UNIVERSITY OF SOUTHAMPTON

On a three-dimensional gait recognition system

by

Richard D. Seely

A thesis submitted in partial fulfilment for the
degree of Doctor of Philosophy

in the
Faculty of Engineering, Science and Mathematics
School of Electronics and Computer Science

July 2010

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

by Richard D. Seely

The University of Southampton Multi-Biometric Tunnel is a high performance data-capture and recognition system; designed with airports and other busy public areas in mind. It is able to acquire a variety of non-contact biometrics in a non-intrusive manner, requiring minimal subject cooperation. The system uses twelve cameras to record gait and perform three-dimensional reconstruction; the use of volumetric data avoids the problems caused by viewpoint dependence — a serious problem for many gait analysis approaches.

The early prototype by Middleton et al. was used as the basis for creating a new and improved system, designed for the collection of a new large dataset, containing gait, face and ear. Extensive modifications were made, including new software for managing the data collection experiment and processing the dataset. Rigorous procedures were implemented to protect the privacy of participants and ensure consistency between capture sessions. Collection of the new multi-biometric dataset spanned almost one year; resulting in over 200 subjects and 2000 samples.

Experiments performed on the newly collected dataset resulted in excellent recognition performance, with all samples correctly classified and a 1.58% equal error rate; the matching of subjects against previous samples was also found to be reasonably accurate. The fusion of gait with a simple facial analysis technique found the addition of gait to be beneficial — especially at a distance. Further experiments investigated the effect of static and dynamic features, camera misalignment, average silhouette resolution, camera layout, and the matching of outdoor video footage against data from the Biometric Tunnel. The results in this theis prove significant due to the unprecedented size of the new dataset and the excellent recognition performance achieved; providing a significant body of evidence to support the argument that an individual's gait is unique.

L. Middleton, D. K. Wagg, A. I. Bazin, J. N. Carter and M. S. Nixon. A smart environment for biometric capture. *Automation Science and Engineering, Proceedings of IEEE International Conference on*, 57–62, 2006.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Richard David Seely, declare that the thesis entitled *On a three-dimensional gait recognition system* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as:

1. S. Samangooei, J. D. Bustard, R. D. Seely, M. S. Nixon and J. N. Carter. On Acquisition and Analysis of a Dataset Comprising of Gait, Ear and Semantic Data. In *Multibiometrics for Human Identification*; upcoming book. B. Bhanu and V. Govindaraju, eds. Cambridge University Press.
2. R. D. Seely, M. Goffredo, J. N. Carter and M. S. Nixon. View Invariant Gait Recognition. In *Handbook of Remote Biometrics: for Surveillance and Security*, Springer. ISBN 978-1-84882-384-6
3. 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 *Biometrics: Theory, Applications and Systems, Proceedings of IEEE Conference on*, September 2008.
4. M. Goffredo, R. D. Seely, J. N. Carter and M. S. Nixon. Markerless view independent gait analysis with self-camera calibration. In *Automatic Face and Gesture Recognition, Proceedings of IEEE International Conference on*, September 2008.
5. R. D. Seely, J. N. Carter and M. S. Nixon. Spatio-temporal 3D Gait Recognition. In *3D Video Analysis, Display and Applications*, February 2008.

Signed: ____________________________________

Date: ____________________________________

Acknowledgements

I would like to thank my supervisors John Carter and Mark Nixon for their support and encouragement throughout my time at the University of Southampton; where I have learnt much about research, academia, and myself. Thanks to my fellow researchers Imed Bouchrika, Cem Direkoglu, Bassam Farran, Michela Goffredo, Aly Hassan and many others in the ISIS research group for their friendship and interesting conversations. I would especially like to thank Sina Samangooei and John Bustard for their assistance in supervising the large dataset's capture sessions, which would have been an extremely time-consuming task to manage without their kind support. I would also like to thank my parents Penny and David, and my brother, Kevin for their support and encouragement throughout my studies.

To my mother and father; Penny and David Seely

Chapter 1

Introduction

The ability to recognise an individual and verify their identity is essential in many areas of modern day life; from authorising financial transactions to preventing certain individuals from travelling into a country. The most common method for a person to prove their identity is through the use of identification documents; most people possess a wide range of such documents; including banking cards, driving licenses, passports, proof of age and workplace identity tags. Usually these documents will feature the name of the associated person, a unique serial number and some means for the individual to prove that they are the rightful holder; this could be a copy of their signature, a photograph of their face or an electronic chip containing a secret password only known to the owner. The use of passwords and personal identification numbers has replaced signatures in many applications due to the convenience and improved security offered — as the entered password is usually not seen or retained by the party requesting proof of identity — unlike signature based authentication, where the signature must be visually checked and is often retained after the transaction. Whilst password based identity verification is popular, it is extremely easy for an unscrupulous individual to impersonate another person once they have obtained the victim's password — either by means of trickery or covert surveillance. In applications where a greater degree of security is required, a photograph of one's face or a copy of their fingerprints may be used, to facilitate either visual or automated verification of one's identity. Facial appearance and fingerprints are both examples of biometrics; physical attributes of a person that exhibit some degree of variation between individuals, which can be measured quantitatively. Unlike signatures and passwords, biometric features are generally much harder for another party to accurately observe and impersonate.

1.1 Biometrics

A wide range of biometrics exist, such as one's DNA, fingerprints, iris patterns and facial appearance. The ideal biometric must provide various key attributes[51]: it must be present for all people; it should provide sufficient variance to ensure no two subjects are indistinguishable; and it should remain stable over time. If a biometric is to be useful, it must be capable of achieving a high correct acceptance rate, able to reject false matches, easy to measure and be deemed acceptable by the general public. Unfortunately, there is no such thing as the *perfect* biometric; all have associated advantages and disadvantages, which means that the best compromise must be chosen according to the specific application.

Biometrics have long been used in the field of forensics, where the identity of a suspect must be ascertained from evidence left at the scene of an criminal offence; such as fingerprints or DNA. In surveillance and security applications, the ability to identify an unknown subject without their knowledge or cooperation is often required; this is an area where non-contact biometrics such as facial appearance, ear characteristics or gait can prove extremely useful. In scenarios where recognition must be performed from a distance whilst covering a large area, face or ear based recognition becomes impractical, due to the insufficient resolution provided by a camera having a wide field of view. This is where the use of a subject's gait — the way in which they walk — is advantageous; a much larger proportion of the subject's body is considered for analysis, meaning that much more of the available information is used for recognition purposes. An individual's gait can be observed from a distance using standard video camera equipment, whilst subject cooperation is not necessarily required — making it attractive for use in surveillance applications.

1.2 Gait

The way in which one walks has been shown to vary between individuals; this was first recorded by Murray et al. [74] in 1964, where it was found that amongst a small group of participants, each exhibited their own unique movement pattern. Research into computer-vision based gait recognition started in the early nineties, with Niyogi and Adelson [80] and Guo et al. [42] the first to announce gait analysis techniques capable of discriminating between individuals. Interest in gait as a biometric gradually increased over the years, with DARPA establishing the Human Identification at a Distance program, to encourage and support research into gait and other non-contact biometrics that could be observed from a distance. As part of the research program, several institutions recorded large gait datasets, each containing in excess of one-hundred unique subjects; these new datasets facilitated the development and evaluation of new state-of-the-art analysis techniques.

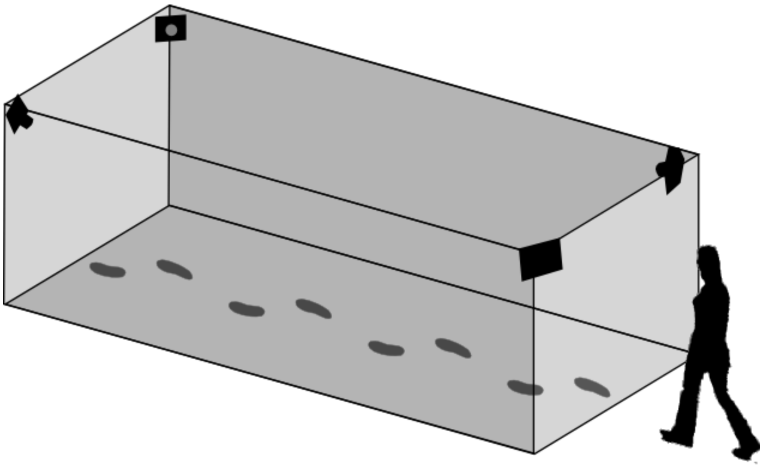


Figure 1.1: Schematic of a controlled environment with fixed cameras, ideal for automatic gait recognition.

Several significant limitations still remain for most gait recognition approaches; no dataset of sufficient size exists to evaluate the applicability of gait for usage in large population environments, and not enough is known about the covariates that affect one's gait; such as footwear, clothing, time, surface type and incline. Another significant limitation is that the signature produced by many gait analysis techniques varies with the orientation of the subject relative to the camera — this is known as viewpoint dependence.

In controlled laboratory experiments, gait has been shown to an extremely effective biometric for distinguishing between individuals[109]; although in real world scenarios it has been found to be much harder to achieve good recognition accuracy[54]. Many popular gait analysis techniques are unable to reliably match samples from differing viewpoints and are also strongly dependant on the ability of the background segmentation algorithm to accurately discriminate between the subject and the surrounding background. Whilst the use of advanced computer vision processing algorithms can yield some improvement, this is often at the expense of computational complexity or the use of additional assumptions, making real-world use difficult.

The performance of any recognition system is ultimately dependant on the quality of the original source data; therefore it is advantageous to consider the lessons learnt from previous laboratory experiments and find ways to introduce a similar degree of control and consistency as found in an experimental setting. By constraining the directions of travel in the target environment using walls or other obstacles, the range of observed subject orientations can be reduced to a small set of likely possibilities. The use of walls as backgrounds for the cameras greatly simplifies the task of distinguishing between subjects and their surroundings, resulting in less background segmentation errors. This makes the walkways and corridors found in most airports ideal environments for gait recognition, where the surroundings and the path taken by individuals is inherently fixed. Figure 1.1 depicts an idealised constrained environment; the path taken by the subject is restricted to a narrow path and once inside, the subject is in an area where lighting and other conditions can be controlled to facilitate the accurate measurement of an individual's gait and facial appearance.

1.3 The Biometric Tunnel

An example of a system operating within a constrained environment is the Biometric Tunnel — a prototype non-contact biometric recognition system — originally constructed by Middleton et al. [70] at the University of Southampton. The Biometric Tunnel was developed with the objective of producing a system capable of automated biometric measurement and recognition. The configuration of the environment was chosen to mimic a walkway or corridor, with walls running either side of the passage to constrain the walking direction of subjects; as shown in Figure 1.2. A network of systems could be installed throughout a large environment — such as an airport — to facilitate the tracking an individual's movement between areas. Alternatively, such systems could be deployed at entrance points to facilitate the accurate measurement of an individual's biometric characteristics at the time of their arrival; these features could then be used for identifying subjects from standard surveillance video camera systems.

The system uses a network of time-synchronised cameras to record video footage of a subject as they walk through the environment. The data from the cameras is processed to separate the individual from their surroundings, which is then used to derive a three-dimensional reconstruction of the subject. The ability to use three-dimensional gait data facilitates new lines of enquiry into novel techniques capable of exploiting such data. Alternatively, the data can be used to synthesise a two-dimensional image from any chosen viewpoint, which could then be used with a standard analysis technique; this means that the same relative viewpoint can always be used for gait analysis; avoiding the problem of viewpoint-dependence.

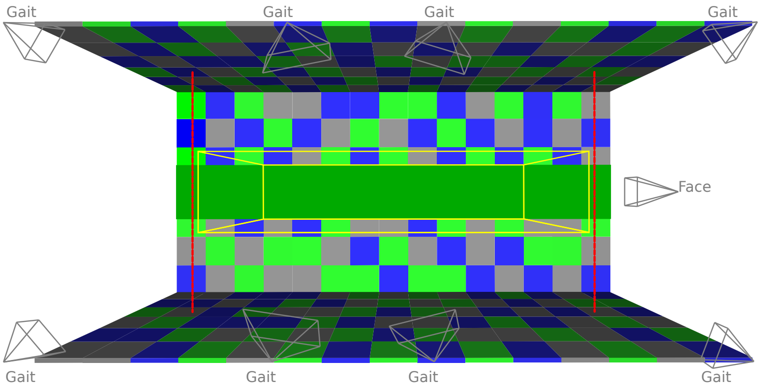


FIGURE 1.2: Layout of the early prototype Biometric Tunnel by Middleton et al. [70]

1.4 Contributions and Original Work

An evaluation of the original prototype Biometric Tunnel is presented in Chapter 2, which was performed using a small dataset collected during the original system's development; the correct recognition rate was found to be greatly below expected. The findings of an initial investigation are presented, where several possible factors affecting performance are found; although no firm conclusions can be made due to the lack of available unprocessed source data. Therefore an audit of the hardware configuration and the system's software was performed, which helped to identify a lack of proper time-synchronisation between the cameras within the system. As discussed in Chapter 4, further revisions were made to the system, to facilitate the collection of a new dataset, which contained raw unprocessed video data from the cameras. Analysis of this dataset identified several additional sources of data degradation; where steps were then taken to address these problems. A large amount of software and hardware development has taken place throughout the duration of this thesis; much of this is only discussed at a summary level in the body of the document, although it is covered in greater detail in the appendices.

Further extensions to the Biometric Tunnel were then performed, with the aim of preparing the system for the collection of a new non-contact multi-biometric dataset of an unprecedented size; this is documented in Chapter 6. The collection of this new dataset proved to be a significant undertaking; spanning in excess of a year and involving over two-hundred unique participants. The resulting dataset is one of the largest non-contact biometric datasets containing gait and one of the only to record gait using high quality three-dimensional volumetric data. Analysis of the recorded dataset found that it was possible to correctly identify every sample within the dataset, this is especially significant considering that this is the largest gait dataset to date. The findings in this thesis provide significant evidence in favour of the argument by Murray et al. [74], that each person's gait is unique.

Using the newly collected dataset, a range of additional investigations are performed in Chapter 7, to further understand the recognition performance and limitations of the revised Biometric Tunnel. It is found that it is possible to perform recognition of an individual against samples collected from an earlier date, although with a significantly reduced accuracy, compared to recognition across a smaller time period. As the Biometric Tunnel system uses video footage from multiple cameras, any change in camera orientation since calibration will result in a distorted reconstruction; therefore an investigation was performed to find out what impact this could have on recognition performance. This experiment confirmed the belief that recognition performance would be severely impacted by any change in camera alignment between samples, where no re-calibration of the cameras had taken place. These findings demonstrate the importance of monitoring the alignment of the cameras and performing calibration on a regular basis.

In Chapter 8, several recognition experiments are performed to investigate the benefits of using both face and gait for recognition, where the signatures generated by gait analysis are combined with an experimental face analysis technique. The results of the experiment show that the use of both biometrics together results in a better accuracy than either face or gait alone. This is especially useful for public environments where an unknown individual's face could be concealed by clothing and it would be impractical or offensive to demand the removal of the attire. A further experiment is performed where data collected from the Biometric Tunnel is used to compare the performance of both gait and face at varying distances from a simulated camera. It is found that gait recognition is still possible at distances where facial recognition proves impossible due to insufficient resolution.

Finally, a set of experiments are conducted in Chapter 9, to investigate the feasibility of matching three-dimensional data from the Biometric Tunnel against video footage from a standard video camera located in an outdoor environment, where there is less control over lighting and the background. It is found that whilst recognition is possible, the accuracy is severely limited by poor background segmentation quality and the presence of strong shadows in the outdoor footage. With the use of an improved background segmentation method or a more robust gait analysis technique, it is expected that the effect of these issues could be minimised.

The collection and analysis of this new multi-biometric dataset, along with the wide range of associated experiments presented in this document all help to reinforce the value of gait as a viable non-contact biometric for use in a variety of scenarios, such as public areas, airports, large events and surveillance applications. The size of this new dataset puts it at the forefront of gait analysis research, facilitating the development and evaluation of cutting edge gait analysis techniques, whilst providing a significant population size, allowing realistic conclusions about recognition performance to be made. The ability to correctly identify every sample within the new dataset, using a simple yet intuitive gait analysis technique demonstrates the potential for gait analysis and suggests that future research into more sophisticated techniques using much larger datasets may yield encouraging results.

Chapter 2

The Biometric Tunnel

2.1 Introduction

The Biometric Tunnel is a purpose built non-contact biometrics acquisition environment, located within the University of Southampton. The system was purpose built to acquire video of a subject from multiple viewpoints, as they walked through the system's measurement area. The recorded video was used to reconstruct a sequence of binary-volumetric frames, which could then be used to characterise one's gait. This compares favourably against most other existing gait measurement systems, which either produced only two-dimensional measurements, or required the participant to wear several sets of retro-reflective markers.

The initial concept system was developed by Middleton et al. [70], as discussed further in Section 2.2. It was designed to record a subject's gait using three-dimensional volumetric data, which could be used to facilitate further research into gait recognition. The system was also intended to demonstrate the effectiveness of gait in a non-contact biometric recognition system. The Biometric Tunnel was constructed in an indoor laboratory, which allowed the use of controlled artificial lighting, helping to reduce the effect of shadows, resulting in better consistency between recordings. The system was built around a pathway spanning the length of the room, which was used to constrain the walking direction of the participants. A pair of purpose constructed walls ran the length of the environment, surrounding the central pathway. The floor either side of the pathway and the two walls were painted with a non-repeating pattern, which was used to assist the camera calibration process. The pattern was comprised of three standard chroma-keying colours; often used in the broadcast industry to ease the process of background segmentation and substitution.

Video footage was simultaneously obtained from nine cameras; where eight were configured to measure the subject's gait and the remaining camera recorded video of the their

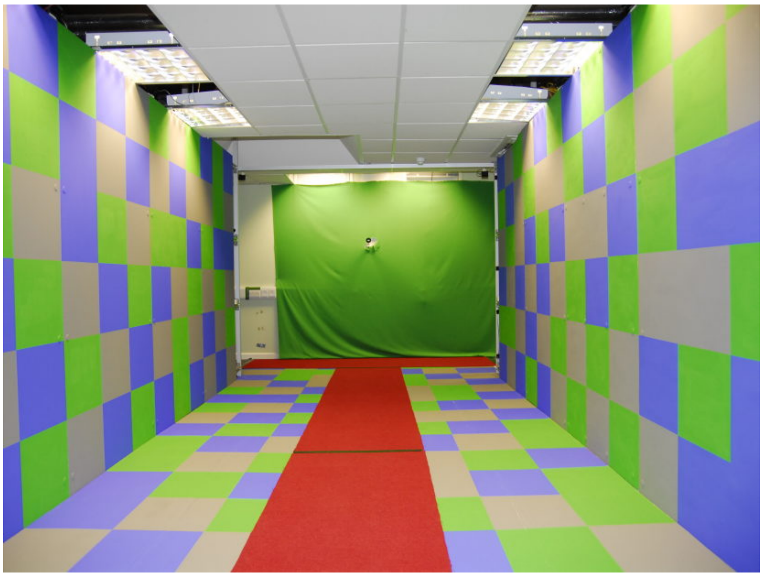


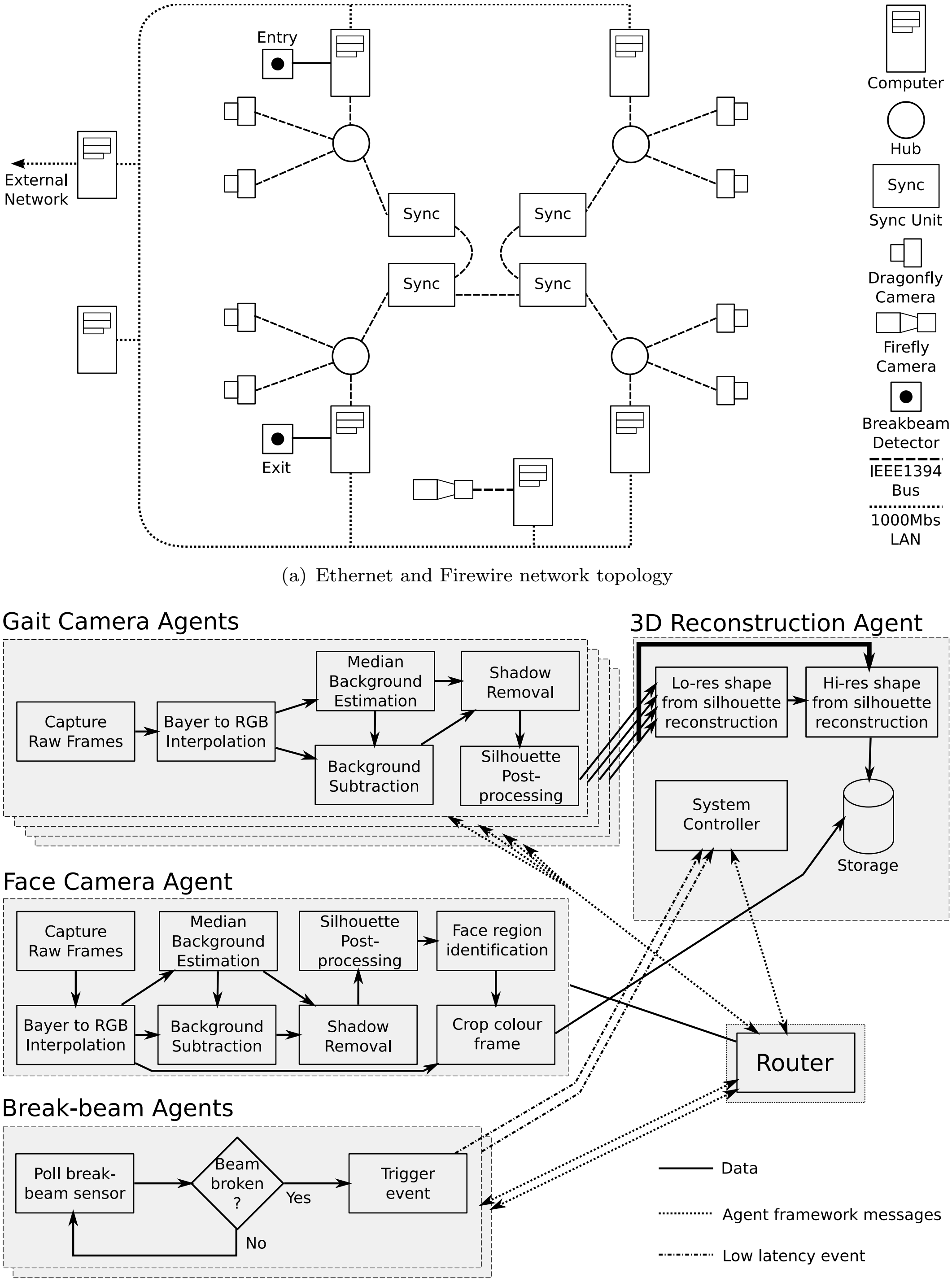
Figure 2.1: View from the entrance of the Biometric Tunnel by Middleton et al. [70]

face and upper body. Infrared break-beam sensors were located at the entrance and exit of the measurement area to control the recording of video data. Whilst recording, the captured video data was saved to random access memory in the connected computers, processed and then saved to disk immediately after recording; requiring approximately five minutes to complete and allow subsequent capture. Section 2.2 provides an in-depth discussion of the software and hardware that formed the basis of the Biometric Tunnel.

During the development of the system by Middleton et al. [70], a small dataset was collected; analysis of the dataset in Section 2.3 showed that the system produced unsatisfactory classification performance. Therefore, an investigation was performed to find the causes of the poor classification performance. It is found that several issues were to blame: data corruption, inconsistent gait cycle labelling and poor reconstruction quality. By removing the affected samples from the evaluation dataset, a substantial improvement in recognition performance was achieved, demonstrating that the removed samples had a significant negative impact on performance.

2.2 The original Biometric Tunnel

The original Biometric Tunnel system was developed by Middleton et al. [70] at the University of Southampton, as a means for demonstrating gait recognition in a controlled environment. The system consisted of a narrow pathway surrounded by nine video cameras; with eight having a wide field of view to record one's gait and the other camera having a smaller coverage for capturing video of one's face. The pathway ran down the centre of the tunnel, spanning from one end of the laboratory to the other, with walls



(a) Ethernet and Firewire network topology

(b) Software for capturing and processing data

FIGURE 2.2: Hardware and Software Layout of Original Biometric Tunnel

running along either side. The layout of the Biometric Tunnel is shown in Figure 2.1 and previously in Figure 1.2.

Eight PointGrey Dragonfly Colour video cameras were mounted along the tops of the two walls, in order to obtain a wide field of view, suitable for observing a subject's gait. These cameras captured VGA (640 × 480 pixels) resolution video footage at a rate of thirty frames per second, which was streamed unprocessed to the host computer over an IEEE1394 bus. A PointGrey Flea camera was chosen for capturing front-on facial imagery, which featured a higher SVGA (1024 × 768 pixels) resolution and also used the IEEE1394 bus to stream video data. The system had four computers configured for capturing gait video footage, with each computer connected to two of the cameras using a IEEE1394 network. The camera networks were interconnected using timing-synchronisation units, to ensure that the video frames from all gait cameras were captured at the same point in time, to facilitate accurate 3D reconstruction of the subject. The face camera was connected to its own dedicated computer, due to its much greater bandwidth requirements. The topology of the Ethernet network and the IEEE1394 networks are shown in Figure 2.2(a).

The various software applications running on the computers were coordinated using a specially developed multi-agent framework[71], which provided the ability to register and locate available resources, and route messages between them. Figure 2.2(b) shows the interaction between the different resources running on the system. A total of seven computers were used in the system, with five of the computers connected to cameras, as mentioned earlier. The sixth computer acted as the controller for the system, running the router for the agent framework, also performing 3D reconstruction and allowing the user to control the system. The final computer was intended for file storage; holding the recorded samples.

The acquisition process was controlled by infrared break-beam sensors, mounted at the entry and exit points of the measurement area. As a subject entered the measurement area, they would trigger the first infrared break-beam sensor; starting the acquisition of video footage by the cameras. The recorded video frames were streamed as unprocessed raw video data back to the host computers and then saved in local memory. Upon leaving the measurement area, the subject would break the second infrared beam; stopping the capture of any further video footage and starting the processing of the video data saved in memory by each computer.

The first stage of processing the recorded data was to convert the captured images into colour from their original raw Bayer format, using nearest-neighbour interpolation; as discussed in Chapter 5.2. Background estimation and segmentation was then performed to find the subject's silhouette; modelling each background pixel with a single Gaussian distribution per colour channel. The distribution for each pixel was found using previously captured video footage, where no subject was present. The background segmentation was performed by calculating the distance between a pixel and its corresponding background distribution, where a pixel would be marked as background if its distance was less than a global threshold; linked to the standard-deviation found by the background estimation. Shadow labelling and removal was performed to reduce the number of pixels incorrectly marked as foreground. Binary morphological post-processing was then performed to reduce noise levels and smooth the silhouette's shape. Finally, all regions except that with the greatest area were removed and any holes in the remaining region were filled. Radial distortion caused by the camera optics was removed by the use of a non-linear transformation. The resulting images from each camera were then streamed from their respective computers to the central control computer; where three-dimensional reconstruction was performed using a basic multi-resolution strategy with a six or more camera criteria; as discussed in Section 5.4. The reconstructed volumetric data was then saved to disk for later analysis. The processing required approximately five minutes for every seven second sample acquired. Recording of subsequent samples was not possible until the processing of the previous sample had completed.

2.3 Analysis of data from the original system

During the development of the system by Middleton et al. [70], a small number of samples were acquired for testing purposes; these were used to construct a dataset for evaluating the recognition performance of the system. For each sample, a single gait cycle was selected by manual inspection. The dataset contained seventy-three samples, from twenty different subjects, with on average four samples per subject; as shown in more detail in Table A.1. The reconstructed volumetric data produced by Middleton et al. [70] was smoothed using binary erosion and dilation morphological operators to reduce the level of noise and reconstruction artefacts. The average silhouette was found for each sample from three orthonormal projections; side-on, top-down and front-on. The resulting average silhouettes were used in a leave-one-out recognition experiment, to find the classification performance of the system by Middleton et al. [70].

The recognition performance was found to be below expected; where average silhouettes from a side-on viewpoint were found to give the best recognition performance on the dataset, only achieving 81.4%. The side-projection average silhouette is known to be a very good classifier, which has been found to achieve almost perfect classification rates on datasets containing in excess of one-hundred subjects[109]. Therefore the recognition performance acheived by system of Middleton et al. [70] was greatly below expected. The full results of the experiment and the system's receiver operating characteristic plot can be found in Appendix A.2. The use of feature-selection techniques such as analysis of variance (ANOVA) and principal component analysis were found to provide marginal

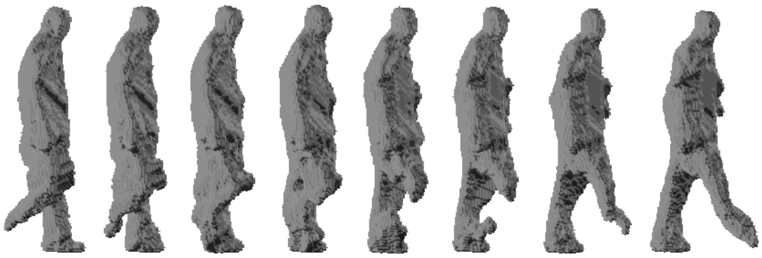


Figure 2.3: Sequence of reconstructed silhouettes with significant areas of the volume missing

classification performance gains; although the performance was still fundamentally limited by the poor quality of the source data.

As a result of the system's poor recognition performance, further investigation was required to discover the causes of the degraded performance. The initial belief was that the use of manual gait cycle labelling had introduced inconsistencies into the dataset, due to human error. Therefore, an automated gait cycle labelling algorithm was devised and implemented; as described in Chapter 5.6. Inspection of the diagnostic output from the automatic labelling revealed that some samples had proved much more difficult to locate gait cycles for; this lead to the discovery that many of the samples with a poor fitting score appeared to have significant parts of the reconstructed volume missing; as shown in Figure 2.3.

Checking the non-post-processed volumetric data confirmed that the same regions were missing, suggesting that the artefacts had been introduced in the processing performed by the system of Middleton et al. [70]. As only the final reconstructed data was saved by the system, it was impossible to confirm which stage of processing had introduced these errors. The most likely cause of the problems was segmentation errors, causing parts of the subject's legs to be incorrectly labelled as background; possibly due to the subjects wearing clothes with similar colours to those used in the tunnel's background; such as blue denim jeans. Some attempts had been made to counter this problem in the system, through the use of a reconstruction algorithm that allowed up to two of the eight cameras to incorrectly label a voxel's corresponding pixels as background. Although this improvement in robustness came at the expense of reconstruction accuracy; resulting in volumes that were larger than the true convex hull of the silhouettes, as discussed in Chapter 5.4.

Many of the samples were also found to be incomplete or corrupted, as a result of programming errors introduced during development. In order to establish whether the aforementioned problems fully accounted for the degradation in performance, a second recognition experiment was conducted, where samples were visually inspected and excluded if deemed of unacceptable quality. Twenty-eight of the samples were considered to be of *good* quality, with very few visible artefacts and no frames where areas of the reconstructed volume were missing. Twenty samples were deemed as *acceptable*; where the samples featured some artefacts; small parts of the volumes were missing or minor shape distortion was present. Twenty-five of the samples were found to be *bad*; where serious artefacts were present; the shape of the subject was heavily distorted or several frames had significant regions missing. Subjects with less than two samples were removed from the dataset, as this would add an unfair negative bias to the recognition performance. The resulting dataset contained 42 samples from 11 subjects, as detailed in Table A.1. As expected, leave-one-out recognition performance on the revised dataset was significantly improved, with a 97.6% correct classification rate for the side-on viewpoint. The full results of the recognition expirement and receiver operating characteristic plot are given in Appendix A.3. The improvement in recognition performance confirmed that one or more of the outlined issues were to blame for the poor performance of the initial system.

2.4 Discussion

Analysis of the previously collected dataset revealed that the correct classification rate was greatly below expected, which was a clear indication that there was serious issues with the quality of the recorded samples. Further investigation found there were many corrupt or empty samples present in the dataset; suggesting that the reliability of the prototype system was an issue. The quality of the reconstructed output was also found to be poor, both visually and in terms of the attainable recognition performance — achieving only a 81.4% correct classification rate — possibly due to the choice of reconstruction algorithm. Many of the samples in the dataset had severe artefacts present in the reconstructed data, where the limbs of subjects were severely distorted or completely missing. Removal of the affected samples from the analysis experiment lead to a significant gain in recognition performance; although some samples were still incorrectly classified as other subjects, which indicated that there was still problems with the quality of the remaining samples.

Investigation into the causes of the degraded reconstruction quality was made extremely difficult, as the unprocessed data from the cameras was not saved by the system during recording; making it impossible to re-process the data and locate the sources of the problems. The time taken to process the acquired data after each sample also added a delay of approximately five minutes between samples; slowing down the rate at which participants could be recorded.

Chapter 3

Background and Literature Review

3.1 Introduction

Gait analysis has become a popular research topic over the last ten years, with groups at many large and prestigious institutions taking interest. Researchers from a medical background were the first to publish studies showing how the manner in which one walks varies amongst a population [74, 39]. At a later stage, psychology experiments were carried out to see if humans could recognise subjects or gender from moving light displays [52, 69].

The earliest research into computer-vision based gait analysis techniques was published in 1994 by Niyogi and Adelson [80], which was based on spatio-temporal analysis and model fitting. Later that year, Guo et al. [42] published an algorithm based upon a 10 stick model and neural network classification. Soon after, Little and Boyd [64] published a gait analysis technique based upon the spatial distribution of optical flow and how it varied over time. Murase and Sakai [72] proposed a technique that compared the eigenspace trajectories between subjects, this concept was later extended by Huang et al. [49] to also use canonical analysis. Cunado et al. [26] published a model based technique that used the Hough transform to fit a model to the video frames; results were published on a small dataset recorded indoors, which was to become the first gait dataset widely used by others. Little and Boyd [65] published results of their previous algorithm[64] applied to a new dataset recorded outdoors; this dataset also became very popular in the research community. By 1998, the number of researchers working on gait analysis had increased massively, with the pace of research increasing year after year. Other significant milestones include the release of the Gait Challenge dataset and baseline algorithm[84] and the University of Southampton's HumanID dataset[93]; these are still some of the largest publicly available datasets and are used extensively by researchers around the world. There are many literature reviews documenting the progress of the gait analysis and human motion analysis research community, this includes [35, 78, 77, 114, 48, 76]; there is also a book[79] that provides an extensive overview of the progress made in gait analysis and recognition. This chapter provides a comprehensive review of the literature relevant to gait analysis, classification techniques, datasets and multi-biometric fusion.

3.2 Gait datasets

This section introduces the various gait datasets that have been produced by members of the computer vision and biometrics community and discusses the various advantages and disadvantages of each. An overview of the datasets is given in Figures 3.1(a) and 3.1(b).

One of the earliest documented datasets was that of Cunado et al. [26] from the University of Southampton, produced in 1997. The dataset was filmed indoors with a static background and controlled lighting; this had the effect of reducing shadows. The presence of shadows could prove problematic for early background subtraction approaches; therefore it was desirable to try and reduce the appearance of shadows in the video data. In order to aid analysis, subjects wore white trousers with a black stripe running down the leg on the near-side. Subjects walked in a straight path, with the camera being perpendicular to the subject's walking direction. A total of ten subjects were recorded, each walking through the field of vision four times.

In 1998, Little and Boyd [65] from the University of California, San Diego published works using a new dataset that they had collected. It was filmed outdoors in the shade to ensure diffuse lighting, which would reduce the effect any shadows. A large wall was used as the background, and subjects walked in a large circular path around the camera. The dataset contained six subjects, each walking through the field of view seven times.

The Georgia Institute of Technology also produced its own dataset[6]; containing twenty subjects, each subject having six samples recorded indoors using a magnetic sensor system to give *ground-truth* data for the subject's joint positions. In addition to the magnetic sensor dataset, a subset of the twenty subjects were recorded walking indoors, with a single video camera placed in three different positions. At a later point in time, fifteen of the original twenty subjects were recorded walking outdoors; from a single camera position.

In the same year, Carnegie Mellon University announced their Motion of Body (MoBo) database[40]; which was recorded indoors using a treadmill. The use of a treadmill allowed them to record subjects walking and running at varying gradients. In some samples, subjects held a large ball to inhibit any motion of their arms. Six cameras were

Dataset	Subjects	Samples	Markers	Indoor	Treadmill	Outdoor	Viewpoints	Simultaneous
Soton 1997 [26]	10	40	N	Y	N	N	1	-
UCSD 1998 [65]	6	42	N	N	N	Y	1	-
GaTech 2001 [6]	15–20	426	Y	Y	N	Y	3	N
CMU MoBo 2001 [40]	100	600	N	N	Y	N	6	Y
MIT 2002 [61]	24	194	N	Y	N	N	1	N
Gait Challenge 2002 [84, 89]	122	1870	N	N	N	Y	2	Y
UMD 2002 [54]	44	176	N	N	N	Y	1	-
Soton 2002 [93]	114	>2500	N	Y	Y	Y	2	Y
CASIA 2003 [117]	20	80	N	N	N	Y	3	N
CASIA 2006 [122]	124	1240	N	Y	N	N	11	N

(a) Dataset composition; where the number of samples refers to independently recorded sequences; therefore the total number of sequences is higher for samples simultaneously recorded from multiple cameras.

Dataset	Time	Speed	Incline	Surface	Footwear	Carrying items	Clothing	Direction
Soton 1997 [26]	Minutes	N	N	N	N	N	N	N
UCSD 1998 [65]	Minutes	N	N	N	N	N	N	N
GaTech 2001 [6]	Days	N	N	N	*	N	*	N
CMU MoBo 2001 [40]	N	**Y**	**Y**	N	N	**Y**	N	N
MIT 2002 [61]	**Months**	N	N	N	*	N	*	**Y**
Gait Challenge 2002 [84, 89]	Months	N	N	**Y**	**Y**	**Y**	*	N
UMD 2002 [54]	Days	N	N	N	*	N	*	**Y**
Soton 2002 [93]	**Weeks/Months**	**Y**	N	N	**Y**	**Y**	**Y**	**Y**
CASIA 2003 [117]	Days	N	N	N	*	N	*	N
CASIA 2006 [122]	Minutes	N	N	N	N	**Y**	**Y**	N

(b) Covariate features included in various gait datasets. Clothing and footwear may vary in datasets recorded over several days, where this is not deliberate, it is marked with a *.

Figure 3.1: Comparison of various gait datasets, information collated from [2, 6, 26, 40, 54, 65, 89, 93, 117, 122]

positioned around the subject, allowing simultaneous multi-viewpoint video capture. The database contained twenty five subjects, each having a total of twenty-four samples.

Lee and Grimson [62] from the Massachusetts Institute of Technology produced a dataset consisting of 24 subjects, which was recorded indoors, with subjects walking in a straight path perpendicular to the camera. Subjects walked in both directions, and the sequences were flipped to result in a constant direction of travel. The number of video sequences per subject varies between 4 and 22 sequences, with a minimum of 3 gait cycles per sequence. Recording was performed on four separate days, spanning two months. The dataset contained a total of 194 sequences. Unlike many of the other indoor datasets, no specialised lighting equipment was used, instead relying on standard fluorescent overhead office lighting, which resulted in quite strong shadows.

In 2002, The National Institute of Standards and Technology and The University of South Florida released the Gait Challenge dataset[84], which is now one of the most commonly used benchmark datasets for gait analysis researchers. The dataset was filmed outdoors, using two video cameras simultaneously recording the subject from differing viewpoints. A calibration target was included in all scenes to allow calibration of the cameras. Subjects walked in an elliptical path around the cameras in a similar manner to Little and Boyd [65]. Each subject was recorded walking on both grass and concrete surfaces, with differing shoe types and partial occlusion from a briefcase. The recording was also repeated six months later to enable the evaluation of gait's temporal variance. There was initially 74 subjects in the dataset, although this was later extended to 122 subjects[89]. Each subject only walked once through the field of view for each combination of covariate measures; this meant that in order to evaluate an algorithm's performance, the same data would be used for both training and evaluation. Phillips et al. [84] also proposed a baseline algorithm, which could be used as the benchmark for comparing a new gait analysis algorithm's performance.

The University of Maryland produced its own database[54, 2], which was distinctly different from the other available datasets at the time; it was designed to have a very close resemblance to real world surveillance data. A single camera was mounted outdoors, at a height of 4.5 metres, similar to a typical outdoor CCTV setup. Forty four subjects were used, walking in a "T" shaped path, to give multiple orientations relative to the camera. Each subject was sampled twice on two different days; this meant that the clothing worn by the subjects may be different between samples.

In the same year as the release of the Gait Challenge dataset, the University of Southampton released the Human ID at a Distance (HID) dataset[93], this was one of the most comprehensive databases, containing over one hundred subjects. The subjects were recorded walking in a variety of scenarios; indoors along a straight path, indoors on a treadmill and outdoors with a non-static background. The indoor video capture setup used carefully controlled lighting and a green chroma-key background, which meant that subjects could be reliably extracted from the background with minimal shadow artefacts. The outdoor setup was recorded with the cameras directed towards a busy road, resulting in a large amount of unwanted motion; this allows the testing of algorithms on “real-world” data. Multiple cameras were used to record the subjects walking from multiple viewpoints, which were manually synchronised after recording had taken place. Each subject walked at least eight times in both directions past the cameras. Having such a large number of samples per subject enabled the use of different samples for development, training and classification of gait analysis algorithms; this ensures that *over-fitting* of the training data does not unfairly affect classification performance. This is in contrast to many of the other datasets, where leave-one-out evaluation is the only option, which means that the algorithms are trained and optimised for that specific dataset; this means that the trained algorithm may not generalise well to new data. A smaller database containing covariates such as shoe type, clothing, carried items and temporal variation was also produced, containing a subset of the subjects from the main dataset.

The National Laboratory for Pattern Recognition, part of the Institute of Automation from the Chinese Academy of Science (CASIA) also created their own dataset for developing and evaluating gait analysis algorithms. It was filmed outdoors in a controlled environment, with a static background. A single camera was used to film subjects walking in three different views. The dataset contained twenty subjects, each walking in a straight path four times through the camera’s field of view. A further dataset was recently released by CASIA[122]; which contained 124 subjects, each walking six times without a coat or bag, then twice wearing a coat and finally twice with a bag. The subjects were captured by eleven USB cameras placed at varying angles relative to the subject, the video data was then saved in a compressed MJPEG format.

As discussed in this section, a variety of datasets have been produced by the computer vision community for evaluating gait analysis techniques, recorded in a range of environments. Most of the datasets were recorded with the subjects walking in a straight or elliptical path on a stationary surface, although there were a few notable exceptions that used treadmills. The use of a treadmill facilitates the capture of a subject walking on an inclined surface or running, which would otherwise require a large area for the subject to accelerate and for their running pattern to stabilise. Datasets have been produced in both indoor laboratories and outdoors; where an outdoor environment could be considered a more realistic scenario, whilst the use of controlled backgrounds and studio lighting can reduce the effect of shadows and improve the quality of background segmentation. Many of the datasets were produced for internal use in their respective institutions and were of a modest size, typically containing less than twenty five subjects. However, as mentioned earlier, several larger datasets have been produced by NIST/USF[84], the University of Southampton[93] and CASIA[122]; of which all contain in excess of one-hundred subjects and have been made available to other researchers.

