Multibiometrics for Human Identification

April/May 1997

[Nixon et. al]
Sina Samangooei
John D. Bustard
Richard D. Seely
Mark S. Nixon
John N. Carter

Chapter 1

On Acquisition and Analysis of a Dataset Comprising of gait, ear and semantic data

1.1 Introduction

1.1.1 Multi-biometrics

With the ever increasing demand for security and identification systems, the adoption of biometric systems is becoming widespread. There are many reasons for developing multibiometric systems, for example, a subject may conceal or lack the biometric a system is based on. This can be a significant problem with non-contact biometrics in some applications (e.g. Surveillance). Many non-contact biometric modalities exist. Of these face recognition has been the most widely studied, resulting in both its benefits and drawbacks being well understood. Others include gait, ear and soft biometrics. Automatic gait recognition is attractive because it enables the identification of a subject from a distance, meaning that it will find applications in a variety of different environments [29]. The advantage of the ear biometric is that the problems associated with age appear to be slight, though enrolment can be impeded by hair [14]. There are also new approaches to using semantic descriptions to enhance biometric capability, sometimes known as soft biometrics [35]. The semantic data can be used alone, or in tandem with other biometrics, and is suited particularly to analysis of surveillance data.

The deployment of mutilbiometric systems is largely still at a research phase [33]. Of the biometrics discussed here, there have been approaches that fuse face with gait [22, 38, 42, 45] and fuse ear with face [8]. There has been no approach which

fuses gait and ear data. One of the first biometric portals was based on iris data and mentioned use of face in early marketing material, but the current literature does not mention this [24]. In order to assess recognition capability, ideally we require a database wherein the biometrics were recorded concurrently, though it appears acceptable to consider disparate sets of data (different biometrics acquired at different times), especially when seeking to establish fusion performance, rather than biometric performance. Naturally, when the effect of age is the target of analysis, then concurrent acquisition of multiple biometrics will be the only practicable way to handle what is otherwise an enormous and challenging metadata labelling and reconciliation approach.

1.1.2 Multibiometric Data

There have been many calls for acquisition of multiple biometric data dating from the inception of biometrics [6]. The major multimodal databases currently available include the XMVTS, BANCA, WVU and MBGC databases.

The XM2VTSDB multi-modal face database project contains four recordings of 295 subjects taken over a period of four months [25]. Each recording contains a speaking head shot and a rotating head shot. Sets of data taken from this database are available including high quality colour images, 32 KHz 16-bit sound files, video sequences and a 3D Model.

The BANCA database is a large, realistic and challenging multi-modal database intended for training and testing multi-modal verification systems [32]. The BANCA database was captured in four European languages in two modalities (face and voice). For recording, high and low quality microphones and cameras were used. The subjects were recorded in three different scenarios, controlled, degraded and adverse over 12 different sessions spanning three months. In total 208 people were captured, half men and half women.

The WVU multimodal biometric dataset collection, BIOMDATA collects iris, fingerprint, palm-print, voice and face data from over 200 people [33]. The data was collected using standard enrolment devices, where possible, such as the SecuGen optical fingerprint biometric scanner, the OKI IRISPASS-h handheld device for the iris, and the IR Recognition Systems HandKey II for hand geometry with image and sound recordings for face and voice, respectively. The dataset also includes soft biometrics such as height and weight, for subjects of different age groups, ethnicity and gender with variable number of sessions/subject.

The Multiple Biometric Grand Challenge (MBGC) data build on the data-challenge and evaluation paradigm of FRGC, FRVT 2006, ICE 2005 and ICE 2006, and address requirements which focus on biometric samples taken under less than ideal conditions. As such, the data includes low quality still images; high and low quality video imagery; and face and iris images taken under varying illumination conditions as well as off-angle and occluded images. There is no established literature yet, but there is an extensive website with many presentations, especially from the early (recent) workshops [1]. The primary goal

[1] http://face.nist.gov/mbgc/

of the MBGC is to investigate, test and improve performance of face and iris recognition technology on both still and video imagery through a series of challenge problems and evaluation. The MBGC seeks to reach this goal through several technology development areas:

1. face recognition on still frontal, real-world-like high and low resolution imagery;

2. iris recognition from video sequences and off-angle images;

3. fusion of face and iris (at score and image levels);

4. unconstrained face recognition from still and video imagery;

5. recognition from Near Infrared (NIR) & High Definition (HD) video streams taken through portals;

6. unconstrained face recognition from still images and video streams.

One of the purposes of the data is for fusion of face and iris as subjects walk through a biometric portal which is a likely deployment scenario for biometrics.

The Biosecure database[2] is now available. The databases include hand, iris, signature, fingerprint, still face, and audio video for around 200 to 300 subjects [11]. The main characteristics of the databases accommodate different application scenarios as:

1. an internet dataset (PC-based, on-line, internet environment, unsupervised conditions) with voice and face data,

2. a desktop dataset (PC-based, off-line, desktop environment, supervised conditions), including voice, face, signature, fingerprint, hand and iris; and

3. a mobile dataset (mobile device-based, indoor/outdoor environment, uncontrolled conditions), including voice, face, signature and fingerprint.

None of these databases include specific concentration on gait and ear. There are separate databases available for these biometrics. In order to advance our research agenda in gait, ear and semantic biometrics, and to further our investigations into the effects of covariates (exploratory variables) on performance, we sought to acquire a database which included face, gait and ear, as well as to investigate, via semantic data, potential relating to surveillance applications.

1.1.3 Non-Contact Biometrics

In this section we discuss the biometrics gathered by our new database.

[2] http://biosecure.it-sudparis.eu/AB

Gait Biometrics

Gait as a biometric can be used alone, to cue acquisition of other biometrics, or fused with other biometric data. It is suited to deployment in portals, since this is where a subject must walk through.

There have been many previous approaches to gait which rely on data where a subject walks in a plane normal to the camera's view[29]. These offer encouragement to the use in a portal arrangement since with laboratory data recognition performance approaches that of many other biometrics. There are several datasets recording such data in indoor and outdoor scenarios, in particular the HumanID [36], CASIA [7] and Southampton dataset [37]. There is also multiview data available, and so there are also view dependent and viewpoint invariant approaches. There has been some work on fusing gait with other biometrics[22, 39], particularly faces though there has been none fusing gait and ear/face. There has been little work on recognition in pure 3D, which our dataset allows.

Ear Biometrics

Ears are a particularly appealing approach to non-contact biometrics because they are unaffected by expressions and vary less with age when compared to faces. Also, reported levels of recognition are promising [13]. Although automated ear biometrics is a relatively recent development, the use of ears for forensics dates back to the 1800s when they formed part of the system developed by Alphonse Bertillon [3]. However, it was not until 1955 that a criminologist, Alfred Iannarelli, developed a practical recognition process based solely on the ear [15]. In developing this process, he gathered and analysed over 10,000 ear photographs to demonstrate they could be used for accurate recognition. Like fingerprints, ear prints have been used in the police service as a forensic tool, and in 1967 their analysis provided key evidence in a criminal prosecution [30]. Ear prints have continued to be used in cases as recently as 2008. However, at least one conviction has been overturned on appeal due to insufficient ear print quality [1].

