal biometric databases can be traced back to benchmarked single biometric modality databases such as the CMU PIE (Sim et al., 2002) face database, the AR Face (Martinez and Benavente, 1998) database, the NIST (Khanna and Weicheng, 1994) fingerprint database, the FVC (Maio et al., 2002) series of fingerprint databases, the NIST (Pallett, 2003) speaker recognition database, the CASIA (Ho et al., 2009) gait database, CMU Mobo (Gross and Shi, 2001) gait database, and so on. As research interest moved away from single modal biometrics, more multimodal biometric databases have been contributed by the research community. Examples of multimodal biometric databases include the BioSec (Fierrez et al., 2007) baseline corpus (which incorporates face, voice, iris and fingerprint modalities), BioSecurid (Fierrez et al., 2009) (speech, iris, face, signature and handwriting, finger prints, hand and key-stroking), MyIDEA (Dumas et al., 2005) (talking face, audio, fingerprints, signature, handwriting and hand geometry), BioSecure (Ortega-Garcia et al., 2010) (fingerprints, hand, iris, voice and face), MCYT (Ortega-Garcia et al., 2003) (a bi-modal database consisting of signatures and fingerprints), MBioID (Dessimoz et al., 2007) (2D and 3D face, fingerprint, iris, signature and speech), BANCA (Messer et al., 2004) (face and voice), FRGC (Phillips et al., 2005) (2D and 3D face) and many others. Applications of biometric databases have seen an evolution from identity authentication and recognition purposes to action recognition in recent times.

Action recognition is an emerging research field wherein the focus moves away from the issue of identity and focuses instead on the recognition of actions. Action recognition is useful for event detection followed by mitigation actions. One of the earliest available action database was reported in (Schuldt et al., 2004) where six different sets of human actions were recognized via the usage of Support Vector Machines (SVM). (Laptev et al., 2008) proposed the idea of using movie scripts as an automated way of annoting human action shown on video. Yifan et al. (2004) reported on human action recognition research in a home setting, with the example of unscrewing caps and turning on switches. Yilma and Shah (2005) took a different direction in recognizing human action when they utilized moving video cameras, as opposed to the conventional still video cameras. Gorelick et al. (2007) reported on the usage of silhouettes as space-time action shapes in the recognition of walk related actions. Our work reported here compares favourably with that of Gorelick et al. as the covariants used for gait action recognition are somewhat similar (walking with bags and walking while wearing a piece of fabric). Shechtman and Irani (2007) also utilized space-time behaviour in detecting action, which interestingly negates the need for background subtraction. Another effort

* Corresponding author. Fax: +60 383125264.

E-mail addresses: ccho@mmu.edu.my (C.C. Ho), nghu@mmu.edu.my (H. Ng), twhaw@mmu.edu.my (W.-H. Tan), kwng@mmu.edu.my (K.-W. Ng), hltong@mmu.edu.my (H.-L. Tong), timothy@mmu.edu.my (T.Tzen-Vun Yap), pfchong@mmu.edu.my (P.-F. Chong), eswaran@mmu.edu.my (C. Eswaran), junaidi.abdullah@mmu.edu.my (J. Abdullah).

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.01.027

utilizing space-time features is reported in (Weinland et al., 2006), which describes how motion history volumes are used to recognize actions. Yan and Essa (2005) reports on the recognition of action in an occluded setting. Their work which utilizes a genetic algorithm, and focused on object tracking which is applied to human action recognition on the surface of a desk.

Schuldt et al.'s database for human action paved the way for many human action databases. The MERL Motion database (Wren et al., 2007) uses the user's infra red tracks as recorded by wireless sensor networks to approximate action. Another sensor enabled action database can be found in (Choudhury et al., 2008) which makes use of a myriad of sensors (temperature, accelerometer, audio, light, pressure, digital compass) to represent human actions. Another effort utilizes radio frequency identification (RFID) to recognize human action, as reported in (Patterson et al., 2005). The MIT Place Lab is another database for human action recognition but with a twist: subjects actually carried out day-to-day action while living in the lab with attached accelerometer and hear rate monitors. Large scale human action databases can be found in (Oh et al., 2011) where the VIRAT video database is described. A multi camera view action database is presented in (Kulathumani, 2012) where 12 different actions are presented from many different camera angles in the WVU Multi View Action recognition database.