Even though these datasets are much larger than the others, they are still not sufficiently comprehensive to prove the applicability of gait recognition in real world large population environments[110]. None of the large datasets contained time-varying three-dimensional volumetric data and were mostly recorded on cameras featuring no time-synchronisation or calibration, which meant that it was extremely difficult to develop and evaluate three-dimensional gait analysis techniques. Therefore, a dataset featuring many more subjects, recorded from multiple time-synchronised cameras, with three-dimensional data would be highly beneficial for further investigating the capabilities of gait recognition.

3.3 Human model-based analysis approaches

This section considers gait analysis techniques that explicitly describe a subject's gait in terms of a model, where the model's parameters are used to create a set of features for recognition. In most cases, the model's parameters are meaningful quantities such as the lengths of body parts, stride length, or dynamic properties such as joint angles. In most cases, the variation of the dynamic properties can be treated as periodic over a sequence of gait cycles; therefore Fourier coefficients are often used to characterise the variation of these parameters. A variety of models have been utilised by the gait community, this includes ellipse based models, stick figures or more complex models comprised of ribbons or three-dimensional primitives.

Two of the earliest published gait analysis algorithms were those of Niyogi and Adelson [80] and Guo et al. [42]; both using human-models as the basis for recognition, although both varied substantially in both model design and fitting strategy. Niyogi and Adelson [80] took a sequence of silhouettes and stacked them along the temporal dimension, resulting in a three-dimensional volume, where vertical slices were then taken through the volume at differing heights, to result in a sequence of images that featured diagonal double-helix patterns. Two pairs of active contours were fitted to the leading and trailing edges of the double-helix patterns, and a five-stick model was then fitted to the double-helices. The use of active contours improved the robustness of the model fitting process, resulting in smoother parameter variation over time. The parameters of the stick model were then extracted and used for recognition. Guo et al. [42] employed a more complex ten-stick model, which was fitted to a silhouette sequence by calculating a cost field for each silhouette, then finding the set of model parameters that minimised the cost accumulated by the model. Classification was then performed with a neural network, using the model parameters' Fourier coefficients. Whilst these two early methods both demonstrated that gait was suitable for recognition purposes, they both utilised relatively complex models for their time, making them computationally expensive on the hardware of the time.

Therefore other early researchers used less complex models to characterise one's gait; Cunado et al. [26] demonstrated that it was possible to perform recognition using a simple model approximating each leg as a single line segment, joint at the hip. The parameters of the model were found by applying the Sobel edge operator to the source images, then using an implementation of the Hough transform[31] to locate the two lines. The angles of the lines were found for each frame, then smoothed and interpolated by fitting high-order polynomial splines to the time varying angular data; recognition was then performed using the the coefficients found by a discrete Fourier transform. Cunado et al. [25] later extended the previous approach to use a more advanced model; where each leg was modelled by a pair of articulated pendulums. The revised model was fitted to the edge image using a new more efficient approach; the Genetic Algorithm Velocity Hough Transform. Cunado et al. [27] later published works claiming a correct classification rate of 100% on a small ten subject database. Yam et al. [120, 121] then further extended the work of Cunado et al. [25] to perform analysis of a subject whilst both walking and running.

Bobick and Johnson [6] also used a simple model, consisting of three line segments, representing the two limbs and the torso, all connected at the pelvis. Unlike many of the other model-based approaches, only static parameters were used; such as the distance between the head and pelvis, the pelvis and feet, and between both feet. The results of the approach were validated against ground-truth data acquired from a magnetic sensor system. BenAbdelkader et al. [3] also proposed an approach using only static features for recognition; where the subject's stride length and cadence were found by analysing the variation in the subject's bounding box width. Davis and Taylor [29] also used a similar three stick model for gait analysis; although unlike Bobick and Johnson [6], used basic dynamic features for recognition instead. The subject's feet are located by finding the principal axis of the pixels in each the leg region, then taking the furthest silhouette pixel's location along the principal axis as the foot position. Basic dynamic features are then taken such as the gait cycle time, the stance to swing ratio and the double support time.

A simple approach proposed by Lee and Grimson [62] approximated a subject's side-on silhouette by splitting it into seven fixed regions, where an ellipse was fitted to each region. It was found that the ellipse parameters exhibited a very poor signal to noise ratio, which meant that only robust features such as the mean, variance, fundamental frequency and phase were used to describe the variation of the parameters. A significant limitation of this approach was the use of fixed region boundaries and that the ellipses were often not joined to neighbours; resulting in an inaccurate model. Lee [61] later extended the ellipse fitting approach to volumetric data to achieve view invariant gait recognition. This was achieved by recording a subject's gait using a multiple camera system; then performing three-dimensional reconstruction to find the convex hull. The trajectory of the subject was estimated and a virtual camera was then placed perpendicular to the subject's walking direction to produce side-on silhouettes of the subject. The synthesised silhouettes were then analysed using the multiple ellipse representation proposed by Lee and Grimson [62].

Wagg and Nixon [112] makes use of a more sophisticated model, further extending work of Cunado et al. [27]; where the head and torso were represented by a pair of ellipses and each leg consisted of two pairs of line segments, for the upper and lower parts of the leg. Fitting such a model with many degrees of freedom is a computationally demanding and difficult task; therefore Wagg and Nixon [112] attempted to solve the parameters of the model over multiple stages of fitting, increasing in complexity with each iteration. First the velocity of the subject was estimated, then a bounding region surrounding the subject was established, which was then refined to consist of three primitives, then finally the complete model was fitted using constraints determined from a clinical study of gait. This approach achieved a correct classification rate of 84%, using the indoor samples from the University of Southampton HID gait database[93]. An alternative approach by Bouchrika and Nixon [8] extracted the subject's heel-strike information from the recorded video footage, which was then used reduce the complexity of fitting a two-dimensional biped model to the video footage.

One of the biggest limitations of the gait analysis techniques mentioned so far in this chapter is the assumption that subjects are walking perpendicular to the camera; whilst practical in a controlled environment, this assumption is unlikely to be reliable in a unconstrained environment. Spencer and Carter [99] proposed a technique that overcomes this problem, by correcting the effects of perspective distortion and the subject's orientation. This was achieved by following several points on the subject, to construct a set of lines that converge at a single point – known as the epipole – which was used to derive a projective transform matrix that aligns the walking direction with the horizontal axis and removes the effect of perspective. Finally an affine transform was calculated using the measurements from the transformed image and those from clinical studies to result in a new coordinate space where angular measurements can be accurately taken. The original work by Spencer and Carter [99] demonstrated that it was possible to reconstruct a subject's joint angle variation with a good degree of accuracy and viewpoint invariance, using manually annotated video footage of a single individual wearing reflective markers. This was extended by Goffredo et al. [37], where video was captured of five subjects wearing markers, walking in six different orientations relative to the camera; it was demonstrated that the joint angles could be accurately reconstructed from multiple viewpoints and subjects. Goffredo et al. [36] later devised a model-based gait analysis algorithm using the same viewpoint invariant reconstruction techniques, but without the need for marker data. The model consisted of articulated pendulums for each leg, which were interconnected by a rigid section representing the hips. A small dataset was collected and analysed, containing video data from three subjects and six viewpoints. It was shown that the normalised angular information could be accurately reconstructed with a low error and that the subjects could be easily distinguished using the first and second principal components.

3.4 Non model-based analysis approaches

One of the earliest gait analysis algorithms to use a non-problem-specific approach was that of Little and Boyd [64]; where the optical flow between frames in a sequence was approximated by fitting ellipses to the calculated optical flow fields. The phase and magnitude for each ellipse parameter's temporal variation were found over the sequence of frames, then Analysis of Variance (ANOVA) was used to remove any of derived features having poor discriminatory abilities. Since then a wide range of non-problem-specific techniques have been proposed by the research community; using a variety of different approaches, such as direct silhouette comparison, pixel distribution modelling, moments, and shape based descriptors.

One of the simplest approaches to gait recognition is to perform a direct comparison between silhouette sequences. By comparing the silhouettes from a sequence against themselves, BenAbdelkader et al. [4] produced a self-similarity matrix, which could be used for recognition purposes and identifying gait cycles. Phillips et al. [84] proposed the use of a direct silhouette comparison technique as a baseline algorithm for the Gait Challenge database. Whilst simple to understand and implement, this approach was extremely inefficient, due to the large number of features required to represent a single gait cycle, resulting in large storage and computational requirements. A more efficient approach by Collins et al. [21] compared only four silhouettes from a gait cycle, known as key-frames, which were taken at fixed points in the gait cycle; making recognition much more practical, due to the smaller number of features required.

Another approach to reducing the number of features required for a sample is to apply a transformation to each silhouette, resulting in a reduced feature-set that approximates the original silhouette. Murase and Sakai [72] used principal component analysis to approximate each silhouette, where sequences were then compared in the derived feature-space, using time-warping to match the sequence lengths. Huang et al. [49] later extended the approach to use both principal and canonical component analysis, resulting in improved separation between different subjects. A different approach was used for measuring the similarity between samples, comparing the mean points in the derived feature-space; this is almost equivalent to the average silhouette approach of Liu and Sarkar [67] and Veres et al. [109]. A clustering technique was used by Tolliver and Collins [106] to find a set of exemplar silhouettes, similar to key-frames. A similar approach by Zhao et al. [124] characterised the exemplar silhouettes using coefficients found by applying a discrete Fourier transform. It was shown by He and Debrunner [45] that a

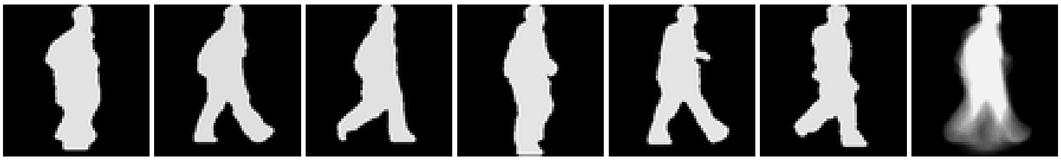


FIGURE 3.2: Several silhouettes from a gait cycle and the resulting average silhouette

Hidden-Markov model could be employed to describe the transition between poses in a gait cycle; where the model was constructed using every other frame in a sequence as a state. Kale et al. [54] used a clustering technique to greatly reduce the number of states required in the Hidden-Markov model. Sundaresan et al. [101] also proposed an analysis framework based upon a Hidden-Markov model, where exceptional recognition results were achieved for the Gait Challenge dataset; this was done by uniformly partitioning a gait cycle sequence into a fixed number of clusters; where the transition between exemplars was controlled by the Hidden-Markov model.

Many of the non-problem-specific approaches characterise a subject's gait by finding a set of features describing the variation of the subject's silhouette over time. Some of these approaches treat every pixel within the subject's image separately; such as that of Boyd [11], Liu and Sarkar [67], Veres et al. [109], or Han and Bhanu [43]. Boyd [11] assumed that each pixel's variation over the sequence of frames was part of a periodic signal, that was parametrised using an array of phase-locked loops. One of the most simple yet effective gait analysis techniques is the average silhouette[67, 109]; calculated by aligning a sequence of silhouettes by their centre of mass, normalising their size, then calculating the mean average for each pixel. An example of an average silhouette is shown in Figure 3.2. Several extensions to the average silhouette exist; such as that of Han and Bhanu [43], where a set of synthesised silhouettes featuring varying levels of occlusion are added to the gallery set; or Lam et al. [59], where the static and dynamic information are characterised separately, by calculating the mean average for the silhouettes' edge images and by finding the intersection of all aligned and scaled silhouettes.

Instead of treating each pixel within a silhouette sequence as an isolated time-varying element, many researchers have chosen to use techniques that characterise the silhouette sequence's distribution in both the spatial and temporal domains; these methods are sometimes referred to as spatio-temporal analysis techniques. Moments provide an efficient method of describing various properties of a distribution within a discrete space of arbitrary dimensionality; they can be calculated by accumulating the product of every point in the space with a moment generating function, which is dependant on the point's location and several additional parameters. Shutler et al. [97, 96] argued that time and space are very different and should not be treated as additional dimensions of one another; as would occur when using basic measures such as Cartesian or Centralised-Cartesian moments. Instead a new type of moment was proposed; the Velocity moment, which extended the Centralised-Cartesian moment to account for the travelling velocity of the subject. Shutler and Nixon [94, 95] also extended Zernike moments to incorporate velocity information, in a similar fashion. Boulgouris and Chi [9] used the Radon transform to characterise silhouette images; where the Radon transform can be expressed as a special type of moment, where the generating function contains the delta Dirac function. A simple approach by Liu et al. [66] calculated frieze patterns by counting the number of pixels in each row and column, for every silhouette. The column and row counts for the sequence of silhouettes were both concatenated to result in two images, each containing repeating patterns. Foster et al. [34] proposed a simple technique based upon area masks, where the temporal variation of area inside a set of masked regions was used as the basis for recognition. In the approach taken by Kobayashi and Otsu [57], the temporal dimension was treated as if it was a third spatial dimension; from this a set of features describing the correlation between the dimensions across the entire volume was found.

Whilst many of the aforementioned analysis techniques make only indirect use of the silhouette's shape, there are a few notable exceptions that describe a silhouette in terms of its shape. Hayfron-Acquah et al. [44] demonstrated that the essential information within a silhouette's shape could be found using a symmetry operator, which identified the axes of symmetry within and surrounding the silhouette. The symmetry images resulting from the silhouette sequence were then combined by averaging, similar to that of Liu and Sarkar [67] and Veres et al. [109]; finally Fourier coefficients were calculated and used for recognition. The shape of a silhouette can be completely described by its set of boundary pixels, which can be closely approximated using a reduced set of points, which must be selected carefully to ensure accuracy. The use of a point-distribution model, such as an active shape model[23], can efficiently encode the typical variation found within a silhouette's shape using a much smaller set of features, by exploiting the correlation between the point locations. Tassone et al. [104] demonstrated that it was possible to adapt a point distribution model to account for the temporal variation present in a sequence of silhouettes from a gait cycle. Wang et al. [115] uses a subset of the boundary points from the silhouettes to calculate the Procrustes mean shape for the gait cycle sequence. The resulting Procrustes mean shape can then be compared against others using the Procrustes distance as a metric for similarity between samples. This approach does not directly make use of the dynamic time-varying information contained within a subject's gait; therefore Wang et al. [116] later combined the Procrustes mean shape of a subject with the parameters found using a human model based gait analysis technique. Another approach by Wang et al. [113] that also used boundary unwrapping, measured the distance from each boundary point to the shape's centroid, sampling a subset of these distances to result in a distance-signal. Principal component analysis was then used to find a new feature-space of reduced dimensionality to encode the distance signal, matching was then performed by comparing the trajectories taken through the derived feature-space by the silhouette sequences.

3.5 Analysis techniques utilising three-dimensional data

In the previous sections an overview of two-dimensional gait analysis techniques has been given; whilst many of these approaches have a lower complexity compared to three-dimensional approaches, most are highly viewpoint dependant; making their use in real environments difficult. On the other hand, the use of a three-dimensional analysis technique can overcome viewpoint dependence — at the expense of computational cost — although with modern computing equipment this is less of an issue. In this section, a variety of techniques are discussed, some using only a single camera, others using multiple cameras, and several using magnetic position sensors attached to the joints of a subject.

Tanawongsuwan and Bobick [103] acquired joint information such as 3D position and orientation from subjects walking through a magnetic marker system, from which joint angle information was derived. Recognition was performed using dynamic time warping on the normalised joint angle information. The use of a magnetic marker system would not be practical in real-world situations, although it gives an insight into the recognition performance attainable from the use of a near-perfect source. Problems were reported with the algorithm being very dependant on the positioning of the markers on the subjects.

Wu et al. [119] also collected gait information from subjects using a marker based system; this consisted of joint angle measurements and their fluctuation over time. The variation of these measurements is likely to be non-linear, meaning that the use of Principal Component Analysis and linear classification techniques will result in sub-optimal performance. Therefore, a non-linear mapping function was applied to the gait data, transforming it to a new feature space of increased dimensionality, where Principal Component Analysis would then able to provide greater separation between subjects. This approach is referred to as Kernel-based Principal Component Analysis (K-PCA). Finally, a Support Vector Machine was used to classify the subjects.

Bhanu and Han [5] made use of a highly sophisticated three-dimensional model consisting of spheres and cones to represent the legs, arms, torso and head. The model had in excess of thirty degrees of freedom, which made fitting the model to two-dimensional silhouette data very difficult. Therefore several assumptions were made to reduce the dimensionality of the problem; this included the camera being stationary, the subject travelling in a straight path, and their limbs swinging parallel to their direction of travel. The static parameters of the model such as the size of body parts, were found using key-frames from the silhouette sequence. Kinematic features were then found by fitting the model to the sequence of frames, where the degrees of freedom were reduced by the use of the previously estimated static parameters. The static and kinematic features were then used for classification.

The use of a single viewpoint often results in problems with self-occlusion; caused by one limb obscuring the view of another, which can complicate the process of fitting a model. The use of a single viewpoint also makes it difficult to fit models containing a large number of degrees of freedom, as the data from a single viewpoint does not provide sufficient constraining properties, which can result in multiple non-optimal solutions being found when model fitting. Therefore, the use of multiple viewpoints can prove beneficial when fitting complex models.

It is possible to derive three-dimensional information from just two cameras using stereo depth reconstruction techniques, where the distance of a point from a pair of cameras can be found from the point's disparity. Stereo vision was used by Urtasun and Fua [108] to aid fitting a sophisticated three-dimensional deformable model[85] to the shape of the human subjects, and a motion model describing the deformation of the shape model was produced from data collected using a marker based computer-vision system to capture four subjects walking. The shape and motion models were then fitted to three-dimensional data collected from a multi-view stereo vision system using the least squares method.

Orrite-Uruñuela et al. [82] proposed a gait analysis technique where point distribution models were fitted to silhouette data from multiple viewpoints. A stick-model was then fitted to the resulting point distribution models. For gait analysis, only the hip, knee and ankle points of the skeleton were considered, and a gait cycle was treated as four discrete states; left support and right foot moving forwards, left support and right foot moving backwards, and the corresponding states with the right foot as support. Linear discriminant analysis was applied to the skeletal point data to improve the separation between the four gait cycle states. The CMU MoBo dataset[40] was used to demonstrate that this gait analysis technique was effective at tracking the skeletal points even with self occlusion in some viewpoints.

A similar method of fitting a three-dimensional model was proposed by Zhao et al. [125], where multiple views were used to improve model fitting performance. A skeletal model was initially fitted to the first frame in a sequence, with the position, orientation, body geometry and joint angles being manually chosen. Tracking was then performed on the subsequent frames to find the variation in the model's parameters, which could then be used for recognition.

The majority of single viewpoint based gait analysis techniques rely on the orientation of the subjects relative to the viewpoint staying relatively constant in order to provide optimum recognition performance. In many real world applications it would prove difficult to control the direction in which subjects walk, which poses a problem for many gait analysis techniques. Shakhnarovich et al. [91] collected video data of subjects walking from several different viewpoints simultaneously, then reconstructed three-dimensional volumes of the subjects. Two-dimensional silhouettes could then be synthesised from a virtual camera, placed perpendicular to the walking direction. This meant that the subject's walking direction did not affect the synthesised silhouette data. The resulting silhouette data could then be used as the input for a viewpoint dependant gait analysis algorithm.

3.6 Discussion

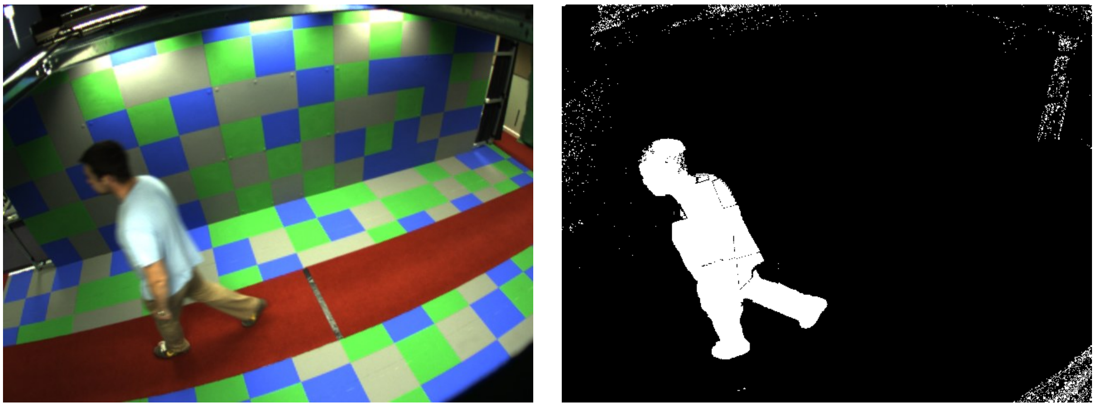
Gait has attracted much interest over the years, with early research being conducted by the medical and psychology community[74, 52]. As computing power became more readily accessible, investigation into automated gait recognition commenced[80]. The pace of research quickly increased, with the collection of several datasets, and the development of many analysis techniques[76]. Two of the largest and most widely known datasets are the Gait Challenge[84] and the University of Southampton dataset[93]. Both contain in excess of one-hundred unique subjects, were recorded from multiple viewpoints and include a limited set of covariates. Consumer grade video cameras were used for both datasets, which meant that the recording equipment was not time-synchronised, making the recovery of three-dimensional data difficult. The use of standard video camera equipment also meant that the process of transferring, editing and labelling the recorded footage would have been extremely time-consuming; limiting the size of dataset that could be collected in a reasonable period of time. Extremely good recognition results have been achieved for both datasets, where Veres et al. [109] was able to achieve 100% correct classification on the Southampton dataset, using the average silhouette analysis technique. Similar performance was also achieved on the Gait Challenge dataset by Sundaresan et al. [101], using a more sophisticated Hidden-Markov model based technique. Whilst excellent recognition results have been demonstrated by several researchers, several fundamental limitations still exist. Viewpoint dependence is a significant problem for many analysis techniques, where it proves difficult to match between different orientations of the subject; also very few results have been published that consider the matching of subjects against samples acquired at an earlier date. These are both important questions that must be answered in order to fully understand the limitations of gait and where the deployment of an automated recognition system could prove beneficial.

Chapter 4

The Revised Biometric Tunnel

4.1 Introduction

Analysis of the development dataset from the original Biometric Tunnel[70] raised several major issues; as discussed earlier in Chapter 2.3. Several of the acquired samples were found to be blank; containing no volumetric frames. Most of the captured samples featured significant noise artefacts, which had to be removed using aggressive post-processing. Another issue that caused great concern was that quite a few samples had sequences of frames where one or both of the subject's legs were missing — as shown in Figure 2.3. In order to find where the artefacts were being introduced, the output from each stage of processing needed to be inspected; unfortunately this information was not stored by the system. It was also impossible to repeat the processing of the collected data, as the original raw camera data was not saved to disk. This only left one option; collect a new dataset containing unprocessed video data. In order to do this, a range of modifications to the original system were necessary; this included significant alterations to the system's software and hardware, with the aim of improving data quality and the maintainability of the system. The capture software for the system was rewritten to facilitate the acquisition of unprocessed video data from the cameras; this led to the discovery that no time-synchronisation was performed between cameras in the previous system. The software was also changed to record video footage of the background before the capture of each sample, to minimize the time period between background estimation and segmentation. A batch processing system was implemented to automate the execution of the various computer-vision algorithms required to perform background segmentation, three-dimensional reconstruction and gait analysis. The new software is discussed further in Section 4.3. The layout of the hardware in the environment was also changed to allow easier access to critical equipment, such as computers and time-synchronisation units.



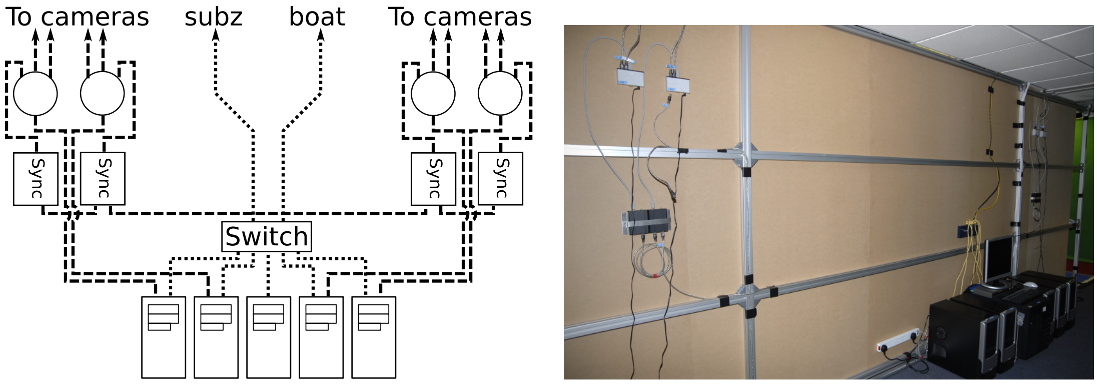
(a) Original video frame

(b) Silhouette

FIGURE 4.1: Example of poor background segmentation, where the subject's head and shoulders have regions missing

Once the Biometric Tunnel had been altered to enable the capture of unprocessed video data, a small dataset was collected to enable the validation and refinement of processing algorithms within the system. The dataset was acquired in a single day over a two hour period. Ten different subjects participated; each walking through the Biometric Tunnel four times; resulting in a total of 40 samples, although one sample was later found to be invalid due to an error in the implementation of the camera synchronisation algorithm. The composition of the dataset is shown in Appendix A. Analysis of the dataset was performed by conducting several leave-one-out recognition experiments, using average silhouettes produced from side-on, top-down and front-on viewpoints. The classification performance of the new dataset was found to be greatly improved over the original system's dataset. The receiver operating characteristic and the classification performance for each viewpoint are both given in Appendix B. It is apparent that the number of "bad" samples was significantly reduced using the new Biometric Tunnel configuration. The improvement in recognition accuracy is most likely due to the addition of time-synchronisation between cameras and the increased frequency of background estimation, as discussed in Section 4.3.

Whilst the classification performance and overall reliability of the data produced by the revised tunnel system was greatly improved, the visual quality of the reconstructed output was still quite poor. Possible causes were the use of a visual hull reconstruction technique with a partial intersection criteria — as discussed in Chapter 5.4 — and also background subtraction errors. The acquisition of background data for each sample is likely to have reduced background subtraction errors; as any drift in the camera and lighting characteristics would have been much smaller. Although, significant background subtraction errors were still present in the processed data; Figure 4.1 shows an example frame and the corresponding erroneous silhouette derived from background subtraction. It was found that the performance of the segmentation was poor when the subject occluded grey regions in the background; this was because the hue of the subject and the



(a) Layout of equipment

(b) Photograph of equipment layout

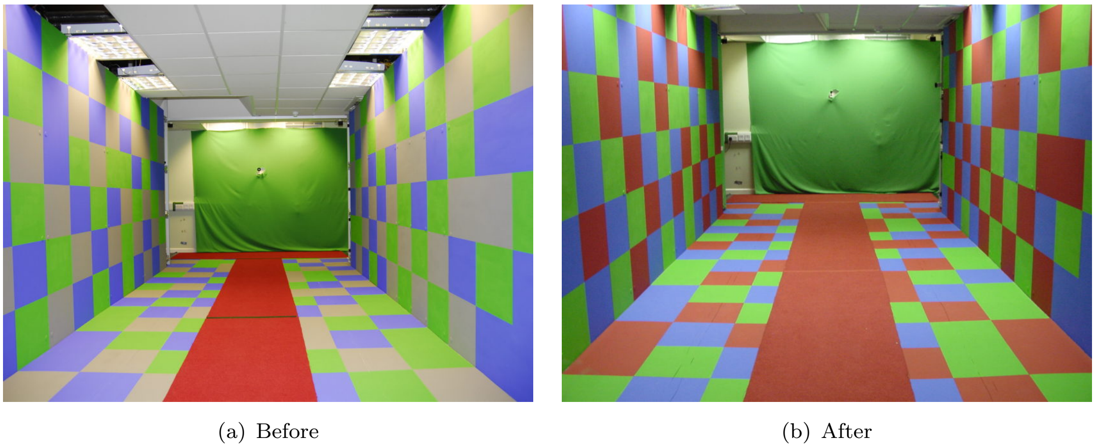
FIGURE 4.2: The modified Biometric Tunnel configuration

background were often similar, causing the background subtraction to mark the region as a shadow instead of foreground. It was decided that grey was an unsuitable colour to use on the tunnel walls and floor, as it often does not provide sufficient separation from the subject's clothing. Therefore a new colour was chosen, as discussed later in Section 4.4.

With the new improvements to the Biometric Tunnel showing promising results during testing, final modifications were made to the system to prepare for the collection of a large multi-biometric dataset. This included the addition of a camera for recording imagery of a subject's ear, and the replacement of the face camera with a better performing model. The batch-processing system was further developed, with additional features added to automate the execution of recognition experiments and data archival. Several key computers were also replaced with more modern equivalents, featuring large disk arrays for storing the collected data.

4.2 Alterations to system configuration

The layout of the equipment in the Biometric Tunnel was significantly revised, as it had been proving difficult to maintain and diagnose problems. The previous placement of the system's hardware around the environment resulted in long runs of cable between equipment; with key components such as timing-synchronisation units located in the suspended ceiling, making access difficult. All the computers were relocated to a single area on one side of the tunnel; greatly simplifying the Ethernet and IEEE1394 network cabling arrangements. This facilitated the move of the IEEE1394 hubs and synchronisation units to a more accessible location, above the computers. The revised system topology is shown in Figure 4.2(a). A monitor, keyboard, mouse and switching unit were installed above the computers; allowing control over any of the computers. The revised hardware layout as shown in Figure 4.2(b), made system diagnostics and maintenance much easier.



(a) Before

(b) After

Figure 4.3: Before and after the repainting of tunnel

Due to the experimental nature of the system, the video camera mounting brackets were fixed loosely, to facilitate adjustment of a camera's orientation during development; however this also meant that the slightest knock could cause the orientation of a camera to change. Therefore, the mounting brackets holding the cameras were secured, to reduce the likelihood of camera movement. Finally, a full recalibration of all cameras was performed, to ensure that the cameras were correctly characterised. The original green fabric path was replaced with a new carpet, which was firmly secured to the floor to improve safety. An intensely coloured red carpet was chosen, as it was the only brightly coloured carpet that was easily available and hard-wearing.

As discussed earlier, the quality of the reconstructed data was found to be poor; the most likely cause of this was the sub-optimal performance of the background segmentation. This was found to be due to the difficulty in separating subjects wearing pale coloured clothing from the grey areas of the background. The analysis in Chapter 4.4 confirmed that grey was a poor choice of colour, and found that red would be a much more suitable colour. Using these findings, the grey regions of the background were repainted red. Figure 4.3 shows the tunnel before and after the repainting of the grey background regions.

The computer responsible for controlling the system and performing volumetric reconstruction was replaced with a more modern and powerful computer, to improve the system's speed and ease development. Due to the substantial amount of data storage capacity required to collect a dataset comparable in size to the other existing gait datasets, a new hard-disk array was added to system. Previously, the data collected for a single sample was spread across multiple computers, with the raw video and silhouette data for each camera saved on the corresponding computer. This approach resulted in extremely fast writing of the recorded data, although it made management extremely difficult and often cumbersome. Three 750GB hard-disks were added to the main computer system, with two additional 1TB disks added during the experiment described in

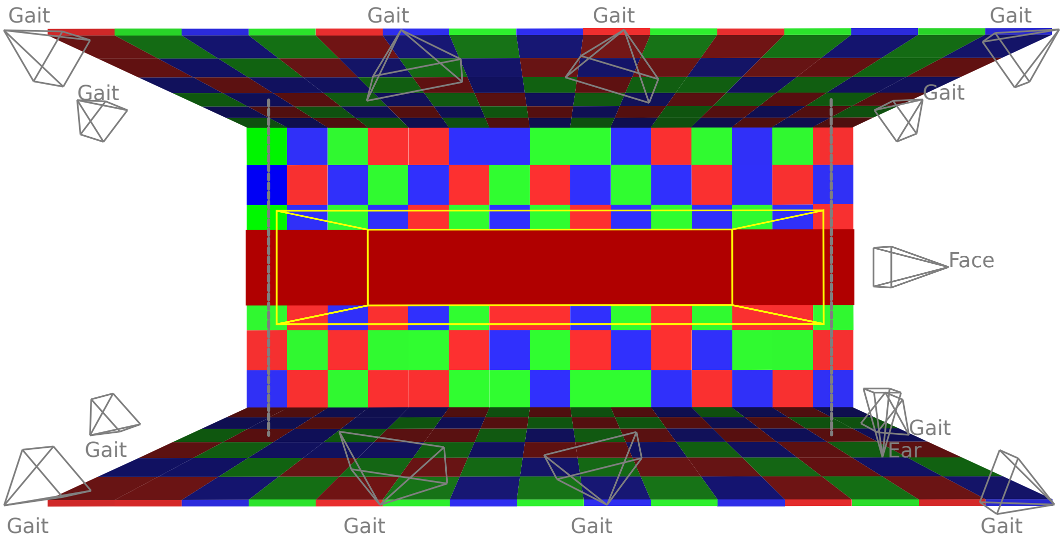


FIGURE 4.4: Final layout of the Biometric Tunnel

Chapter 6. Caching of the captured video data was performed to improve performance; where the data was initially saved to the local camera computers, then transferred to the main storage system in the background. A network attached storage system was used as a backup server for the collected data, which was located in a different building to the main system. The backup server employed a RAID-5 array of hard-disks, to protect the integrity of the archived data. The backup process was automated through the use of a special task in the batch-processing system, which transferred new samples to the backup server outside of office hours. Unfortunately, serious stability problems were experienced with both the main and backup storage systems, due to serious technical problems. As a result of this, the collection of the large dataset discussed in Chapter 6 was temporarily suspended until the problems were resolved and extensive reliability testing had been completed.

Multiple biometrics can be used in automated recognition systems to improve classification performance and make forgery attempts much more difficult. Therefore it was desirable to include additional sensors in the system for recording other biometrics, which could prove useful to other research projects. It was decided to add cameras for two additional non-contact biometrics to the system: face and ear; as both could be captured whilst the subject was walking through the measurement area. As discussed in Chapter 4.5, the original system's face camera was upgraded and a camera for recording ear imagery was added.

Four additional gait cameras were added to the system, during the collection of the large dataset described in Chapter 6. The new cameras were placed one metre from the floor in each corner of the Biometric Tunnel area. A large multi-colour LED was also added; located near the entrance of the Biometric Tunnel, which lit up red when the tunnel was busy and changed to green once the tunnel was ready. This provided a simple yet

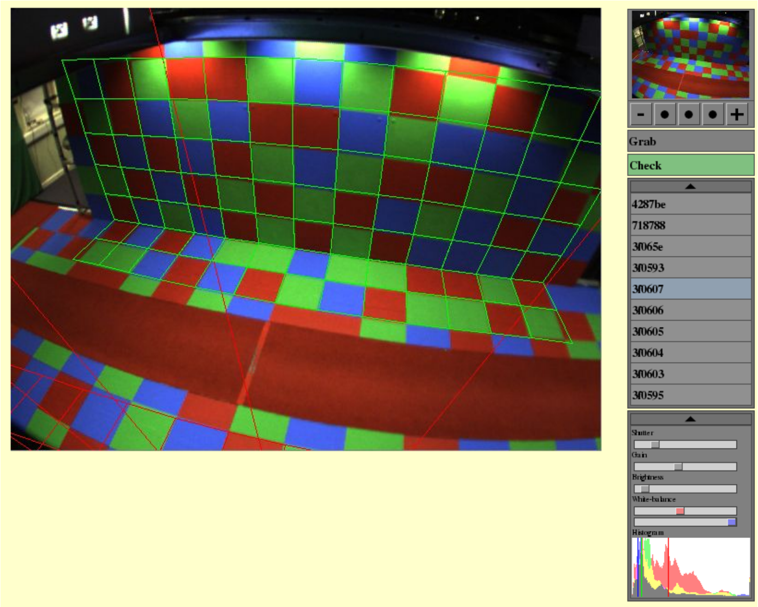
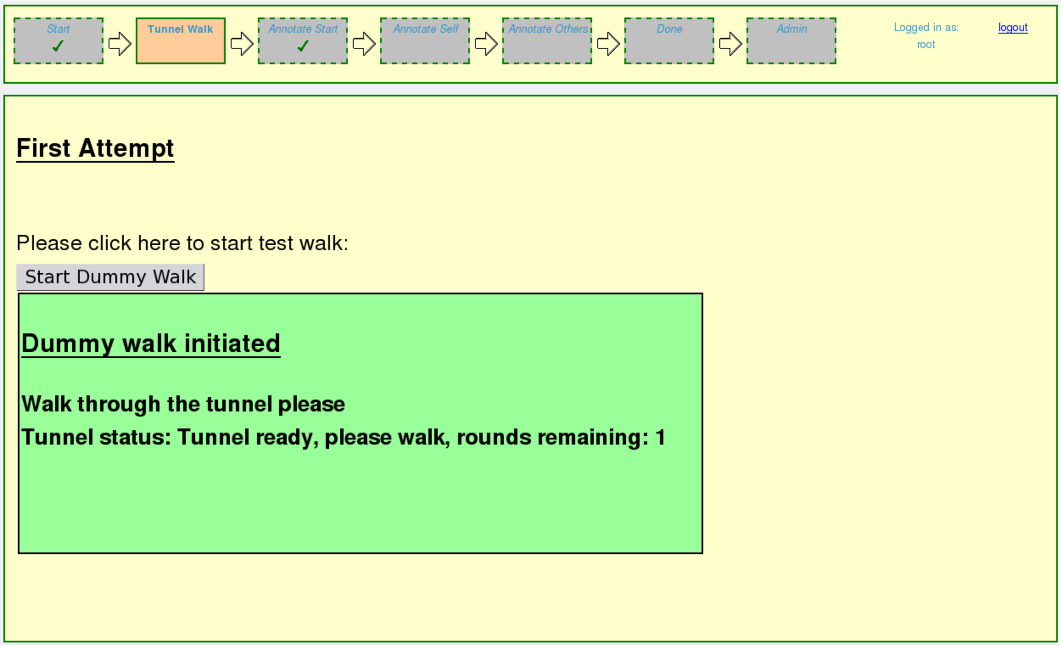


Figure 4.5: Screen-capture showing web-based application for viewing and controlling cameras within the Biometric Tunnel

intuitive status indicator for participants. The layout of the final system is shown in Figure 4.4.

4.3 Software overview

New camera agent software was written to capture and save video footage, without any processing of the data. The removal of the processing algorithms from the code-base greatly simplified both the camera and controller applications, facilitating the discovery of improper timing-synchronisation between camera agents. This issue resulted in an average timing-synchronisation error of 60 milliseconds between camera agents. It was also discovered that the background images for each camera were only acquired at the beginning of each capture session, allowing any drift in the lighting conditions to cause problems. The new camera and controller agent software rectified the timing-synchronisation issues using the time-stamp data saved by the cameras, and also performed background acquisition before each sample to minimise lighting variation. The capture control software was re-implemented; where its key functionality was exposed through a web-service, allowing the control of the system through a web-site. A sophisticated live view web-application was also written, to allow the monitoring of cameras in the tunnel area and the adjustment of their parameters; such as exposure and white-balance. It was designed to allow the use of a small hand-held internet tablet device to monitor the system's cameras; this allowed the interactive adjustment of focus and orientation. The application could also overlay a wire-frame model of the tunnel; making it easy to check for camera misalignment; as shown in Figure 4.5.



First Attempt

Please click here to start test walk:

Dummy walk initiated

Walk through the tunnel please

Tunnel status: Tunnel ready, please walk, rounds remaining: 1

FIGURE 4.6: Screen-capture showing the data collection experiment website; which controls the Biometric Tunnel and provides instructions for the participants

As discussed in Chapter 6, the collection of the large multi-biometric dataset was performed in collaboration with Sina Samangooei[88]; who also needed to conduct a similar large scale experiment. A common website was created that provided a front-end for both experiments, which allowed participants to enrol into the system, perform data capture, and complete the other experiment. A screen-capture of the website is shown in Figure 4.6.

In order to ensure that the Biometric Tunnel could be operated by users of varying ability, the procedure for managing the system's software needed to be as easy as possible. Therefore the system's website was extended to feature an administration area, which provided simplified controls and diagnostics for the entire system. This required the addition of remote procedure call interfaces to several of the system's key software applications. A simple web-service for controlling the running of programs on each computer was implemented; allowing the administrator to start and stop key applications on all the system's computers, from a single area in the administration section of the website. The administration interface also provided controls to allow the supervisor to easily select and create new capture sessions and datasets. An application for batch producing identity cards for the data collection experiments was produced, where each card had a bar-code containing a unique identifier. This meant that large numbers of identity cards could be produced before the experiment, reducing the time taken for each participant.

The processing algorithms contained within the original system were removed from the camera and controller software and then placed in their own independent executable programs; where image data was passed between algorithms using files. This meant that the performance of each algorithm in the processing system could be evaluated in isolation from the other algorithms. After the collection of the small unprocessed dataset, many of the processing algorithms were replaced with more sophisticated versions, which were evaluated against the new dataset. The colour interpolation method was changed from a Bilinear algorithm to a Cok based implementation, as discussed in Section 5.2. This resulted in colour frames with finer detail and reduced colour error along edges. The background segmentation and shadow suppression algorithms were replaced with a more robust background segmentation algorithm, which used a semi-normalised RGB colour space to reduce the effect of shadows, as described in Section 5.3. The shape from silhouette algorithm was replaced with a faster multi-resolution implementation, using a full camera intersection criterion to improve the accuracy of the reconstructed data; which is covered in Section 5.4. The calibration software was re-written, using connectivity analysis to locate world points within the camera images; as described in Section 5.5. The resulting reconstructions proved more accurate and had less problems due to segmentation errors.

In order to process a sample, each processing algorithm would have to be executed in turn on the sample. Manually performing this on a large set of samples would have proved to be a very time consuming and inefficient task; therefore a system was required to automate the execution of the processing applications. The system needed to be capable of managing the processing of a large number of samples, whilst keeping track of what processing stages had been performed on each sample, to avoid redundant processing operations. A certain degree of flexibility was desired from the system, such as the ability to prioritise the execution of certain tasks. In order to develop and troubleshoot processing algorithms, a means for inspecting any errors raised was also required.