In 1998, Burge and Burger [4] proposed the first computerised ear recognition system. Although their paper had no recognition results, it led to a range of further studies into the effectiveness of ears as a biometric. Many approaches have been used to achieve accurate recognition on small collections of ear images taken under controlled conditions. Recent work has focused on improving the robustness to achieve recognition in less constrained environments which contain, background clutter, occlusion and lighting and pose variation [5].

Semantic Biometrics

The description of humans based on their physical features has been explored for several purposes including medicine[34], biometric fusion [17], eyewitness analysis [20] and human identification [16]. Descriptions chosen vary in levels of visual granularity and include visibly measurable features but also those measurable

only using specialised tools. One of the first attempts to systematically describe people for identification based on their physical traits was the anthropometric system developed by Bertillon [3] in 1896. His system used eleven precisely measured traits of the human body including height, length of right ear and width of cheeks. This system was quickly superseded by other forms of forensic analysis such as fingerprints. More recently, description of anthropometric traits have been used along side primary biometrics in *soft biometric fusion* to improve recognition rates [18, 28, 41, 44]. Jain et al. [17] present an example where, using a general bayesian framework, they fuse fingerprints with the soft features of gender, ethnicity and height, achieving improved identification rates.

Meaningful words (semantic terms) humans use to describe one another by their visually discernible traits can also be used as a soft biometric. In the Southampton Multi-Biometric Tunnel, selected semantic terms describing visual traits of subjects are collected from human observers. The traits described are those discernible by humans at a distance, complementing the primary biometrics gathered in the Multi-Biometric Tunnel (i.e. gait, face and ear). Furthermore, the traits and descriptive terms are chosen for their *consistent* and *accurate* mention by humans in various scenarios [35].

1.1.4 On our New Database

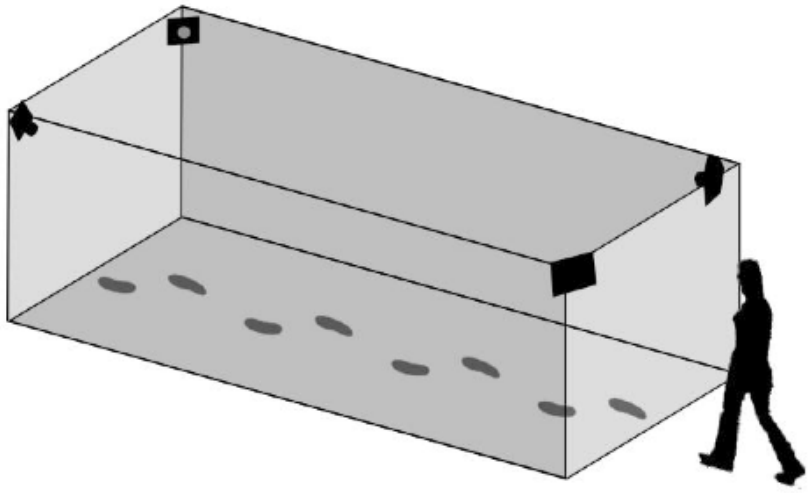

Figure 1.1: A controlled environment with fixed cameras provides an ideal scenario for automatic gait recognition. The subject is constrained to walk through the middle; controlled lighting and background facilitate analysis.

In outdoor scenarios such as surveillance where there is very little control

over the environments, complex computer vision algorithms are often required for analysis. However constrained environments, such as walkways in airports where the surroundings and the path taken by individuals can be controlled, provide an ideal application for such systems. Figure 1.1 depicts an idealised constrained environment. The path taken by the subject is restricted to a narrow path and once inside is in a volume where lighting and other conditions are controlled to facilitate biometric analysis. The ability to control the surroundings and the flow of people greatly simplifies the computer vision task, compared to typical unconstrained environments. Even though biometric datasets with greater than one hundred people are increasingly common, there is still very little known about the inter and intra-subject variation in many biometrics. This information is essential to estimate the recognition capability and limits of automatic recognition systems. In order to accurately estimate the inter- and the intra- class variance, substantially larger datasets are required [40]. Covariates such as facial expression, headwear, footwear type, surface type and carried items are attracting increasing attention; although considering the potentially large impact on an individuals biometrics, large trials need to be conducted to establish how much variance results.

This chapter is the first description of the multibiometric data acquired using the University of Southampton's Multi-Biometric Tunnel [26, 37]; a biometric portal using automatic gait, face and ear recognition for identification purposes. The tunnel provides a constrained environment and is ideal for use in high throughput security scenarios and for the collection of large datasets. We describe the current state of data acquisition of face, gait, ear, and semantic data and present early results showing the quality and range of data that has been collected. The main novelties of this dataset in comparison with other multi-biometric datasets are:

1. gait data exists for multiple views and is synchronised, allowing 3D reconstruction and analysis;
2. the face data is a sequence of images allowing for face recognition in video;
3. the ear data is acquired in a relatively unconstrained environment, as a subject walks past; and
4. the semantic data is considerably more extensive than has been available previously.

We shall aim to show the advantages of this new data in biometric analysis, though the scope for such analysis is considerably greater than time and space allows for here.

1.2 Data Collection

The main components of our new multibiometric database are separate, synchronised and integratable sample databases of gait, face, ear and semantic

descriptions. Gait samples are from 12 (and some early experiments using 8) multiview overhead cameras and suitable for per camera analysis as well as 3D reconstruction. Face samples are taken as the user walks down the tunnel; resulting in a sequence of frames per sample, where the subject's face is automatically extracted and the background removed. Example face samples can be seen in Figure 1.2. Ear samples are comprised of a single snapshot, one taken per gait sample. Finally, semantic descriptions of subjects in the form of self annotations and observed descriptions are captured on a subset of the subjects. The exact contents of these datasets is summarised in Tables 1.1− 1.4. Note, in the description of the gait database contents, the total number of unique subjects is 192 which is less than the subjects recorded. This is due to some subjects providing repeat samples.