The motivation for creating MMU GASPFA (Gait–Speech–Face) database is to address both long standing research issues as well as to create an avenue for future and novel research. Previously, researchers who worked on multimodal biometric research involving the fusion of gait, face and speech were limited to using chimera combinations of modalities from different data sets. MMU GASPFA is a unique dataset that captures gait, speech and face data from the same set of subjects, thus negating the need for chimera subjects. At the same time, advances in sensor technology (miniaturized accelerometers in smart phones) and new COTS sensors such as the Kinect camera have provided the means to record and analyze gait data from a new point of view. The combination of video analysis (from conventional digital video cameras and the Kinect video camera) and sensor analysis presents a new area of research particularly for action recognition (in the form of gait-related recognition). Thus we believe that the MMU GASPFA database is relevant for supporting ongoing research as well as future research in action recognition. A thorough discussion of the novelty of the MMU GASPFA database will be presented in the discussion section of this paper.

2. Related works

2.1. Commercial off the shelf (COTS) biometric sensors and databases

COTS biometric databases are defined by both Indovina et al. (2003) and Snelick et al. (2005) as biometric databases which were gathered using state-of-art biometric sensors on a large population of subjects. Sensors used in the recording of the MMU GASPFA database are available commercially off the shelf, and are not purpose built for the acquisition of biometric data. COTS biometric sensors are attractive due to the following reasons:

- *Cost:* COTS sensors used in the recording of the MMU GASPFA database are readily available on the open market, and are priced competitively as many of these sensors are targeted at the consumer market.
- *Support:* The COTS sensors used in the recording of the MMU GASPFA database are widely supported by reputable vendors, thus ensuring minimal downtime in the event of a sensor failure.
- *Availability:* The COTS sensors used in this work are easily available, and there is no significant lag time for vendors to fulfil the purchase order. This might not be the case for specialized biometric sensors, which may have limited production capability and scope of availability. Readily available COTS sensors also allow the repetition of the experiments conducted as described by this paper.
- *Ease of use and integration:* The consumer targeted sensors used in this work are easy to operate, and supports commonly used data and file formats. This allows for data that are gathered to be quickly stored and processed. COTS sensors can also be easily integrated for a complete acquisition system.

The factors listed above contributed tremendously in the choice of sensors selected for the acquisition of the MMU GASPFA database. This offsets the inherent weaknesses of using non specialized biometric sensors.

2.2. The case for face, speech and gait multimodal biometric combination

Face, speech and gait biometric modality can all stand on their own as robust biometric. Face biometric is arguably the most mature of the three, having seen very good results being achieved in all manners of applications. It is also the most natural biometric, appealing to the visual sense of the human perception system. Face as a biometric has found its way into consumer products, ranging from face detectors in digital cameras, all the way to face authentication systems deployed on computers and smart phones. Speech as a biometric modality is almost as mature as the face biometric modality. Similarly, speech biometric has seen evidence of high usability in many different application domains. The usage of speech as a biometric is less pervasive in consumer products, with more focus on speech recognition as compared to speaker identification or recognition. Gait as a biometric is the modality with the most potential for research to be done, as it is a relatively less mature field compared to face and speech modality. Video based analysis of gait has yielded high recognition rates even as the trend turns towards sensor driven gait recognition. Gait as a biometric has also not been widely implemented in consumer devices, as the preference has been more of tracking of a person for feedback and gaming purposes.

Face and speech as a bimodal combination work well due to the inherent relationship that exists between face and speech. This is especially true when local features of the face are used as features, as the shape of the mouth when a person speaks varies over the time it takes for the speech utterance to be over. Millar et al. (2004) summarizes the protocols to be followed during the acquisition of a face and speech corpus, whilst also highlighting the availability of popular benchmarked bimodal face and speech databases.

Face and gait biometrics have been used extensively in an outdoor setting, where environmental factors such as glare may impede on the ability to recognize a person based on facial features as shown in (Liu and Sarkar, 2007). Much research work has not been done on the Gait–speech combination and the practical usefulness of such a combination is yet to be proven. One possible research direction lies in medical informatics, as shown in (Cantiniaux et al., 2009) where possible correlations between slurred speech and gait disorders in response to stimuli and medication were investigated.

The motivation for combining gait, speech and face biometric modalities thus lies in how these three modalities can compensate for one another in situations where performance of any one of the modalities is lacking. The combination of two strong and mature modalities with one less proven modality offers further research