A new batch processing system was developed, to satisfy the above requirements. The system was implemented using Python; a powerful yet flexible high-level scripting language. Sample meta-data was stored using a MySQL database, providing robust and scalable storage for the meta-data. The MySQL database engine provides many useful features that were used in the new system to ensure data integrity and stability, such as table locking, to avoid concurrence issues; foreign key constraints, to enforce data integrity; and transaction level processing, to ensure that failed tasks do not cause data corruption. Backups of the entire database and its underlying structure were performed using off-the-shelf software on a regular basis. The processing of a sample was split up into a set of tasks, each executing a single image processing algorithm. Each task consisted of a simple Python script, which could either call an external program to carry out the processing, or implement the processing directly using Python code. Upon the completion of a task on a sample, any down-stream tasks were automatically added to the execution queue; unless specified by the user or the task code. The database was used to keep track of pending and failed processing operations, which meant that

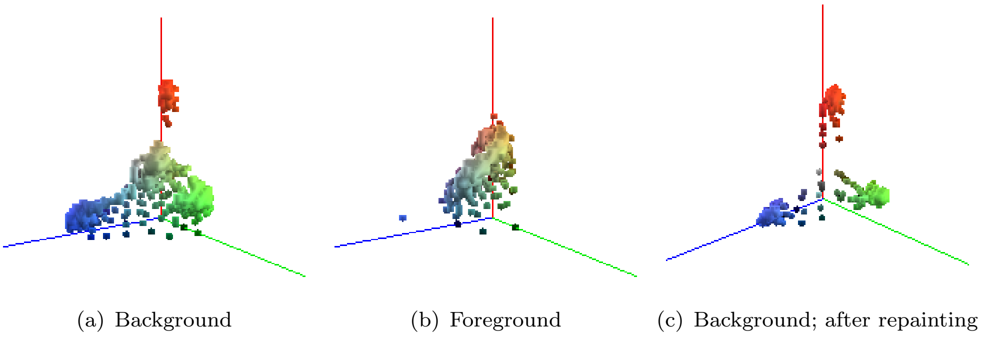
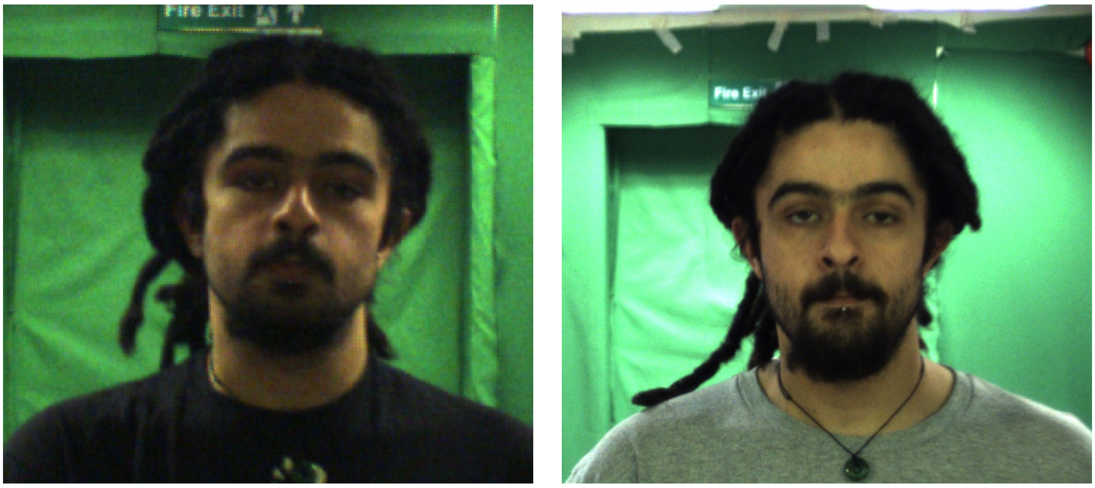

(a) Background (b) Foreground (c) Background; after repainting

FIGURE 4.7: Analysis of colour occurrence in segmentation regions

processing could be resumed after a software failure or system shut-down. For each operation, a priority level and status value could be assigned to allow management of pending tasks. Any failed operations would remain in the database until cleared and could be easily identified by their status value. A special bootstrap task was created to start the processing of a sample by adding all relevant tasks to the process queue. Gait analysis techniques such as the average silhouette were also written as processing tasks. The system made it easy to process and perform gait analysis on an individual sample or an entire dataset with minimal effort. This approach provided much greater flexibility compared to the use of shell scripts. The batch-processing system was designed to work safely with multiple instances running, meaning that large processing runs could be performed using a Linux based compute cluster.

4.4 Analysis of background colours

In order to find a more suitable colour to replace the grey regions in the Biometric Tunnel, the evaluation dataset collected from the revised system was analysed to find the range of colours present in the foreground pixels. The post-processed silhouette data was used to mask the original colour frames; where the masked pixels were added to a three-dimensional luminance-normalised RGB histogram. A similar histogram was also calculated for the background pixels from the dataset. The colour distribution for the foreground and background pixels can be seen in Figure 4.7(a); it can be seen that the background has four predominant colours, red, blue, green and grey; the colours used on the walls, floor and carpet. Figure 4.7(b) shows the colour distribution in the foreground pixels; the prevalent foreground colours are mostly centred around the monochromatic point, this suggests that most people wear clothes with large areas of white, grey, black or pale colours. The figures confirm that there is good separation between the typical foreground colours and the red, green and blue colours used in the background, and also reiterate that grey is a poor choice of colour to use in the background.



(a) Face image from previous dataset, which features poor lighting and high noise levels

(b) Improved face image, using additional lighting and new camera

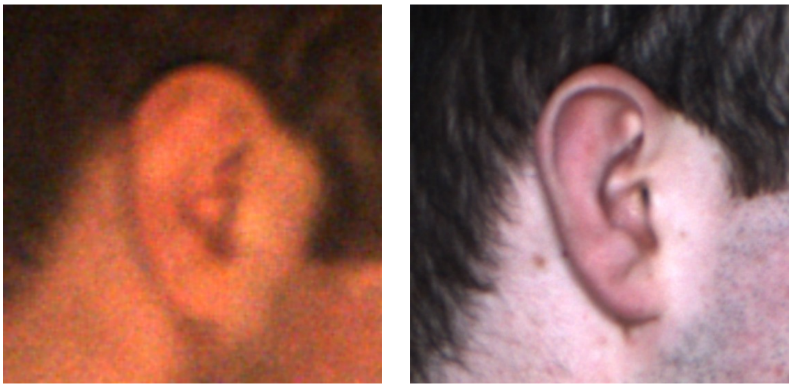
Figure 4.8: Images from previous and improved face camera configurations

Using the findings of the colour analysis, it was decided to repaint the grey regions of the background in a highly saturated shade of red, similar to the carpet. Figure 4.3 shows the tunnel before and after the repainting of the grey background regions. Colour analysis of images captured from the tunnel after repainting shows that only saturated colours remain in the background, as shown in Figure 4.7(c)

4.5 The addition of cameras for face and ear

The original system by Middleton et al. [70] was configured to capture video of a subject's face and upper body using an additional dedicated camera, which was separate to the gait cameras. The recorded video data was found to be of poor quality, as shown in Figure 4.8(a); this was partly due to insufficient frontal illumination and also the chosen camera's poor signal to noise ratio. As a result of this, experimentation was carried out into improving the subject's illumination using a variety of lighting sources, in different positions. The best compromise between illumination quality, practicality and the comfort of participants was achieved using using a pair of point light sources positioned either side of the tunnel, which were pointed inwards towards the subject's face. The camera was replaced with an improved model; featuring an improved resolution and signal to noise ratio. These measures provided a significant improvement in the attainable image quality, as shown in Figure 4.8(b).

A high resolution camera was placed at the side of the tunnel to record imagery of the subject's ear; as it was decided that this data could prove useful for other research projects. Initial experimentation found that by using the existing lighting in the tunnel, it was impossible to achieve images of a suitable quality. This was due to the extremely



(a) Continuous lighting; not even the most powerful units could provide enough illumination to permit fast shutter speeds, resulting in motion blur

(b) Strobe lighting; the high intensity burst provides sufficient illumination to use extremely fast shutter speeds.

Figure 4.9: Cropped section of image from ear camera, using continuous and strobe lighting.

high speed in which the subject passed through the frame of the camera, meaning that a very fast shutter speed was required to avoid motion blur. This in turn required an extremely high camera gain or very powerful illumination to achieve such shutter speeds. Attempts were made using various continuous lighting systems to provide the illumination required for the participant's ear; although no practical solution could be found. Figure 4.9(a) shows the best results achieved using continuous lighting, which is unusable due to the poor signal to noise ratio and the high degree of motion blur present. The continuous lighting systems evaluated were unable to provide sufficient light output; with the most powerful lighting system causing participants to complain of discomfort. Therefore it was decided to use photographic strobe units, which produce an intense burst of light for a very short duration of time, essentially freezing any motion present in a still image. The use of strobe units provided sufficient illumination with minimal discomfort to the participants, as shown in Figure 4.9(b). Unfortunately, this meant that only a single frame could be captured, instead of a video sequence. An electronic strobe control circuit was constructed, allowing the ear camera to trigger the flash using its strobe output. New camera agent software was written specially for the ear camera, which initialised the camera for use with an external flash and fired the flash upon image acquisition. The system was configured to trigger the ear camera and strobe units when the participant crossed through the exit break-beam sensor; this also meant that recording from the other cameras would have stopped beforehand, which meant that the flash would not be recorded by any of the other cameras in the system.

Chapter 5

Video Processing and 3D Reconstruction

5.1 Introduction

The Biometric Tunnel is a complex system, featuring a wide range of image-processing algorithms, to convert the raw video footage from the cameras into three-dimensional volumetric data, suitable for gait analysis and recognition. The processing sequence for a typical sample recorded using the Biometric Tunnel is shown in Figure 5.1. Image data recorded by the cameras within the Biometric Tunnel system was streamed to the connected computers over a IEEE1394 network, in a raw unprocessed format. Upon arrival at the computer, the digitised images were converted to colour from their native format using Cok interpolation, discussed in Section 5.2. From the derived colour images, the subject was identified from the background, through a process known as background segmentation; where previously recorded video footage of the Biometric Tunnel area was used as a reference. The acquired background imagery was modelled using a unimodal normal distribution for each colour channel, where the colour-space was partially luminance normalised to reduce the effect of shadows. Binary morphological operators were then applied to the segmented images to smooth the shape of the silhouettes and reduce segmentation noise. The post-processed silhouette images from all gait cameras were then combined, using the shape from silhouette three-dimensional reconstruction technique. Finally, gait cycle analysis was performed to find the most likely gait cycle within the recorded sample. In this chapter, a detailed review of the techniques used to process the data acquired from the Biometric Tunnel is given.

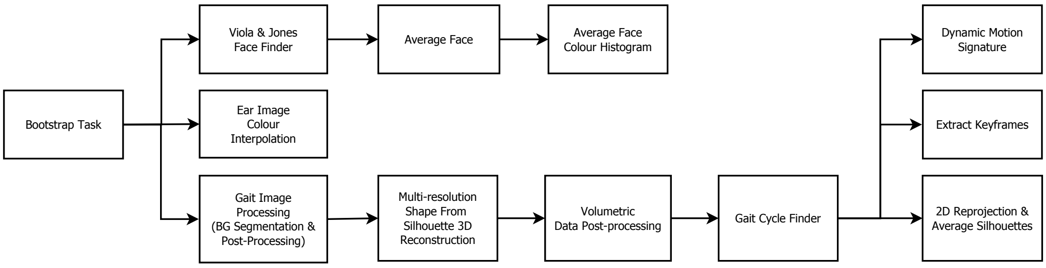

Bootstrap Task
Viola & Jones Face Finder
Average Face
Average Face Colour Histogram
Ear Image Colour Interpolation
Gait Image Processing (BG Segmentation & Post-Processing)
Multi-resolution Shape From Silhouette 3D Reconstruction
Volumetric Data Post-processing
Gait Cycle Finder
Dynamic Motion Signature
Extract Keyframes
2D Reprojection & Average Silhouettes

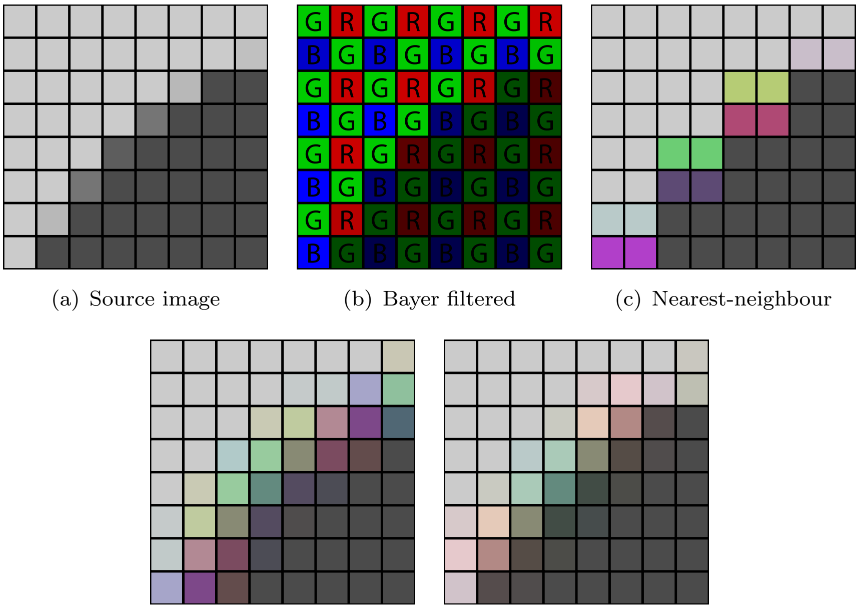
FIGURE 5.1: Execution tree for processing a sample

5.2 Recovery of colour information

Most digital video cameras use a charge-coupled device (CCD) to produce a colour image from the observed scene, where the light falling upon each photo-site of the sensor is converted to a charge, measured and then digitised. The photo-sites are unable to distinguish between the different visible wavelengths of light arriving on an individual site, which means that a CCD can only measure luminance and not colour. The most popular solution to this problem is to apply colour sensitive filters to each photo-site, where each is sensitive to different light wavelengths compared to its neighbours. The most commonly used arrangement of colour filters is known as the Bayer pattern; a 2×2 pattern, with two sites sensitive to green light, one red, and the other blue. The result of filtering the test image in Figure 5.2(a) by a Bayer array is shown in Figure 5.2(b). A full colour image can then be reproduced by interpolation, using a variety of techniques.

The most basic form of colour interpolation is to take the nearest neighbouring colour value for each pixel where the colour is unknown. Applying nearest-neighbour interpolation to the image shown in Figure 5.2(b) results in a reconstructed image with strong colour artefacts surrounding boundary pixels, as shown in Figure 5.2(c). Nearest-neighbour colour interpolation has the advantage of being extremely easy to implement and requires very little computational time.

The use of bilinear interpolation results in a smoother colour image, with less colour artefacts compared to the nearest-neighbour technique, as shown in Figure 5.2(d). More sophisticated colour interpolation methods make use of assumptions regarding the properties of a typical colour scene; such as the colour remaining locally constant. One such approach proposed by Cok [20], assumes that the ratios of green to red and green to blue remain stable over the immediate neighbourhood. It can be seen in Figure 5.2(e) that when put alongside nearest-neighbour or bilinear interpolation, Cok interpolation results in lower levels of colour distortion artefacts. The approach by Kimmel [55] uses improved gradient calculations, resulting in a marginal reduction of artefacts; at the cost of computational complexity[73]. State of the art interpolation methods such as Variable Number of Gradients[16], Pixel Grouping[53] and ADH[46] provide further improvements



(a) Source image (b) Bayer filtered (c) Nearest-neighbour

(d) Bilinear (e) Cok

Figure 5.2: Filtering of image by Bayer array and recovery of colour information by interpolation

to reconstruction quality, but are even more demanding in terms of computational requirements. It was decided to use Cok interpolation in the revised Biometric Tunnel, as it provided a good compromise between accuracy and complexity.

5.3 Background estimation and segmentation

In order to carry out 3D reconstruction as described in the next section, it is necessary to identify all the pixels in the camera images that are occupied by the subject. This is achieved in two stages; background estimation, where the distribution for each pixel in the background is modelled; and background segmentation, the labelling of pixels as to whether they are likely or unlikely to belong to the background.

The most basic form of background estimation is to take a single snapshot of the scene when no foreground objects are present. Segmentation can then be performed by measuring the difference for each pixel between the current frame and the previously acquired reference frame; if the distance exceeds a predefined threshold, the pixel is marked as foreground. This approach leads to sub-optimal performance; as the acquired reference frame is distorted by sensor noise and the use of a global threshold value does not account for the varying noise characteristics of the pixels across the sensor. By recording a sequence of frames without any foreground objects present, it is possible to characterise the statistical distribution of the background for each pixel in the image. In this case, segmentation can be performed by calculating the distance between the test pixel and the mean of the background distribution; if the distance is greater than a predetermined threshold, the pixel is marked as foreground. The segmentation threshold is chosen on a per-pixel basis, as a multiple of each background pixel's standard-deviation. For an indoor environment where the background is fixed, the only source of colour intensity deviation will be the measurement noise from the sensor, which means that each colour channel can be sufficiently approximated by a single Gaussian distribution. In an outdoor scene where some fluctuation in the background is present, the use of a single Gaussian approximation may be insufficient to accurately model the background, resulting in reduced segmentation accuracy[118].

A more sophisticated approach is to approximate the background as a mixture of multiple Gaussian distributions; this is ideal for scenes with fluctuating objects, where a pixel's colour may vary between several colours, such as the green of a tree's leaves and blue from the sky behind the tree[100]. By continuously updating the background model, it is possible to track any drift in the scene's ambient surroundings, such as lighting variation or the addition of a parked vehicle.

Other techniques to account for background fluctuation include Kalman filtering[86] and localised motion compensation[33]. Another notable approach utilises both colour and depth information, which is derived from a stereo camera pair[38]. It is also possible to identify the subject within the image by labelling pixels where movement is present, which can be achieved by comparing the difference between subsequent frames. This approach has minimal computational requirements, although it is unable to reliably label slow moving or stationary objects. The use of a dense optical flow based technique[15] provides more robust segmentation at the expense of complexity.

Many background segmentation approaches are strongly affected by shadows, resulting in the labelling of false-positives. A luminosity normalised colour space can be used to reduce the segmentation algorithm's sensitivity towards shadows[47]. Another approach is to apply a separate shadow removal algorithm to the segmented data, comparing the colour difference between foreground pixels and the background model[19]. Whilst the use of shadow removal processing improves the segmented output, it is very difficult to completely remove false-positive matches caused by shadows, as the ambient reflection of light off other nearby surfaces causes the shadow regions to have a slightly different colour. In the Biometric Tunnel, a partially normalised colour space was used in conjunction with a uni-modal Gaussian distribution to model the background.

5.4 Three-dimensional reconstruction

The process of approximating an object's three-dimensional shape from two-dimensional data, such as photographs or video, is known as three-dimensional reconstruction. A wide range of techniques for reconstruction exist, some building upon concepts found in the human visual system.

The variation in depth across an observed surface can be estimated from the change in tone over the object; this is known as Shape from Shading[123]. Unlike most other techniques, only a single image from one viewpoint is required; although several assumptions are required; the surface's colour does not vary in tone; and also the material exhibits Lambertian properties, meaning that no specular reflections occur.

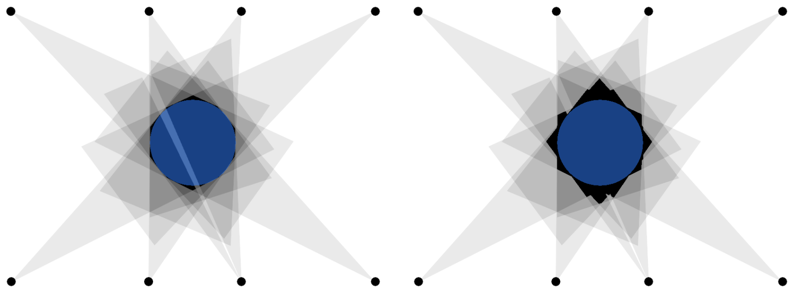
It is also possible to estimate the shape of an object from a single camera, provided that the object moves relative to the camera in the recorded footage; using a technique called structure from motion[102]. This is achieved by identifying landmark points on the object and tracking their movement throughout the sequence of frames. The movement of the points relative to each other can be used to determine the location and trajectory of the points in three-dimensional space. Such techniques require surfaces featuring distinctive and non-repeating patterns, otherwise it is difficult to reliably identify and track landmark points. By recording an object placed on a rotating platform, it is possible to reconstruct the object with full coverage; with the assumption that the target object is rigid[83].

The use of two cameras facilitates the calculation of stereo disparity, which can be used to determine the depth of distinctive features in a scene[14]. Similar to depth from motion, the presence of repetitive patterns or smooth non-detailed regions can adversely affect the accuracy of the reconstructed output. The previously discussed single-camera reconstruction techniques can be used to supplement the stereo techniques [41]. Stereo camera systems are typically unable to provide a complete reconstruction of the observed object, instead only providing information on the surface facing the cameras.

By using three or more cameras, it is possible to produce an approximation of the entire object — instead of just the front surface. The volume occupied by the intersection of the re-projected silhouettes is known as the convex hull[60], which can be found using solid-geometry techniques to find a polyhedron formed by the volume-intersection of the re-projected silhouette cones[68]. Another very popular approach is to divide the reconstruction volume into a three-dimensional grid of equally spaced cubic elements; known as voxels. Each element is then tested for occupancy by establishing whether the corresponding location in all camera images is occupied. This technique is commonly referred to as Shape from Silhouette reconstruction[19].

$$\mathbf{V}(x,y,z) = \begin{cases} 1 & \text{if } \Sigma_{i=n}^{N} \mathbf{S}_n \left(M_n \left(x,y,z \right) \right) = N \\ 0 & \text{otherwise} \end{cases} \tag{5.1}$$

Where V is the reconstructed 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. S_n is the silhouette image from camera n where $I_n(u,v) \in \{0,1\}$, and $M_n\left(x,y,z : u,v\right)$ is a function that maps



(a) Complete Intersection

(b) Relaxed Criteria

FIGURE 5.3: Effect of relaxing shape from silhouette intersection criteria

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 occupied if all cameras observe foreground pixels at the corresponding locations; this means that a single false non-silhouette pixel will have a significant impact on the reconstruction.

Modifying the shape from silhouette algorithm to accept voxels where k or more cameras observe silhouette pixels adds a certain degree of robustness against background segmentation false-negative errors, at the expense of reconstruction accuracy, as shown in Figure 5.3.

$$\mathbf{V}\left(x,y,z\right) = \begin{cases} 1 & \text{if } \Sigma_{i=n}^{N}\mathbf{S}_n\left(M_n\left(x,y,z\right)\right) \geq k \\ 0 & \text{otherwise} \end{cases} \tag{5.2}$$

In order to evaluate the effect of false-negative background segmentation errors on reconstruction quality a simple experiment was conducted; where a volumetric sphere was synthesised and projected to eight different camera viewpoints; resembling those of the Biometric Tunnel. The derived images were distorted by non-correlated false-negative noise and then used for Shape from Silhouette reconstruction, with the results compared against the ground-truth sphere. This process was repeated for varying levels of noise and using different visibility criteria for the Shape from Silhouette reconstruction. Figure 5.4 shows the effect of background segmentation error against reconstruction error, for differing values of k. It can be seen that without any segmentation errors, the full eight camera visibility criterion resulted in the most accurate reconstruction; although with the introduction of segmentation error, the accuracy degraded rapidly. The use of a more relaxed criterion resulted in a poor reconstruction accuracy when low segmentation noise levels were present; although with greater levels of noise it proved more robust. It can be seen that the performance for the non-strict criterion reconstructions show improved accuracy when a certain level of segmentation noise is present; this is because the presence of false-negative segmentation noise reduces the likelihood of false-positive reconstruction errors. This effect is unlikely to be observed with real-world data, where the segmentation noise is often highly correlated.

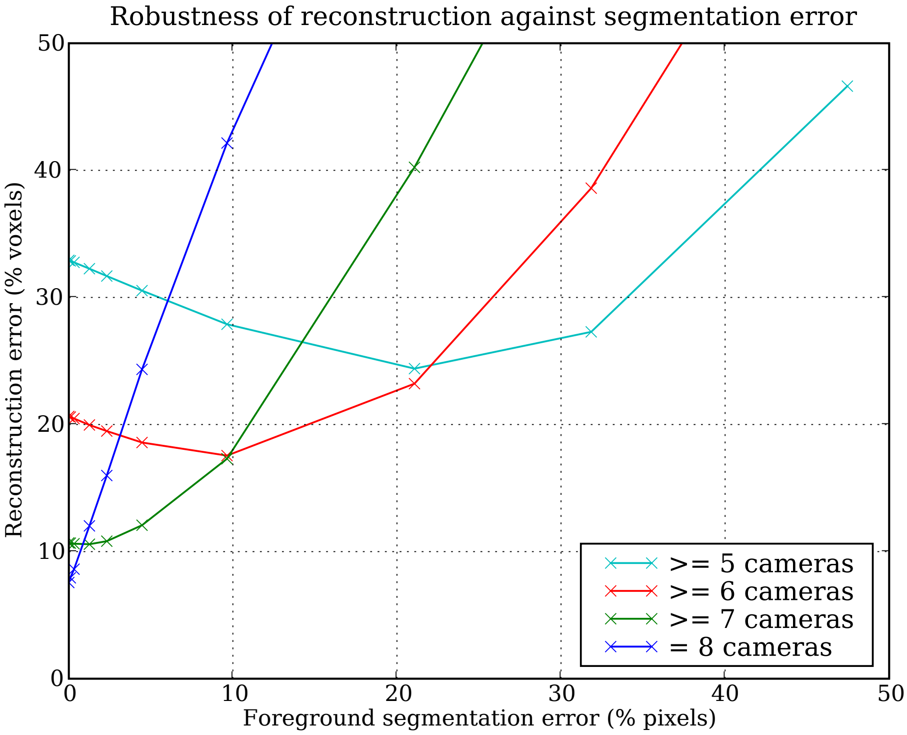
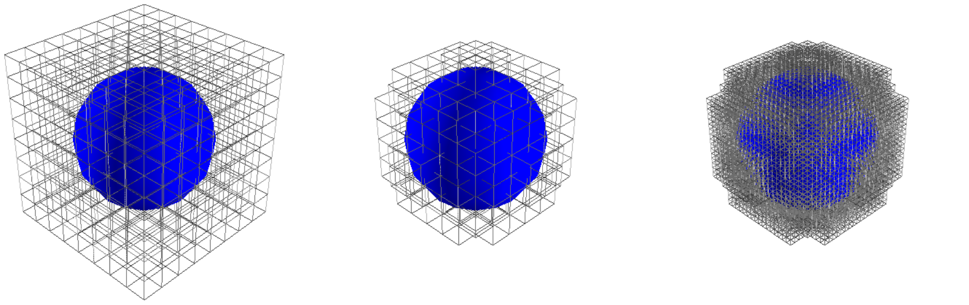


Figure 5.4: Effect of uncorrelated false-negative segmentation errors on reconstructed volume, with varying reconstruction criteria

Silhouette based reconstruction techniques such as Shape from Silhouette do not make use of the colour information present in a scene, which can be used to further constrain the shape of a reconstructed volume. If a surface does not exhibit specular reflections, then any given point on the surface will appear on all visible cameras as the same colour; this means that if given a particular voxel, the colour observed at the corresponding location on each unoccluded view should be similar, otherwise the voxel can be removed. Colour-consistency based reconstruction techniques typically provide superior quality reconstructed output compared to Shape from Silhouette reconstruction, and are also capable of correctly handling convex surfaces. Also, background segmentation is not required for some methods, meaning that reconstruction artefacts due to segmentation errors are no longer an issue.

Unfortunately the removal of inconsistent voxels is a non-trivial task, as the visibility of a voxel must be ensured before it can evaluated; upon removal, the visibility of other voxels must be updated. A variety of different strategies for the removal of inconsistent voxels exist, as described in the review papers by Slabaugh et al. [98], Dyer [32]. When reconstructing objects with no variation in colour and minimal surface detail, colour-consistency based techniques will provide very little benefit over Shape from Silhouette based reconstruction. As with many of the other reconstruction techniques, objects featuring repeating patterns or fine detail beyond the resolving power of the camera's



(a) Initial division of volume (b) Removal of empty regions (c) Sub-division of partially occupied regions

Figure 5.5: Multi-resolution reconstruction strategy

sensor will lead to reconstruction inaccuracies. Such techniques are also extremely sensitive to errors in camera calibration, where areas of detail may no longer be aligned for all cameras; potentially resulting in the removal of valid regions. Compared to Shape from Silhouette based reconstruction, colour consistency based approaches are much more computationally expensive, due to the additional complexity involved in the evaluation of voxels and visibility tracking.

In the Biometric Tunnel, shape for silhouette reconstruction is used, due to its simplicity and robustness to camera mis-calibration. A naive shape from silhouette implementation without any optimisation would prove to be very slow, as the calculations required to map the three-dimensional world coordinates to image coordinates prove to be very costly. Assuming that the position and orientation of the cameras remains constant, the mappings from world-coordinates to image-coordinates can be pre-computed and stored in look-up tables; this replaces the slow floating-point calculations with faster memory access operations. The use of lookup tables achieves a significant reduction in processing time, although further increases in efficiency are required in order to facilitate real-time processing.

Typically in the entire reconstructed volume, only a small proportion of it is occupied; therefore a large amount of time is spent evaluating large empty regions. To improve the efficiency of the algorithm a multi-resolution strategy is often employed; several passes are made of the volume at differing levels of resolution; where only regions determined to be of interest are then processed at a finer resolution. This is shown in Figure 5.5. At a coarse resolution, each test region will consist of many voxels; whilst at the finest level of detail, a region will consist of a single voxel. When evaluating a region's occupancy, there exist three possible cases; completely empty, complete occupancy and partial occupancy. The original Biometric Tunnel by Middleton et al. [70] used a simple dual-resolution reconstruction approach, where a low-resolution reconstruction was performed to find a bounding box for the volume occupied by the subject. Full-resolution reconstruction was then performed inside the bounding box. The low-resolution reconstruction algorithm only evaluated a single pixel location within the corresponding area for each camera,

Optimisation	Time/Frame	Additional memory
None	30s	-
Look-up tables	187ms	174 MB
Multi-resolution	25ms	232 MB

Table 5.1: Comparison of optimisation strategies for shape from silhouette algorithm

located at the centre of the test region. This meant that small areas of detail could be missed; therefore the bounding box was grown by a predetermined amount to reduce the likelihood of this happening. The full-resolution reconstruction algorithm used a six or more camera criterion for labelling a voxel as occupied. Whilst the algorithm featured some improvements in efficiency, it was still a comparatively simple multi-resolution approach.

The revised Biometric Tunnel used a new multi-resolution implementation of the shape from silhouette algorithm, which was able to establish whether regions were empty, fully occupied or partially occupied; unlike the previous implementation that could only establish that a region was occupied. Only partially occupied regions were evaluated at finer levels of detail; reducing the number of redundant voxel evaluations, compared to the previous implementation where all voxels inside the bounding box were evaluated. The algorithm operated at three resolution levels, where the reconstruction volume was split into ten regions along the y axis, then split into sub-regions by dividing the regions along the x, y and z axes to result in $6 \times 3 \times 17$ (306) sub-regions per region. Finally, each sub-region was divided into individual voxels. For each region and sub-region, a list of occupied pixels for each camera was pre-computed, to ensure that each pixel was only evaluated once.

When evaluating a region, it was considered empty if any camera observed no foreground pixels within its corresponding image regions. Alternatively, if all cameras observed only foreground pixels within their corresponding image regions, then the test region was marked as fully occupied. If neither of these cases was true, then the region was considered to be partially occupied; leading to finer-grained evaluation. Each sub-region was evaluated for occupancy in the same manner as their parent regions. If a sub-region was classified as partially occupied, a final full-resolution Shape from Silhouette reconstruction was performed for the sub-region; using the pre-calculated look-up tables to improve speed.

All the data needed to perform reconstruction, such as region data, pixel occupancy lists and lookup tables was held in a single linear forward-read only memory structure to provide the best possible performance. The multi-resolution reconstruction algorithm provided a significant boost in processing speed, as shown in Table 5.1.

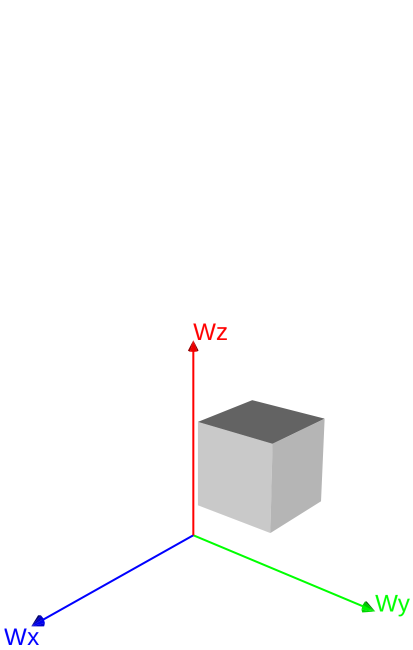


Figure 5.6: An object in the world-space must first be translated and rotated to align it with the coordinate system of the target camera; before scaling and perspective transformation can occur

5.5 The calibration of cameras

In order to perform three-dimensional reconstruction as discussed in the previous section, each camera must be precisely characterised; such that a mapping is found that transforms any given position in the 3D world to a position on the corresponding camera's image. To do this, an appropriate coordinate system must first be chosen for the environment, including the units of scale, origin and alignment of the primary, secondary and tertiary axes.

Each camera can be approximated using a pin-hole model, where there exists a mapping from the world's chosen coordinate system to the image plane, determined by a combination of parameters that can be grouped into two categories; extrinsic and intrinsic parameters. The first describing the translation and rotation required to align the origin and axes of the world's coordinate system to that of the camera; as shown in Figure 5.6. The camera's intrinsic properties describe the scaling and translation due to the camera's optics and the conversion from the world's units to the sensor's units; pixels.

The position of a point in world space can be expressed in the camera's coordinate system by the transformation described by the parameters **T** and **R**; where **T** is the translation of the camera relative to the world's origin, whilst **R** is a 3×3 matrix that rotates the world-space's axes onto those of the camera. The translation and rotation matrices are combined into a 3×4 transformation matrix, known as the extrinsic matrix;

where the world coordinate $\mathbf{W}$ is a 4D homogeneous vector.

$$\mathbf{E} = [\mathbf{R}\,|{-\mathbf{RT}}] \tag{5.3}$$

The factor required to scale the units of the world's coordinate system to that of the images, is defined as the ratio of the lens' focal-length to the camera sensor's pixel size:

$$\alpha = \frac{f}{P} \tag{5.4}$$

If the pixels on the sensor are not square, then separate values of α must be used for the x and y axes of the sensor.

The location of the lens' principal point on the sensor is given by P_x and P_y, in units of pixels. Most cameras exhibit square pixels and no skewing; thus a single value for α will suffice, and the skew factors can be ignored; therefore, the intrinsic matrix is:

$$\mathbf{I} = \begin{bmatrix} \alpha & 0 & P_x \\ 0 & \alpha & P_y \\ 0 & 0 & 1 \end{bmatrix} \tag{5.5}$$

As both are linear, the intrinsic and extrinsic transformations may be combined into a single transformation, described by a 3 × 4 matrix:

$$\mathbf{P} = \mathbf{IE} \tag{5.6}$$

Using the derived transformation matrix, any point in the world-space can be expressed as a homogeneous pixel location within the camera image:

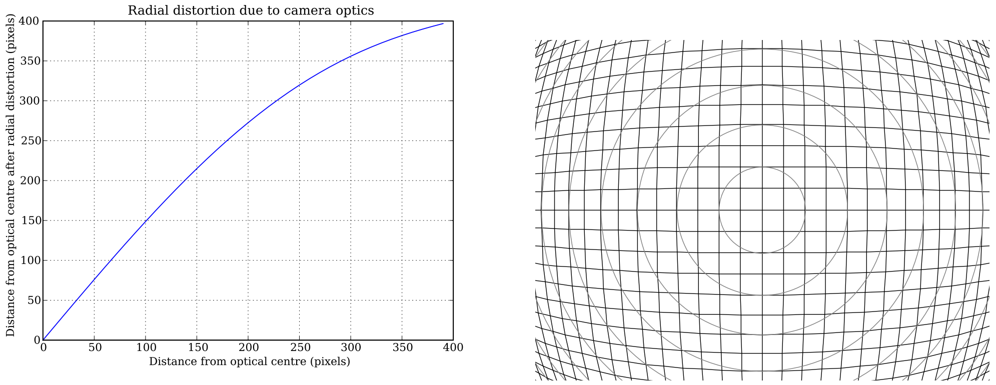
$$\begin{bmatrix} C_x \\ C_y \\ C_z \end{bmatrix} = \mathbf{P} \begin{bmatrix} W_x \\ W_y \\ W_z \\ 1 \end{bmatrix} \tag{5.7}$$

The final two-dimensional image coordinates are found by applying a perspective transformation; normalising the coordinates by C_z.

$$C_u = \frac{C_x}{C_z} \tag{5.8}$$

$$C_v = \frac{C_y}{C_z} \tag{5.9}$$

Unfortunately camera lenses are imperfect, as they often exhibit high levels of radial distortion; a non-linear transformation of the image centred around the principal point of the lens. The result of radial distortion is shown in Figures 5.7(a) and 5.7(b). This



(a) Resulting response curve

(b) Effect of radial distortion on straight lines and circles

Figure 5.7: The optics in most cameras introduce non-linear radial distortion

distortion can be modelled by converting the image coordinates to a polar coordinate system, described by C_r and C_θ. then applying non-linear scaling S to the radial distance C_r. Finally the coordinates are converted back to a Cartesian system. As the angular component remains unaffected by the distortion, it is not required in the calculation:

$$C_r = \sqrt{(C_u - P_x)^2 + (C_v - P_y)^2} \tag{5.10}$$

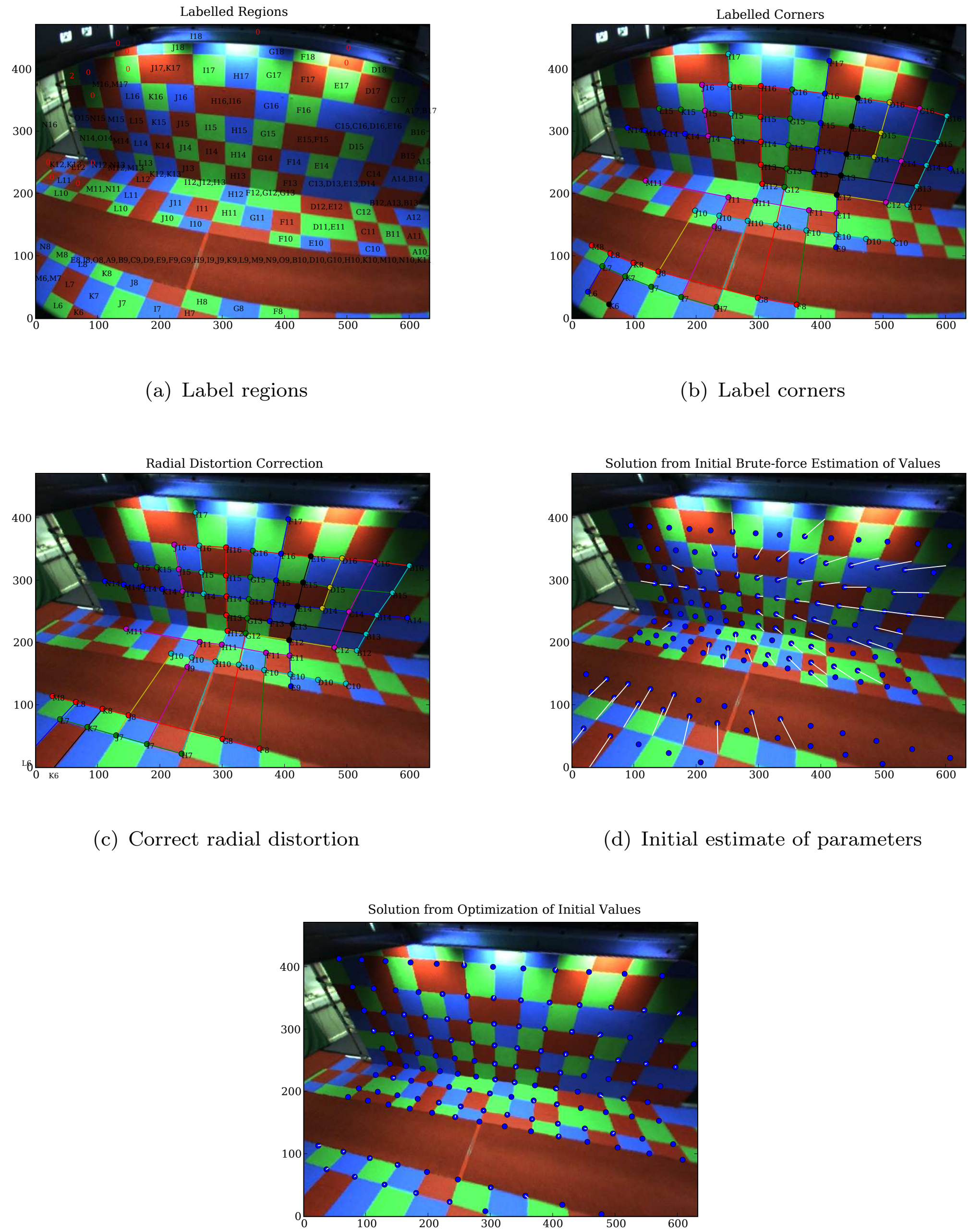
$$S = 1 + k1C_r^2 + k2C_r^4 \tag{5.11}$$

$$C'_u = S\left(C_u - P_x\right) + P_x \tag{5.12}$$

$$C'_v = S\left(C_v - P_y\right) + P_y \tag{5.13}$$

Finding the correct values for the camera's extrinsic, intrinsic and radial distortion parameters is a non-trivial task. It is possible to directly measure many of these properties, although it is often not practical. The process of estimating these parameters is known as camera calibration; there are a wide variety of approaches, all making different assumptions about the scene observed by the camera. Many utilise one or more reference objects in the scene where their geometry is already known; this can be used to directly solve some parameters. More sophisticated techniques can utilise the movement of rigid objects within the scene to approximate some of the camera's parameters[83]. Without prior knowledge of one or more fixed points in the scene, it is impossible to define an absolute origin or axes, instead using camera's position as the frame of reference.

The radial distortion of the camera's optics can be measured by finding lines within the image that should be straight, then determining the correction parameters needed to straighten the lines. Straight lines can also be grouped into sets travelling in the same direction within the scene; each set of lines will converge at a separate locations within the image, known as the vanishing points. If the direction of these lines is known in the scene, then the rotation matrix of the camera and some of its intrinsic parameters can be approximated. Finally, the use of ground-truth points can be used to estimate



(a) Label regions

(b) Label corners

(c) Correct radial distortion

(d) Initial estimate of parameters

(e) Optimization of parameters

Figure 5.8: The steps taken to calibrate a camera

or refine the camera's transformation matrix; if the position for several points is known in both the world and image coordinate systems, then the matrix **P** can be solved, by treating it as an over-complete system.

In the Biometric Tunnel, the walls and floor surrounding the walkway are painted with a three-coloured non-repeating pattern of squares. This pattern can be used to assist the camera calibration process in several ways. The edges of the squares in the pattern form a set of straight lines, travelling in one of three perpendicular directions. These lines can be used to characterise the radial distortion, calculate the intrinsic matrix and estimate the rotation component of the extrinsic matrix. As the pattern is not repetitive, it is possible to uniquely identify each region in the pattern, allowing for the accurate calculation of the **P** matrix.

The original camera calibration algorithm used in the system by Middleton et al. [70] performed calibration in several stages; first a colour image was acquired from the chosen camera, a Sobel filter was then applied to produce an image containing only the edge information. The radial distortion correction parameters were found using a Nelder-Mead based optimisation technique, where the overall curvature of the lines in the image was minimised. The curvature was derived using an estimate of the straightness; found by taking the maximum value from a Hough transform of the edge image. The output of the Hough transform was also used to identify the three sets of lines and their respective epipoles. Using the three epipole positions, the location of the camera lens' principal point and the focal length was calculated. In addition to this, the rotation matrix could be partially estimated, such that the polarity for each dimension in the matrix was unknown. A five-dimensional brute force search was performed at a low resolution, to find the location of the camera and the correct polarity of the rotation matrix's axes. The estimated parameters were then used as an initial estimate for a final Nelder-Mead optimisation of all parameters, minimising the total distance between the projected ground-truth corners and the nearest matching candidates. This approach required several minutes per camera; due to the use of the Hough transform for radial distortion correction. The corner matching algorithm proved inaccurate; only requiring corners to have the same surrounding colours, meaning that there were many matching candidates for a given point. This meant that completely incorrect calibration results could be chosen, as the corner matching algorithm facilitated the existence of many local minima.