Table 1.1: Gait Dataset Samples

Total Sequences	2070 *(∼ 84 invalid)*
Total Subjects	192 *(∼ 5 invalid subjects)*
Average Sequences/Subject	10
12 sensors	895 samples across 89 subjects *(∼ 31 invalid sequences from 4 subjects)*
8 sensors	1175 samples across 117 subjects *(∼ 53 invalid sequences from 5 subjects)*
repeat walks	120 samples across 12 subjects

1.2.1 Gait Tunnel

The Multi-Biometric Tunnel is a unique research facility situated in the University of Southampton, it has been specifically designed as a non-contact biometric access portal [26], providing a constrained environment for people to walk through, whilst facilitating recognition. The system has been designed with

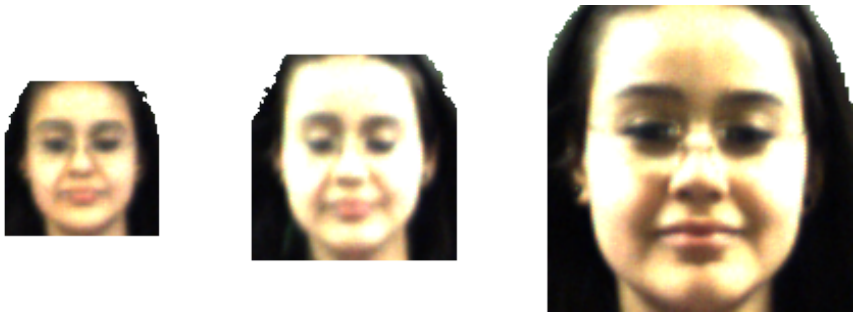

Figure 1.2: Example face samples. All taken from the same gait sequence showing the 1st (76×76), 11th (101×101) and 21st (150×150) samples out of 24.

Table 1.2: Ear Dataset Samples

Total Samples	2070
Total Subjects	192
Average Samples/Subject	10
Completely Occluded Ears	49
Occlusion due to hair	45
Occlusion due to hats	4

Table 1.3: Face Dataset Samples

Total Sequences	2070
Total Subjects	192
Average Sequences/Subject	10
Average Face Frames/Sequence	31.8 *(min = 6 and max = 48 depending of speed of walk and subject height)*

Table 1.4: Semantic Dataset Samples

Total annotations	2828
Total Self annotations	193
Total observed annotations	2635
In set 1	1367 *(∼ 93 users of 15 subjects)*
In set 2	845 *(∼ 59 users of 15 subjects)*
In set 3	288 *(∼ 22 users of 15 subjects)*
In set 4	135 *(∼ 9 users of 15 subjects)*

airports and other high throughput environments in mind, where contact based biometrics would prove impractical. Such a system could be setup in a very unobtrusive manner where individuals might not even be aware of its presence. It also enables the automated collection of large amounts of non-contact biometric data in a fast and efficient manner, allowing very large datasets to be acquired in a significantly shorter timeframe than previously possible.

The Multi-Biometric Tunnel is able to detect the entry and exit of a subject, allowing a high degree of automation. Whilst a subject is inside the tunnel their gait is recorded by 12 Point Grey Dragonfly cameras, allowing the reconstruction of 3D volumetric data. The gait cameras all have a resolution of 640×480 and capture at a rate of 30 FPS (frames per second), they are connected together over an IEEE1394 network employing commercial synchronisation units [3] to ensure accurate timing between cameras. Figure 1.3 shows a single frame as captured by the cameras. Video is also captured of the subject's face and upper body using a high resolution (1600×1200) IEEE1394 camera, enabling face

[3] PTGrey Camera Synchronization unit, Part No. SYNC http://www.ptgrey.com

Figure 1.3: Synchronised images captured by gait cameras

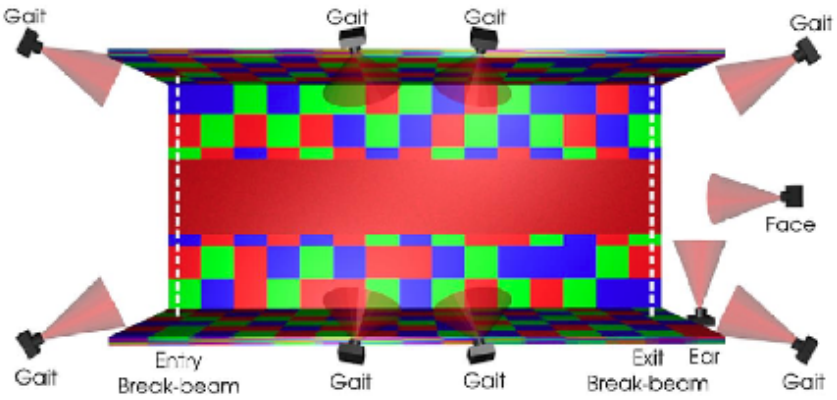

Figure 1.4: Placement of cameras and break-beam sensors in system

recognition. A single snapshot is taken as the subject exits the tunnel, of the side of the subject's head, for ear biometrics. As shown in Figure 1.4, the facility has a central region that participants walk along, with the face and ear cameras placed at the end of the walkway and the gait video cameras positioned around the upper perimeter of the tunnel. The walls of the tunnel are painted with a non-repeating rectangular pattern to aid automatic camera calibration. Saturated colours have been chosen to ease background/foreground separation. These choices are mandated by the nature of the facility. After the subject has walked ten times through the tunnel, taking on average five minutes per person, they are asked to record semantic data associated with the database which included questions about gender, age, ethnicity and physical parameters.

Upon arrival, the purpose and procedure for database acquisition was explained to each potential participant and on agreement they signed a consent form to confirm that they were willing to participate in the experiment. In

order to ensure privacy, the consent forms had no unique identifiers, and as such they are the only record of the participant's identity. Each participant was asked to choose a unique identifier at random, which could then be associated with any data collected from that individual. Before commencing the experiment, each subject was asked to walk through the tunnel as a trial run, this was not recorded and was watched by the supervisor to ensure that the subject understood their instructions. Normally, subjects were not supervised, aiming to collect a natural gait and facial expression.

The tunnel is equipped with a status light mounted outside of the visible area; participants were asked to wait until it indicated that the system was ready. Before each sample the gait and face cameras captured one second of video footage whilst the tunnel area was empty; this was used later for the background estimation and subtraction. Upon entering the tunnel, the subject would walk through a break-beam sensor, starting the capture process. Towards the end of the tunnel another breakbeam sensor stopped the capture process. After capture, the recorded data was saved (unprocessed) to disk. The entire process of induction, walking through the tunnel and answering questions took on average 30 minutes per participant. The result is that data is collected for 3 non-contact biometrics from camera sensors synchronised using commercial IEEE-1394 bus synchronisation devices. Before describing analysis of the biometric data and its fusion, we shall describe some especial considerations of the separate biometrics.

Gait Data

The volume of data acquired when subjects walk through the tunnel currently forces the acquisition procedure to store the unprocessed data straight to disk and further processing is conducted afterwards. With modern processors and storage, the recognition process can actually be complete a few steps after the subject has exited the tunnel, and we have performed this for BBC (UK, 2009). Our purpose here is more the use of the tunnel to acquire a database, and for this purpose, the images from the multiple gait cameras are reconstructed into a 3D silhouette which can then be viewed from any angle. There are several stages to processing the gait video data. Separate background and foreground images are used to facilitate background subtraction and shadow suppression. This is followed by some post-processing using simple morphological operators to clean up the silhouette data. The resulting silhouettes are corrected for radial distortion and then used as the basis for shape from silhouette reconstruction. Shape from silhouette reconstruction is simply the calculation of the intersection of projected silhouettes, see Figure 1.5, and it can be expressed mathematically as:

$$V(x,y,z) = \begin{cases} 1 & \text{if } \Sigma_{i=n}^{N} I_n\left(M_n(x,y,z)\right) \geq k \\ 0 & \text{otherwise} \end{cases}$$

Where V is the derived 3D volume, k is the number of cameras required for a voxel to be marked as valid and N is the total number of cameras. I_n is the silhouette image from camera n where $I_n(u,v) = 0, 1$, and $M_n(x,y,z:u,v)$ is a

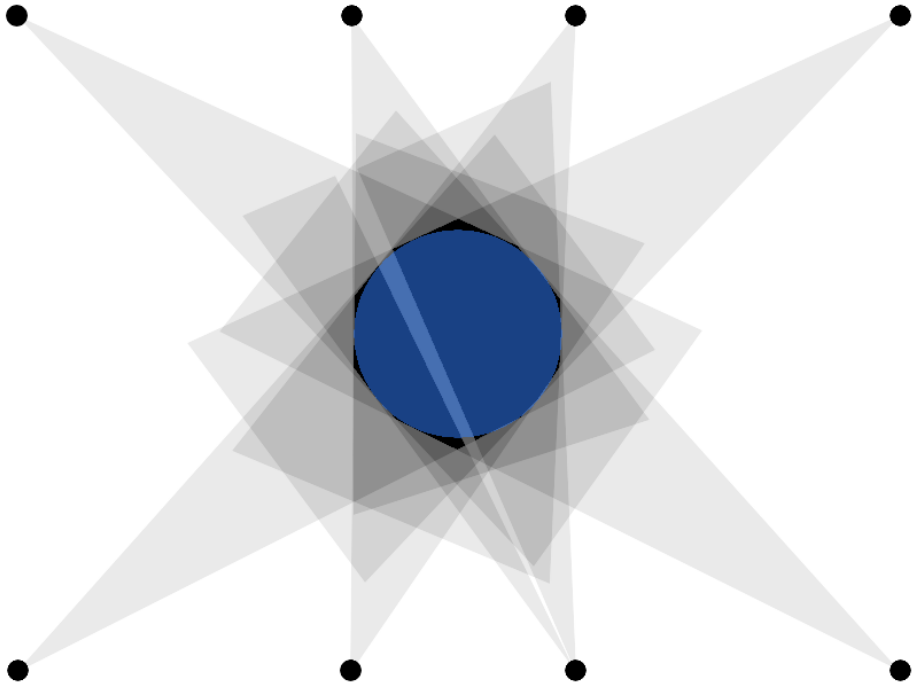

Figure 1.5: 3D reconstruction is performed by taking the intersection of the re-projected silhouettes from each camera

Figure 1.6: A three-dimensional volumetric frame created by the Multi-Biometric tunnel using shape from silhouette reconstruction

function that maps the three-dimensional world coordinates to the coordinate system of camera n. M_n is calculated using the calibration information derived for each camera. In a conventional implementation of shape from silhouette, a voxel may only be considered valid if all cameras have silhouette pixels at its location; therefore $k = N$ must be satisfied. Using a value of k that is lower than N results in a less selective criteria, which adds a degree of resilience against background segmentation errors; although the reconstructed shape is not as accurate. The use of high intensity colours for the background means that very little segmentation error occurs, allowing the use of a $k = N$ criteria. A small amount of post-processing is carried out on the resulting 3D volumes using binary morphology to improve the accuracy of the reconstructed volumes. An example volume created by shape from silhouette reconstruction is shown in Figure 1.6.

Figure 1.7: Example ear sample

Ear Data

To record the ear a digital photograph was taken when a subject passed through the light beam at the end of the tunnel. The camera uses a wide field of view to ensure ears were visible with a large range of subject heights and walking speeds. The photograph was taken with a high shutter speed to minimise motion blur. In addition, two flash cameras were used to provide sufficient light for a high shutter speed and reduce any shadows caused by hair or headgear. The flash guns were positioned to point from above and below the ear.

It should also be noted that subjects were not instructed to explicitly reveal

their ears. This was in order to record subjects with a realistic degree of ear occlusion, representative of a real usage scenario. Of the 187 subjects recorded by the system, 6% walked too fast to be captured by the ear camera. A further 26% had their ears completely obscured by hair or headgear. We intend to include metadata in the database indicating where this occurs. An example ear from the dataset can be seen in Figure 1.7.

Semantic Data

The collection of semantic terms is integrated with the Southampton Multi-Biometric Tunnel. Participants are asked to annotate themselves and a set of 15 other subjects according to a set of traits using a set of predefined terms, listed in Table 1.5.

The annotation gathering process was designed carefully to avoid (or allow the future study of) inherent weaknesses and inaccuracies present in human generated descriptions. The error factors that the system was designed to deal with include:

- **Memory[10] -** Passage of time may affect a witness' recall of a subject's traits. Memory is affected by variety of factors e.g. the construction and utterance of featural descriptions rather than more accurate (but indescribable) holistic descriptions. Such attempts often alter memory to match the featural descriptions.

- **Defaulting[21] -** Features may be left out of descriptions in free recall. This is often not because the witness failed to remember the feature, but rather that the feature has some default value. Race may be omitted if the crime occurs in a racially homogenous area, Sex may be omitted if suspects are traditionally Male.

- **Observer Variables[12][31] -** A person's own physical features, namely their self perception and mental state, may affect recall of physical variables. For example, tall people have a skewed ability to recognise other tall people but will have less ability when it comes to the description shorter individuals, not knowing whether they are average or very short.

- **Anchoring[9] -** When a person is asked a question and is initially presented with some default value or even seemingly unrelated information, the replies given are often weighted around those initial values. This is especially likely when people are asked for answers which have some natural ordering (e.g. measures of magnitude)

The data gathering procedure employed in the tunnel was designed to account for all these factors. Memory issues are addressed by allowing annotators to view videos of subjects multiple times, also allowing them to repeat a particular video if necessary. Defaulting is avoided by explicitly asking individuals for each chosen trait meaning that even values for apparently *obvious* traits are captured. This style of interrogative description where constrained responses

Table 1.5: Physical traits and associated semantic terms

Body	
0. Arm Length	(0.1) Very Short (0.2) Short (0.3) Average (0.4) Long (0.5) Very Long
1. Arm Thickness	(1.1) Very Thin (1.2) Thin (1.3) Average (1.4) Thick (1.5) Very Thick
2. Chest	(2.1) Very Slim (2.2) Slim (2.3) Average (2.4) Large (2.5) Very Large
3. Figure	(3.1) Very Small (3.2) Small (3.3) Average (3.4) Large (3.5) Very Large
4. Height	(4.1) Very Short (4.2) Short (4.3) Average (4.4) Tall (4.5) Very Tall
5. Hips	(5.1) Very Narrow (5.2) Narrow (5.3) Average (5.4) Broad (5.5) Very Broad
6. Leg Length	(6.1) Very Short (6.2) Short (6.3) Average (6.4) Long (6.5) Very Long
7. Leg Direction	(7.1) Very Bowed (7.2) Bowed (7.3) Straight (7.4) Knock Kneed (7.5) Very Knock Kneed
8. Leg Thickness	(8.1) Very Thin (8.2) Thin (8.3) Average (8.4) Thick (8.5) Very Thick
9. Muscle Build	(9.1) Very Lean (9.2) Lean (9.3) Average (9.4) Muscly (9.5) Very Muscly
10. Proportions	(10.1) Average (10.2) Unusual
11. Shoulder Shape	(11.1) Very Rounded (11.2) Rounded (11.3) Average (11.4) Square (11.5) Very Square