A new approach was devised to address the speed and reliability concerns of the previous calibration algorithm; the major advancement of the new approach was that it explicitly identified the regions and corners within the camera image; greatly reducing the likelihood of an incorrect solution being found due to the presence of a local minima in the cost function. Classification was performed on the colour input image, so that each pixel was identified as one of the three colours in the pattern. Connected component analysis was performed on the classified image, to label regions of pixels all belonging to the same colour.

In order to identify each labelled colour region, each region was assigned two descriptor codes; a simple code describing only the region and a complex code describing the region and its neighbourhood. The simple descriptor was constructed using the corresponding region's colour code and the number of surrounding regions for the two other colours. By only encoding the number of regions in the descriptor and not utilising any positional information; the resulting descriptor was affine invariant; although not sufficiently detailed to uniquely identify a region. Therefore, a more detailed descriptor was constructed by concatenating the basic descriptor and a sorted set containing the basic descriptors of the surrounding regions. The resulting descriptor could uniquely identify all regions within the tunnel, whist still remaining affine invariant. An initial attempt was made at matching the calculated complex region codes to the ground-truth codes; although due to colour classification errors, it was likely that some regions would not be resolved. Therefore, an iterative dependency solving algorithm was used to infer the matches for unsolved regions. An example of the region matching is shown in Figure 5.8(a).

With the knowledge of the matched regions, it is possible to identify and label the corners between the regions, as shown in Figure 5.8(b). Using the labelled corners, three sets of lines were constructed; one for each direction of travel. The three line-sets were then used to estimate the radial distortion correction parameters, by attempting to minimise the curvature of the lines; as shown in Figure 5.8(c). The epipoles were found for the three sets of lines, then used to produce an initial estimate of the camera's intrinsic parameters and the rotation matrix.

Similar to the previous approach, a brute force search was performed to find the approximate position of the camera and the correct orientation of the rotation matrix, as shown in Figure 5.8(d). Direct optimisation of the final **P** matrix proved unreliable, due to the matrix having twelve degrees of freedom. Therefore a progressive approach was taken, where a small initial subset of the camera's parameters was optimised, with subsequent passes featuring an increasing number of parameters, until the final pass where all parameters were optimised simultaneously. Optimising the rotation matrix parameters directly proved unreliable, as the constraints of a rotation matrix were not enforced — all axes must be perpendicular and of unit length. To ensure that these constraints remained, the optimiser only provided two out of the three axes; which were renormalised, before finding the third axis using the cross-product. In order to ensure that the first and second axes were perpendicular, the second axis was recalculated using cross-product of the first and the third axes. By enforcing the constraints of the rotation matrix, the optimisation process proved much more robust. The final results of fitting the ground-truth points to the image are shown in Figure 5.8(e).

5.6 Gait cycle labelling

In order to perform gait analysis, it is necessary to identify the beginning and end of a single gait cycle within the set of captured frames. A complete gait cycle is comprised of time periods where the left leg is transit, the right leg is in transit and two periods where both legs are in contact with the ground; known as double-support. Figure 5.9 depicts

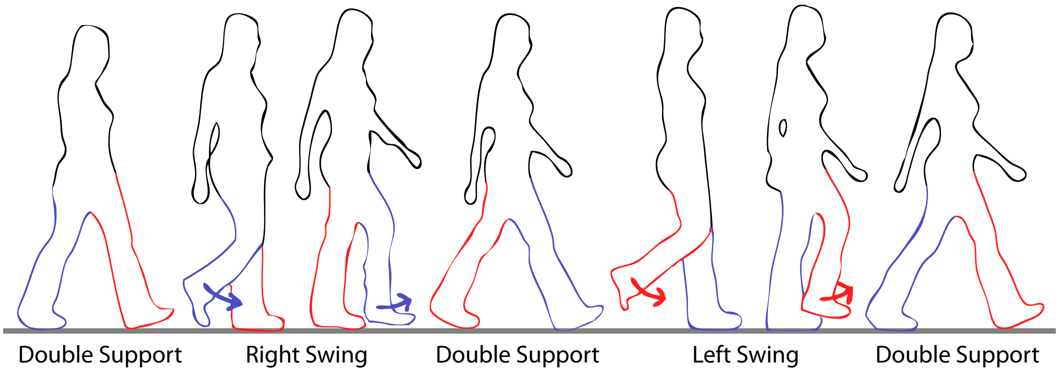


FIGURE 5.9: A complete cycle is comprised of a left and right swing phase and two double-support stances

a complete gait cycle. As previously discussed in Section 2.3, it is possible to perform this manually by hand; however this can introduce human error and is not suitable for a fully automated system. Therefore an automated technique for locating gait cycles within a sequence is required.

One approach is to fit a bounding box to the subject and measure the variation in the bounding box's length over the sequence. This should result in a sinusoidal signal with the maxima corresponding to the double-support stance and the minima occurring when the travelling limb crosses the supporting limb.

It is also possible to locate a gait cycle using a measure of silhouette self-similarity, such as that of BenAbdelkader et al. [4]. Although in order to ensure that a gait cycle always starts from a double-support or mid-swing stance requires additional information, such as the bounding box length data. Another technique as demonstrated by Bouchrika [7] located a subject's gait cycle by finding the position of their footsteps from the silhouette data.

It was decided to use the variation in bounding box length, in order to determine the beginning and end of a gait cycle. As previously stated, the distance between a person's legs is at its maximum when they are in a double-support stance; therefore double-support stances can be detected by finding the instances where the length of the bounding box encompassing a subject is maximal in the direction of travel. This is shown in Figure 5.10, and several peaks can be easily identified; although there are several small erroneous peaks, caused by random fluctuations in the bounding box size, which are a result of noise in the reconstructed volumes. Therefore the time-varying sequence of distance values were low-pass filtered to minimise the effects of noise, using a linear phase-response finite impulse response filter[30], to ensure that the relative positions of the minima and maxima were preserved. It was decided to label gait cycles using the local minima points of the bounding box lengths, instead of the local maxima, as it was found that the positions of the local minima were often more stable.

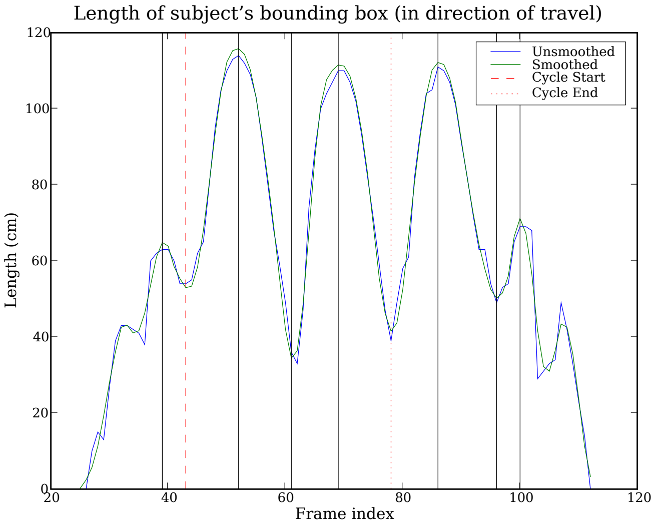


FIGURE 5.10: Double-support stances can be identified by finding the peaks in the variation of their bounding box length.

A cost function was used to establish the likelihood of each local minimum being the start of a gait cycle; the function considered factors such as the timing deviation between minima, the length deviation between minima, the length deviation between maxima and how close the cycle was to the centre of the tunnel. The cost value for the winning minima selection was saved as an estimate of fitting confidence for the sample.

Chapter 6

The Large Multi-Biometric Dataset

6.1 Introduction

Currently, there are no datasets containing three-dimensional volumetric gait data, and the largest two-dimensional datasets only contain approximately 120 subjects — not enough to accurately estimate the inter and intra-class variation within subjects. It was therefore decided to collect a new dataset; with the intention of having at least 300 subjects — making it the largest dataset containing gait data in the world. This would also be the first large dataset to contain both multi-viewpoint silhouette data and three-dimensional reconstructed volumetric data. Finding a sufficiently large number of participants from a wide range of backgrounds is a difficult task; in order to persuade individuals to take part, the experiment must appear straightforward, not take much time and ensure the privacy of participants.

As mentioned in Chapter 4, the tunnel was modified to collect data from two other non-contact biometrics; face and ear, to allow the investigation into recognition from multiple non-contact biometrics; without the need to repeat such an experiment. A further four cameras were added to the system, during the collection of the dataset.

Collecting a large dataset is a substantial undertaking, therefore collection of the dataset was a collaborative effort with Sina Samangooei[88], who also had an interest in running a biometrics experiment, where a large number of participants were required to describe the appearance of others featured in video footage.

6.2 Experimental setup

Collecting a dataset with a large number of subjects is not an easy task; a lot of planning is required in order to ensure that the experiment runs smoothly and that consistency is maintained throughout the duration of the experiment. This section provides an overview of the experiment and process undertaken by the participants.

A session check-list was created, which required the supervisor to ensure that a variety of steps were taken at the beginning and end of a session. This included setting up the system's hardware and software, and ensuring that the laboratory environment was clean and safe. An accompanying instruction sheet was also produced, to help inexperienced supervisors start the system with minimal assistance from others. A copy of the session check-list and instructions are given in Appendices D.3 and D.4 respectively. The use of a check-list for each session was intended to reduce the likelihood of mistakes by the session supervisor and ensure consistency between data capture sessions.

The ultimate goal of the experiment was to collect biometric data from in excess of three-hundred subjects; although it was expected that obtaining a sufficiently large number of individuals from a diverse range of backgrounds would prove extremely difficult. In order to attract a large number of people to the experiment, an incentive was required to persuade individuals to participate. For this reason, it was decided to offer participants a gift voucher with a value of ten pounds sterling. Vouchers were chosen that could be spent at a large variety of high-street retailers, to ensure that the incentive's appeal was as wide ranging as possible.

For both ethical and safety reasons it was necessary to ensure that a strict induction procedure was implemented. As potential participants arrived, the supervisor would inform them of important safety information; such as the laboratory's evacuation procedure and the safety hazards present in the environment; such as the strobe lighting. The supervisor then explained the experiment to the participants; covering important aspects such as the project's purpose, aims and the procedure to be carried out by the participant. Participants were then reassured that the data collected was completely anonymous and could not be traced back to them as an individual, they were also informed that the data might be shared with other research institutions in the future. The potential participants were then asked whether they would like to continue with the experiment; if satisfied, they were given a consent form to read and sign. Upon completion of the consent form, the supervisor ticked off a form on the reverse of the consent form to verify that the induction procedure had been carried out. A copy of the consent form is included in Appendix D.1. For financial audit reasons, participants were required to write their name on the consent form, to ensure that all vouchers could be accounted for. The forms contained no information related to the experiment's collected data, making it extremely difficult to link the individual to their captured biometric data. The anonymity of participants was further ensured by placing the completed forms in a ballet box, where the order of the forms was randomised to remove any temporal link to the collected data. Participants were asked to chose a bar-coded identity card at random from a container holding a large number of pre-printed cards. This identity card was used to enrol the subject and provides the only link between the individual and their biometric data.

Upon successful induction, the participant was assigned one of four computers in the Biometric Tunnel area, to use for the duration of the experiment. Each computer ran an instance of the tunnel system's website, which provided a simple graphical user interface for carrying out the experiment. The supervisor would then log the individual into the website by scanning their identity card using a bar-code scanner connected to the computer.

Assuming that the Biometric Tunnel was not in use by another participant, the supervisor would initiate the tunnel system for the participant. The supervisor would then explain the procedure for walking through the tunnel; before letting the participant conduct a trial walk through the tunnel whilst being watched. The capture process was then started, and the participant was instructed to walk through the tunnel and then wait for the status light to turn green again before starting another walk. The participant was asked to walk through the tunnel ten times — once the system had collected ten valid samples, the status light would extinguish to indicate completion.

Once the participant had completed their walks through the tunnel, they were asked to sit down and enter their personal information, which covered aspects such as their gender, age, ethnicity and weight. The participants then took part in another experiment on the computers, devised by Sina Samangooei; where they were shown videos of other people walking and asked to describe the subject's appearance. If the tunnel was already in use before the subject arrived, they could start the second experiment and then continue with it after walking through the tunnel. Upon completion of all aspects of the experiments, participants were given a gift voucher, their identity card and a leaflet summarising the experiment (included in Appendix D.2); they were then asked to tick a box on their consent form to confirm that they had received a gift voucher, as required for auditing purposes.

6.3 Composition of the dataset

The dataset collected with the Biometric Tunnel consisted of a total of 2705 samples from 227 subjects; including samples from when a participant returned to provide temporal samples. Not including samples from when participants have returned; the dataset contained 2288 samples. Thirty-six of the participants returned one or more times to provide additional temporal samples, where there were 414 samples from the subsequent days; making a total of 780 samples where temporal variation could be analysed. The

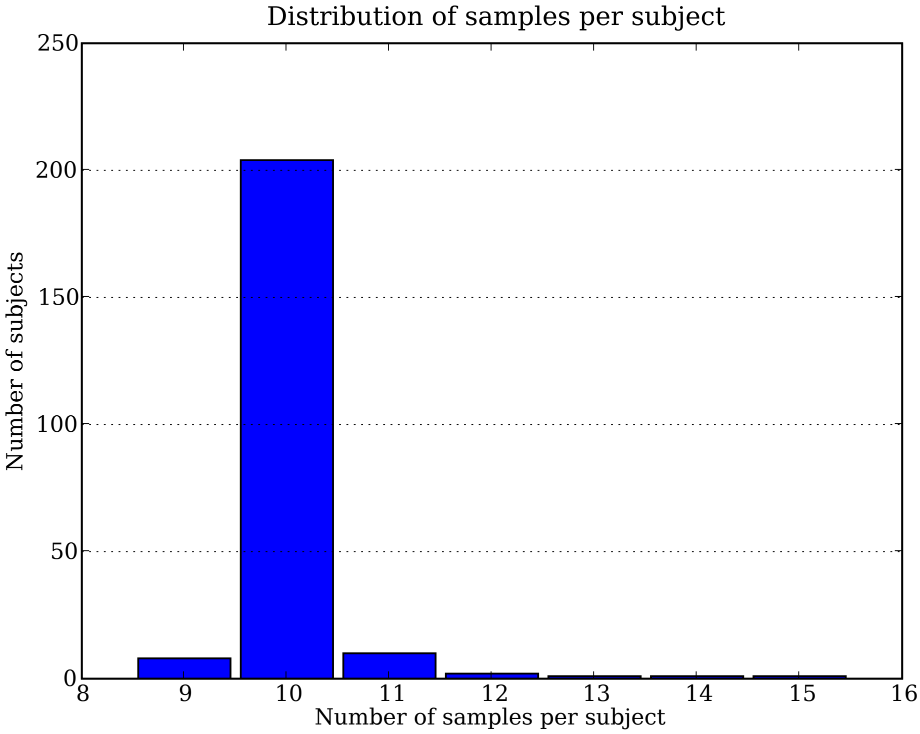


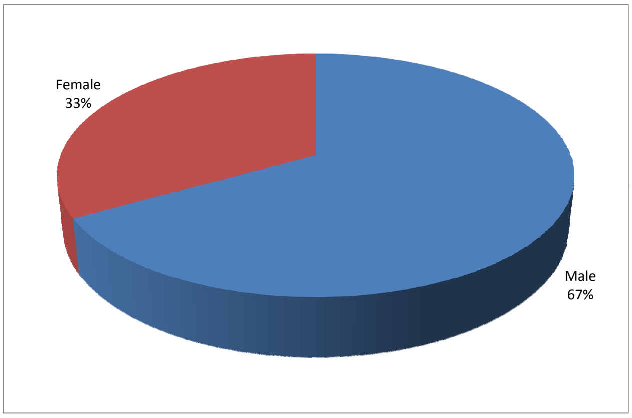
Figure 6.1: Distribution of number of samples/subject/session

original target of three-hundred unique subjects was not met, due to the difficulty in recruiting participants. For most subjects the dataset contained ten samples for each walking session; although there were some cases where this varied due to system failure or human error. Figure 6.1 shows the distribution of the number of samples per subject.

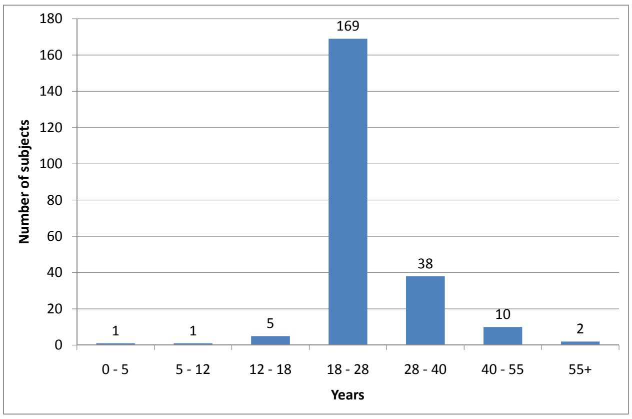
Of the two-hundred and twenty-seven subjects, 67% were male; the majority were aged between 18 – 28 years old and 70% were of European origin. These biases in the demographic of the dataset were expected, as this closely represents the make-up of the student population. Attempts were made at getting University staff to participate, although it proved more difficult to find convenient times for the staff. The ethnic distribution of dataset is quite reasonable, considering that the Office of National Statistics found that over 90% of people in the United Kingdom were of a white skin colour[81]; this shows the exceptional diversity of the staff and students at the University of Southampton.

6.4 Average silhouette based gait analysis

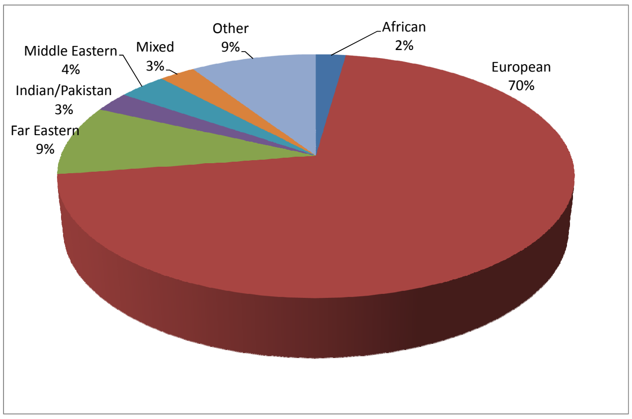
Two recognition experiments were initially performed on the dataset, both using only samples from each participant's first walking session; meaning that minimal temporal



(a) Gender



(b) Age



(c) Ethnicity

Figure 6.2: Demographics of the multi-biometric dataset

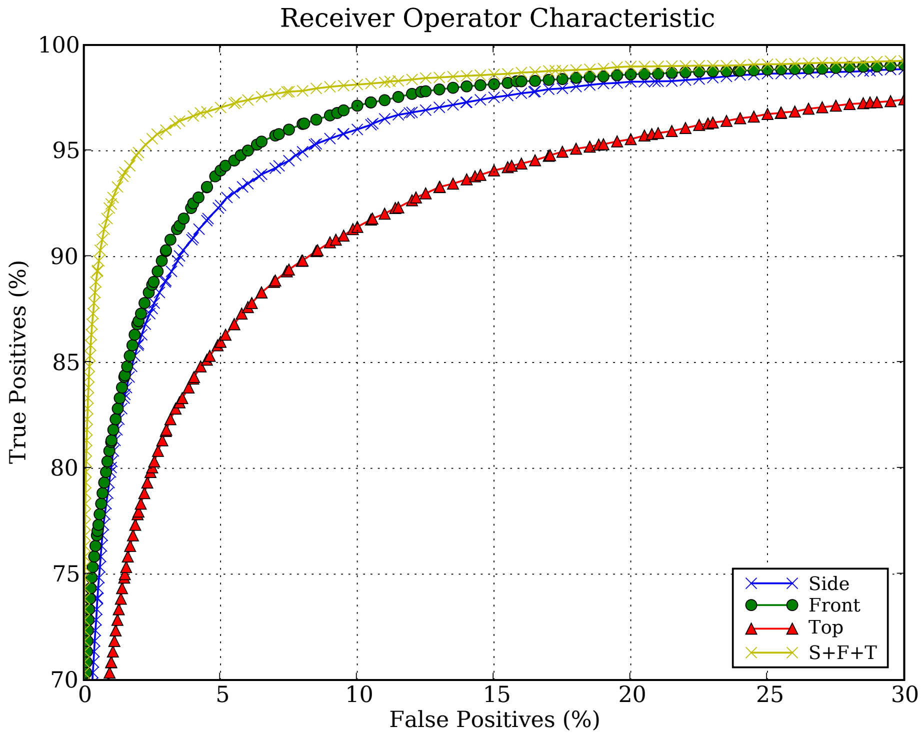
variation existed between samples. The first experiment performed leave-one-out recognition on the complete 2288 samples, using the average silhouette to characterise one's gait. A three-fold experiment was also performed by splitting the dataset into three smaller datasets, where each subject had the same number of samples in each dataset to avoid bias. The classification results of both experiments are shown in Table 6.1. Three different orthonormal viewpoints were evaluated; side-on, front-on and top-down; recognition was also performed by concatenating the feature vectors from the three viewpoints to result in one combined feature vector. The combination of the three viewpoints resulted in an improved recognition rate, showing that the additional information contained within the signatures was beneficial. The three-fold correct classification rate was found to be lower than the single dataset rate, this is because the correct classification rate is dependant on the make-up of the associated dataset, especially how many samples each subject has in the gallery set. In the full recognition experiment each subject typically had ten samples; which meant that when a leave-one-out experiment was performed there were nine samples available for gallery data. For the three-fold experiments, the dataset was split into three subsets, giving only three samples per subject; therefore only two samples were available in the gallery per subject.

Although the correct-classification rate is an intuitive and popular metric for measuring the performance of a recognition system, it is an extremely unreliable metric for comparing differing approaches; as shown above it is dependant on many factors apart from the algorithm; such as the composition of the evaluation dataset. Many published works have included other metrics[24, 105] for evaluation purposes; this includes the Neyman-Pearson receiver operating characteristic (ROC) curve, the Equal Error Rate and the decidability[28], which measures the separation between the in-class and inter-class distributions. The receiver operating characteristic for each viewpoint and the combination of all three is shown in Figure 6.3(a). The equal error rate and decidability is presented alongside the correct-classification rate for each viewpoint in Table 6.1. The receiver operating characteristic, equal error rate and decidability were unaffected by the splitting of the dataset into three; therefore only one set of results are included. The intra-class and inter-class distributions for the combination of all three viewpoints are shown in Figure 6.3(b).

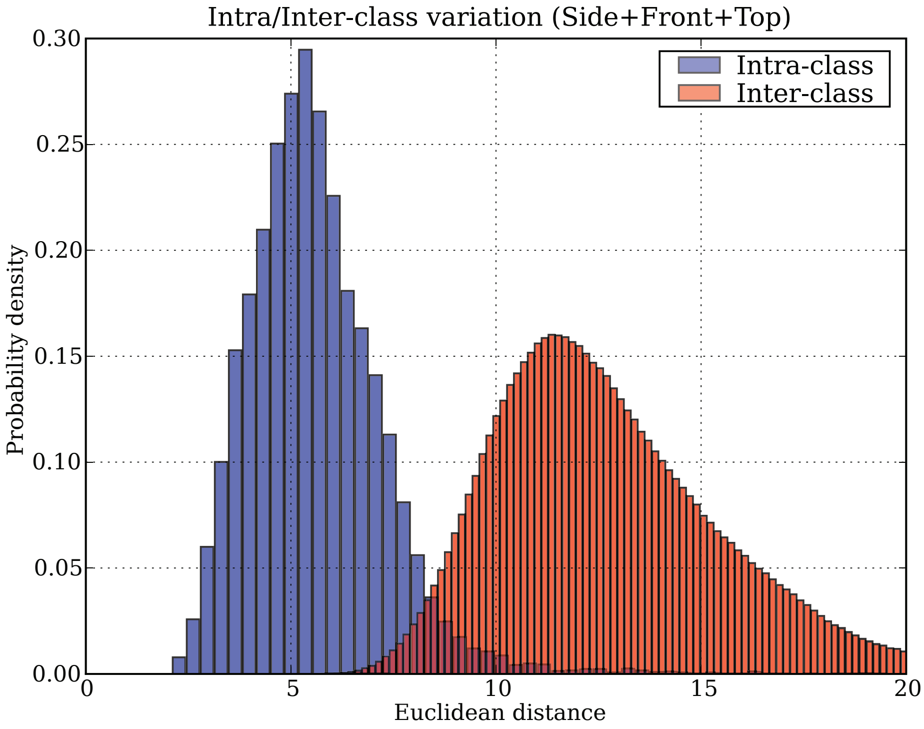
Name	CCR k=1	CCR k=3	CCR k=5	CCR k=1 (3 fold)	EER	d
Side	97.38%	96.59%	96.15%	88.50% ± 0.42%	6.28%	2.80
Front	97.81%	97.42%	97.12%	92.71% ± 0.66%	5.47%	2.76
Top	94.62%	93.84%	92.57%	77.24% ± 1.38%	9.21%	2.33
S+T+F	99.52%	99.52%	99.61%	97.21% ± 0.42%	3.58%	3.12

Table 6.1: Recognition performance of leave-one-out experiment using average silhouettes from dataset

The classification results achieved on the new dataset fell broadly in line with expectation; as the increase in the number of subjects was expected to reduce the separation between classes, meaning that correctly classifying subjects was more difficult. The



(a) Receiver Operating Characteristic



(b) Intra/Inter-class variation for all viewpoints combined

FIGURE 6.3: Recognition performance for large dataset using scale-normalised average silhouette

recognition rate for individual viewpoints falls somewhat behind that of Veres et al. [109], where side-on average silhouettes were used to classify subjects from the University of Southampton HID database; the excellent recognition performance achieved was most likely due to the use of linear-discriminant analysis to increase the separation between subjects; therefore improving the correct classification rate. It is also possible that the re-projected silhouettes used in the new dataset exhibited some shape distortion due to the process of three-dimensional reconstruction and re-projection; resulting in degraded recognition performance — this will have not been an issue for the other aforementioned approaches where two-dimensional video footage was used directly.

6.5 The non-normalised average silhouette

The average silhouette is well regarded for its simplicity and excellent performance; as demonstrated in the previous section. For a standard two-dimensional implementation, the images resulting from background-segmentation are cropped to the subject's silhouette, rescaled to a fixed size, then finally combined by averaging the rescaled images. The scale-normalisation is performed as the size of the subject's silhouette varies depending on their distance from the camera; although this means that one of the most useful features for recognition — their height — is discarded.

The volumetric data from the Biometric Tunnel has the advantage that it can be re-projected to any arbitrary camera view; including orthonormal viewpoints. The use of an orthonormal viewpoint results in the subject's height remaining constant, regardless of their distance from the viewpoint's origin. Therefore the process of scale-normalisation is no longer required; retaining the subject's key characteristics such as their height and body mass.

A non-normalised average silhouette algorithm was implemented; where the centre of mass was found for each silhouette, which was then cropped to a 200×200 voxel region, where one voxel was one centimetre in size for each dimension. For the side-on and front-on average silhouettes, the vertical dimension spanned from the top of the reconstruction area to the bottom, whilst the horizontal spanned 100 voxels either side of the centre of mass. For the top-down viewpoint, both the horizontal and vertical dimensions spanned 100 voxels either side of the centre of mass. The cropped silhouettes were then combined by summing the individual pixel values and dividing by the number frames, to calculate the mean. The resulting average silhouette was then down-sampled by a factor of four, to result in a 50×50 pixel image. It could be argued that one of the key strengths of this approach is that it does not just describe an individual's gait; it also characterises the overall appearance of their entire body over a short period of time.

Name	CCR k=1	CCR k=3	CCR k=5	CCR k=1 (3 fold)	EER	d
Side	99.52%	99.39%	99.26%	96.82% ± 0.62%	2.18%	2.91
Front	99.65%	99.48%	99.26%	97.90% ± 0.50%	2.29%	2.97
Top	90.47%	90.30%	88.55%	73.96% ± 1.21%	9.14%	2.19
S+F+T	100.00%	99.96%	99.96%	99.56% ± 0.24%	1.58%	3.07

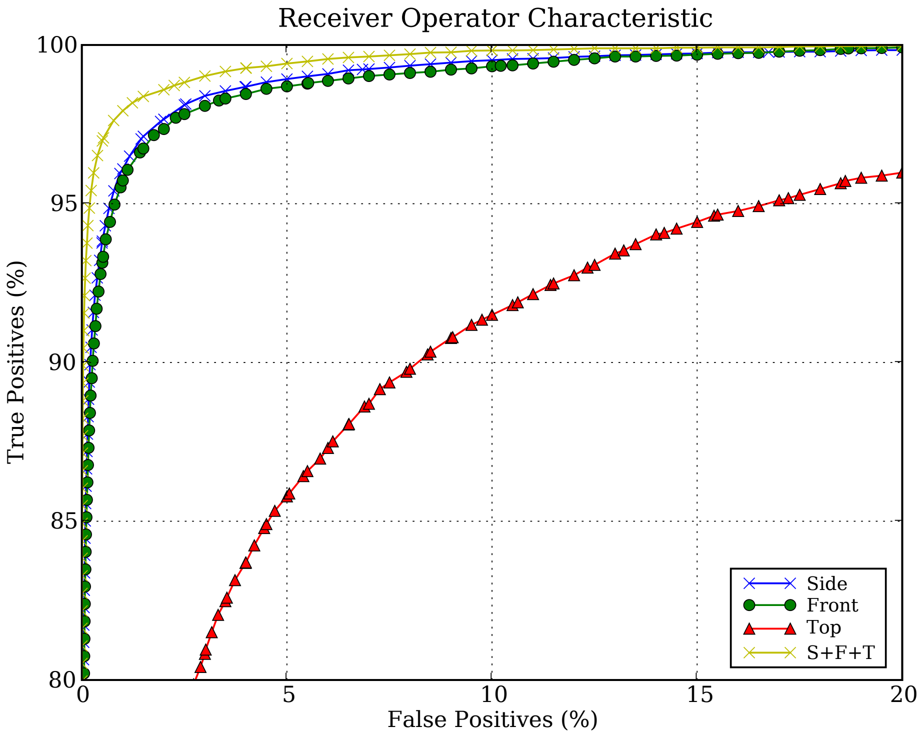
Table 6.2: Recognition performance of leave-one-out experiment using non-normalised average silhouettes from dataset

The non-normalised average silhouette was calculated for all samples in the newly collected dataset, then two recognition experiments identical to those of the previous section were performed — instead using the non-normalised silhouettes. The first was a leave-one-out recognition experiment using the entire dataset; whilst the second was a three-fold leave-one-out experiment, to measure the variation in recognition rates. As shown in Table 6.2, the new technique produced excellent recognition performance and a good degree of separation between subjects, as shown in Figure 6.4(b). The receiver operating characteristic is also improved, as shown in Figure 6.4(a). Combining all three viewpoints results in every sample in the dataset being classified correctly.

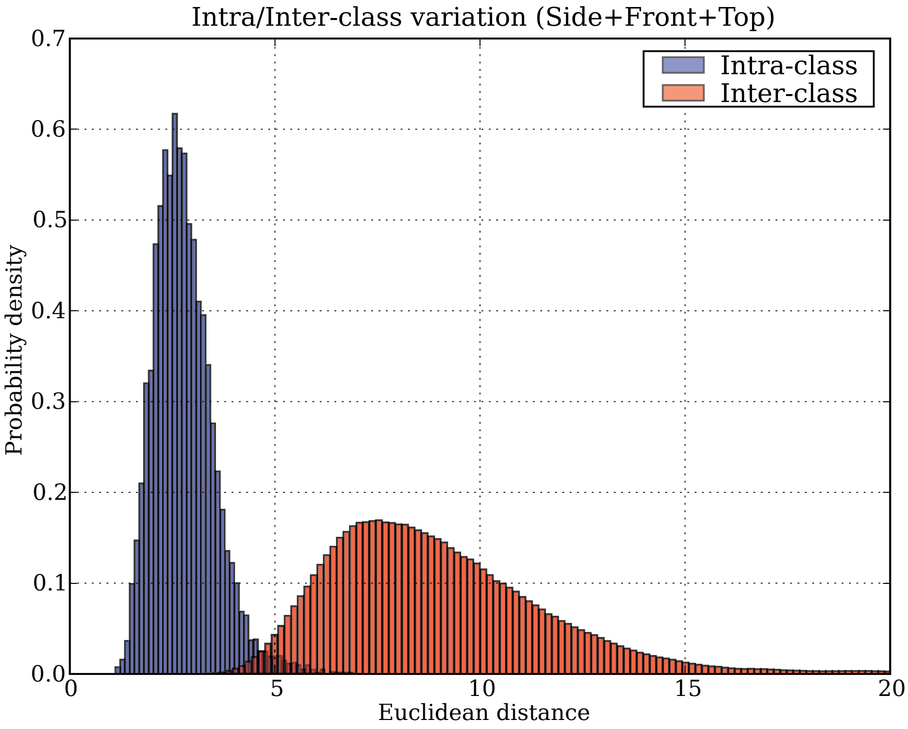
These new results show that the removal of the scale-normalisation stage from the calculation of a subject's average silhouette results in an improvement in recognition performance and inter-class separation. This leads to the fact that static information such as the subject's height and build is extremely important for accurately discriminating between individuals. A significant limitation of this technique is the need for three-dimensional data in order to remove the requirement for scale-normalisation. An experimental approach that is able to make use of two-dimensional video is presented later in Chapter 9.5.3.

6.6 Discussion

The results of analysing this new dataset demonstrate that an individual's gait remains stable over short periods of time and that it is an ideal biometric for recognising individuals from a distance without the need for their cooperation; whilst providing reasonable accuracy for many scenarios. It is difficult to predict how stable one's gait is over longer periods of time using the aforementioned analysis, due to short period of time between a participant's samples. In order to further evaluate the effectiveness of gait recognition and investigate its limitations, a much larger dataset is required; featuring much longer periods of time between samples — although it is recognised that such an undertaking would require a substantial amount of time and resources. Other factors that may affect an individual's gait such as their choice of clothing and footwear have not been considered in the newly collected dataset; as all samples for a participant were collected in a single session. The collection of additional samples featuring covariate data was



(a) Receiver Operating Characteristic



(b) Intra/Inter-class variation for all viewpoints combined

Figure 6.4: Recognition performance for large dataset using average silhouette without scale-normalisation

deemed impractical given the time-frame of the project; although there is no reason why this could not be done in the future, provided that there is sufficient time and resources available.

An interesting finding from the analysis of the new dataset was that recognition performance was improved by combining the average silhouettes from all three viewpoints. This gain in performance suggests that there is some degree of independence between the information contained within the three viewpoints; otherwise no improvement in performance would have been observed. Therefore it can be suggested that the use of three-dimensional data for gait analysis is beneficial to the overall recognition performance.

Other researchers such as Veres et al. [109] and Sundaresan et al. [101] have previously managed to achieve good recognition results using reasonably large gait datasets. What makes the results in this chapter significant is that excellent recognition performance has been achieved using only simple gait analysis and classification techniques on a dataset much larger than any of the previous attempts by others. It is expected that further improvements to the recognition performance and inter-subject separation could be made through the use of more complex analysis and classification methods; although these results serve the extremely important role of providing a baseline for evaluating new gait analysis techniques against. Using the findings in this chapter, it has been demonstrated that it possible to accurately distinguish between individuals using only their gait and body shape; this helps to confirm the statement made by Murray et al. [74] in 1964, that an individual's gait is *unique*.

Chapter 7

Further Analysis of the Dataset

7.1 Introduction

In this chapter we use the large dataset collected in the previous chapter to investigate some of the major factors that affect the performance of gait analysis using the Biometric Tunnel system. Whilst this chapter is by no means exhaustive; it attempts to cover as many covariates as possible using the data available. Covariates not discussed here include fatigue; clothing and footwear; walking surface inclination and material; the presence of music; the carrying of items or walking in groups. Whilst many of these covariates are interesting and will almost certainly have some degree of effect on one's gait, they are not possible to evaluate with the existing data, and collecting a dataset with these covariates would in most cases be impractical.

Instead, we concentrate on factors that need to be understood in order to design and evaluate a system for real-world usage. This includes the average silhouette resolution required; the optimal camera configuration; the effect of temporal variation and the potency of various types of information contained within a gait signature.

7.2 Resolution of average silhouette signature

The original average silhouette algorithm by Liu and Sarkar [67] produced an average silhouette with a 64 × 64 pixel resolution; resulting in 4096 features. Veres et al. [109] utilised analysis of variance and principal component analysis to reduce the number of features needed to characterise a subject's gait; this was possible due the the source features exhibiting a high degree of correlation. The number of source features is determined by the resolution of the calculated average silhouette; surprisingly, there is very little information on the optimal resolution for a subject's average silhouette.

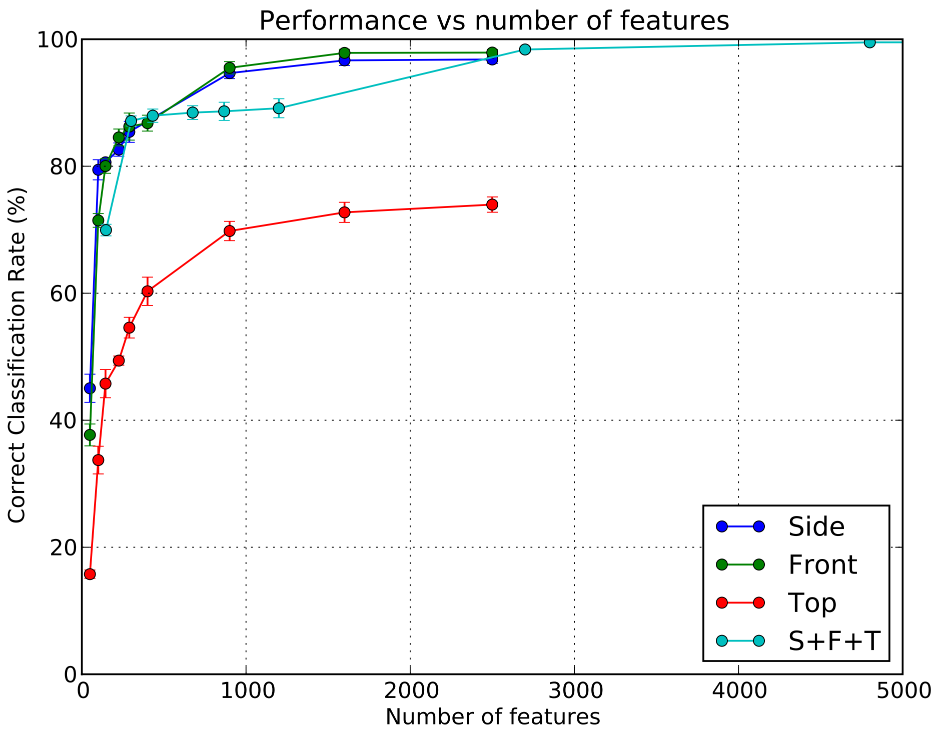


Figure 7.1: Correct classification rate vs. number of features in average silhouette signature; using non-temporal dataset

An experiment was devised to investigate the effect of average silhouette resolution against recognition performance, using the non-normalised variant discussed in Chapter 6. To do this, average silhouettes from the large dataset were rescaled to various sizes and three-fold recognition experiments were performed for each resolution. The original non-normalised average silhouettes had a resolution of 50×50, which were then rescaled to resolutions of 40×40, 30×30, 20×20, 17×17, 15×15, 12×12, 10×10 and 7×7 pixels. The same effect could have been achieved by reducing the resolution at which the three-dimensional reconstruction was performed; although performing the reconstruction at multiple resolutions would have proved extremely time consuming.

The relationship between the number of features and classification rate is shown in Figure 7.1; where the concatenated signatures have three times the number of features compared to a single viewpoint signature. It can be seen that modest performance is achieved for even the smallest of signatures, with very little degradation occurring until the number of features falls below a thousand.

The results from this experiment reveal that gait recognition can performed effectively using average silhouettes with a resolution much lower than the original 64×64 pixel average silhouette implementation. This is investigated further in Section 8.4, where the performance of the average silhouette is compared against a basic facial analysis technique. Whilst the concatenated average silhouette was found to provide the highest overall recognition performance compared to that of a single viewpoint; these results show that for a lower number of features, the use of a single viewpoint is more efficient. This is because the three viewpoints all share a single underlying information source, meaning that the combined signature contains a significant amount of redundant information. It is expected that the use of principal-component or linear-discriminant analysis would greatly reduce the amount of information redundancy, resulting in a more compact and efficient representation.

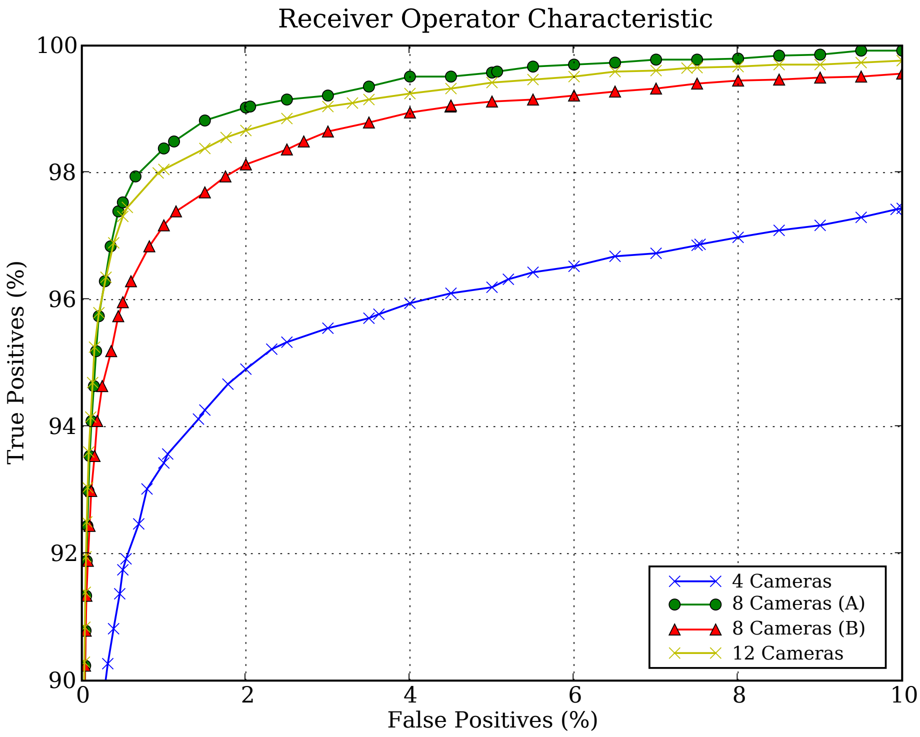
7.3 The configuration of cameras

The Biometric Tunnel originally constructed by Middleton et al. [70] initially contained only eight ceiling mounted cameras; four placed at the far corners of the tunnel area and the other four mounted centrally. Very little is known about what investigation was carried out to determine the optimal number and placement of the cameras in the environment. Therefore four additional cameras were added to the system, located one metre off the ground at the far corners of the tunnel. An experiment was then performed to investigate what effect the number of cameras and their placement had upon recognition performance. Four different camera configurations were evaluated; the first featuring the four top far-placed cameras, the second containing all eight far-placed cameras, the third configuration consisting of the eight top cameras, and finally all twelve cameras. Figure 4.4 shows the layout of the tunnel and all twelve cameras. Non-normalised average silhouettes were generated for each camera configuration on a subset of the large dataset containing the samples recorded from all twelve cameras. Analysis was performed on a total of 1388 samples from 137 subjects. Leave-one-out recognition experiments were performed for each camera configuration, with the results shown in Figure 7.2.