Global	
12. Weight	(12.1) Very Thin (12.2) Thin (12.3) Average (12.4) Big (12.5) Very Big
13. Age	(13.1) Infant (13.2) Pre Adolescence (13.3) Adolescence (13.4) Young Adult (13.5) Adult (13.6) Middle Aged (13.7) Senior
14. Ethnicity	(14.1) European (14.2) Middle Eastern (14.3) Indian/Pakistan (14.4) Far Eastern (14.5) Black (14.6) Mixed (14.7) Other
15. Sex	(15.1) Female (15.2) Male

Head	
16. Skin Colour	(16.1) White (16.2) Tanned (16.3) Oriental (16.4) Black
17. Facial Hair Colour	(17.1) None (17.2) Black (17.3) Brown (17.4) Red (17.5) Blond (17.6) Grey
18. Facial Hair Length	(18.1) None (18.2) Stubble (18.3) Moustache (18.4) Goatee (18.5) Full Beard
19. Hair Colour	(19.1) Black (19.2) Brown (19.3) Red (19.4) Blond (19.5) Grey (19.6) Dyed
20. Hair Length	(20.1) None (20.2) Shaven (20.3) Short (20.4) Medium (20.5) Long
21. Neck Length	(21.1) Very Short (21.2) Short (21.3) Average (21.4) Long (21.5) Very Long
22. Neck Thickness	(22.1) Very Thin (22.2) Thin (22.3) Average (22.4) Thick (22.5) Very Thick

are explicitly requested is more complete than free-form narrative recall but may suffer from inaccuracy, though not to a significant degree [43]. Subject

variables can never be completely removed so instead we allow the study of differing physical traits across various annotators. Users are asked to self annotate based on self perception, also certain subjects being annotated are themselves annotators. This allows for some concept of the annotator's own appearance to be taken into consideration when studying their descriptions of other subjects. Anchoring can occur at various points of the data capture process. We have accounted for anchoring of terms gathered for individual traits by setting the default term of a trait to a neutral "Unsure" rather than any concept of "Average".Table 1.4 shows the current annotations collected in this manner from the Southampton Multi-Biometric Tunnel

1.3 Recognition

There is considerable scope afforded by this data for analysis of recognition potential. We have yet to analyse performance of all the data in a fusion schema, and we have yet to analyse face recognition performance alone. In concert with our research agendas in new approaches to gait, ear and semantic biometrics, we have addressed:

1. gait recognition in 3D
2. robust ear recognition using a planar approximation
3. recognition and recall by semantic labels

We have also shown how fusion of this data can achieve significantly improved performance over the single data thus demonstrating the capability of this new dataset to support fusion as well as individual biometric recognition.

1.3.1 Gait

Since gait is a periodic signal, we only consider one period for analysis; this is the image samples taken between heel strike of one foot until the next heel strike of the same foot. An automatic process was used to locate a complete gait cycle, this was achieved by analysing the variation in the size of the subject's bounding box. Several different variants of the average silhouette gait analysis technique were used to evaluate the dataset collected from the Multi-Biometric tunnel; the normalised side-on average silhouette, the (non-normalised) side-on average silhouette and the combination of side-on, front-on and top-down average silhouettes. The dataset used for analysis comprised of 187 subjects, where 85 subjects were viewed by 12 cameras and 103 subjects were viewed viewed by 8 cameras [37]. The set contained 2070 samples, of which 1986 were valid. Reasons for a sample being invalid include: clipping from where the subject was outside of the reconstruction area and the automatic gait cycle finder being unable to reliably identify a complete cycle. The database is made up of 76% male and 24% female subjects and the average age was 27 years. All three gait analysis techniques discussed below have some similarities with

the work of Shakhnarovich et al. [39], in that the 3D volumetric data is used to synthesise silhouettes from a fixed viewpoint relative to the subject. The resulting silhouettes are then analysed by using the average silhouette approach. The advantage of using three-dimensional data is that silhouettes from any arbitrary viewpoint can be synthesised, even if the viewpoint is not directly seen by a camera. For example, silhouettes from an orthogonal side-on viewpoint can synthesised from the volumetric data by:

$$J_i(y,z) = \bigcup_{x=x_{MIN}}^{x_{MAX}} V_i(x,y,z)$$

In other words, the side-on orthogonal viewpoint J_i for frame i is synthesised by taking the union of voxels in volume V_i along the x axis, where the x axis spans left to right, y spans front to back and z spans from the top to the bottom. In a similar manner, the front-on and top-down orthogonal viewpoints can be synthesised by taking the union of the voxels along the y or the z axis respectively. In the first analysis, silhouettes are taken from a side-on orthogonal viewpoint so that normal gait recognition could be assessed. This view is not seen by any camera and so can only be synthesised. The use of a side-on viewpoint facilitates comparison with previous results. The average silhouette is calculated wherein the centre of mass $C_i = (C_{i,x}, C_{i,y})$ is found for each frame i. This is calculated by rendering the 3D reconstruction to an image and estimating its center of mass by defining each image pixel within the silhouette to have a unit mass. The average silhouette is then found by summing silhouettes after they have been aligned using this value:

$$A(xy) = \frac{1}{M}\Sigma_{i=0}^{M-1} J_i(x - C_{i,x}, y - C_{i,y})$$

where A is the average silhouette and M is the number of frames in the gait cycle. The derived average silhouette is normalised in size so that it is 64 pixels high, whilst preserving the aspect ratio. The average silhouette is treated as the feature vector and used for leave-one-out recognition, using nearest-neighbour classification and the Euclidean distance as the distance metric between samples. A recognition rate of 97.9% was achieved. No feature-set transformation or selection was performed in this and subsequent analysis. This result is then similar in performance to current state-of-art approaches to gait biometrics, yet allows other views to be analysed in future. Because the silhouette data can be synthesised from an orthogonal viewpoint, the subject's distance from the viewpoint will not affect the silhouette size, thus meaning that scale normalisation is unnecessary and removes valuable information. For this reason a second analysis was conducted using non scale-normalised average silhouettes, the average silhouettes were downsampled by a factor of four to reduce the computational workload. The non-normalised average silhouette retains information such as the subject's build and height. The same viewpoint as the previous normalised variant was used, achieving an improved recognition rate of 99.8%.

The above analysis methods only utilise one viewpoint, meaning that very little of the additional information contained within the three-dimensional data

Table 1.6: Performance of Various Average Silhouette Signatures Measured using Equal Error Rate (EER) and Correct Classification Rate (CCR)

Average Silhouette	CCR	EER
Side (Scale-normalised)	97.9%	6.8%
Side	99.8%	1.8%
(Side, Front, Top)	100%	1.9%

was exploited. Therefore one additional analysis technique was performed, using non-normalised average silhouettes derived from three orthogonal viewpoints; side-on, front-on and top-down. The features from the three average silhouettes were simply concatenated and the resulting feature vector used for recognition, achieving an even better recognition rate of 100%. Again this is comparable with state-of-art approaches. Several different analysis methods have been carried out to evaluate the quality of the collected data. The correct classification rate and equal error rate was found for each analysis method and a summary of the results are presented in Table 1.6. The respective cumulative match scores are shown in Figure 1.8; it can be seen that the normalised average signature yields relatively poor performance, most likely due to the loss of information such as height and build. This is confirmed by the much improved classification performance of the non-normalised average silhouette. Classification performance using the concatenated average silhouettes proves better than both other methods, although the improvement in the equal error rate is marginal; this suggests that the additional information contained within three-dimensional data is useful for recognition.

In addition, ROC (receiver operating characteristic) curves demonstrating the system's capability to verify identity are shown in Figure 1.9. These confirm that normalised side-on average silhouettes are clearly inferior. However the situation is less clear between the other two cases, where the method using multiple viewpoints proves more selective than that of a single viewpoint. These results together suggest that the gait data alone is as worthy as a contender for evaluation of gait as a biometric, as it is in a multibiometric system.

1.3.2 Ear

The recorded ear images were used to recognise the subjects using the technique developed by Bustard and Nixon [5]. The technique uses SIFT feature points [23] to detect and align known samples of subject's ears with an image to be identified. SIFT points are a highly robust means of matching distinctive points between images. They define both a location, which includes a position, scale and orientation, and an associated signature calculated from the image region around the point. SIFT points have been shown to retain similar signatures under a wide range of variations, including pose, lighting, field of view and resolution [27]. Any four matching points between a gallery and probe image are sufficient to align and recognise an ear. This enables the technique

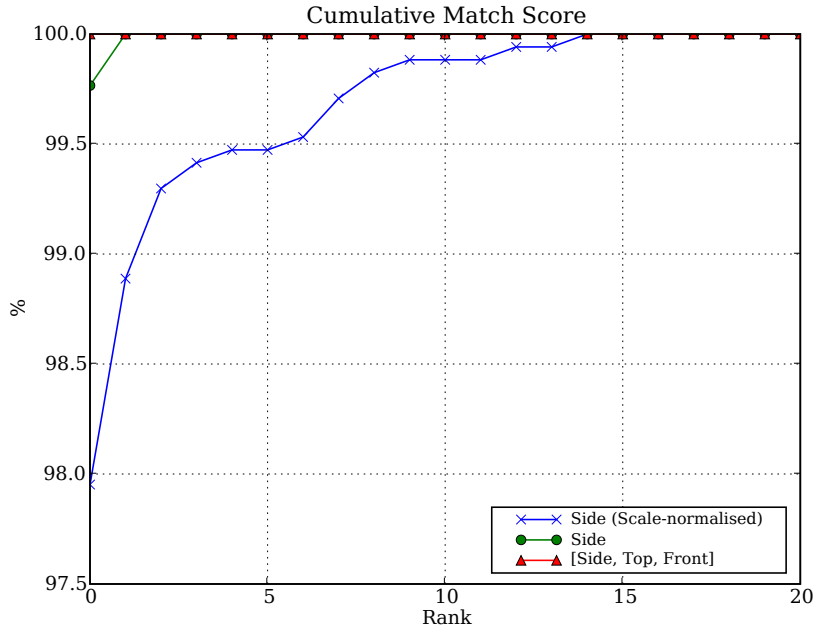

Figure 1.8: Cumulative match score plots for gait silhouettes derived from 3D

to remain accurate even when significantly occluded. In addition, by requiring that each point's relative location conforms to the same configuration as that in an existing ear gallery the detection is precise. Therefore, non ear images are rarely misclassified as ears. This precision is further enhanced by using the points to align the sample image with the unknown ear and robustly comparing the two images. If the images are not similar the match is rejected. The difference between the images then forms an estimate of similarity for matching ears and enables the most likely identity to be determined. These steps enable the algorithm to recognise ears accurately, even in the presence of brightness and contrast differences, image noise, low resolution, background clutter, occlusion and small pose variation. The ear recognition accuracy was evaluated using a "leave one out" strategy, with each image removed from the gallery and tested against the rest of the dataset in turn.

As can be seen in Figure 1.10, the rank 1 recognition performance for visible ears was 77%. This is lower than the performance in previous publications and reflects the less constrained ear images, which include a greater degree of occlusion (Figure 1.11) than the original publication.

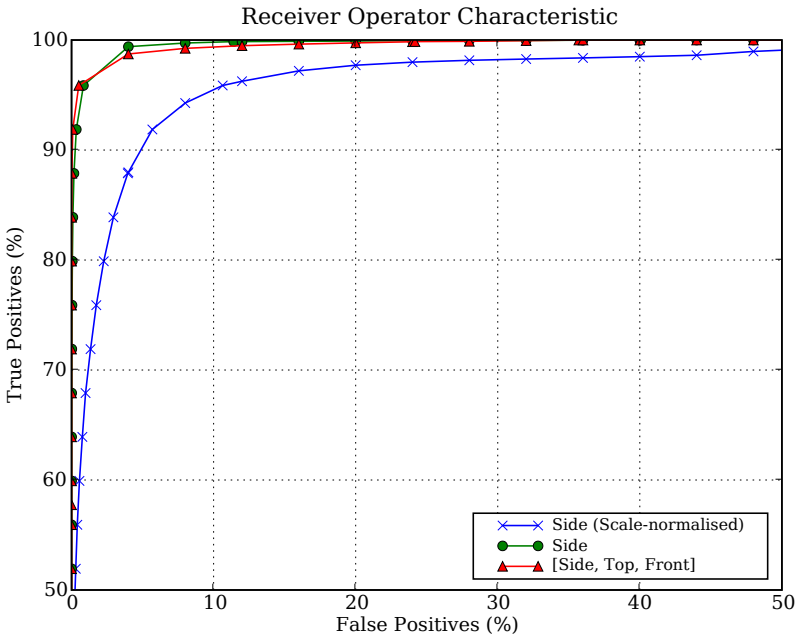

Figure 1.9: Receiver operating characteristic plots for gait silhouettes derived from 3D

1.3.3 Score Based Fusion

We have also performed a fusion approach to investigate the suitability of this data for fusion purposes. Score fusion was used to combine the multiple biometric recognition results. When high quality camera data is available, gait recognition provides almost perfect recognition performance making fusion unnecessary. This is useful when subjects are both recorded and recognised using the tunnel. When subjects are recorded by the tunnel but recognised using existing security cameras, the quality of recordings will be reduced. This can be simulated in the dataset by degrading the camera data to produce lower quality gait signatures. Under these circumstances performance is reduced making fusion desirable.