The results presented show that the inclusion of the four centrally mounted cameras had a negative impact on the system's classification performance; this was an unexpected result, as it was assumed that a greater number of cameras would result in a more accurate three-dimensional reconstruction of the subject, leading to improved results. Inspection of the reconstructed data found that the combination of all twelve cameras produced the most visually pleasing reconstruction, although this did not correspond with the recognition rate found. The most likely cause for the performance degradation was the significant radial distortion present in the central cameras' images, which proved difficult to accurately compensate for using radial distortion correction and calibration. It is expected that the use of better quality optics and more sophisticated camera calibration routines would result in an improved performance for the twelve camera configuration. The strategy currently used to calibrate the cameras within the Biometric Tunnel only uses points found on the two planes formed by the floor and whichever side-wall is visible; the use of a removable calibration structure and solving the intrinsic properties of

Name	CCR k=1	EER	Decidability (d)
Side - 4 Cameras	99.57%	4.14%	2.73
Front - 4 Cameras	99.28%	4.86%	2.80
Top - 4 Cameras	74.35%	15.08%	1.73
S+F+T - 4 Cameras	99.42%	4.06%	2.74
Side - 8 Cameras (A)	100.00%	1.24%	3.05
Front - 8 Cameras (A)	99.42%	2.15%	2.92
Top - 8 Cameras (A)	83.93%	11.89%	1.92
S+F+T - 8 Cameras (A)	100.00%	1.34%	2.97
Side - 8 Cameras (B)	99.35%	2.19%	2.82
Front - 8 Cameras (B)	99.57%	2.92%	2.81
Top - 8 Cameras (B)	92.51%	10.28%	2.10
S+F+T - 8 Cameras (B)	100.00%	1.93%	2.90
Side - 12 Cameras	99.42%	2.22%	2.86
Front - 12 Cameras	99.28%	2.12%	2.92
Top - 12 Cameras	93.16%	9.72%	2.10
S+F+T - 12 Cameras	100.00%	1.57%	2.98

(a) Summary of classification performance, equal error rate and decidability



(b) Receiver Operating Characteristic

FIGURE 7.2: The effect of camera configuration on recognition performance; using non-temporal dataset

the cameras separately is likely to result in a more accurate calibration. As discussed in Section 9.5.1, global optimisation of the camera calibration matrices was found to provide a significant improvement in the quality of the reconstructions.

7.4 The effect of camera calibration error

In a system such as the Biometric Tunnel, where three-dimensional reconstruction is performed from an array of cameras; accurate characterisation of the cameras is essential. Poor calibration will often result in misalignment between the cameras; in turn significantly degrading the quality of the reconstructed output. It is expected that any deterioration in the quality of the reconstructed volumes will have an impact on the system's recognition performance; especially when using the average silhouette, which is extremely sensitive to the static elements of a subject's shape. Calibration error can be introduced into the system by a variety of means; from initial inaccuracies caused by the calibration algorithm to the orientation of the cameras changing over time, which could be caused by vibration, sudden knocks or mechanical creep.

In a deployed system, it is reasonable to expect that the orientation of the cameras will change over time; therefore it is essential to understand the system's sensitivity to the effects of camera misalignment. Whilst it is expected that such a system would have a maintenance schedule and possibly some form of camera alignment tracking; it is still important to understand the system's tolerance levels. Therefore an experiment was devised to measure the effect of camera misalignment on recognition performance. Instead of physically altering the orientation of the cameras in the system and collecting new data; existing recorded data from the large dataset was re-processed using distorted calibration data exhibiting angular error. This was achieved by applying a random three-axis rotation to each camera's projection matrix; where the angular error on each axis was determined by a separate normal distribution. The standard-deviation for all three distributions was equal and varied for each experiment, which was then performed multiple times. No translational error was added during the experiments, as it would have added further complexity to the experiment and the effect of translational error is very similar to that of rotational error at low levels.

The drop in the system's correct classification rate against the standard deviation of the angular error is shown in Figure 7.3. As expected, the system's recognition rate is severely impacted by almost any level of calibration error, which is due to average silhouette based recognition being most sensitive to pixels around the boundary of a subject's silhouette[109]. These results outline the importance of accurately calibrating the cameras within the system, ensuring that they are firmly mounted and that a regular inspection and maintenance program is implemented. The development of a robust online incremental calibration algorithm would be highly beneficial for such a system, or

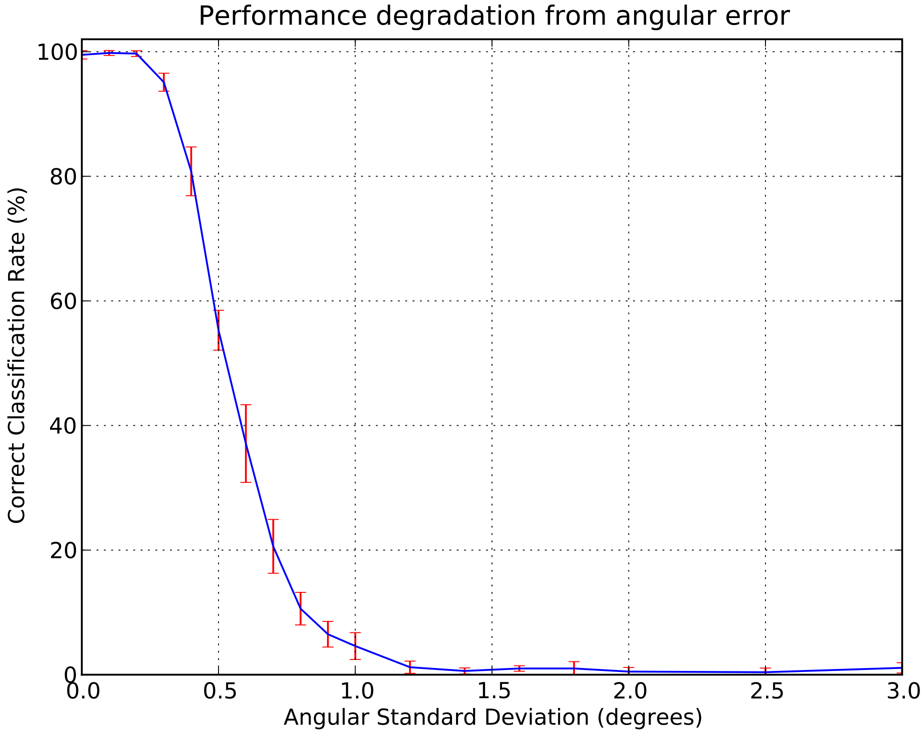


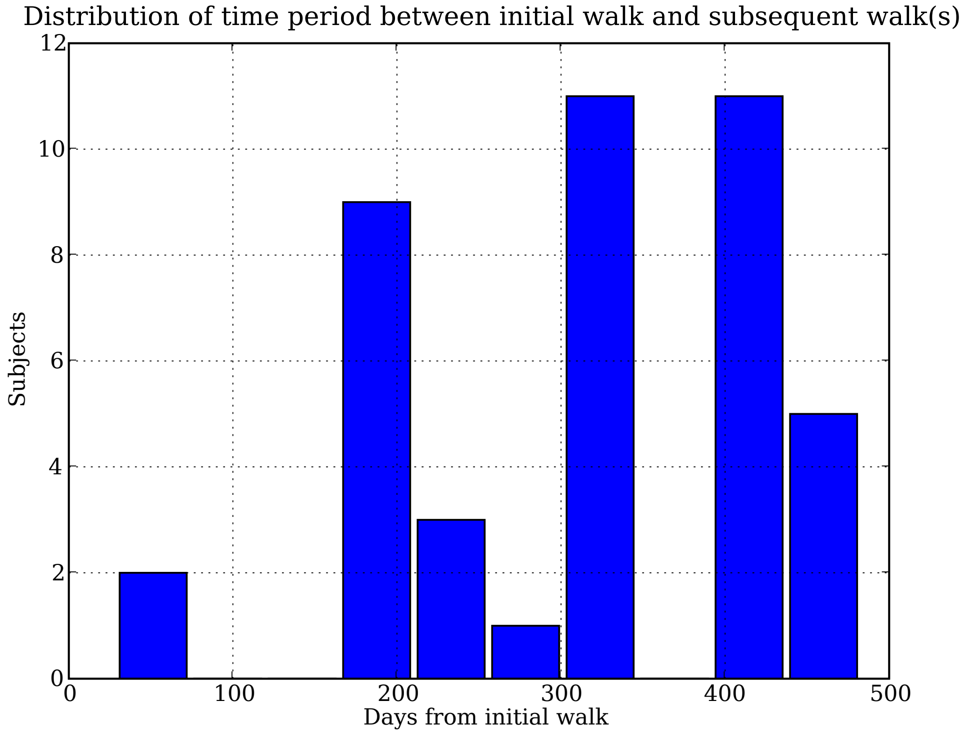
Figure 7.3: The impact of angular error upon recognition performance

any other permanent installation where multi-camera three-dimensional reconstruction is employed.

7.5 Variation over time

During the collection of the large multi-biometric dataset discussed in Chapter 6, participants were encouraged to come back again at a later date to provide subsequent samples. Thirty-six of the original participants returned at a later date to take part in the experiment again. For the thirty-six subjects, there were 366 samples from their first recording sessions and 414 samples from subsequent sessions. The time duration between a participant's first recording session and subsequent sessions varied greatly between subjects, as shown in Figure 7.4(a); although it was ensured that the duration was at least one month.

A recognition experiment was performed by splitting the samples into two sets; a gallery and a probe set, where the gallery set contained samples from each subject's first session, and the probe set consisted of samples from the subjects' further sessions. Non-normalised average silhouettes were used for characterising each subject's body shape and gait; classification was performed using a simple nearest-neighbour matching technique. Classification performance was found to be significantly better than expected;



(a) Distribution of time period between a subject's first and subsequent sessions

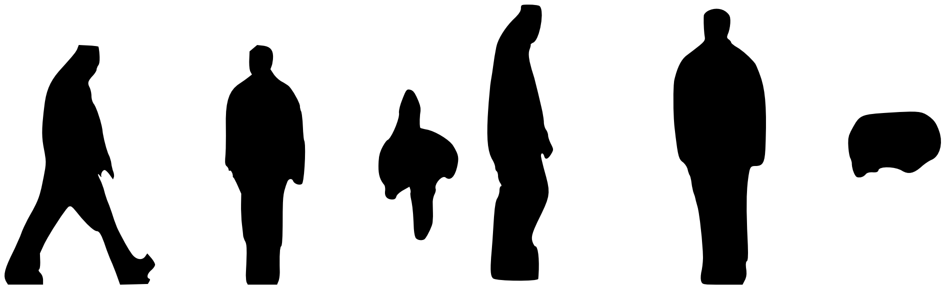
Name	CCR k=1	EER	Decidability (d)
Side	72.22%	12.90%	1.64
Front	53.86%	15.14%	1.56
Top	44.69%	23.90%	1.25
S+F+T	69.08%	11.76%	1.63

(b) Summary of recognition performance results

FIGURE 7.4: Results from recognition experiment matching samples against those from an earlier date; using non-normalised average silhouette analysis technique

with the side-on average silhouette giving a correct classification rate of 72.22%; as shown in Table 7.4(b). The use of all three viewpoints resulted in an improved separation between subjects, similar to the other recognition experiments presented in this thesis; this is shown by Figures C.1(a) and C.1(b). One of the most interesting findings is that the side-on viewpoint acheives the greatest recognition performance, this is unlike the short-term analysis experiments presented in the previous sections. The movement of the subject's limbs is difficult to observe from the frontal and top-down views, due to self occlusion. Therefore, these results raise the question of whether dynamic information regarding one's gait is more important for gait recognition over longer time periods.

The resulting recognition rate for the temporal experiment is quite reasonable, considering that one's body shape and weight can vary considerably over such a time period. It is also unlikely that the subjects were wearing the same clothing and footwear in



(a) Keyframe 1

(b) Keyframe 2

Figure 7.5: The two keyframes used for gait analysis

subsequent recording sessions, meaning that their outline and gait may change as a result. As with the previous recognition experiments in this document, good classification performance has been achieved using only simple analysis techniques. It is expected that with the use of a larger temporal dataset, the overall recognition performance will degrade. Unfortunately, collecting large datasets incorporating temporal variance is an extremely challenging task; especially when attempting to maintain complete anonymity for all subjects; as this makes it difficult to communicate with participants, requesting further sessions.

7.6 The components of gait and their potency

It has been suggested in published works that one's gait can be described as a combination of both static and dynamic features[116, 1, 59]; where static features remain constant over time and the dynamic features account for any variation over a gait cycle. In a model-based analysis approach, the static features would include the subject's upper and lower leg lengths and shape; whilst the angular variation of their joints would be considered as dynamic information. For a non-model approach, static and dynamic features are often found at the pixel level, taking the mean or intersection of a pixel to represent the static information; whilst any motion or variation in pixel value can be considered as the dynamic component of a subject's gait. This categorisation of the information contained within a subject's gait leads to the question; what contribution does each component make to to a system's discriminatory performance?

7.6.1 Static features

In order to investigate the importance of static information for gait analysis a recognition experiment was performed, where only the static information in a gait sequence was considered. This was achieved using key-frames[21] for the feature-vectors. Two different key-frames were evaluated; from multiple viewpoints. The first key-frame was taken when the subject was in a double support stance, the second when their legs crossed

Viewpoint	Metric	Double support	Legs crossing	Concatenated
Side-on	CCR	82.8%	70.4%	85.8%
	EER	15.1%	22.6%	12.6%
	Decidability (d)	1.97	1.53	2.13
Front-on	CCR	86.2%	88.3%	90.9%
	EER	15.1%	16.7%	12.2%
	Decidability (d)	2.00	1.93	2.18
Top-down	CCR	65.0%	47.6%	64.9%
	EER	33.0%	33.0%	30.2%
	Decidability (d)	1.01	0.91	1.13
S + F + T	CCR	90.4%	83.7%	91.6%
	EER	18.1%	16.4%	12.8%
	Decidability (d)	1.93	1.90	2.22

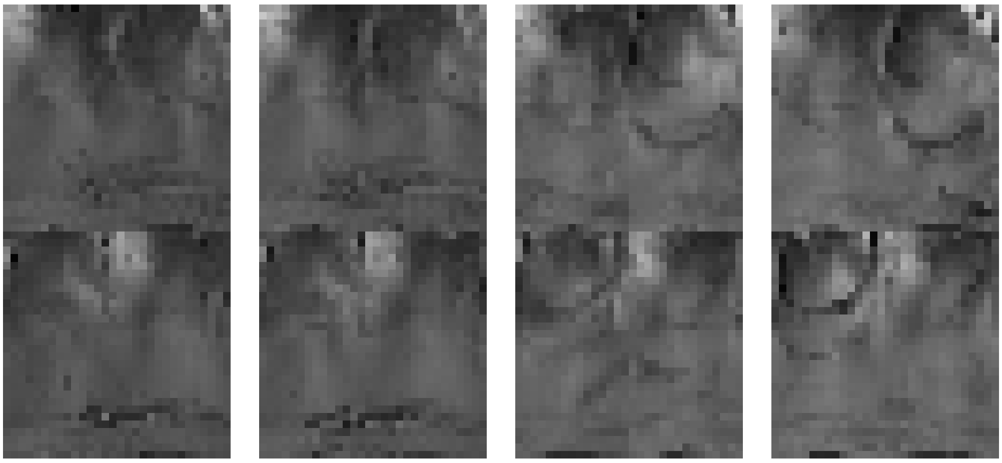
Table 7.1: Recognition performance using key-frames from differing viewpoints

over. Classification performance was also measured for the concatenation of both the key-frames and also the three viewpoints.

As shown in Table 7.1, the recognition performance for the key-frames was found to be insufficient for stand-alone use in a real system. The double-support stance key-frame was found to be more effective for discriminating between subjects compared to the other key-frame; this was unsurprising, as the double-support stance encapsulates more information about the subject; such as their stride length and the geometry of their legs when fully extended. Concatenation of the key-frames and viewpoints led to a predictable improvement in recognition performance; due to the increased amount of information available. Similar to the results shown in Figure 9.2, the front-on viewpoint provided the best recognition performance. When compared against the results of the average silhouette based technique used earlier in this chapter, it is clear that the static information contained within the key-frames is not sufficient on its own, to accurately identify individuals.

7.6.2 Time-varying dynamic information

The results given in the previous section demonstrate that the static component of one's gait is separable and can be used to discriminate between individuals with a limited accuracy. This raises the question of how much useful information is contained within the dynamic component of one's gait. In order to answer this question, an analysis algorithm was devised that attempted to characterise only the dynamic information contained within an individual's gait. This was done by considering the movement of the subject's legs separately, using an approach similar to optical flow; where only the leading edges of the limbs were considered — to allow for loose clothing or skirts. The average displacement for each leg was taken at several different heights for each volume



(a) A,1 (b) A,2 (c) B,1 (d) B,2

Figure 7.6: Dynamic motion signatures for two different subjects (A and B), where the horizontal axis for each signature represents time, the vertical axis corresponds to the sampling height and the intensity represents the travelling velocity.

in a gait cycle sequence. This resulted in a two-dimensional image for each leg; with the horizontal and vertical axes corresponding to time and position respectively. Each pixel within the image was a measure of velocity for a specific point in time and position along the subject's leg. The images for each leg were concatenated to result in a single feature vector, which was horizontally shifted to ensure that the image always started with the right leg in a support stance. Figure 7.6 shows the signatures for two different subjects.

A recognition experiment was conducted to evaluate the new analysis technique proposed above, using the large non-temporal dataset discussed in Chapter 6. The correct classification rate was found to be reasonable, at 84.5%; although the equal error rate and decidability were poor, with values of 22.7% and 1.36 respectively, indicating that there was little separation between subjects. As the technique only characterises the motion present within the subject's silhouette, these results demonstrate that the dynamic component of one's gait does contain some information of discriminatory ability; although techniques such as the average silhouette achieve much better performance without distinguishing between static and dynamic information.

7.6.3 The combination of static and dynamic information

As demonstrated by the results from the two previous sections, it is possible to classify a subject using only the static or dynamic components of their gait in isolation; although the recognition performance for both approaches was found to be much lower than the average silhouette.

Metric	Double Support	Legs Crossing	Concatenated
CCR	93.0%	89.8%	92.5%
EER	15.3%	15.0%	22.9%
Decidability (d)	2.03	1.98	2.29

Table 7.2: Recognition performance achieved by fusing motion signature with various key-frame types

A simple fusion experiment was performed to investigate whether the two analysis techniques described earlier fully encompassed all the information contained within one's gait. To do this, the features derived from the concatenated viewpoint key-frames were combined with the features from the dynamic motion signature. The overall standard-deviation for the features in the key-frame was calculated, along with the standard-deviation for the dynamic motion signature. The two feature-vectors were then normalised using the calculated standard deviation values, then concatenated to achieve the final feature vector for classification. By normalising both feature vectors before concatenation, this helped to ensure an equal contribution from both feature-sets towards the recognition performance.

The recognition performance using the large non-temporal dataset described in Chapter 6 was marginally improved by the fusion of the two modalities, as shown in Table 7.2. From these results, it can be seen that both dynamic and static information play their part in gait recognition; although it is clear that the combination of the proposed static and dynamic analysis techniques does not fully account for the information found within the average silhouette, which is capable of achieving much better recognition rates and separation of subjects. It is expected that the use of a more sophisticated technique to extract dynamic information could lead to improvements in the overall recognition performance of a static and dynamic information fusion strategy.

Chapter 8

The Fusion of Gait and Face

8.1 Overview

The use of multiple feature sources results in a greater range of information available to use for recognition or identity verification, which leads to improved accuracy and a higher degree of confidence in the decision outcome. The use of multiple biometrics also makes it much harder for an unscrupulous individual to avoid recognition or imitate another person. The use of multiple biometrics can also enable a subject to be correctly identified when one of their biometric features is obscured or occluded. The recognition performance of a single biometric modality can also be improved using fusion; where the feature vectors from several different analysis techniques or sensors can be combined, such as the fusion of 2D facial images with three-dimensional scans[17]. Examples of multiple-biometrics include face and iris[18]; palm-print and hand-geometry[58]; or face and gait[92, 91].

Combining information from more than one biometric can be achieved in a variety of ways in a recognition system[51]. One of easiest methods to implement is decision-level fusion; using the classification results from the different modalities to “vote” for the most likely match. This approach is sometimes the only option; when proprietary systems are used for one or more of the biometrics, the derived feature-vectors and match-likelihood scores are often unavailable. Decision-level fusion is unlikely to provide a substantial performance improvement over the use of a single strong biometric, as it does not consider for any of the given biometrics the certainty of a match or the proximity of other close matches.

It is possible to achieve improved recognition performance in most cases by combining biometrics using their match scores; instead of final decision outcomes[56]. This is because it is possible to consider the match likelihood for all outcomes, meaning that other similar candidates are also considered when determining the final decision. Match-score fusion can also take factors such as uncertainty and conflict between biometrics into account using techniques such as that of Shafer [90].

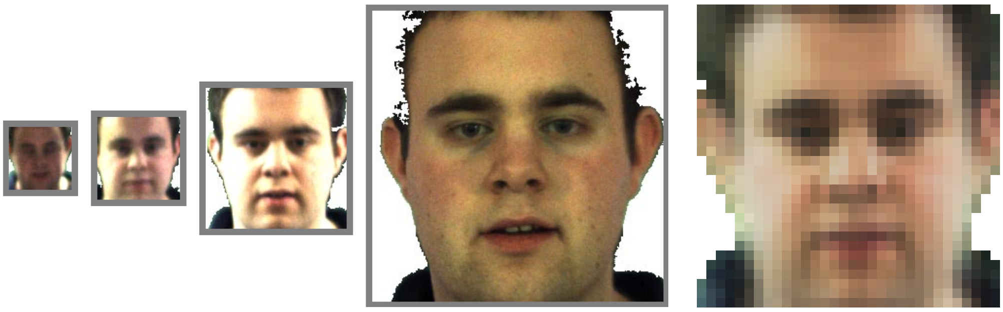
The features derived from the analysis of several different biometrics can also be directly combined, which is often referred to as feature-level fusion. Directly combining the extracted biometric features means that all possible information is used to determine the most likely match; therefore it should be possible to obtain the greatest accuracy fusing biometric data at the feature level. Unfortunately, feature-level fusion is not always practical, as the resulting feature vector can be excessive in size and great care must be taken to ensure that the features from each biometric are weighted correctly. As mentioned earlier, many off-the-shelf biometric recognition systems do not expose their feature-vectors, or the features are incompatible with those from other biometrics. Comprehensive reviews of the literature associated with multiple-biometric fusion are given by Ross and Jain [87] and Jain and Ross [50].

One of the simplest methods for combining multiple biometrics is to concatenate the feature-vectors from the various biometrics; this is a form of feature-fusion, as discussed above. Such an approach can potentially add bias to a particular biometric if the overall variance of the corresponding feature-vector is much greater than the others; by normalising the variance for each biometric, any bias can be compensated for. This approach is used in the following experiments, where recognition is performed using the combination of face and gait. Whilst imagery of ears was recorded in the dataset discussed in Chapter 6, it was primarily intended for future projects, as research into ear recognition is currently at a relatively early stage. In this chapter, a simple algorithm is devised for characterising one's facial appearance, which is then combined with non-normalised average silhouettes to evaluate the potential for a multiple-biometric recognition system based around facial appearance and gait.

8.2 A simple descriptor for facial appearance

Biometric recognition based upon facial appearance is an extremely popular research area, with a wide range of available analysis techniques. Some of the most popular and widely accepted methods include the use of principal component analysis to calculate the eigenface[107] and the active appearance model[22], which accounts for the facial shape using an active shape model[23] and texture using the principal component analysis — similar to the eigen-face. More sophisticated techniques using video, near-infrared or three-dimensional facial scan data also exist, as reviewed by Zhao et al. [126] and Bowyer et al. [10].

One of the most intuitive methods for comparing two images is to simply rescale them to an equal size and then calculate the Euclidean distance between the images. This



(a) Face images of varying size are found by a face-finder algorithm (b) The images are all rescaled to 32 × 32 pixels and averaged

FIGURE 8.1: The calculation of the average face

approach can be used for measuring the similarity between two registered face images; although it is extremely inefficient due to the large number of features. The use of principal component analysis to calculate the eigenface greatly reduces the number of features, without any significant impact on recognition performance.

The recognition performance of a direct comparison or eigenface based technique would be very sensitive to any registration error or temporary changes in facial appearance, such as talking or blinking. This could be rectified by comparing multiple frames from each face; although this would result in large feature vectors and would therefore not be practical for a large-scale recognition system. An almost equivalent result can be achieved by summing the multiple face frames together to result in an *average face*. By using multiple frames, the effect of random registration error and temporary facial appearance changes can be reduced. As the average face analysis technique is essentially performing a direct image comparison, it will be extremely sensitive to any variation in lighting orientation, colour cast and intensity; any changes in the subject's orientation are also likely to severely impact recognition performance. The use of this technique is only possible due to the use of a highly controlled and consistent data capture environment — the Biometric Tunnel. It is therefore expected that the performance of the average face analysis technique will quickly degrade outside of such an environment; making its use impractical in most other scenarios. This approach was chosen due to its simplicity and similarity to the average silhouette gait analysis technique; where neither make use of statistical or multi-variate techniques to improve performance.

Using the video data from the face camera in the Biometric Tunnel dataset, the OpenCV implementation[12] of the Viola Jones face detector[111] was used to locate the image region containing the subject's face. The background pixels were removed from the face image and it was rescaled to a resolution of 32×32 pixels. The sequence of rescaled face images was summed and divided by the number of frames, to result in the average face signature; as shown in Figure 8.1.

As an extension to the new average face technique, a three-dimensional histogram of the red, green and blue values found within the average face image was calculated. The intensity values for each colour channel were quantised into 16 discrete values, resulting in the histogram having 4096 bins to describe colour. The average face histogram contains information on the frequency of occurrence for different colours and is relatively robust against facial expression. The histogram was combined with the average face by concatenating the two corresponding feature vectors.

Using the non-temporal samples from the dataset collected in Chapter 6, a leave-one-out recognition experiment was performed; achieving an extremely high correct-classification rate. A similar experiment using the temporal data resulted in a reasonable correct classification rate, which was better than the performance achieved in Chapter 7.5 using gait. The addition of the colour histogram resulted in an improvement for both experiments; at the expense of having more features. The correct classification rate, equal error rate and decidability, for the two experiments are given in Figures 8.2(a) and 8.3(a). The receiver operating characteristic plots are given in Figures 8.2(b) 8.3(b), which also confirm that the technique is able to deliver reasonable performance when using data from a controlled environment.

The average face provides a simple and intuitive approach for facial recognition, providing acceptable classification performance for both time-varying and non-time-varying applications; making it an ideal baseline analysis algorithm for the data collected with the Biometric Tunnel. Like the average silhouette, the results achieved by this new technique demonstrate that simple techniques can often be surprisingly effective.

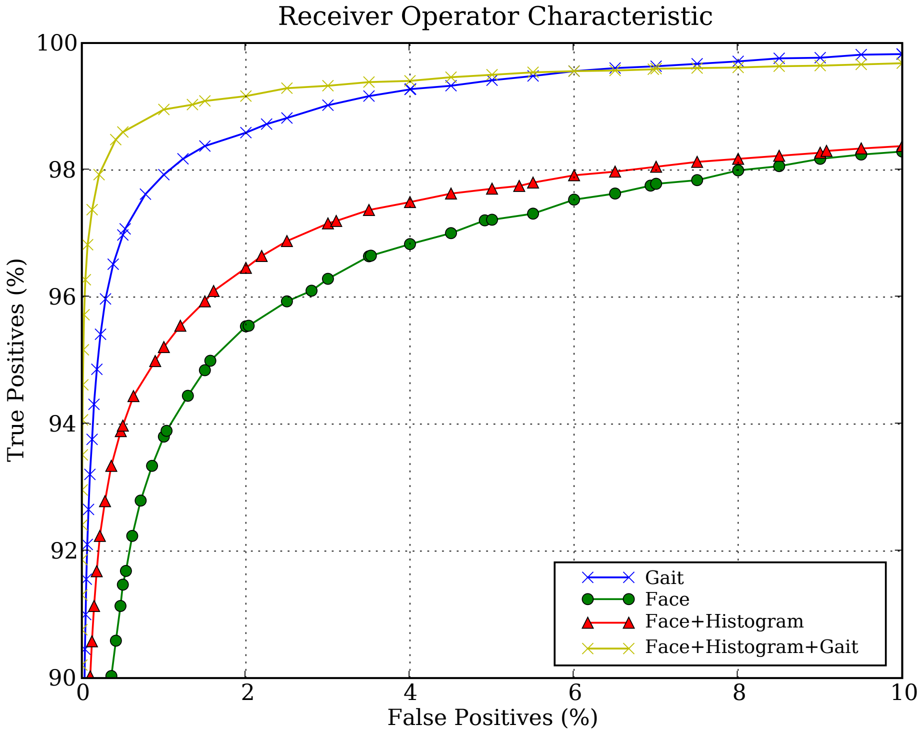
8.3 Evaluation of performance using large and temporal datasets

In order to evaluation the effectiveness of the proposed gait and face fusion approach, a leave-one-out experiment was performed using the non-temporal data collected using the Biometric Tunnel, discussed in Chapter 6. The dataset used for analysis consisted of 2288 samples from 227 different subjects. The non-normalised average silhouette was found for the side-on, front-on and top-down viewpoints, which was combined with the average face and corresponding colour histogram. The overall variance for each feature-set across the entire dataset was calculated, and used to normalise the feature-sets before concatenation.

Using all three gait signatures, the average face and colour histogram, every sample in the dataset was correctly classified and a low equal error rate of 1.04% was acheived. The classification performance was also found for gait, the average face and also the average face fused with the colour histogram; these results are shown in Table 8.2(a).

Name	CCR k=1	EER	Decidability (d)
Gait	100.00%	1.58%	3.07
Face	99.39%	3.42%	3.31
Face+Histogram	99.52%	2.89%	3.38
Face+Histogram+Gait	100.00%	1.04%	3.83

(a) Recognition performance for different combinations of features



(b) Receiver Operating Characteristic

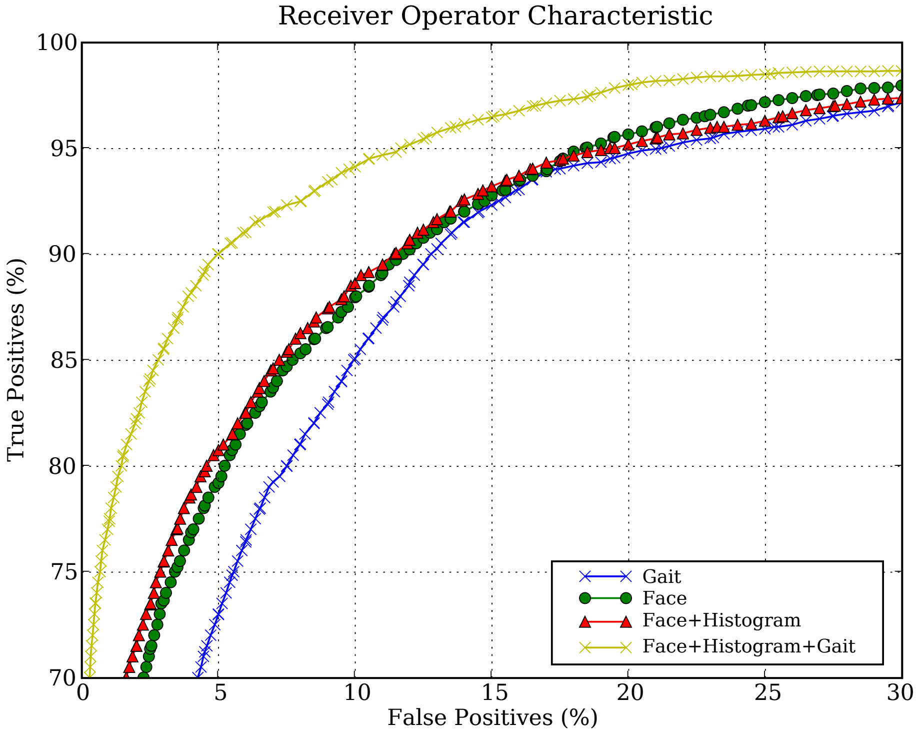
FIGURE 8.2: Results of leave-one-out multi-biometric experiment using non-temporal dataset

As shown in Figures 8.2(b) and C.2(a), the combination of more of than one biometric results in a greater separation between subjects and therefore improved discriminatory ability. These results show that the combination of gait and facial appearance is extremely potent; proving ideal for applications where the identity of subjects needs to be ascertained without close contact or subject cooperation. Whilst many attempts have been made at fusing facial appearance with other biometrics such as fingerprint, voice or iris; this is the first large-scale experiment to consider the combination of face and gait; arguably the two biometrics best suited for recognition at a distance.

A similar experiment was performed using only the samples within the collected dataset where temporal variation was present. This distribution of this temporal dataset was discussed earlier in Chapter 7.5. As expected, the recognition performance was below that of the non-temporal dataset discussed above. Table 8.3(a) gives the recognition

Name	CCR k=1	EER	Decidability (d)
Gait	69.08%	11.76%	1.63
Face	77.54%	10.95%	1.93
Face+Histogram	79.47%	10.67%	2.04
Face+Histogram+Gait	88.65%	7.60%	1.99

(a) Recognition performance for different combinations of features



(b) Receiver Operating Characteristic

FIGURE 8.3: Results of leave-one-out multi-biometric experiment using temporal dataset

performance for gait alone; facial appearance; face fused with its corresponding colour-histogram; and all three combined. The receiver operating characteristic for all four different signatures is shown in Figure 8.3(b) and Figure C.2(b), located in the Appendices, shows the intra and inter-class variation of the the fused signatures. Although these results do not compare to those of the non-temporal variation dataset they are still significant, as recognition over such a time-period is an extremely difficult task for such biometrics.

It is interesting to note that gait provides a better performance compared to the average face algorithm when used on the non-temporal dataset; although this is not the case for the time-varying dataset; where the samples are matched against samples from a previous date. This is not unexpected, as the facial analysis algorithm is extremely basic compared against most modern approaches. This is further explained by the average silhouette being very sensitive to any changes around the silhouette's boundary, which is likely to be the case if the subject is wearing different clothing, or they have lost or gained weight. Although the discriminatory ability of the gait signature may be degraded by temporal deviation, its addition to the facial appearance signature results in a significant reduction in the recognition error rate.

These results show that it is beneficial to combine face and gait in a non-contact recognition environment, especially in situations where one's face could be obscured or hidden. It is clear that further investigation is needed; including the use of state of the art analysis techniques for both face and gait, along with more sophisticated fusion and classification algorithms. Although most importantly, these results serve to demonstrate the potential of a multi-biometric recognition system based around gait and facial appearance.

8.4 Evaluation of face and gait at varying distances

In the previous section, we demonstrated that the addition of gait to a facial recognition system provides a worthwhile improvement in recognition performance. The fusion of gait and face data could lead to even greater benefits when used as part of a single camera system where a subject's distance from the camera is not as well controlled. Many gait analysis techniques make use of a subject's entire body shape, whilst a facial analysis technique only uses a small region of one's body; the head. For example, if a subject has a height of 100 pixels in an observed camera image, their head would be typically twelve pixels high. This make it difficult to provide facial recognition over a wide coverage area — unlike gait analysis, which can still function at these greater distances. The use of a system employing fusion can provide the best aspects of both biometrics; when the subject is close to the camera, both the subject's face and gait can be considered; whilst only gait is used for recognition if the subject is too far away for accurate facial recognition.

An experiment was performed to simulate the effect that a subject's distance from the camera would have on the recognition performance in a multi-biometric system. The front-on scale-normalised average silhouette was chosen to represent gait; as this is likely to be a realistic representation of the data available from a single camera system; where it is assumed that the subject is walking towards the camera — essential if their face is to be fully visible. The temporal data from the dataset collected in Chapter 6 was used in the recognition experiment; as it is of greater difficulty and is more likely to resemble that of a realistic scenario. A virtual camera with a 7.4μm pixel size and 6mm focal length was used to calculate the size of the observed average silhouette and average face images. Leave-one-out recognition was performed for varying distances from the camera, using only facial appearance, gait and the combination of both modalities. The results shown in Figure 8.4 demonstrate that the fusion of the two biometrics always proves beneficial and that using gait, it is possible to identify individuals at a much

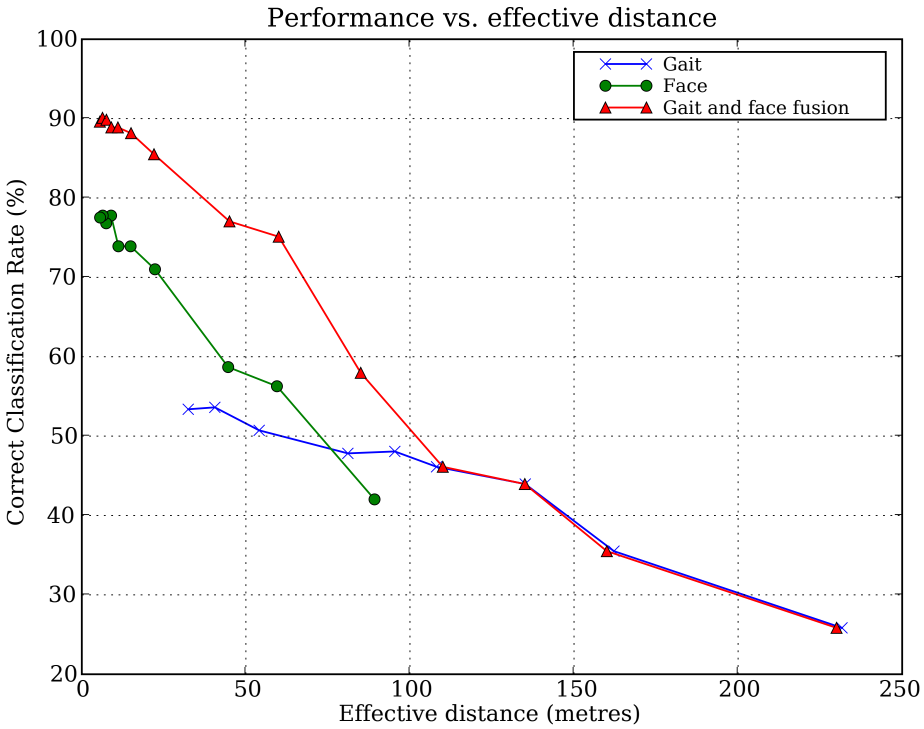


FIGURE 8.4: Classification performance of face and gait vs. distance using temporal dataset; from simulated camera with 6mm focal length and 7.4μm pixel size

greater distance than facial appearance; with a better likelihood than chance alone. The use of the frontal average-silhouette and average-face results in very little motion being captured in the signatures, which means that such a recognition system is almost exclusively using static information to classify the subjects.

Chapter 9

The Matching of Volumetric Data Against a Single Viewpoint

9.1 Introduction

In a large environment such as an airport, where access monitoring and identity verification is required, there are typically many entrances and exits, which means there will be many areas that will require recognition systems. Whilst the Biometric Tunnel system discussed in Chapter 2 is capable of providing excellent recognition performance as a standalone system, it would prove impractical deploying such a system at every entrance and exit of a large building. A more practical solution would locate the Biometric Tunnel system at only the primary entrances to the building and use single cameras to provide coverage for all other areas. As a subject enters the building, they would walk through a corridor configured as a Biometric Tunnel, which would enrol them into the recognition environment using the acquired volumetric data. When an unknown subject walks into the coverage of one of the single cameras, their gait signature would be compared against their 3D volumetric data re-projected to the same position and orientation as the subject.

In order for this approach to work, the re-projected silhouette of the subject must be sufficiently accurate to calculate a matching average silhouette. This depends on the calibration of the single camera, the quality of the background segmentation, along with the accuracy of the three-dimensional reconstruction from the Biometric Tunnel, which is reliant on the accurate calibration of the cameras within the system.

9.2 Reprojection of three-dimensional data

Using three-dimensional volumetric data, it is possible to re-project the volume to any arbitrary camera view. The resulting synthesised silhouette images can then be used with any standard two-dimensional gait analysis techniques. By determining the walking direction of a subject, their silhouettes can be synthesised from a viewpoint perpendicular to their walking direction; this means for most gait analysis algorithms, viewpoint dependence is no longer an issue.

The virtual camera can be characterised in the same manner as described in Chapter 5.5, where a 3 × 4 transformation matrix is used to map a three-dimensional coordinate to a two-dimensional image coordinate, by rotation, translation, scaling and the application of perspective. A coordinate on the target camera image (C_x, C_y) can be expressed as a line originating from the virtual camera in the world's three-dimensional coordinate system:

$$W = \mathbf{R}^{\mathsf{T}} \begin{bmatrix} C_z \alpha^{-1} (C_x - P_x) \\ C_z \alpha^{-1} (C_y - P_y) \\ Cz \end{bmatrix} + \mathbf{T} \tag{9.1}$$

When synthesising the target image, a ray is projected from the camera for each pixel. If one of the rays intersects an occupied voxel in the volumetric space, its corresponding pixel is marked as occupied. This can be done efficiently by using a three-dimensional version of Bresenham line algorithm[13], to walk along each of the rays in the volume. Once an occupied voxel has been encountered, it is no longer necessary to continue along the line.