The distance measures returned by each algorithm were normalised using an estimate of the offset and scale of the measures between different subjects. For each algorithm these values were calculated using the mean and standard deviation of the distance between subjects in the gallery set. In addition, missing distance values were also estimated. If subjects walked too quickly to be captured by the ear camera their distance to each gallery probe was estimated to be the mean distance for all matched ears. Also when the recognition algorithm did not find a match between a gallery image and a probe, the distance measure

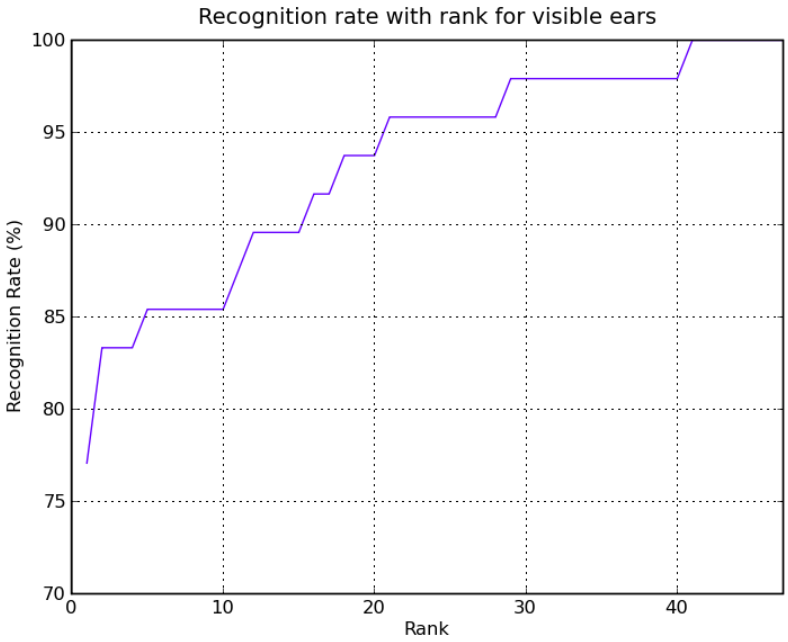

Figure 1.10: Ear Recognition Performance

Figure 1.11: Occluded Ears

was estimated to be the maximum value of the recorded distances. The normalised gait, semantic and ear data scores were then combined using the sum rule [19].

This fusion approach was evaluated using a "leave one out" recognition test. It was applied to a subset of the database that contained biometric data across all modalities. Figure 1.12 shows the recognition rates for each biometric and

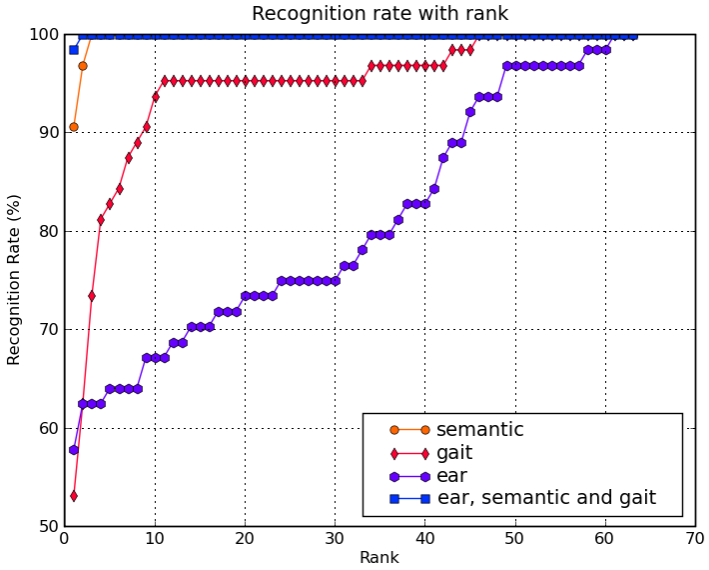

Figure 1.12: Results for individual techniques and complete fusion

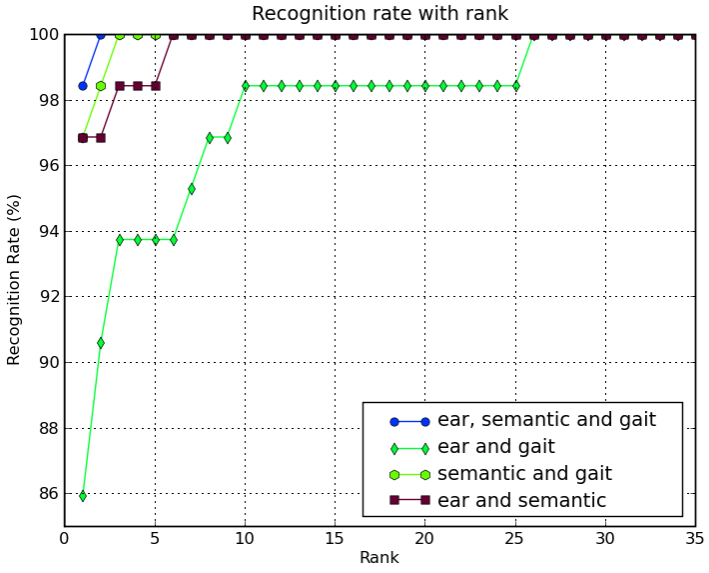

Figure 1.13: Results for all fusion combinations

the results obtained when their results are fused. Using the fusion algorithm a 98% rank 1 recognition rate can be obtained. Figure 1.13 shows the results of separately fusing each of the modalities, in all cases the recognition results improve significantly with fusion. As such, the semantic data can be used on its own or in fusion.

1.4 Conclusions

We have developed a new database for non-contact biometrics. The data is largely complete and we are currently finalising its distribution which will be arranged via our current gait database[4]. We already distribute one of the worlds largest gait databases [2] and we shall make our new database available there. The database comprises data recorded as subjects pass through the portal. The data includes sequences of face images, sequences of gait recorded from multiple synchronised cameras affording 3D data, ear images and semantic descriptions. We have shown already that this data is a promising avenue for investigation for non-contact biometrics, either alone or fused.

[4] www.gait.ecs.soton.ac.uk

References

[1] State vs David Wayne Kunze. Court of Appeals of Washington, Division 2., 1999.

[2] Aug. 2003. URL `http://www.gait.ecs.soton.ac.uk/`.

[3] A. Bertillon. *Instructions For Taking Descriptions For The Identification Of Criminals And Others, By Means Of Anthropometric Indications*. American Bertillon Prison Bureau, 1889.

[4] M. Burge and W. Burger. *Biometrics, Personal Identification in Networked Society: Personal Identification in Networked Society*, chapter Ear Biometrics. Kluwer Academic Publishers, 1998.