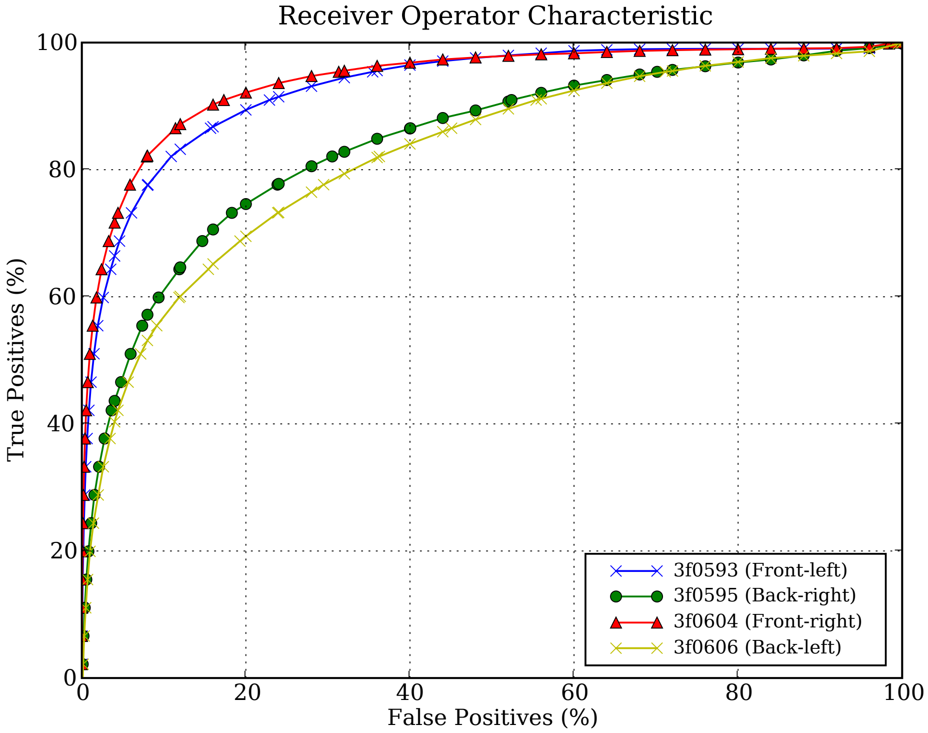
9.3 Volumetric data matched against camera in Biometric Tunnel

An initial feasibility study was performed to see whether it was possible to match a subject using their re-projected volumetric data. To do this, a re-projection test was performed for each camera in the Biometric Tunnel. The silhouette images from the dataset collected in Chapter 6 were reconstructed with all cameras except the test camera, then the resulting volume was re-projected to the test camera's viewpoint. The scale-normalised average silhouette was calculated for the test camera's original silhouette data and also the re-projected silhouette data. A recognition experiment was performed for each camera, using the average silhouettes from the re-projected data for the gallery set, and the original camera's average silhouettes for the probe set.

It was found that recognition was impossible using the four wide angle cameras placed near the centre of the Biometric Tunnel, this was because of the difficulty in obtaining

Camera	Location	CCR k=1	EER	Decidability (d)
3f0593	Front-left	74.76%	14.53%	1.70
3f0604	Front-right	84.40%	12.36%	1.83
3f0606	Back-left	53.37%	25.45%	1.21
3f0595	Back-right	54.34%	22.93%	1.17

(a) Summary of recognition performance



(b) Receiver Operating Characteristic

FIGURE 9.1: Results of recognition experiment matching average silhouettes generated from a single camera against reprojected volumetric data

an accurate camera calibration, due to significant radial distortion from the wide-angle lenses. A similar problem was experienced in Section 7.3, where it was found that the inclusion of the wide-angle cameras degraded the system's overall recognition performance. Therefore only the results from the four top far-mounted cameras are included in Table 9.1(a) and Figure 9.1(b). An improvement is found from using the front-on cameras, compared to the rear-view cameras; although this could be due to minor variations in camera calibration quality.

The recognition performance from the preliminary tests demonstrated that the concept had sufficient potential to warrant further investigation; some degree of recognition was possible, even when using very basic gait analysis techniques and removing cameras from the reconstruction; resulting in significant distortion to the reconstructed volumes' shape.

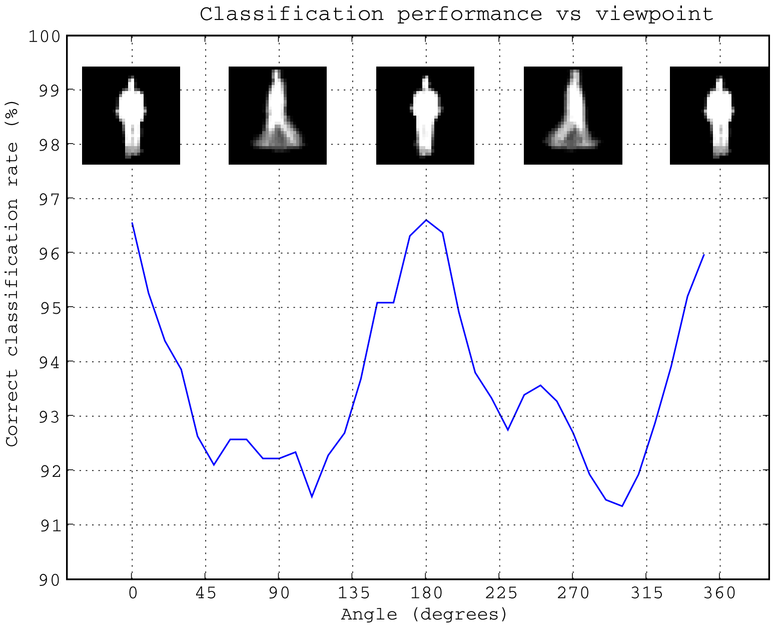


Figure 9.2: Correct classification rate for 32 × 32 average silhouette at different viewpoints

9.4 The effect of viewpoint on performance

Using the large dataset collected in Chapter 6, it is possible to re-project the volumetric data to any arbitrary viewpoint. This allows the evaluation of differing viewpoints, which will help to identify the optimal viewpoint for the average silhouette gait analysis technique. The non-temporal samples from the collected dataset were re-projected from 36 different viewpoints; where a virtual camera was moved around the subject in 10 degree increments. Low-resolution 32 × 32 scale-normalised average silhouettes were produced for each viewpoint and sample; where a reduced resolution was used to reduce the time taken to run the experiment and ease storage requirements. A leave-one-out recognition experiment was performed for each viewpoint to find its recognition performance. The results are shown in Figure 9.2; it can be seen that the front and rear viewpoints provide the best recognition performance, with almost identical results due to the symmetry about the origin. These results suggest that the static information within the average silhouette is more important for recognition than the dynamic components.

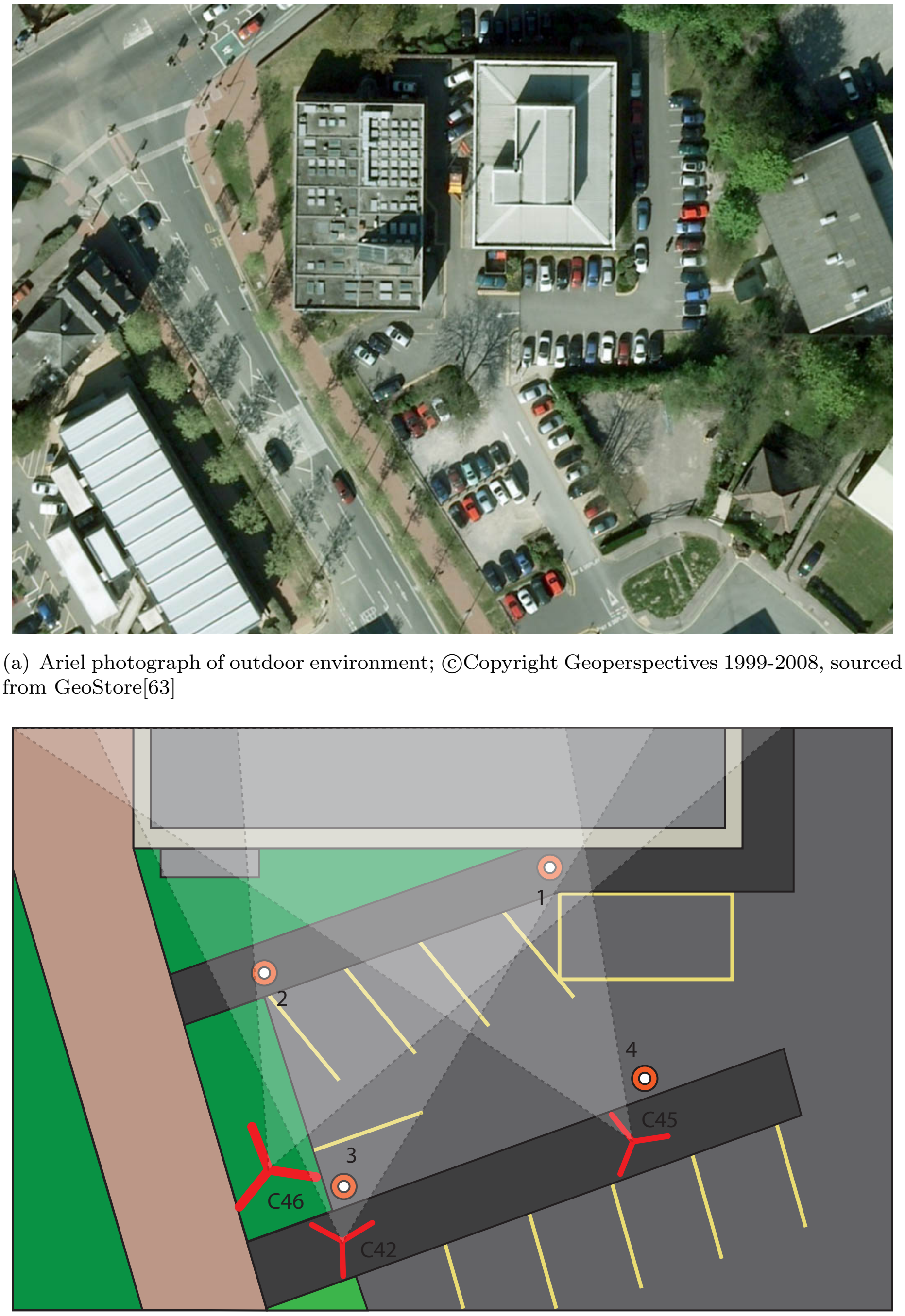
9.5 Volumetric data matched against outdoor camera

9.5.1 Experimental Setup

Although the results given in the previous section were encouraging, the experiment did not represent a realistic scenario, as the silhouette data produced by cameras within the tunnel is of extremely high quality, and do not suffer from the same degree of segmentation error as realistic data captured outside of a controlled environment. Artefacts such as shadows are also much more likely to occur in silhouette data captured from an uncontrolled environment. In order to test the proposed concept in a more realistic manner, the single viewpoint probe data must be captured in an uncontrolled environment, outside of the Biometric Tunnel. Therefore a small dataset was recorded outdoors, on the University of Southampton Highfield campus, as shown in Figure 9.3(a). The dataset was recorded in a single morning and contained video data from seventeen different participants, who were all asked to walk through the Biometric Tunnel in the afternoon. Unfortunately, of the original participants, only eleven returned to walk through the tunnel later that day.

The participants were recorded walking with three ProSilica GC-750E Gigabit network attached cameras. All the cameras were mounted on tripods, with two mounted at approximately two metres above ground level, the other approximately three metres above ground level. Four traffic cones were placed in the area and participants were asked to walk between the cones in a pattern. Figure 9.3(b) shows a top-down view of the area used, where the traffic cones are labelled one, two, three and four. The subjects were asked to walk from cone one to two, then back to cone one, and then repeat by walking to cone two and back. Next the subjects walked diagonally across the area from cone one to three, before returning to one and walking to cone two; then walking diagonally to cone four, back again to two and finally leaving the recording area after walking past cone one.

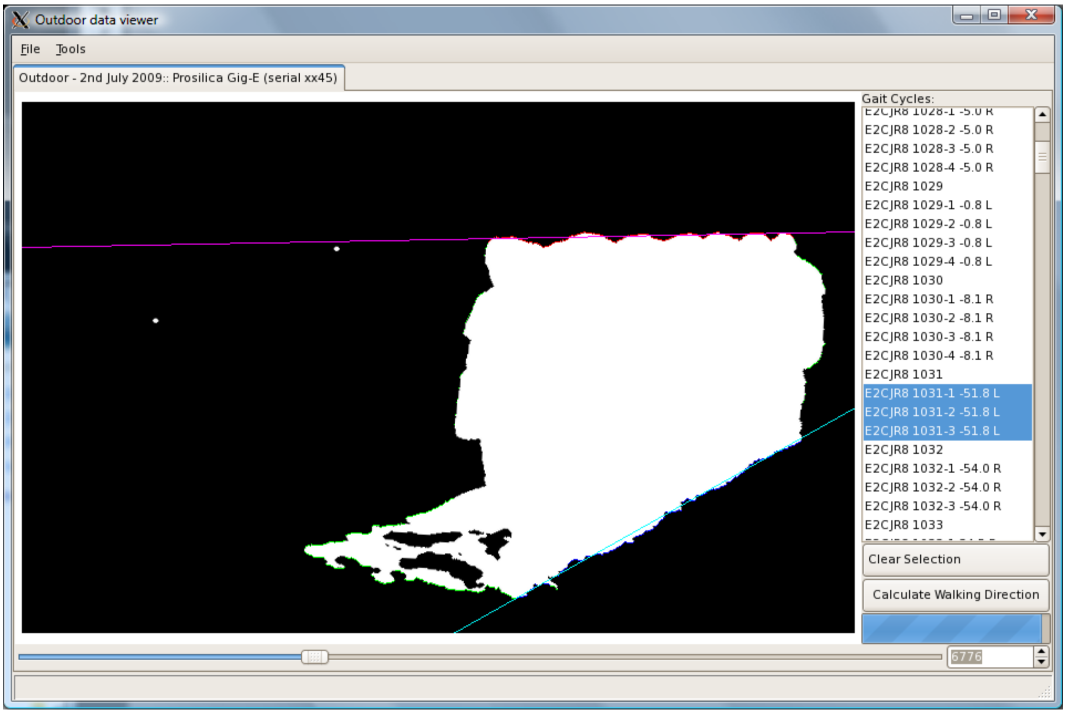
The three cameras were connected to a single computer, using a specialist Intel network interface card, featuring four Gigabit Ethernet ports. The raw unprocessed data from each camera was saved to a separate hard-drive to ensure minimal resource contention on the host computer, in order to reduce the likelihood of dropped frames. After capture, software was written to read the video data files and extract colour video frames for viewing. A segmentation algorithm was applied to the camera data, using a normalised RGB colour space and a per-pixel infinite impulse response (IIR) filter to estimate the background model. Connected component analysis was applied to the segmented images, and regions of interest were found. A viewer written for the camera data displayed the regions of interest overlaid on top of the camera images; regions of interest containing subjects were then labelled using their Biometric Tunnel identifiers. Gait cycles were manually identified for each subject. The extracted silhouette data was found be of



(a) Ariel photograph of outdoor environment; ©Copyright Geoperspectives 1999-2008, sourced from GeoStore[63]

(b) Plan-view of experiment's layout, where the tripod-mounted cameras are shown in red, the traffic cone markers in orange and the car parking bays in yellow.

Figure 9.3: Layout of the outdoor recognition experiment


Outdoor data viewer
File
Tools
Outdoor - 2nd July 2009:: Prosilica Gig-E (serial xx45)
Clear Selection
Calculate Walking Direction
6776

Gait Cycles:
E2CJR8 1028-1 -5.0 R
E2CJR8 1028-2 -5.0 R
E2CJR8 1028-3 -5.0 R
E2CJR8 1028-4 -5.0 R
E2CJR8 1029
E2CJR8 1029-1 -0.8 L
E2CJR8 1029-2 -0.8 L
E2CJR8 1029-3 -0.8 L
E2CJR8 1029-4 -0.8 L
E2CJR8 1030
E2CJR8 1030-1 -8.1 R
E2CJR8 1030-2 -8.1 R
E2CJR8 1030-3 -8.1 R
E2CJR8 1030-4 -8.1 R
E2CJR8 1031
E2CJR8 1031-1 -51.8 L
E2CJR8 1031-2 -51.8 L
E2CJR8 1031-3 -51.8 L
E2CJR8 1032
E2CJR8 1032-1 -54.0 R
E2CJR8 1032-2 -54.0 R
E2CJR8 1032-3 -54.0 R
E2CJR8 1033

FIGURE 9.4: Walking direction is found by merging the silhouettes and finding the vanishing point

extremely poor quality, due to the poor contrast present in the scene. It is expected that the use of a better quality camera and background segmentation algorithm would have rectified the silhouette quality issues; although this is outside of the scope of this thesis, which has a primary focus on gait — not state of the art background segmentation. Therefore, some of the worst silhouettes, where substantial regions of the subject were incorrectly identified were manually corrected. Due to time constraints, data from only cameras 45 and 46 was processed, and manual silhouette retouching was only performed on camera 45, whilst the other camera was retained as a control, with no retouching performed.

A partial calibration of the cameras was performed, where each camera's intrinsic parameters were estimated using a separate scene to find the focal length, principal point and radial distortion for the camera. The extrinsic parameters were found using the convergence of the horizontal and vertical lines formed by the bricks in the background of the scene; the vanishing points were used to estimate the camera's rotation around its principal axis. The vanishing point of the vertical lines formed by the bricks was used to calculate the camera's elevation angle.

For each sequence where a subject was walking in a straight line, their silhouettes were merged by finding the union of all the images. Lines were fitted to the top and bottom of the silhouette trails, to find a vanishing point, as shown in Figure 9.4. This vanishing point was used to estimate the subject's walking direction relative to the camera. Using

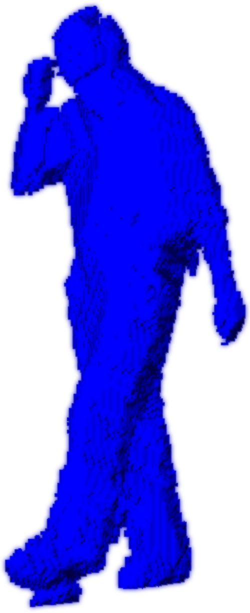


FIGURE 9.5: Three-dimensional reconstruction after global optimisation of camera calibration

this information, it was possible to re-project the volumetric data from the tunnel to a similar orientation to that of the subject; allowing the corresponding average silhouette to be calculated.

Due to the poor results encountered earlier in this chapter whilst attempting to re-project the volumetric data, all cameras within the Biometric Tunnel were recalibrated to ensure accuracy. Following this, it was decided to further refine the calibration of the cameras within the system by attempting to globally optimise the calibration for all cameras. This was achieved by refining each camera in turn, with the intent to maximise the volume of the reconstructed target. A set of frames were used with the target subject located at several different points in the tunnel in order to avoid local over-fitting issues. This process proved to be computationally expensive, requiring a substantial amount of time to complete. The global optimisation of the cameras resulted in a significant improvement in reconstruction quality, with much finer details resolved within the volume, as shown in Figure 9.5.

9.5.2 Average silhouette based recognition

The first experiment to test the viability of re-projection based matching was performed by comparing subjects from the outdoor dataset against re-projected silhouettes from the Biometric Tunnel; where the scale-normalised average silhouette was used to characterise and classify the subjects. A leave-one-out validation experiment was performed using

Probe Source	Gallery Source	CCR k=1	EER	Decidability (d)
Camera 45	Camera 45	80.37%	48.05%	0.28
Camera 45	Biometric Tunnel	68.10%	27.96%	1.14
Camera 46	Camera 46	91.97%	40.12%	0.49
Camera 46	Biometric Tunnel	55.47%	31.05%	0.84

Table 9.1: Results of matching reprojected volumetric data against single camera outdoor footage, using the average silhouette analysis technique

only the outdoor recorded data, to evaluate the quality of the collected video footage. The recognition performance for the footage recorded by the two cameras was found to be poor, achieving correct classification rates of only 80.37% and 91.97%; demonstrating the difficulty of performing gait recognition using data from an outdoor environment. In contrast, the previous results by Veres et al. [109] have shown that it is possible to achieve extremely high correct classification rates using the average silhouette on a large indoor dataset. The poor quality of the remaining unedited silhouettes is likely to have had a serious impact on the recognition performance; this is outlined by the difference in classification rates between the two cameras, where some editing of the silhouettes was performed for the former. The results are summarised in Table 9.1, whilst the receiver operating characteristic is shown in Figure C.3(a).

The classification performance of the experiment matching the outdoor data against re-projected gallery data was found to be comparatively poor, where the retouching of silhouette data for camera 45 clearly resulted in an improved performance against the Biometric Tunnel data, indicating that the segmentation errors had a severe impact on recognition performance. It is interesting to note that although the verification experiment achieved a higher correct classification rate, the equal error rate, decidability and the receiver operating characteristic were all significantly worse than that of the experiment matching tunnel data against the outdoor data. This also suggests that the poor background segmentation quality was having a strong influence on the degraded recognition performance, it is also possible that human error was introduced during the manual editing of the silhouette data. The equal error rate for both cameras was substandard compared to the rates obtained in earlier Chapters, suggesting that the average silhouette was unable to provide sufficient inter-class variation to facilitate accurate matching. The poor recognition performance of the average silhouette in this scenario reaffirms the suggestion that the average silhouette is extremely sensitive to any error in the silhouette's shape, which can be caused by poor calibration of the cameras within the tunnel, inaccurate estimation of the subject's orientation or walking direction, or as mentioned earlier; segmentation errors. Another factor that may have had a minor impact on the results was the time delay between the probe and gallery samples; where the probe data was recorded in the morning and the gallery data was recorded in the afternoon of the same day — this could have lead to some variation the subjects' gait between probe and gallery recordings. It is expected that the use of better quality silhouette data would have resulted in a greatly improved recognition rate.

9.5.3 Non-normalised average silhouette based recognition

The disappointing results presented in the previous section confirmed that it was extremely difficult to match samples from a single outdoor viewpoint against those re-projected from a three-dimensional dataset. The poor quality of the silhouettes extracted from the outdoor data was believed to be a major cause of the degraded recognition performance, as confirmed by the difference in performance between the two cameras. Whilst it is expected that the use of a more carefully controlled scene or a better quality background segmentation algorithm would result in improved performance, it is also likely that the use of an improved gait analysis technique would yield an improvement.

As discussed in Section 6.5, it was found that the removal of the scale-normalisation stage from the average silhouettes resulted in a significant improvement in recognition performance. This was possible by re-projecting the three-dimensional data to one or more orthonormal viewpoints. Unfortunately it is impractical to obtain a true orthonormal view of a subject using only a single camera; as this would require a telecentric lens with a front-element larger than the subject. Using a standard non-telecentric lens, the magnification of the subject in the image will be dependant on their distance from the camera; this means that the distance of the subject must be known in order to remove the effects of perspective scaling. The distance of a person from a partially calibrated camera can be estimated by finding the position of their feet in the image; assuming that their feet are on the ground plane. Once a distance estimate has been obtained, the position of the top of the subject's head can be found and used to calculate the subject's height. In order to calculate the subject's distance from the camera and their height, the camera's height from the ground-plane, elevation angle and intrinsic parameters must be known. The average silhouette can then be calculated in the same manner as the scale-normalised variant; although the resulting average silhouette is scaled the by subject's measured height. The height for the subjects in the Biometric Tunnel gallery data is found by simply taking the height of the three-dimensional bounding box surrounding each subject.

The verification experiment showed some improvement for both cameras in terms of the equal error rate and decidability d; although the classification performance was still lower than expected, which was likely due to the segmentation errors discussed earlier. The use of non-normalised average silhouettes resulted in a reasonable improvement in recognition performance, where the classification rate for the new approach was much closer to that of the corresponding verification experiment. The recognition results are shown in Table 9.2 and Figure C.3(b). The results confirm that this new technique is

Probe Source	Gallery Source	CCR k=1	EER	Decidability (d)
Camera 45	Camera 45	83.44%	39.94%	0.61
Camera 45	Biometric Tunnel	79.75%	18.24%	1.66
Camera 46	Camera 46	89.78%	31.35%	1.00
Camera 46	Biometric Tunnel	70.07%	24.20%	1.32

Table 9.2: Results of matching reprojected volumetric data against single camera outdoor footage, using the non-normalised average silhouette analysis technique

capable of recognising an individual by their gait with a reasonable degree of accuracy in a viewpoint-invariant manner.

In this section a simple yet effective technique has been proposed to facilitate viewpoint-invariant gait recognition, by re-projecting three-dimensional enrolment data to the same orientation as an unknown subject in a camera image. This approach avoids the usual problems encountered with two-dimensional gait analysis techniques, where incorrect matching may occur if the orientation of a subject differs between their enrolment and test samples. By re-projecting the enrolment gallery samples to the same viewpoint as the probe samples, almost any two-dimensional gait analysis technique may be used for recognition. It is expected that the use of a more sophisticated gait analysis technique may lead to significantly better recognition performance; although poor quality background segmentation will always be an issue for any gait analysis technique that uses silhouette data.

Chapter 10

Conclusion and Recommendations

In this thesis the University of Southampton of Biometric Tunnel has undergone a significant transformation, to result in a system capable of acquiring non-contact biometric measurements in a fast and efficient manner. More importantly, the issues experienced with the early prototype of Middleton et al. [70] have been investigated in a rigorous manner and duly rectified; resulting in a system capable of achieving excellent recognition performance on populations of a significant size. The system has been used to collect the largest to date multi-biometric dataset featuring two and three-dimensional gait data. Correct classification of all samples within the dataset can be achieved with the use of a simple average silhouette derivative, proposed in these works. These results provide significant weight behind the argument of Murray et al. [74], that each individual's gait is unique. Further experiments have helped to show that the system could be deployed in real-world scenarios, providing that good quality background segmentation is possible and that all cameras are accurately calibrated.

The initial prototype Biometric Tunnel was designed and constructed by Middleton et al. [70]; evaluation of data produced by this system found the recognition performance to be significantly below expected. This prompted an in depth examination and evaluation of the system's underlying hardware and software. An in depth inspection of the acquisition software revealed several issues with how captured video data was handled. A new system was devised, using lessons learnt from the previous prototype. The hardware configuration and layout was substantially modified, to improve access and reliability. The underlying software was split up so that the captured video data was saved to disk without any processing; each stage of the processing was then implemented as a separate application. This meant that each computer-vision algorithm could be executed and evaluated in isolation. The previous manual gait cycle labelling strategy was found to result in inconsistent and error prone labelling; therefore, an automated gait cycle finding algorithm was devised.

A batch processing system was implemented to automate the execution of the processing applications and other tasks, such as gait cycle labelling and sample backup. The batch processing system was capable of managing the processing of samples and entire datasets. A variety of tools were written for administering the batch processing system, including a web-based interface, which facilitated live progress monitoring and the debugging of failed processing operations.

In order to test the revised system and evaluate the underlying processing algorithms, a small dataset was collected that contained unprocessed video data, to enable the evaluation of the previous system's processing algorithms. Analysis of the new dataset revealed that the system's stability had improved greatly, along with the visual quality of the three-dimensional reconstructions; most likely as a result of the newly implemented camera synchronisation algorithm. This was reflected in the significantly improved correct classification rates achieved, which demonstrated that the revised system was capable of producing data of a sufficient quality for gait recognition on a large population.

Evaluation of the video processing algorithms using the newly collected dataset revealed that the background segmentation was producing erroneous results; further inspection found that the grey regions of the background did not provide sufficient discriminatory ability against the subjects and their clothing. Analysis of the colours present in the video of the previously collected dataset confirmed that grey was a poor choice of background colour; therefore, these areas were repainted with an intense red colour, which provided better separation between participants and the background. This resulted in cleaner silhouettes with less artefacts from erroneous background segmentation.

An experimental real-time shape from silhouette reconstruction algorithm was implemented, which was capable of achieving rates in excess of 30 reconstructions per second, whilst only using one processor core. This reconstruction algorithm was later integrated into an experimental system using graphics hardware to accelerate the processing of the camera images, to result in a system capable of almost real-time recognition[75]. The ability to perform processing in real-time system demonstrates that it is possible to produce a deployable system capable of low-latency identification of subjects.

The collection of a large dataset containing measurements from multiple non-contact biometrics was carefully planned and prepared for. The dataset was primarily focused on three-dimensional gait, although also included face and ear imagery. The Biometric Tunnel and its associated systems were prepared for the collection of this dataset, which included the streamlining of the capture process, new storage and backup systems and a rigorous experimental procedure to ensure consistency between participants. An advanced web-interface was produced for collecting the participant's personal information; such as age, gender and ethnicity. The web-interface also featured a dedicated administration section for managing the system and the experiment. Collection of the dataset commenced, with data from over one-hundred participants collected in twelve one-day sessions. Collection of data was temporarily halted for several months due to serious technical problems; during this period, four additional additional cameras were added to the system, to allow the evaluation of multiple camera configurations. Once the technical issues had been resolved and thorough testing had been performed, data collection recommenced. The completed dataset contained in excess of two-hundred unique participants and two-thousand samples, of which some individuals participated in more than one capture session during the experiment.

Conducting a leave-one-out recognition experiment on the entire dataset revealed excellent results, with a correct classification rate of 99.52% — meaning that almost every sample within the dataset was correctly identified. The equal error rate for the combination of all three viewpoints was found to be 3.58%, the point where both the false rejection and false acceptance rates are equal. These results were achieved using the well known gait analysis technique; the average silhouette. A new derivative of the average silhouette was devised, where no scale-normalisation was used. This technique is ideally suited to three-dimensional data, where artificial views can be created without the effect of perspective. The new analysis method resulted in improved performance, with every sample being correctly classified, and a reduced equal error rate of 1.58%. At the previous 3.58% false reject rate, the false accept rate significantly drops to 0.4%. The results of both experiments were acheived without the aid of feature selection or transformation techniques, whilst only using nearest-neighbour classification. This proves to be a very significant finding; demonstrating that there is a significant variation in gait between individuals, which remains sufficiently stable over short periods of time to facilitate accurate recognition. Another recognition experiment was conducted, where an individual's samples were matched against those from a previous data-capture session; whilst recognition performance was not comparable to the short-term variation results, it still exhibited significant discriminatory abilities, acheiving a 72.22% correct classification rate — greatly above the rate of chance alone. The side-on viewpoint was found to perform significantly better than the other viewpoints when matching samples from different dates; this suggests that the motion of the subject's limbs revealed by the side-on viewpoint is of significant discriminatory value.

Of the many additional experiments conducted using the newly collected dataset, many were focused on gaining a greater understanding of the system's performance and more importantly — its limitations. The resolution of the average silhouettes used for analysis was shown to be appropriate, with a reduction in the number of features resulting in a decreased recognition rate; where the decline accelerated rapidly when reducing the number of features by 50% or more. It was also found that the simple concatenation of the three viewpoints was a comparatively inefficient representation, compared against that of a single viewpoint, where on a per feature basis the single viewpoint was capable of better recognition performance. The recognition performance of the system was found to be extremely sensitive to camera calibration errors, where the removal of the central wide field-of-view cameras resulted in an improved performance. This was most likely due to the difficulty in accurately calibrating the central cameras using only the two planes formed by the corresponding visible side-wall and the floor. An investigation into the effects of camera calibration error on the classification performance revealed the extent of the system's sensitivity to calibration error. This is mostly expected, as almost all multi-viewpoint reconstruction techniques exhibit sensitivity to calibration error, where a given point on one camera no longer correctly corresponds to the equivalent projection on another. An attempt was also made at evaluating the discriminatory abilities of the static and dynamic components of gait in isolation, whilst reasonable recognition performance was achieved for both components, the performance of both elements combined was not comparable to that achieved by the average silhouette; raising questions as whether the two extracted signatures truly accounted for the dynamic and static variation within one's gait.

The practicality of using both gait and facial recognition in a single system was evaluated, using an extremely simple facial analysis technique to characterise the appearance of one's face. The performance of the fused biometrics was found to be greater than either biometric in isolation; where the equal error rate was 1.04%, meaning that the false accept rate was less than 0.4% for the 1.58% false reject rate acheived by gait alone. The recognition rates of both biometrics were evaluated against the distance from a simulated camera, where it was found that gait was able to provide useful information at distances where facial recognition had ceased to function.

Finally a set of experiments were conducted to assess the possibility of using the Biometric Tunnel in an enrolment only scenario, where matching was then performed against video footage from single cameras placed in a less controlled environment. Initial experimentation where single cameras inside the Biometric Tunnel were used for matching purposes proved relatively unsuccessful, due to calibration problems and the additional inaccuracy of the three-dimensional reconstructed data when a camera was removed. An experiment was also conducted to investigate the effect of camera position on recognition performance, where the recognition performance of a large number of viewpoints was evaluated. It was found that the best recognition performance could be achieved using a front-on viewpoint, suggesting that the static information contained within a subject's gait and body shape provides some of the most important identifying features. A small outdoor dataset was collected and then analysed. It was found that it was possible to estimate a subject's walking direction and orientation using the vanishing point formed by the bounding trails of their silhouette; this information combined with basic camera calibration information facilitated the projection of the three-dimensional Biometric Tunnel data to a similar pose to that of the observed sample. Serious problems with the quality of the background segmentation resulted in extremely poor recognition performance. Manual retouching of the silhouette data from one of the cameras resulted in an improvement in matching accuracy, increasing from 55.47% to 68.10%. By estimating the camera's orientation, the height of the observed subjects could be estimated; allowing the creation of average silhouettes where the effects of scale-normalisation had been removed. Matching the new gait signatures against the data from the Biometric Tunnel yielded a significant increase in the accuracy of matches — achieving a 79.75% correct classfication rate — although poor background segmentation quality still had a substantial impact on recognition performance.

In this thesis it has been shown that it is possible to correctly identify an individual from a population of over two-hundred others. The inevitable question from these findings is whether the the recognition performance of gait and the Biometric Tunnel will remain acceptably high when used in an environment with a much larger population. In Chapter 6, several different measures were used to evaluate the recognition performance of the Biometric Tunnel; this included the correct classification rate, the equal error rate and the decidability. As discussed in earlier chapters, the correct classification rate is strongly dependant on the composition of the gallery set used for matching; this means that it is difficult to accurately estimate with a larger gallery size. Although it is expected that the error rate will increase with the number of subjects, as the increased density of subjects within the feature space will reduce the separation between classes. Whilst the correct classification rate is of little use for assessing the scalability of such a system, the equal error rate and decidability are likely to prove more useful, as they are both determined by the shape and separation of the inter and intra-class distributions. As the size of the dataset increases, the inter and intra-class distributions will stabilise, which means that the equal error rate and decidability will converge towards their correct values. With a dataset containing in excess of two-hundred subjects and two-thousand samples, the change in the the equal error rate and decidability is unlikely to be significant for a larger dataset.

The dataset collected for the purposes of this thesis does not attempt to account for covariates such as changes in clothing or footwear, walking surface or speed, although a limited study of temporal variation was performed. Most of these factors are believed to affect an individual's gait, and therefore could have a significant impact on the ability to recognise an individual. In order to fully evaluate the real world performance of a recognition system, an analysis of these covariates and their effects is necessary. From the experience gained collecting the dataset discussed in this thesis, it is believed that it will be extremely difficult to collect a dataset of a significant size containing such covariates, as it would substantially increase the time required of each participant, making the recruitment of individuals much harder without much larger financial incentives. Whilst collecting the new dataset, recruiting participants was found to be extremely difficult and time-consuming, with very few people responding to advertisements placed around the University of Southampton; the most effective method of recruitment was found to be by word of mouth and persistence. An efficient way to collect a much larger dataset would be to organise a collaborative project involving several different research institutions, with each recruiting from their own population of potential participants.

The development of a more thorough calibration strategy will prove greatly beneficial to the long term practicality of the system, reducing the impact of camera registration error. It is suggested that a system of continual monitoring and calibration refinement is implemented, meaning that any change in camera orientation over time can be tracked and compensated for. As shown in Chapter 9, the use of a volume maximisation bundle-adjustment algorithm on a recorded video sequence resulted in an improved calibration, with results that were a significant improvement in visual quality. The use of this calibration optimisation strategy on the entire dataset is likely to result in further improvements to the system's recognition performance, especially when matching samples from different dates. Using the high quality three-dimensional data produced by the Biometric Tunnel, it should be possible to develop a system capable of accurately fitting a three-dimensional model to one's gait. This could have a wide range of uses, including biometric recognition, computer animation and medical gait analysis, where it could be used instead of marker-based system, to help rehabilitate patients who have difficulties walking.

Bibliography

[1] Alex I. Bazin and Mark S. Nixon. Probabilistic combination of static and dynamic gait features for verification. In *Biometric Technology for Human Identification II*, volume 5779 of *Proceedings of SPIE*, pages 23–30, March 2005.

[2] Chiraz BenAbdelkader, Ross Cutler, and Larry S. Davis. Motion-based recognition of people in eigengait space. In *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 254–259, May 2002.

[3] Chiraz BenAbdelkader, Ross Cutler, and Larry S. Davis. Stride and cadence as a biometric in automatic person identifcation and verification. In *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 357–362, May 2002. ISBN 0-7695-1602-5.

[4] Chiraz BenAbdelkader, Ross Cutler, Harsh Nanda, and Larry S. Davis. Eigengait: Motion-based recognition of people using image self-similarity. In *Proceedings of Third International Conference on Audio- and Video-Based Biometric Person Authentication*, volume 2091 of *Lecture Notes in Computer Science*, pages 284–294, June 2001.

[5] Bir Bhanu and Ju Han. Human recognition on combining kinematic and stationary features. In *Proceedings of Audio- and Video-Based Biometric Person Authentication*, volume 2688 of *Lecture Notes in Computer Science*, pages 600–608. Springer-Verlag, 2003.

[6] Aaron F. Bobick and A. Y. Johnson. Gait recognition using static, activity-specific parameters. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 423–430, 2001.

[7] Imed Bouchrika. *Gait Analysis and Recognition for Automated Visual Surveillance*. PhD thesis, University of Southampton, June 2008.

[8] Imed Bouchrika and Mark S. Nixon. Model-based feature extraction for gait analysis and recognition. In *Proceedings of Mirage: Computer Vision / Computer Graphics Collaboration Techniques and Applications*, pages 150–160, 2007.

[9] Nikolaos V. Boulgouris and Zhiwei X. Chi. Gait representation and recognition based on radon transform. In *Proceedings of IEEE International Conference on Image Processing*, pages 2665–2668, October 2006.

[10] K. W. Bowyer, K. Chang, and P. Flynn. A survey of approaches to three-dimensional face recognition. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 1, 2004.

[11] Jeffrey E. Boyd. Video phase-locked loops in gait recognition. In *Proceedings of Eighth IEEE International Conference on Computer Vision*, volume 1, pages 696–703, July 2001.

[12] G. Bradski and A. Kaehler. *Learning OpenCV: Computer vision with the OpenCV library.* O'Reilly Media, Inc., 2008.

[13] J. E. Bresenham. Algorithm for computer control of a digital plotter. *IBM Systems Journal*, 4(1):25–30, 1965.

[14] M. Z. Brown, D. Burschka, and G. D. Hager. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8): 993–1008, 2003.

[15] H. Bülthoff, J. Little, and T. Poggio. A parallel algorithm for real-time computation of optical flow. *Nature*, 337(6207):549–553, 1989.

[16] E. Chang, S. Cheung, and D.Y. Pan. Color filter array recovery using a threshold-based variable number of gradients. In *Proceedings of SPIE*, volume 3650, page 36, 1999.

[17] K. I. Chang, K. W. Bowyer, and P. J. Flynn. Multi-Modal 2D and 3D Biometrics for Face Recognition. In *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures.* IEEE Computer Society Washington, DC, USA, 2003.

[18] C. H. Chen and C. Te Chu. Fusion of face and iris features for multimodal biometrics. *Lecture notes in computer science*, 3832:571–580, 2005.

[19] German K. M. Cheung, Takeo Kanade, Jean-Yves Bouguet, and Mark Holler. A real time system for robust 3d voxel reconstruction of human motions. *CVPR*, 02:2714, 2000. ISSN 1063-6919.

[20] David R. Cok. Signal processing method and apparatus for producing interpolated chrominance values in a sampled color image signal. US Patent 4,642,678, February 1987.

[21] Robert T. Collins, Ralph Gross, and Jianbo Shi. Silhouette-based human identifcation from body shape and gait. In *Proceedings of the Fifth IEEE*

International Conference on Automatic Face and Gesture Recognition, pages 351–356, May 2002. ISBN 0-7695-1602-5.

[22] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In H. Burkhardt and B. Neumann, editors, *European Conference on Computer Vision, Proceedings of*, volume 2, pages 484–498. Springer, 1998.

[23] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Training models of shape from sets of examples. In *Proc. British Machine Vision Conference*, pages 266–275, Berlin, 1992. Springer.

[24] P. Courtney and N. A. Thacker. Performance characterization in computer vision: The role of statistics in testing and design. In J. Blanc-Talon and D. C. Popescu, editors, *Imaging and Vision Systems: Theory, Assessment and Applications*, Nova Science, 2001.

[25] David Cunado, Jason M. Nash, Mark S. Nixon, and John N. Carter. Gait extraction and description by evidence-gathering. In *Proceedings of the International Conference on Audio and Video Based Biometric Person Authentication*, pages 43–48, 1999.

[26] David Cunado, Mark S. Nixon, and John N. Carter. Using gait as a biometric, via phase-weighted magnitude spectra. In *Proceedings of 1st International Conference on Audio- and Video-Based Biometric Person Authentication*, pages 95–102. Springer-Verlag, 1997.

[27] David Cunado, Mark S. Nixon, and John N. Carter. Automatic extraction and description of human gait models for recognition purposes. *Computer Vision and Image Understanding*, 90(1):1–41, April 2003.

[28] J. Daugman. Biometric decision landscapes. Technical Report 482, University of Cambridge Computer Laboratory, jan 2000.

[29] James W. Davis and Stephanie R. Taylor. Analysis and recognition of walking movements. In *Proceedings of 16th International Conference on Pattern Recognition*, volume 1, pages 315–318, 2002.

[30] Philip Denbigh. *System Analysis & Signal Processing.* Addison-Wesley, 1998. ISBN 0-201-17860-5.

[31] Richard O. Duda and Peter E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15:11–15, January 1972.

[32] Charles R. Dyer. *Volumetric Scene Reconstruction From Multiple Views*, chapter 16, pages 469—489. Foundations of Image Understanding. Kluwer, Boston, 2001.

[33] A. Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. *Lecture Notes in Computer Science*, 1843:751–767, 2000.

[34] Jeff P. Foster, Mark S. Nixon, and Adam Prügel-Bennett. Automatic gait recognition using area-based metrics. *Pattern Recognition Letters*, 24:2489–2497, 2003.

[35] D. M. Gavrila. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1):82–98, January 1999.

[36] Michela Goffredo, Richard D. Seely, John N. Carter, and Mark S. Nixon. Markerless view independent gait analysis with self-camera calibration. In *Proceedings of the Eighth IEEE International Conference on Automatic Face and Gesture Recognition*, September 2008.

[37] Michela Goffredo, Nicholas Spencer, Daniel Pearce, John N. Carter, and Mark S. Nixon. Human perambulation as a self calibrating biometric. *Lecture Notes in Computer Science*, 4778:139, 2007.

[38] G. Gordon, T. Darrell, M. Harville, and J. Woodfill. Background estimation and removal based on range and color. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 459–464. IEEE, Los Alamitos, CA, USA, 1999.

[39] T. A. Gore, G. R. Higginson, and J. Stevens. The kinematics of hip joints: normal functioning. *Clinical Physics and Physiological Measurements*, 5(4): 233–252, 1984.

[40] Ralph Gross and Jianbo Shi. The cmu motion of body (mobo) database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, June 2001.

[41] E. Grosso, G. Sandini, and M. Tistarelli. 3-D object reconstruction using stereo and motion. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6): 1465–1476, 1989.

[42] Yan Guo, Gang Xu, and Saburo Tsuji. Understanding human motion patterns. In *Proceedings of 12th International Conference on Pattern Recognition*, volume 2, pages 325–329, oct 1994.

[43] Ju Han and Bir Bhanu. Statistical feature fusion for gait-based human recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 842–847, June 2004.

[44] James B. Hayfron-Acquah, Mark S. Nixon, and John N. Carter. Recognising human and animal movement by symmetry. In *Proceedings of International*

Conference on Image Processing, volume 3, pages 290–293, 2001. ISBN 0-7803-6725-1.

[45] Qiang He and Chris Debrunner. Individual recognition from periodic activity using hidden markov models. In *Proceedings of Workshop on Human Motion*, pages 47–52, 2000.