[5] J. D. Bustard and M. S. Nixon. Towards Unconstrained Ear Recognition from 2D Images. *Accepted for publication, IEEE Trans. SMC(A)*, 2009.

[6] J. Carter and M. Nixon. An integrated biometric database. In *IEE Colloq. on Electronic Images and Image Processing in Security and Forensic Science*, pages 4/1—4/5, 1990.

[7] CASIA. Casia gait database, http://www.sinobiometrics.com.

[8] K. Chang, K. Bowyer, S. Sarkar, and B. Victor. Comparison and combination of ear and face images in appearance-based biometrics. *IEEE Trans. PAMI*, 25(9):1160–1165, 2003.

[9] G. B. Chapman and E. J. Johnson. *Incorporating the irrelevant: Anchors in judgments of belief and value*, pages 120–138. Heuristics and Biases: The Psychology of Intuitive Judgment. Cambridge University Press, 2002.

[10] H. D. Ellis. *Practical aspects of facial memory*, section 2, pages 12–37. Eyewitness Testimony: Psychological perspectives. Cambridge University Press, 1984.

[11] J. Fierrez-Aguilar, J. Ortega-Garcia, D. T. Toledano, and J. Gonzalez-rodriguez. Biosec baseline corpus: A multimodal. *Pattern Recognition*, 40 (4):1389–1392, 2007.

[12] R. H. Flin and J. W. Shepherd. Tall stories: Eyewitnesses' ability to estimate height and weight characteristics. *Human Learning*, 5, 1986.

[13] D. J. Hurley, M. S. Nixon, and J. N. Carter. Force field feature extraction for ear biometrics. *Computer Vision and Image Understanding*, 98:491–512, November 2005.

[14] D. J. Hurley, B. Arbab-Zavar, and M. S. Nixon. *The Ear as a Biometric*, chapter 7. Handbook of Biometrics. Springer, 2008.

[15] A. Iannarelli. *Ear Identification*. Paramount Publishing Company, 1989.

[16] Interpol. Disaster Victim Identification Form (Yellow). booklet, 2008.

[17] A. Jain, S. Dass, and K. Nandakumar. Can soft biometric traits assist user recognition. In *Proc. SPIE*, 2004.

[18] A. K. Jain, K. Nandakumar, X. Lu, and U. Park. Integrating faces, fingerprints, and soft biometric traits for user recognition. In *Proc. BioAW*, pages 259–269, 2004.

[19] A. K. Jain, P. Flynn, and A. A. Ross. *Handbook of Biometrics*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.

[20] P. V. Koppen and S. K. Lochun. Portraying perpetrators; the validity of offender descriptions by witnesses. *Law and Human Behavior*, 21(6): 662–685, 1997.

[21] R. Lindsay, R. Martin, and L. Webber. Default values in eyewitness descriptions. *Law and Human Behavior*, 18(5):527–541, 1994.

[22] Z. Liu and S. Sarkar. Outdoor recognition at a distance by fusing gait and face. *Image Vision Comput.*, 25(6):817–832, 2007.

[23] D. G. Lowe. Object recognition from local scale-invariant features. *Computer Vision, IEEE International Conference on*, 2:1150, 1999.

[24] J. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D. LoIacono, S. Mangru, M. Tinker, T. Zappia, and W. Zhao. Iris on the move: Acquisition of images for iris recognition in less constrained environments. *Proceedings of the IEEE*, 94(11):1936–1947, 2006.

[25] K. Messer, J. Matas, J. Kittler, and K. Jonsson. XM2VTSDB: The extended M2VTS database. In *AVBPA*, pages 72–77, 1999.

[26] L. Middleton, D. K. Wagg, A. I. Bazin, J. N. Carter, and M. S. Nixon. Developing a non-intrusive biometric environment. In *IEEE Conf. IROS*, 2006.

[27] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. PAMI*, 27(10):1615–1630, 2005.

[28] K. Nandakumar, S. C. Dass, and A. K. Jain. Soft biometric traits for personal recognition systems. In *Proc. ICBA*, pages 731–738, 2004.

[29] M. S. Nixon, T. N. Tan, and R. Chellappa. *Human Identification based on Gait*. International Series on Biometrics. Springer, 2005.

[30] J. W. Osterburgh. *Crime Laboratory*. Paramount Publishing Company, 1989.

[31] A. J. O'Toole. Psychological and Neural Perspectives on Human Face Recognition. In *Handbook of Face Recognition*. Springer-Verlag, 2004.

[32] V. Popovici, J. Thiran, E. Bailly-Bailliere, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariethoz, J. Matas, K. Messer, B. Ruiz, and F. Poiree. The BANCA Database and Evaluation Protocol. In *Proc. AVBPA*, volume 2688, pages 625–638, 2003.

[33] A. A. Ross, K. Nandakumar, and A. K. Jain. *Handbook of Multibiometrics (International Series on Biometrics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[34] C. Rosse and J. L. V. Mejino. A reference ontology for biomedical informatics: the foundational model of anatomy. *J. of Biomed. Informatics*, 36 (6):478–500, 2003.

[35] S. Samangooei, B. Guo, and M. S. Nixon. The use of semantic human description as a soft biometric. In *Proc. IEEE BTAS*, 2008.

[36] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2): 162–177, 2005.

[37] R. D. Seely, S. Samangooei, L. Middleton, J. N. Carter, and M. S. Nixon. The University of Southampton Multi-Biometric Tunnel and introducing a novel 3D gait dataset. In *Proc. IEEE BTAS*, Sep. 2008.

[38] G. Shakhnarovich and T. Darrell. On probabilistic combination of face and gait cues for identification. In *IEEE Proc. FGR*, page 176, 2002.

[39] G. Shakhnarovich, L. Lee, and T. Darrell. Integrated face and gait recognition from multiple views. In *IEEE CVPR*, pages 439–446, 2001.

[40] G. Veres, M. Nixon, and J. Carter. Is enough enough? what is sufficiency in biometric data? *LECTURE NOTES IN COMPUTER SCIENCE*, 4142: 262, 2006.

[41] J. L. Wayman. Benchmarking Large-Scale Biometric System: Issues and Feasibility. In *Proc. CTST*, 1997.

[42] K. Yamauchi, B. Bhanu, and H. Saito. Recognition of walking humans in 3d: Initial results. *Computer Vision and Pattern Recognition Workshop*, pages 45–52, 2009.

[43] A. D. Yarmey and M. J. Yarmey. Eyewitness recall and duration estimates in field settings. *J. of App. Soc. Psych.*, 27(4):330–344, 1997.

[44] R. Zewail, A. Elsafi, M. Saeb, and N. Hamdy. Soft and hard biometrics fusion for improved identity verification. *MWSCAS*, 1:225–8, 2004.

[45] X. Zhou and B. Bhanu. Integrating face and gait for human recognition at a distance in video. 37(5):1119–1137, October 2007.