[46] K. Hirakawa and TW Parks. Adaptive homogeneity-directed demosaicing algorithm. *IEEE Transactions on Image Processing*, 14(3):360–369, 2005.

[47] T. Horprasert, D. Harwood, and L.S. Davis. A statistical approach for real-time robust background subtraction and shadow detection. In *IEEE ICCV*, volume 99, 1999.

[48] Weiming Hu, Tieniu Tan, Liang Wang, and Steve Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics*, 34(3):334–352, August 2004.

[49] Ping S. Huang, Chris J. Harris, and Mark S. Nixon. Canonical space representation for recognizing humans by gait and face. In *Proceedings of IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 180–185, April 1998.

[50] Anil K. Jain and Arun Ross. Multibiometric systems. *Communications of the ACM*, 47(1):40, 2004.

[51] Anil K. Jain, Arun Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, January 2004. ISSN 1558-2205.

[52] G. Johannson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14:201–211, October 1973.

[53] Chuan kai Lin. Pixel grouping for color filter array demosaicing. Online, April 2003.

[54] Amit Kale, A. N. Rajagopalan, Naresh P. Cuntoor, and Volker Krüger. Gait-based recognition of humans using continuous hmms. In *Proceedings of Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 321–326, May 2002.

[55] R. Kimmel. Demosaicing: image reconstruction from color CCD samples. *IEEE Transactions on Image Processing*, 8(9):1221–1228, 1999.

[56] J. Kittler and F. M. Alkoot. Sum versus vote fusion in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 110–115, 2003.

[57] Takumi Kobayashi and Nobuyuki Otsu. Action and simultaneous multiple-person identification using cubic higher-order local auto-correlation. In *Proceedings of the 17th International Conference on Pattern Recognition*, volume 4, pages 741–744, August 2004.

[58] A. Kumar, D.C.M. Wong, H.C. Shen, and A.K. Jain. Personal verification using palmprint and hand geometry biometric. *Lecture notes in computer science*, pages 668–678, 2003.

[59] Toby H. W. Lam, Raymond S. T. Lee, and David Zhang. Human gait recognition by the fusion of motion and static spatio-temporal templates. *Pattern Recognition*, 40(9):2563–2573, September 2007.

[60] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2): 150–162, February 1994. ISSN 0162-8828.

[61] Lily Lee. Gait analysis for classification. Technical Report AITR-2003-014, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, June 2003.

[62] Lily Lee and W. E. L. Grimson. Gait analysis for recognition and classification. In *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'02)*, pages 155–161. IEEE, 2002.

[63] Infoterra Limited. Geostore. Online, 2010.

[64] James J. Little and Jeffrey E. Boyd. Describing motion for recognition. In *Proceedings of International Symposium on Computer Vision*, pages 235–240, November 1995.

[65] James J. Little and Jeffrey E. Boyd. Recognizing people by their gait: The shape of motion. *Videre: Journal of Computer Vision Research*, 1(2), 1998.

[66] Yanxi Liu, Robert T. Collins, and Yanghai Tsin. Gait sequence analysis using frieze patterns. In *Proceedings of 7th European Conference on Computer Vision*, volume 2351 of *Lecture Notes in Computer Science*, pages 657–671, May 2002.

[67] Zongyi Liu and Sudeep Sarkar. Simplest representation yet for gait recognition: averaged silhouette. In *Proceedings of the 17th International Conference on Pattern Recognition*, volume 4, pages 211–214, August 2004.

[68] Worthy N. Martin and J. K. Aggarwal. Volumetric descriptions of objects from multiple views. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 5(2):150–158, March 1983. ISSN 0162-8828.

[69] George Mather and Linda Murdoch. Gender discrimination in biological motion displays based on dynamic cues. In *Biological Sciences*, volume 258, pages 273–279. The Royal Society, December 1994.

[70] Lee Middleton, David Kenneth Wagg, Alex I. Bazin, John N. Carter, and Mark S. Nixon. Developing a non-intrusive biometric environment. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 723–728, October 2006. ISBN 1-4244-0259-X.

[71] Lee Middleton, Sylvia C. Wong, Michael O. Jewell, John N. Carter, and Mark S. Nixon. Lightweight agent framework for camera array applications. In *Proceedings of 9th International Conference on Knowledge-Based Intelligent Information and Engineering Systems*, volume 3684 of *Lecture Notes in Computer Science*, pages 150–156. Springer-Verlag, 2005. ISBN 3-540-28897-X.

[72] Hiroshi Murase and Rie Sakai. Moving object recognition in eigenspace representation: gait analysis and lip reading. *Pattern Recognition Letters*, 17(2): 155–162, February 1996.

[73] D. D. Muresan, S. Luke, and T. W. Parks. Reconstruction of color images from ccd arrays. In *Texas Instruments DSP Fest*, Houston, TX, USA, August 2000.

[74] M. Pat Murray, A. Bernard Drought, and Ross C. Kory. Walking patterns of normal men. *The Journal of Bone & Joint Surgery*, 46:335–360, 1964.

[75] J. Musko. Enabling real-time gait recognition through the use of graphics processing hardware. Master's thesis, School of Electronics and Computer Science, University of Southampton, UK, 2009.

[76] Mark S. Nixon and John N. Carter. Automatic recognition by gait. *Proceedings of the IEEE*, 94(11):2013–2024, November 2006. ISSN 0018-9219.

[77] Mark S. Nixon, John N. Carter, Michael G. Grant, Layla Gordon, and James B. Hayfron-Acquah. Automatic recognition by gait: progress and prospects. *Sensor Review*, 23(4):323–331, 2003. ISSN 0260-2288.

[78] Mark S. Nixon, John N. Carter, Jason M. Nash, Ping S. Huang, David Cunado, and S. V. Stevenage. Automatic gait recognition. In *IEE Colloquium on Motion Analysis and Tracking*, volume 3, pages 1–6, 1999.

[79] Mark S. Nixon, Tieniu Tan, and Rama Chellappa. *Human Identification Based on Gait*. International Series on Biometrics. Springer, 2006. ISBN 0-387-24424-7.

[80] Sourabh A. Niyogi and Edward H. Adelson. Analyzing and recognizing walking figures in xyt. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 469–474, June 1994.

[81] The Office of National Statistics. The one number census, 2001.

[82] Carlos Orrite-Uruñuela, Jesús Martínez del Rincón, J. Elías Herrero-Jaraba, and Grégory Rogez. 2d silhouette and 3d skeletal models for human detection and tracking. In *Proceedings of the 17th International Conference on Pattern Recognition*, volume 4, pages 244–247, 2004.

[83] Q. Pan, G. Reitmayr, and T. Drummond. ProFORMA: Probabilistic Feature-based On-line Rapid Model Acquisition. In *Proc. 20th British Machine Vision Conference (BMVC)*, London, September 2009.

[84] P. Johnathon Phillips, Sudeep Sarkar, Isidro Robledo, Patrick Grother, and Kevin W. Bowyer. The gait identification challenge problem: data sets and baseline algorithm. In *Proceedings of The 16th International Conference on Pattern Recognition*, volume 1, pages 385–388, 2002.

[85] R. Plänkers and P. Fua. Articulated soft objects for video-based body modeling. In *Proceedings. Eighth IEEE International Conference on Computer Vision*, volume 1, pages 394–401, July 2001.

[86] C. Ridder, O. Munkelt, and H. Kirchner. Adaptive background estimation and foreground detection using kalman-filtering. In *Proceedings of International Conference on recent Advances in Mechatronics*, pages 193–199, 1995.

[87] Arun Ross and Anil K. Jain. Multimodal biometrics: An overview. In *Proceedings of 12th European Signal Processing Conference*, pages 1221–1224, 2004.

[88] Sina Samangooei, Baofeng Guo, and Mark S. Nixon. The use of semantic human description as a soft biometric. In *Procedings of IEEE Conference on Biometrics: Theory, Applications and Systems, BTAS 08*, September 2008.

[89] Sudeep Sarkar, P. Johnathon Phillips, Zongyi Liu, Isidro Robledo Vega, Patrick Grother, and Kevin W. Bowyer. The humanid gait challenge problem: data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2):162–177, February 2005.

[90] G. Shafer. *A mathematical theory of evidence*. Princeton university press Princeton, NJ, 1976.

[91] G. Shakhnarovich, Lily Lee, and T. Darrell. Integrated face and gait recognition from multiple views. In *Computer Vision and Pattern Recognition*, volume 1 of *Proceedings of the 2001 IEEE Computer Society Conference on*, pages 439–446, 2001. ISBN 0-7695-1272-0.

[92] C. Shan and S. Gong. Fusing gait and face cues for human gender recognition. *Neurocomputing*, 71(10-12):1931–1938, 2008.

[93] Jamie D. Shutler, Michael G. Grant, Mark S. Nixon, and John N. Carter. On a large sequence-based human gait database. In *Proceedings of Fourth International Conference on Recent Advances in Soft Computing*, pages 66–72, 2002.

[94] Jamie D. Shutler and Mark S. Nixon. Zernike velocity moments for description and recognition of moving shapes. In *British Machine Vision Conference*, pages 705–714, 2001.

[95] Jamie D. Shutler and Mark S. Nixon. Zernike velocity moments for sequence-based description of moving features. *Image and Vision Computing*, 24 (4):343–356, 2006.

[96] Jamie D. Shutler, Mark S. Nixon, and Chris J. Harris. Statistical gait description via temporal moments. In *Image Analysis and Interpretation*, Proceedings of the 4th IEEE Southwest Symposium, pages 291–295, 2000.

[97] Jamie D. Shutler, Mark S. Nixon, and Chris J. Harris. Statistical gait recognition via velocity moments. In *Visual Biometrics*, volume 11 of *Proceedings of IEE Colloquium*, pages 1–5, 2000.

[98] G. Slabaugh, B. Culbertson, T. Malzbender, and R. Schafer. A survey of methods for volumetric scene reconstruction from photographs. In *Volume Graphics 2001: Proceedings of the Joint IEEE TCVG and Eurographics Workshop*, page 81. Springer, Birkhäuser, June 2001. ISBN 321183737X.

[99] Nick Spencer and John N. Carter. Towards pose invariant gait reconstruction. In *IEEE International Conference on Image Processing*, volume 3, pages III–261–4, Sept. 2005.

[100] Chris Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 246–252, June 1999.

[101] Aravind Sundaresan, Amit K. Roy-Chowdhury, and Rama Chellappa. A hidden markov model based framework for recognition of humans from gait sequences. In *Proceedings of International Conference on Image Processing*, volume 2, pages 93–96, September 2003.

[102] R. Szeliski and S. B. Kang. Recovering 3D shape and motion from image streams using nonlinearleast squares. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 752–753, 1993.

[103] Rawesak Tanawongsuwan and Aaron F. Bobick. Gait recognition from time-normalized joint-angle trajectories in the walking plane. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 726–731, 2001.

[104] Ezra Tassone, Geoff West, and Svetha Venkatesh. Temporal PDMs for Gait Classification. In *Proceedings of 16th International Conference on Pattern Recognition*, volume 2, pages 1065–1068, August 2002.

[105] N. A. Thacker, A. F Clark, J. Barron, R. Beveridge, C. Clark, P. Courtney, W. R. Crum, and V. Ramesh. Performance characterisation in computer vision: A guide to best practices. Tina Memo 2005-009, Image Science and Biomedical Engineering Division, Medical School, University of Manchester, United Kingdom, May 2005.

[106] David Tolliver and Robert T. Collins. Gait shape estimation for identification. In *Proceedings of Audio- and Video-Based Biometric Person Authentication*, volume 2688, pages 734–742, 2003.

[107] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, Proceedings of IEEE Conference on*, volume 591, pages 586–591, 1991.

[108] R. Urtasun and P. Fua. 3d tracking for gait characterization and recognition. In *Proceedings. Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 17–22, May 2004.

[109] Galina V. Veres, Layla Gordon, John N. Carter, and Mark S. Nixon. What image information is important in silhouette-based gait recognition? In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 776–782, 2004.

[110] Galina V. Veres, Mark S. Nixon, and John N. Carter. Is enough enough? what is sufficiency in biometric data? *Lecture Notes in Computer Science*, 4142:262, 2006.

[111] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 511–518, 2001.

[112] David Kenneth Wagg and Mark S. Nixon. On automated model-based extraction and analysis of gait. In *Proceedings of 6th International Conference on Automatic Face and Gesture Recognition*, pages 11–16, 2004.

[113] Liang Wang, Weiming Hu, and Tieniu Tan. A new attempt to gait-based human identification. In *Proceedings of 16th International Conference on Pattern Recognition*, volume 1, pages 115–118, August 2002.

[114] Liang Wang, Weiming Hu, and Tieniu Tan. Recent developments in human motion analysis. *Pattern Recognition*, 36(3):585–601, March 2003.

[115] Liang Wang, Huazhong Ning, Weiming Hu, and Tieniu Tan. Gait recognition based on procrustes shape analysis. In *Proceedings of International Conference on Image Processing*, volume 3, pages 433–436, 2002.

[116] Liang Wang, Huazhong Ning, Tieniu Tan, and Weiming Hu. Fusion of static and dynamic body biometrics for gait recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(2):149–158, February 2004.

[117] Liang Wang, Tieniu Tan, Weiming Hu, and Huazhong Ning. Automatic gait recognition based on statistical shape analysis. *IEEE Transactions on Image Processing*, 12(9):1120–1131, September 2003.

[118] CR Wren, A. Azarbayejani, T. Darrell, and AP Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780–785, 1997.

[119] Jianning Wu, Jue Wang, and Li Liu. Kernel-based method for automated walking patterns recognition using kinematics data. In *Proceedings of Second International Conference on Advances in Natural Computation*, volume 4222 of *Lecture Notes in Computer Science*, pages 560–569, 2006.

[120] Chew-Yean Yam, Mark S. Nixon, and John N. Carter. Extended model-based automatic gait recognition of walking and running. In *Proceedings of Third International Conference on Audio- and Video-Based Biometric Person Authentication*, volume 2091 of *Lecture Notes in Computer Science*, pages 278–283. Springer-Verlag, June 2001.

[121] Chew-Yean Yam, Mark S. Nixon, and John N. Carter. Automated person recognition by walking and running via model-based approaches. *Pattern Recognition*, 37(5):1057–1072, 2004.

[122] Shiqi Yu, Daoliang Tan, and Tieniu Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *International Conference on Pattern Recognition*, volume 4, pages 441–444, 2006.

[123] R. Zhang, P. S. Tsai, J. E. Cryer, and M. Shah. Shape-from-shading: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8): 690–706, 1999.

[124] Guoying Zhao, Rui Chen, Guoyi Liu, and Hua Li. Amplitude spectrum-based gait recognition. In *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 23–28, May 2004.

[125] Guoying Zhao, Guoyi Liu, Hua Li, and Matti Pietikäinen. 3d gait recognition using multiple cameras. In *Proceedings of the Seventh IEEE International Conference on Automatic Face and Gesture Recognition (FG '06)*, pages 529–534, Los Alamitos, CA, USA, 2006. IEEE Computer Society. ISBN 0-7695-2503-2.

[126] W. Zhao, Rama Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003. ISSN 0360-0300.

Appendix A

Evaluation of Original System by Middleton et al.

A.1 Composition of dataset

Session	Subject	Initial samples	Retained samples
17/08/2006–18/08/2006	0	4	4
	1	4	4
	2	4	0
	3	4	0
	4	4	3
	5	4	4
	6	4	4
	7	4	4
	8	4	4
	9	4	4
31/10/2006–01/11/2006	10	5	0
	11	4	4
	12	2	0
	13	4	0
	14	5	3
	15	2	0
	16	3	0
	17	1	0
	18	4	4
	19	3	0

Table A.1: Composition of dataset collected by Middleton et al. [70] and revised version

A.2 Analysis of initial data

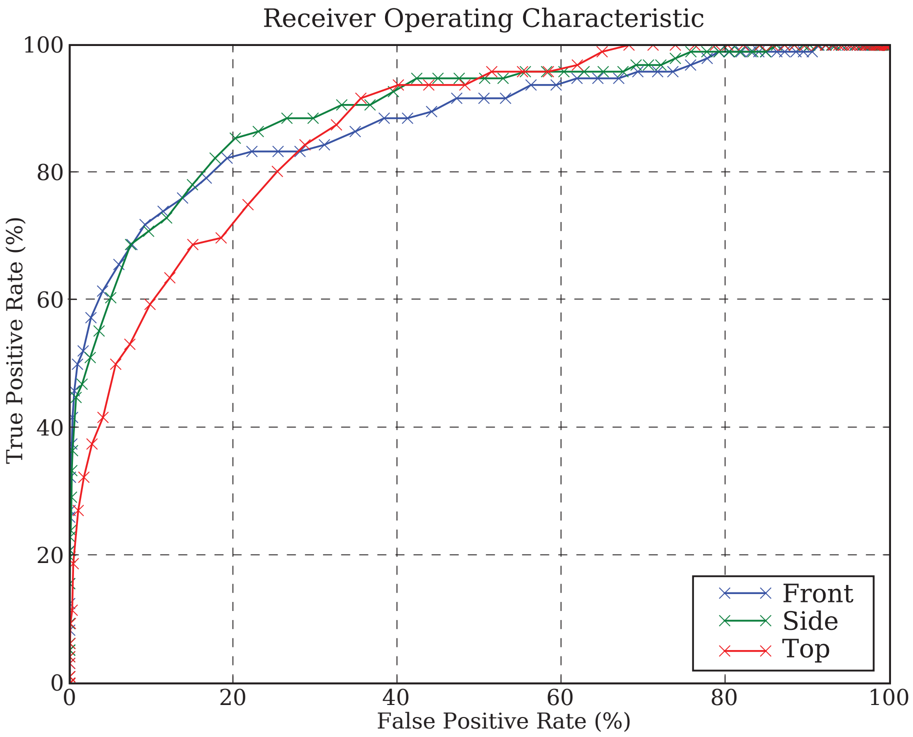


Figure A.1: Gait classification performance of average signature representation using initial dataset

Projection	$k = 1$	$k = 3$	$k = 5$
Top-down	60.0%	51.4%	41.4%
Front-on	75.7%	71.4%	57.1%
Side-on	81.4%	75.7%	61.4%

Table A.2: Performance of gait classification from average silhouette signature

A.3 Analysis of rectified data

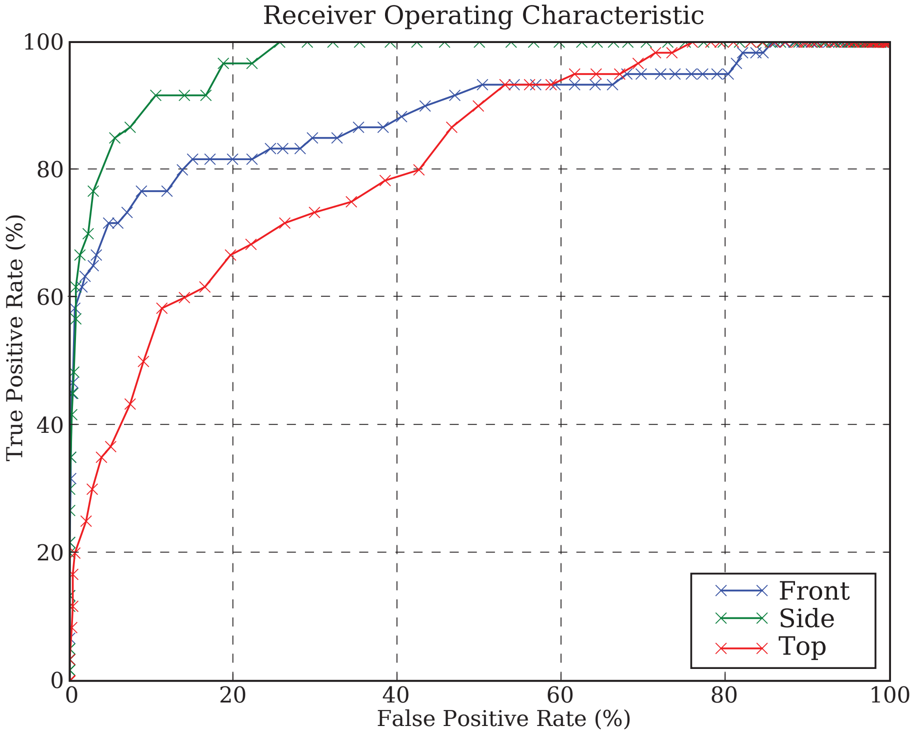


Figure A.2: Gait classification performance of average signature representation using revised initial dataset

Projection	$k = 1$	$k = 3$	$k = 5$
Top-down	66.7%	57.1%	42.9%
Front-on	81.0%	69.0%	66.7%
Side-on	97.6%	88.1%	85.7%

Table A.3: Performance of gait classification using average silhouette from multiple viewpoints; using revised dataset

Appendix B

Analysis of Development Dataset

Session	Subject ID	Number of Samples
29/06/2007	17	4
	19	4
	22	4
	23	4
	24	4
	25	4
	26	4
	20	4
	21	4
	9	3

Table B.1: Composition of testing dataset collected from revised Biometric Tunnel configuration

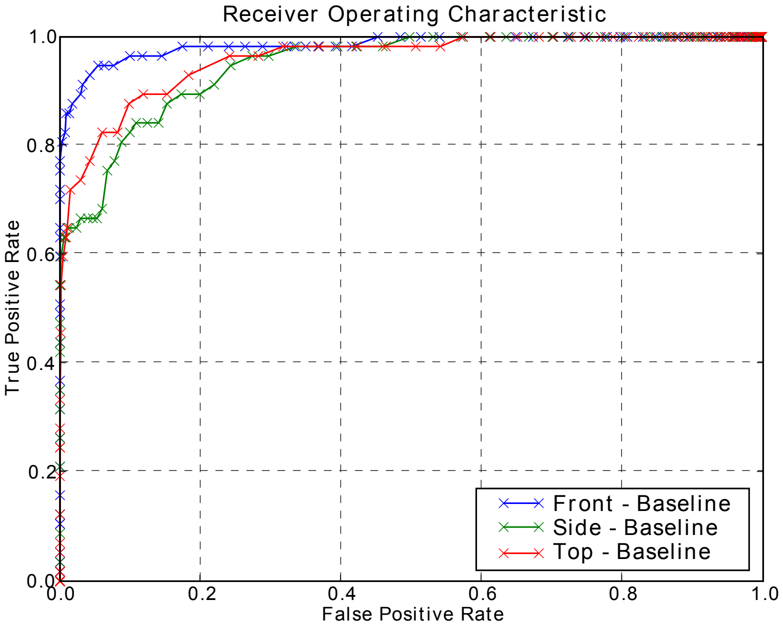


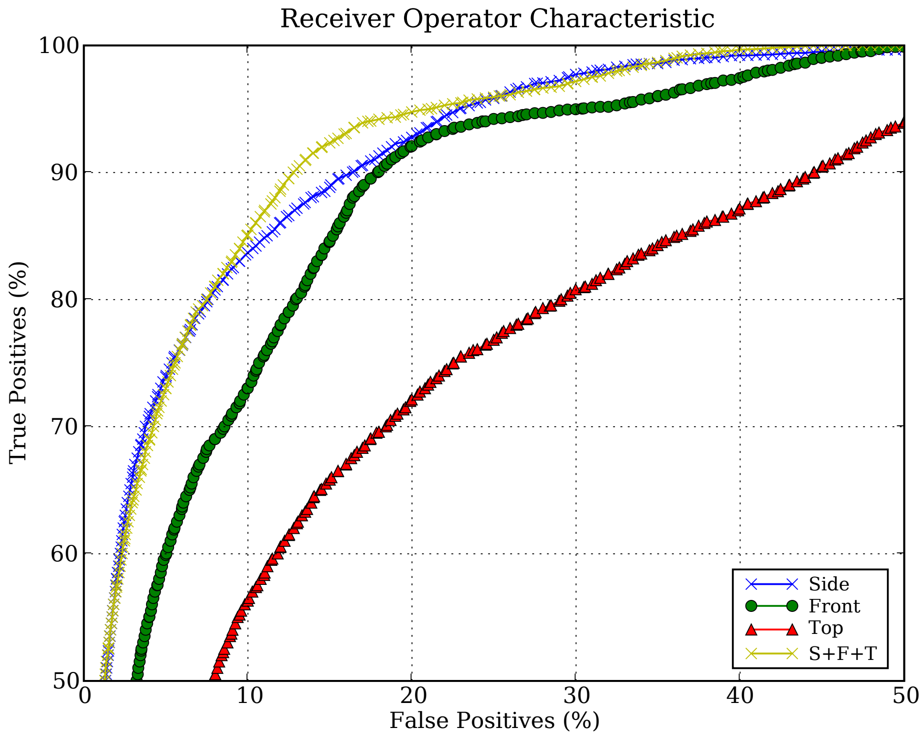
Figure B.1: Receiver Operating Characteristic, using development dataset

Projection	$k = 1$	$k = 3$	$k = 5$
Front-on	100.0%	97.4%	87.2%
Side-on	94.9%	94.9%	74.4%
Top-down	94.9%	94.9%	84.6%

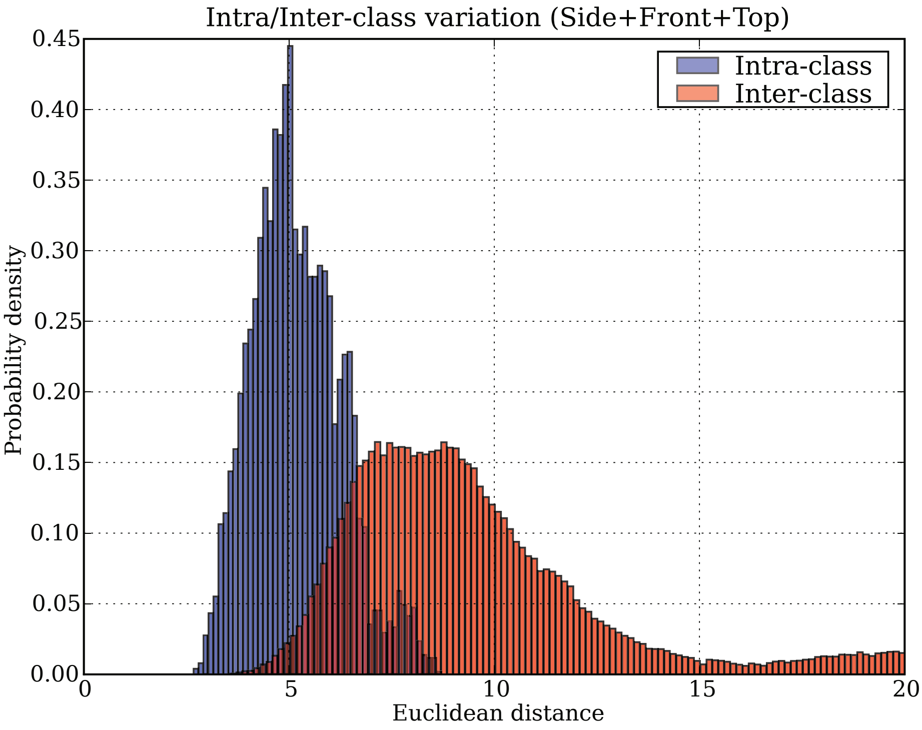
Table B.2: Correct classification rate for various viewpoints, using development dataset

Appendix C

Additional Results from Analysis of Large Multi-Biometric Dataset

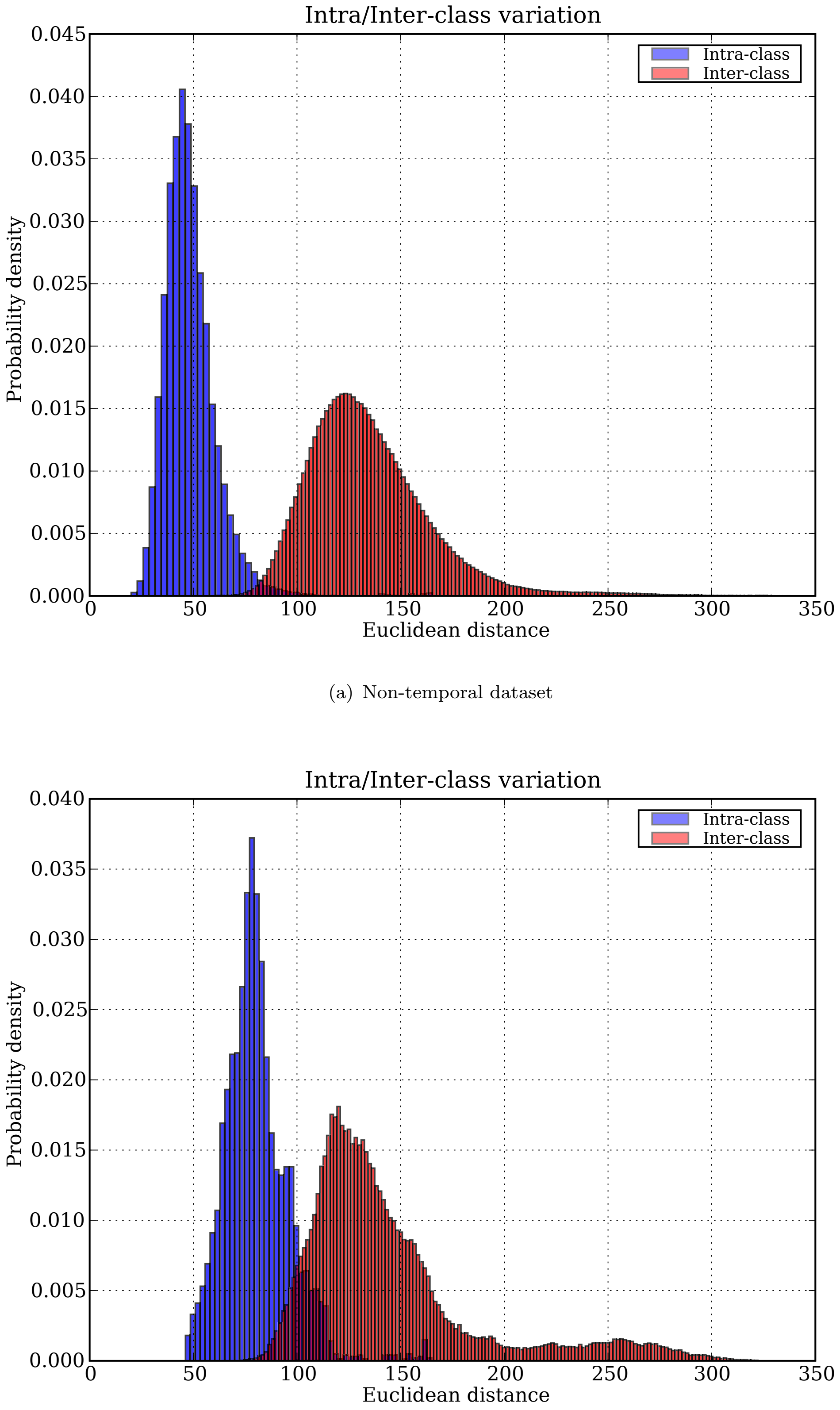


(a) Receiver Operating Characteristic



(b) Intra/Inter-class variation

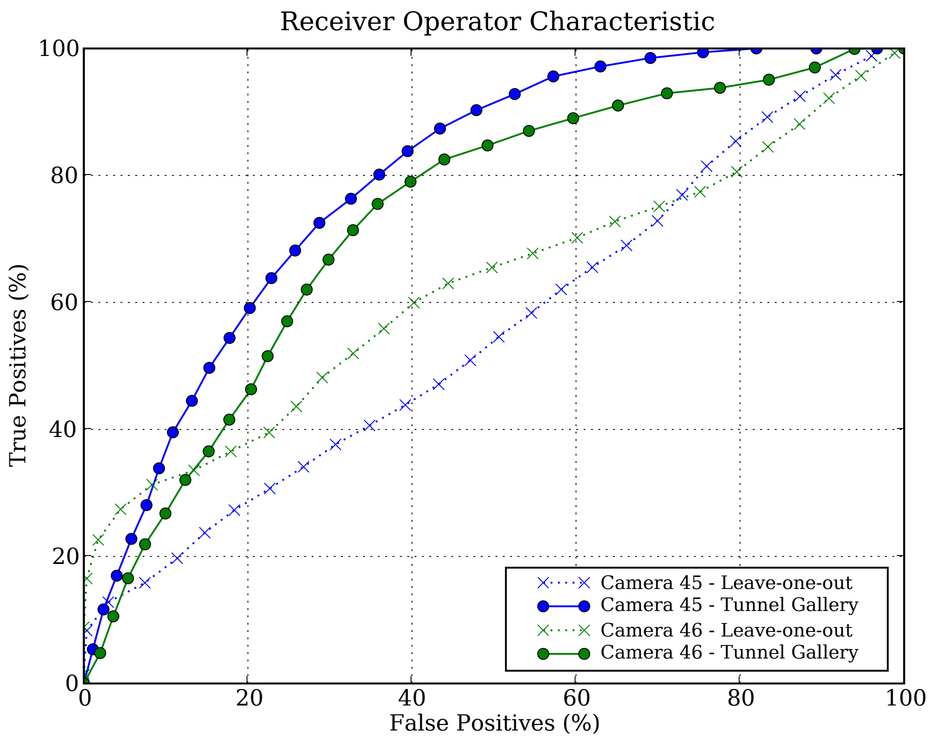
Figure C.1: Results of leave-one-out experiment matching non-normalised average silhouettes against those from a previous date



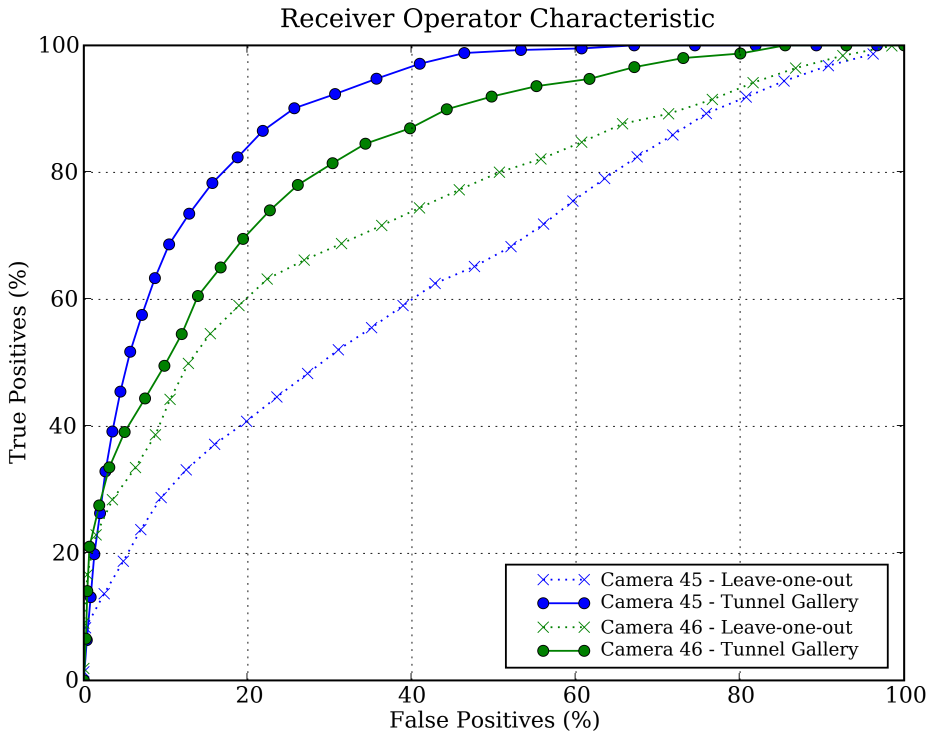
(a) Non-temporal dataset

(b) Temporal dataset

Figure C.2: Intra/Inter-class variation for combination of average face, colour histogram and gait



(a) Scale-normalised average silhouette



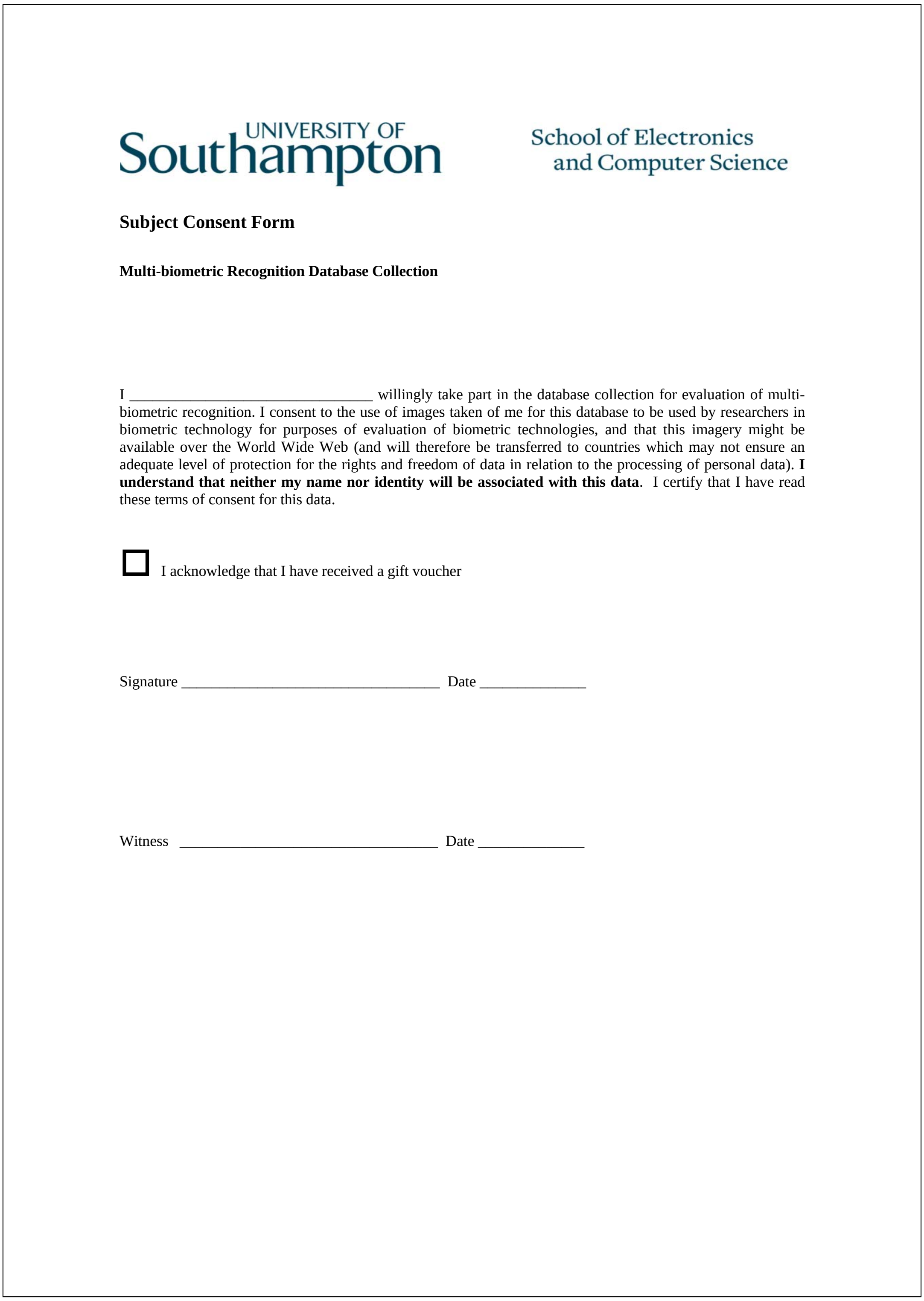
(b) Non-normalised average silhouette

Figure C.3: Receiver Operating Characteristic curves for reprojection matching experiments

Appendix D

Administrative Paperwork for Large Multi-Biometric Dataset

D.1 Subject Consent Form



UNIVERSITY OF
Southampton

School of Electronics
and Computer Science

Subject Consent Form

Multi-biometric Recognition Database Collection

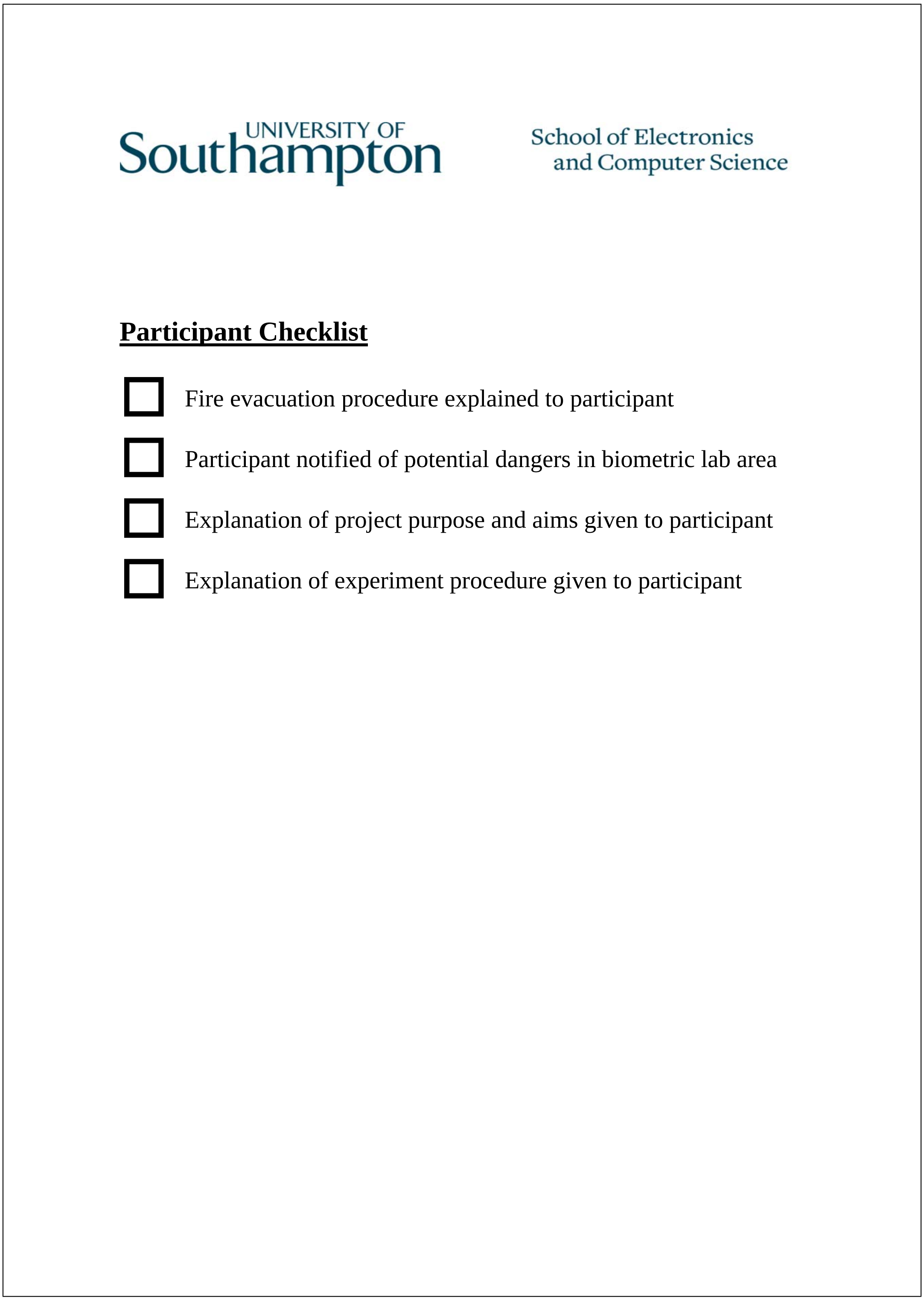
I ________________________________ willingly take part in the database collection for evaluation of multi-biometric recognition. I consent to the use of images taken of me for this database to be used by researchers in biometric technology for purposes of evaluation of biometric technologies, and that this imagery might be available over the World Wide Web (and will therefore be transferred to countries which may not ensure an adequate level of protection for the rights and freedom of data in relation to the processing of personal data). **I understand that neither my name nor identity will be associated with this data**. I certify that I have read these terms of consent for this data.

□ I acknowledge that I have received a gift voucher

Signature __________________________________ Date ______________

Witness __________________________________ Date ______________

Front



UNIVERSITY OF Southampton

School of Electronics
and Computer Science

Participant Checklist

□ Fire evacuation procedure explained to participant

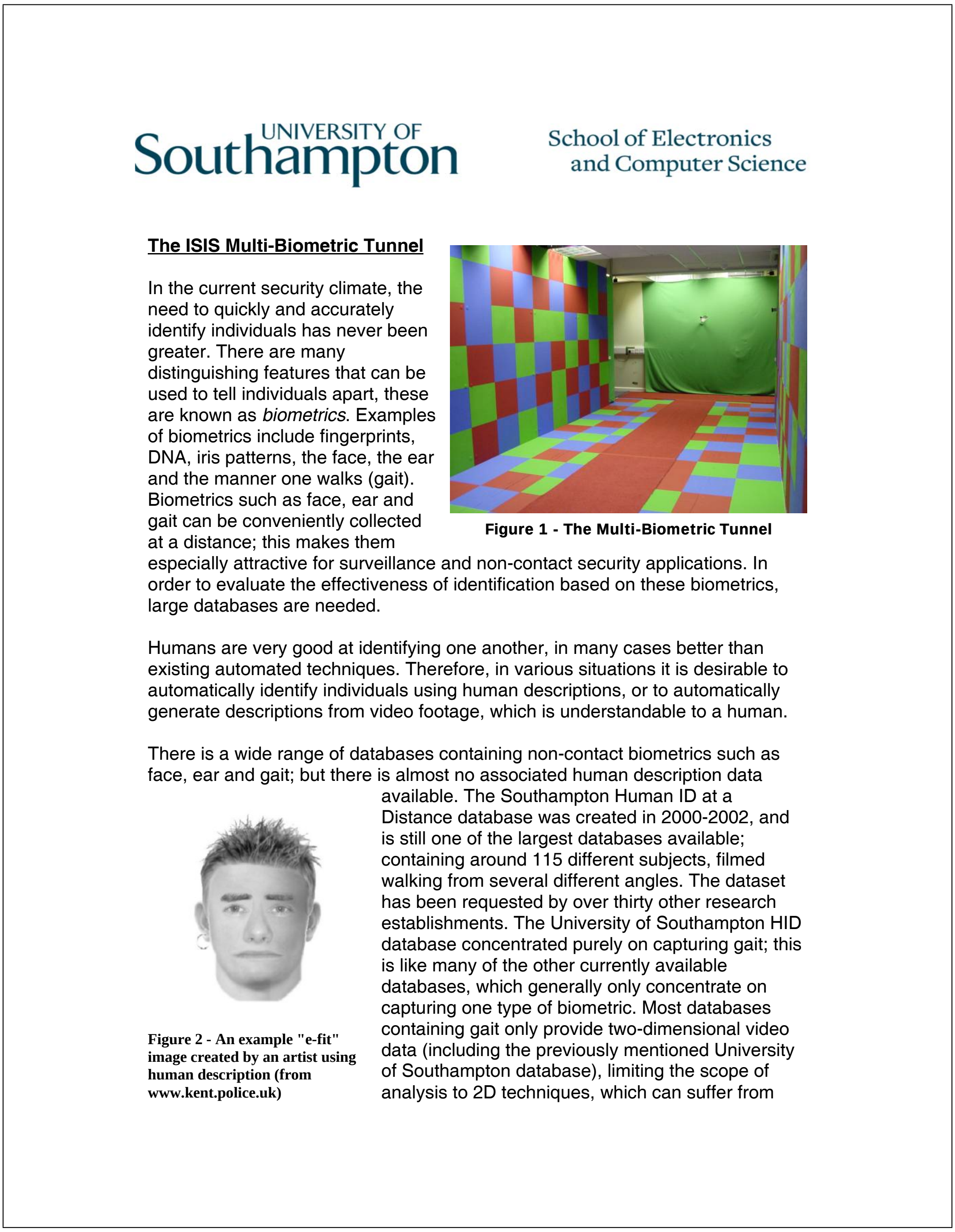
□ Participant notified of potential dangers in biometric lab area

□ Explanation of project purpose and aims given to participant

□ Explanation of experiment procedure given to participant

Reverse

D.2 Project Information Sheet



UNIVERSITY OF Southampton

School of Electronics
and Computer Science

The ISIS Multi-Biometric Tunnel

Figure 1 - The Multi-Biometric Tunnel

In the current security climate, the need to quickly and accurately identify individuals has never been greater. There are many distinguishing features that can be used to tell individuals apart, these are known as *biometrics*. Examples of biometrics include fingerprints, DNA, iris patterns, the face, the ear and the manner one walks (gait). Biometrics such as face, ear and gait can be conveniently collected at a distance; this makes them especially attractive for surveillance and non-contact security applications. In order to evaluate the effectiveness of identification based on these biometrics, large databases are needed.

Humans are very good at identifying one another, in many cases better than existing automated techniques. Therefore, in various situations it is desirable to automatically identify individuals using human descriptions, or to automatically generate descriptions from video footage, which is understandable to a human.

There is a wide range of databases containing non-contact biometrics such as face, ear and gait; but there is almost no associated human description data available. The Southampton Human ID at a Distance database was created in 2000-2002, and is still one of the largest databases available; containing around 115 different subjects, filmed walking from several different angles. The dataset has been requested by over thirty other research establishments. The University of Southampton HID database concentrated purely on capturing gait; this is like many of the other currently available databases, which generally only concentrate on capturing one type of biometric. Most databases containing gait only provide two-dimensional video data (including the previously mentioned University of Southampton database), limiting the scope of analysis to 2D techniques, which can suffer from

Figure 2 - An example "e-fit" image created by an artist using human description (from www.kent.police.uk)

Front

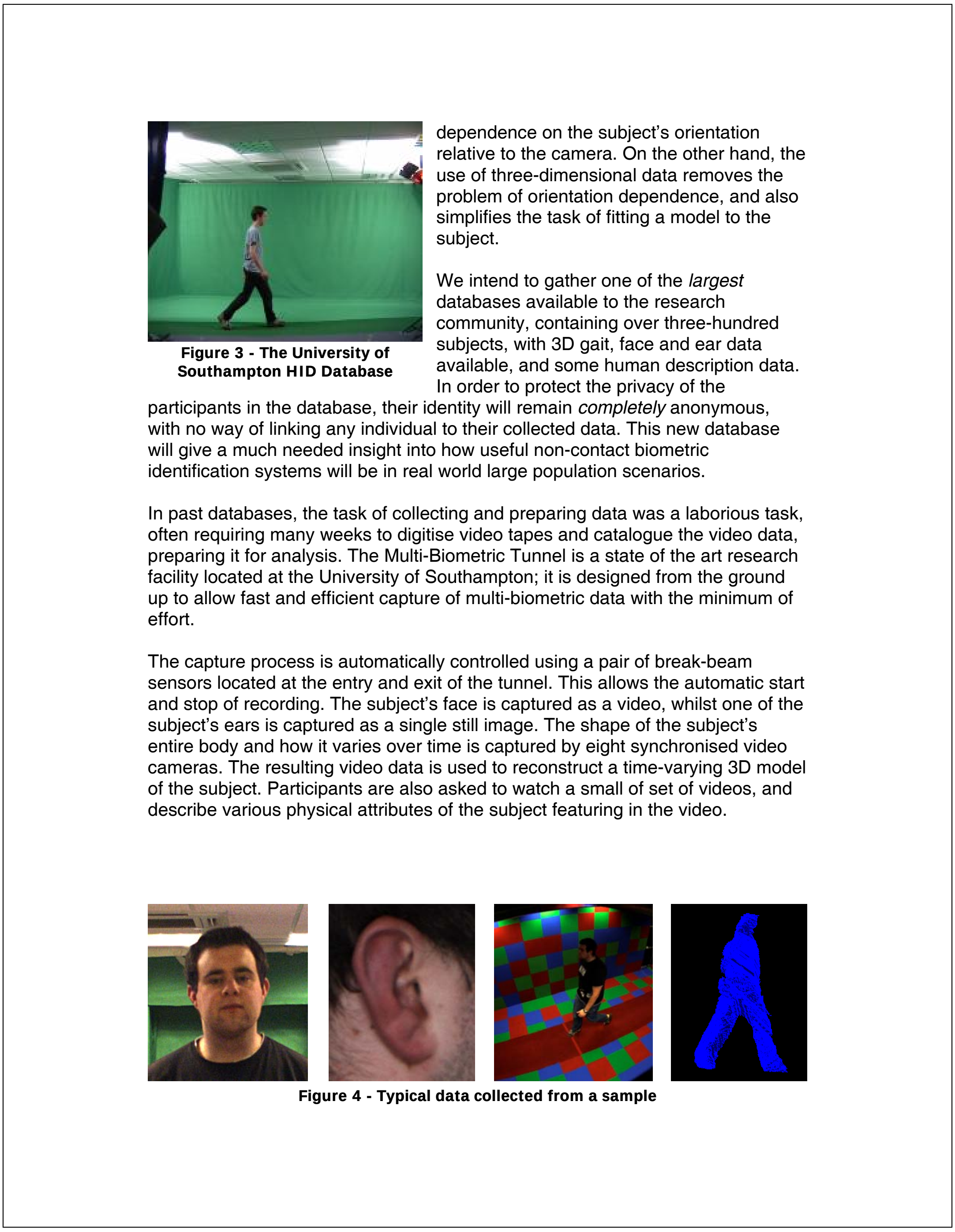


Figure 3 - The University of Southampton HID Database

dependence on the subject's orientation relative to the camera. On the other hand, the use of three-dimensional data removes the problem of orientation dependence, and also simplifies the task of fitting a model to the subject.

We intend to gather one of the *largest* databases available to the research community, containing over three-hundred subjects, with 3D gait, face and ear data available, and some human description data. In order to protect the privacy of the participants in the database, their identity will remain *completely* anonymous, with no way of linking any individual to their collected data. This new database will give a much needed insight into how useful non-contact biometric identification systems will be in real world large population scenarios.

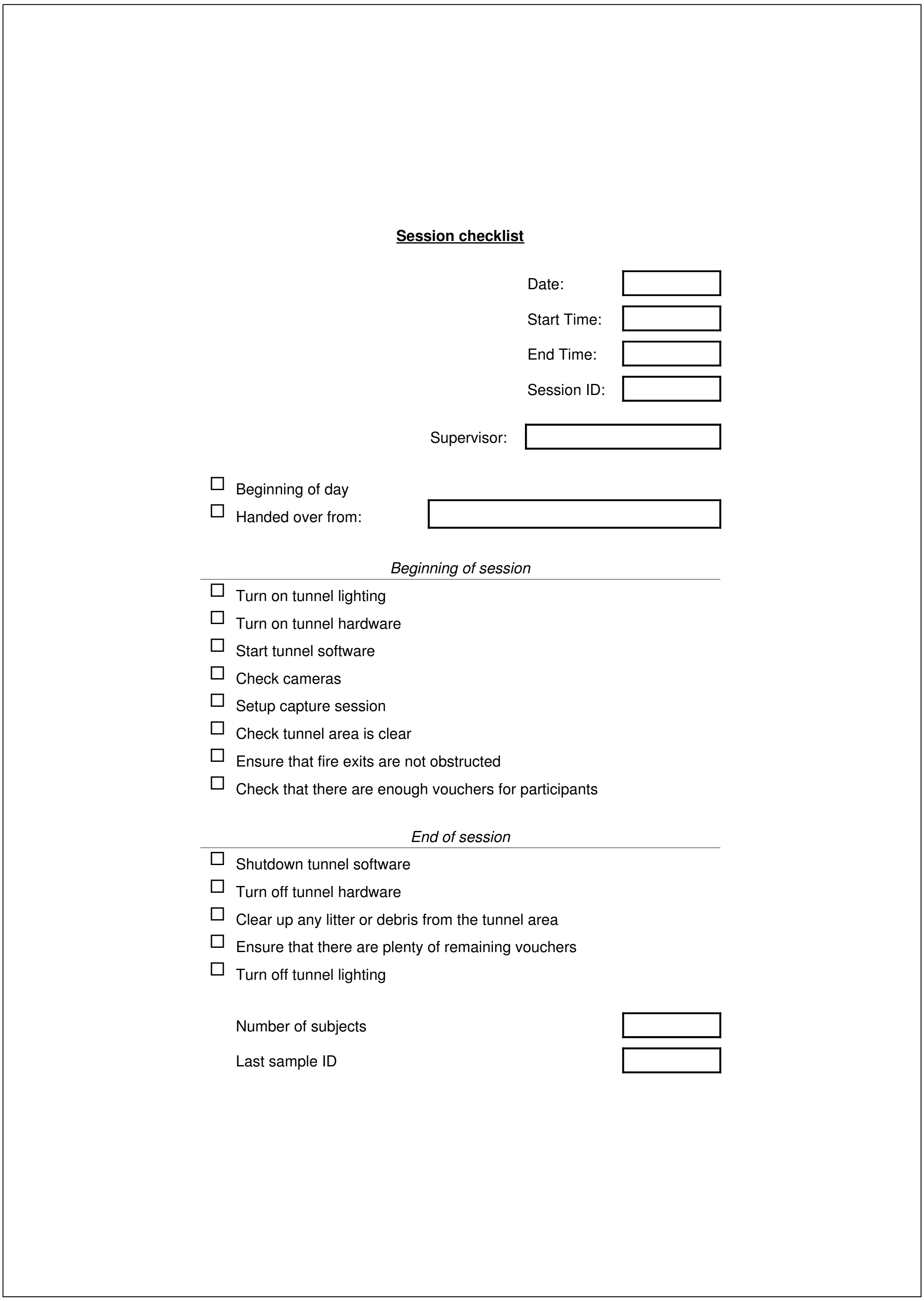
In past databases, the task of collecting and preparing data was a laborious task, often requiring many weeks to digitise video tapes and catalogue the video data, preparing it for analysis. The Multi-Biometric Tunnel is a state of the art research facility located at the University of Southampton; it is designed from the ground up to allow fast and efficient capture of multi-biometric data with the minimum of effort.

The capture process is automatically controlled using a pair of break-beam sensors located at the entry and exit of the tunnel. This allows the automatic start and stop of recording. The subject's face is captured as a video, whilst one of the subject's ears is captured as a single still image. The shape of the subject's entire body and how it varies over time is captured by eight synchronised video cameras. The resulting video data is used to reconstruct a time-varying 3D model of the subject. Participants are also asked to watch a small of set of videos, and describe various physical attributes of the subject featuring in the video.

Figure 4 - Typical data collected from a sample

Reverse

D.3 Session Checklist



Session checklist

Date:

Start Time:

End Time:

Session ID:

Supervisor:

□ Beginning of day

□ Handed over from:

Beginning of session

□ Turn on tunnel lighting

□ Turn on tunnel hardware

□ Start tunnel software

□ Check cameras

□ Setup capture session

□ Check tunnel area is clear

□ Ensure that fire exits are not obstructed

□ Check that there are enough vouchers for participants

End of session

□ Shutdown tunnel software

□ Turn off tunnel hardware

□ Clear up any litter or debris from the tunnel area

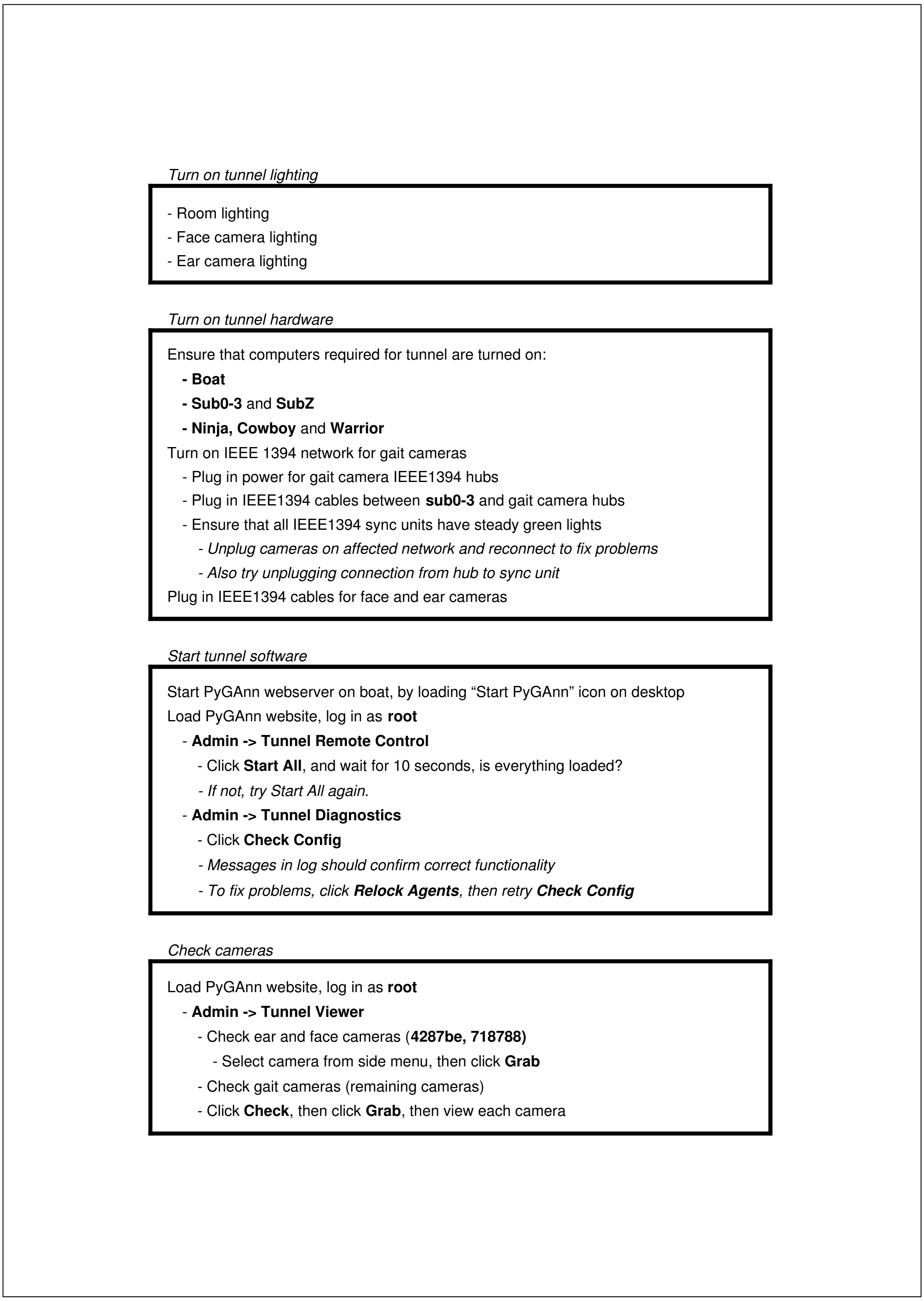
□ Ensure that there are plenty of remaining vouchers

□ Turn off tunnel lighting

Number of subjects

Last sample ID

D.4 Instructions for Session



Turn on tunnel lighting

- Room lighting
- Face camera lighting
- Ear camera lighting

Turn on tunnel hardware

Ensure that computers required for tunnel are turned on:

- **Boat**
- **Sub0-3** and **SubZ**
- **Ninja, Cowboy** and **Warrior**

Turn on IEEE 1394 network for gait cameras

- Plug in power for gait camera IEEE1394 hubs
- Plug in IEEE1394 cables between **sub0-3** and gait camera hubs
- Ensure that all IEEE1394 sync units have steady green lights
    - *Unplug cameras on affected network and reconnect to fix problems*
    - *Also try unplugging connection from hub to sync unit*

Plug in IEEE1394 cables for face and ear cameras

Start tunnel software

Start PyGAnn webserver on boat, by loading “Start PyGAnn” icon on desktop

Load PyGAnn website, log in as **root**

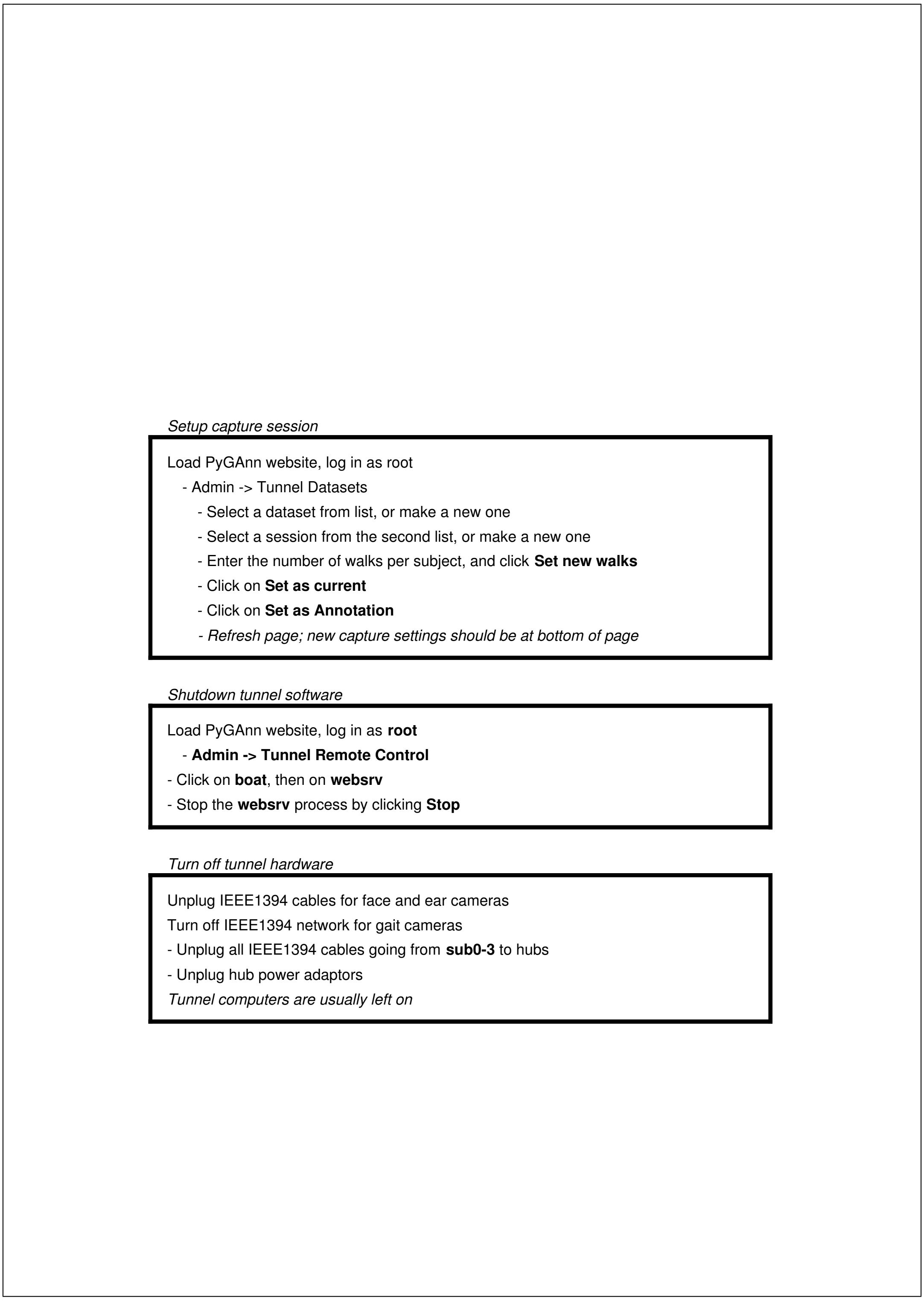
- **Admin -> Tunnel Remote Control**
    - Click **Start All**, and wait for 10 seconds, is everything loaded?
    - *If not, try Start All again.*
- **Admin -> Tunnel Diagnostics**
    - Click **Check Config**
    - *Messages in log should confirm correct functionality*
    - *To fix problems, click **Relock Agents**, then retry **Check Config***

Check cameras

Load PyGAnn website, log in as **root**

- **Admin -> Tunnel Viewer**
    - Check ear and face cameras (**4287be, 718788)**
        - Select camera from side menu, then click **Grab**
    - Check gait cameras (remaining cameras)
    - Click **Check**, then click **Grab**, then view each camera

Front



Setup capture session

Load PyGAnn website, log in as root
- Admin -> Tunnel Datasets
  - Select a dataset from list, or make a new one
  - Select a session from the second list, or make a new one
  - Enter the number of walks per subject, and click **Set new walks**
  - Click on **Set as current**
  - Click on **Set as Annotation**
  - *Refresh page; new capture settings should be at bottom of page*

Shutdown tunnel software

Load PyGAnn website, log in as **root**
- **Admin -> Tunnel Remote Control**

- Click on **boat**, then on **websrv**
- Stop the **websrv** process by clicking **Stop**

Turn off tunnel hardware

Unplug IEEE1394 cables for face and ear cameras

Turn off IEEE1394 network for gait cameras
- Unplug all IEEE1394 cables going from **sub0-3** to hubs
- Unplug hub power adaptors

Tunnel computers are usually left on

Reverse

Appendix E

Installation of the Tunnel Software

E.1 Introduction

In this chapter the installation of the Biometric Tunnel system is discussed; it is assumed that the reader has a reasonable level of competency using a Linux based system. These instructions have been written for a Redhat RPM/YUM based distribution, although installing on Debian based distributions (such as Ubuntu) should not prove too difficult — except that `apt-get` or `synaptics` is used instead and package names will differ slightly. Before starting the installation of the tunnel software, ensure the Linux distribution is correctly configured, with the correct accelerated graphics drivers installed. It is recommended that security features such as SELinux and any firewalls are initially disabled whilst installing the system; these can be enabled after installation and then configured to allow correct access. A location for the tunnel data on the computer's filesystem should be decided upon (`/data` is recommended) and prepared to allow access:

```
[user@computer ~]$ su -c 'mkdir /data'
[user@computer ~]$ chown user:user /data
[user@computer ~]$ cd /data
[user@computer ~]$ mkdir analysis datasets misc samples sessions
```

E.2 Base Packages

First we must install a range of "base" packages on the target system; these packages are needed for compiling and running the various subsystems of the tunnel software.

This is done using **yum** on a Fedora/Redhat system. The system used with this document was Fedora 11 x86_64.

```
[user@computer ~]$ su -c 'yum install ipython numpy scipy Cyrex
    mod_python gcc gcc-c++ cmake zlib-devel libpng-devel freeglut-
    devel subversion mysql-devel mysql-server mysql-query-browser
    mysql-administrator MySQL-python'
```

The MySQL server should be started for the first time and configured:

```
[user@computer ~]$ su -c '/etc/init.d/mysqld start'
[user@computer ~]$ su -c 'mysql_secure_installation'
[user@computer ~]$ mysql-administrator
```

From the graphical user interface, create a database (schema) for the tunnel, this is usually called *tunnel*. Also a user must be created for accessing the database. Grant the user full privileges over the newly created database. We will populate the database with tables and data later; in section E.6.

E.3 The Vis4D Package

The Vis4D code contains a C++ library for handling images and volumetric data along with a suite of applications for 3D reconstruction, manipulation of the 3D data and also viewing the data. The code can be either obtained from an archive on the DVD or from the ECS subversion server.

E.3.1 Copying from DVD sources

To obtain from DVD:

```
[user@computer ~]$ cp -rv /mnt/cdrom/Vis4D .
```

E.3.2 Downloading from subversion

To obtain from subversion:

```
[user@computer ~]$ svn co svn+ssh://ecsuser@forge.ecs.soton.ac.uk/
    projects/rds06r/Vis4D
```

E.3.3 Compiling/installing the base library

As many of the tools in the Vis4D suite require the library, it must be built and installed first. The library uses the CMake system to simplify the install process. Here we will use the ccmake tool, which provides a user interface

```
[user@computer ~]$ cd ~/Vis4D/lib/
[user@computer lib]$ mkdir build
[user@computer lib]$ cd build
[user@computer build]$ ccmake ..
```

Once the ccmake program has loaded, press **c** to configure the make process. Hopefully no serious errors or warnings should be displayed. Exit the messages by pressing **e**, now set *CMAKE_BUILD_TYPE* to *Release*. Also set all entries starting with *TUNNEL_* to the correct values. Press **c** again to reconfigure and exit the messages again. In order to get the best performance out of the library, it is recommended that you set the advanced build settings to reflect your computer's processor. This is achieved by pressing **t** to toggle the advanced settings. Then add *-march=core2* to the *CMAKE_CXX_FLAGS_RELEASE* and *CMAKE_C_FLAGS_RELEASE* fields (replace *core2* with the most appropriate type for your system — see the gcc manual for more information). Finally, press **c** to reconfigure, exit the info screen with **e**, then press **g** to generate the make files. Now compile and install the library:

```
[user@computer build]$ make
[user@computer build]$ su -c 'make install'
```

E.3.4 Compiling/installing the viewer

The viewer can be compiled in a similar manner to the main library; using the CMake system.

```
[user@computer ~]$ cd ~/Vis4D/viewer
[user@computer viewer]$ mkdir build
[user@computer viewer]$ cd build
[user@computer build]$ ccmake ..
```

The CMake configuration process is the same as in section E.3.3, making sure that *CMAKE_BUILD_TYPE* is set to *Release*. Generate the make files by pressing **g**. Now compile and install:

```
[user@computer build]$ make
```

```
[user@computer build]$ su -c 'make install'
```

E.3.5 Compiling/installing the other tools

The Vis4D suite contains a variety of tools for processing the 3D data and performing 3D reconstruction. Most of these tools reside in the apps sub-folder, whilst the reconstruction code is in its own folder. First we will build the small applications:

```
[user@computer ~]$ cd ~/Vis4D/apps
[user@computer viewer]$ mkdir build
[user@computer viewer]$ cd build
[user@computer build]$ ccmake ..
```

Configure and generate the makefiles, ensuring that all *TUNNEL_* variables are correct and that the *CMAKE_BUILD_TYPE* is set to *Release*. Adding *-march=XXX* to the compiler flag entries will also help to improve performance. Now make and install:

```
[user@computer build]$ make
[user@computer build]$ su -c 'make install'
[user@computer ~]$ cd ~/Vis4D/reconstruction
[user@computer viewer]$ mkdir build
[user@computer viewer]$ cd build
[user@computer build]$ ccmake ..
```

Configure and generate the makefiles in the same manner as above.

```
[user@computer build]$ make
[user@computer build]$ su -c 'make install'
```

E.4 Installing the tunnel toolchain

The tunnel toolchain is formed by a variety of different applications, and is mostly written in Python. It's primary purpose is to control the capture of data from the tunnel and to automate the processing of the data. The toolchain can either be decompressed from the DVD, or downloaded from the ECS subversion server.

E.4.1 Copying from DVD sources

To obtain from DVD:

```
[user@computer ~]$ cp -rv /mnt/cdrom/Tunnel-Toolchain .
```

E.4.2 Downloading from subversion

To obtain from subversion:

```
[user@computer ~]$ svn co svn+ssh://ecsuser@forge.ecs.soton.ac.uk/
    projects/rds06r/Tunnel-Toolchain
```

E.4.3 Building and installing the Python package

One of the toolchain's core components is the Python **Tunnel** package; this contains a variety of modules for database access, data processing, calibration and much more. The toolchain contains a mixture of pure python modules, Cython modules and hand-written C modules. The package can be built using the standard python distutils setup.py file:

```
[user@computer ~]$ cd ~/Tunnel-Toolchain/Python-Package
[user@computer Python-Package]$ ./setup.py build
```

Upon running the setup.py script, a variety of configuration questions will be asked, such as the MySQL server location (*localhost* if on the same computer), username and password, along with the location to save the tunnel data. Once all questions have been answered and the build is complete, install the package:

```
[user@computer Python-Package]$ su -c './setup.py install'
```

E.5 Installing the web-interface

The tunnel toolchain's web-interface provides a powerful way of controlling the processing of data, developing new processing algorithms and performing recognition experiments. The web-interface requires an apache webserver, with the mod_python extensions loaded. It is recommended the SELinux is disabled, as it can make the running of the web-interface problematic. The web-interface is located in *Tunnel-Toolchain/v4/ProcessSite*. Either copy the web-interface directory into apache's document folder, or modify the apache configuration to access the site from its current location.

E.6 Populating the tunnel toolchain with data

The tunnel toolchain can import data from two sources; either from a CD/DVD or by downloading the data from *boat.ecs.soton.ac.uk*. A special script *InstallDB.py* can be used to automate the process. Firstly, the script checks whether any of the tables need creating. Secondly, the script downloads the database skeleton file from the webserver *boat.ecs.soton.ac.uk* (or off the DVD/CD). Finally, the script will copy any samples data off the DVD/CD and install them into the toolchain. If using a DVD or CD as the data-source, please ensure that it is mounted in a subfolder under either /mnt or /media; the script will check these locations for the data-files.

```
[user@computer ~]$ cd ~/Tunnel-Toolchain/v5/Utils
[user@computer Utils]$ ./InstallDB.py
```

If you would like to download samples from the webserver *boat.ecs.soton.ac.uk*, this can be achieved by adding the *Sample Downloader* task for the corresponding samples. The process of running tasks in the toolchain is described in section G.

Appendix F

Processing Data Using the Toolchain

F.1 Introduction

The toolchain has the ability to automate the processing of samples — this is done by writing *tasks*; small Python scripts designed to run within a special environment provided by the toolchain. The tasks and the corresponding code is all saved in the tunnel's MySQL database under the *tasks* table. The *task_queue* table manages the execution of the tasks, where each row specifies a task to be run, the sample to be run on (optional), the priority, additional data needed by the task and also the execution status of the task. Most tasks are either written to be run on a per sample basis, or as a standalone task. The web-interface provides extensive facilities for creating, editing and debugging tasks. Also, the Tunnel Python package provides a module called DummyEnvironment, which creates an environment almost identical to that of the toolchain; this is useful for developing under IPython. Below is an example of a simple task, which loads all the colour face images (found by the face-finder task), resizes them to 32 × 32 and calculates the average; the result is then saved as a feature.

```
# Average Face
import numpy
import Image

UpdateStatus("Fetching metadata")
imageFiles = GetFiles("face image")

UpdateStatus("Calculating Average Face")
InitProgress(len(imageFiles))
avgImage = numpy.zeros( (32, 32, 4), 'f' )
for imageFile in imageFiles:
    pilim = Image.open( imageFile.Location )
    pilim = pilim.resize( ( 32, 32 ) )
    fv = numpy.fromstring( pilim.tostring(), 'u1' ).reshape( 32, 32, 4 ) / 255.0
    avgImage += fv * fv[:,:,3:4]
    UpdateProgress()

mask = avgImage[:,:,3:4] > 5 # threshold above zero to ignore noise
avgImage /= avgImage[:,:,3:4] + 0.0000001 # add a small constant to avoid NaN
avgImage = (avgImage * mask) + ((1 - mask) * numpy.array([0, 0, 1, 1.0])) # add a background colour to unoccupied regions
avgImage = numpy.array( avgImage[:,:,:3], copy=True ) # remove the alpha channel
SetFeature( "Average Face", "sensors", 101, avgImage )
```

In order to run these tasks outside of the tunnel toolchain (such as in IPython), the following two lines must precede any task code:

```
from Tunnel.DummyEnvironment import *
InitEnvironment(Sample_ID=1000)
```

The tunnel toolchain task execution environment provides a range of variables and functions for manipulating metadata within the database, simple threading, progress feedback and handling the execution of tasks.

F.2 Variables/Objects

TQ_ID The ID of the (task,sample) pair in the task_queue

Task_ID The ID of the task code

Sample_ID The ID of the sample being operated on

Py_Args Additional parameters passed to the task; often used for analysis tasks. Py_Args is usually a dictionary of variable name, value pairs. These variables are also imported into the name space of the task, allowing direct access to the variables.

Version What version of the tunnel toolchain is being used

DB An object containing an open connection to the MySQL database (DB.db) and a cursor object (DB.c). DB.db.commit() is called after successful completion of the task

F.3 Classes/Objects

F.3.1 Success

This is an exception class, which can be raised to stop execution of the task and mark it as successful

```
raise Success(”a success message”)
```

F.3.2 Error

This is an exception class, which can be raised to stop execution of the task and mark it as unsuccessful

```
raise Error(”a success message”)
```

F.3.3 File

Constructor: `File(ID=`*`None`*`, Location=`*`None`*`, Type=`*`None`*`, Format=`*`None`*`, Frame=`*`None`*`, Comment=`*`None`*`, Sensors=`*`None`*`)`

File attributes:

ID The ID of the file as used in the database (not needed for new files)

Location The location of the file on the filesystem

Type A string giving the type of the file

Format A string giving the format of the file (ie png or jpeg)

Frame The frame number of the file (optional)

Comment (optional)

Sensors A list containing the IDs of the file’s associated sensors (optional)

F.3.4 Job

Constructor: `Job(Function, arg1, arg2, ...)`

```
j=Job(shutil.copy, ”/tmp/afile”, ”/tmp/bfile”)
```

Constructor: `Job(Program_Name, arg1, arg2, ...)`

```
j=Job("cp", "/tmp/afile", "/tmp/bfile")
```

The job class is used to define a single operation to be done, it is used in conjunction with the Jobs class

F.3.5 Jobs

Constructor: `Jobs(jobs_list)`

The Jobs class is used for executing a number of repetetive tasks, it automatically makes use of multiple threads if the executing process client allows and also provide progress updates whilst running. The Jobs class is a sub-class of the standard Python list class.

F.3.5.1 Jobs functions

append(item) Appends an item to the Jobs list. The item should be an instance of the Job class.

run() Runs all the jobs in the list, returns a list containing the return values of each job

raise_errors() Raises an exception if any of the jobs failed

F.3.5.2 Jobs example

```
jobs = Jobs( [ Job(shutil.copy, "/somewhere/%d". "/else/%d" % (i,i)) for i in range(10) ] )
jobs.run()
jobs.raise_errors()

def a_func(n):
    time.sleep(1)
    return n*n + n + 1

jobs = Jobs( [ Job( a_func, i ) for i in range(10) ] )
rets = jobs.run()
# rets = [ 1, 3, 7, 13, 21, 31, 43, 57, 73, 91 ]
jobs.raise_errors()
```

F.4 Global Functions

AddFiles(files) Adds files to database - where files is a list of File objects.

SuggestLocation(filename, filetype, sensor=*None*) Constructs a string containing the correct location for a file with name filename, type filetype and associated sensor to be stored.

GetFiles(types) Retrieves a list of File objects from the database for the current sample, returned files will all be of types contained in list types, or a single type given as a string.

DeleteFiles(types) Deletes files from disk and from DB, list files contains File objects to be deleted.

GroupFilesBySensor(files) Groups the given list of File objects by sensor ID, and returns a dictionary mapping (sensor =¿ sensor_files)

GroupFilesByFrame(files) Groups the given list of File objects by frame index, and returns a dictionary mapping (frame =¿ frame_files)

SortFilesByFrame(files) Sorts the given list of File objects by frame index, returns sorted list of File objects

GetAttribute(attr) Get the first value of the given attribute

GetAttributes(attr) Gets all values for the given attribute

UpdateStatus(status) Updates the status of the running task to the given string

InitProgress(maxsteps, reset=*False*) Initialises the progress counter for the current task, if reset is *True* then the progress is reset to 0

UpdateProgress(nsteps=*1*) Increment the progress counter

SetFeature(Feature_Name, Source, Source_ID, Value, Feature_Help=””) Updates or creates a feature for the current sample, where Source is the name of the source table and Source_ID is the ID of the feature generator in the source table. Value may either be a list of values (for a vector) or a string or number.

GetFeature(Feature_Name, Source=*None*, Source_ID=*None*, Sample_ID=*None*) Retrieves single or multiple features of Feature_Name for Sample_ID (will default to the task's associated Sample ID). If Source and Source_ID are specified, then the single respective value is returned if found, or *None* otherwise. If Source_ID or Source is not specified, then a list of dictionary objects are returned, containing the missing source and source_id data along with the value.

SetAnalysisResult(Result_Name, Result_Value) Saves a result for the current analysis task - Result_Value can be a number, list, or numpy array

GetAnalysisResult(Result_Name) Retrieves a previously saved result for the current analysis task

Appendix G

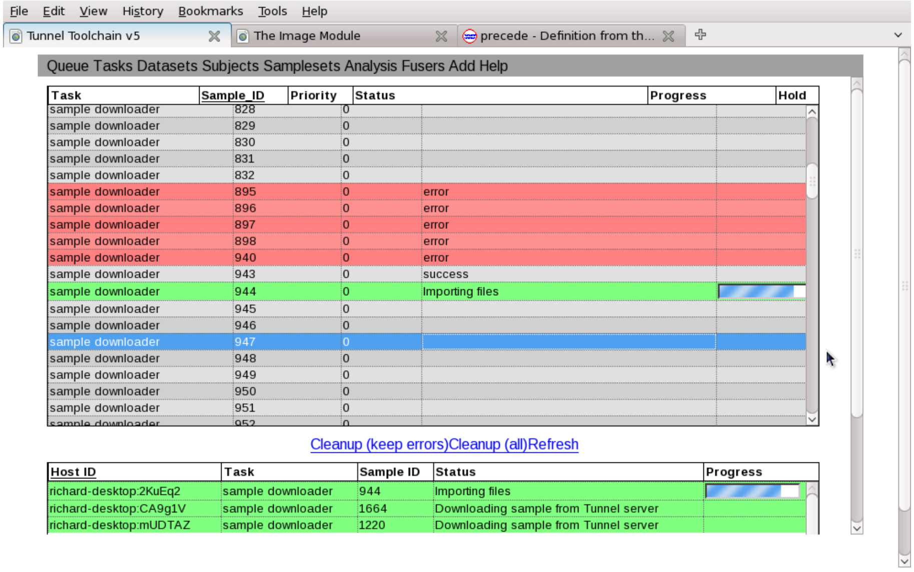
The Web Interface

G.1 Introduction

The Web-interface is a powerful tool allowing the control of processing tasks, the editing and debugging of task code, the managing of datasets, sessions and samples and the control of recognition experiments and other similar analysis tasks. The web-interface is written in Python and runs from Apache using the mod_python extensions. The web-interface requires the tunnel PushChannel to be active, in order to provide real-time processing status. Without the PushChannel running, the web-interface will still function, but no feedback will be given whilst a task is running. The PushChannel server is located in `Tunnel-Toolchain/v5/ProcessConnector/`. Finally, one or more ProcessClients should be running, so that any tasks in the queue can be executed; the client is located in `Tunnel-Toolchain/v5/ProcessClient/`. The default behaviour of the client is to quit once the task_queue becomes empty, this can be prevented with the `--persist` option. The main web-interface is located at `http://boat.ecs.soton.ac.uk/process/`.

G.2 The processing queue

The main page of the web-interface is the processing queue, as shown in Figure G.1. The processing queue shows all the entries in the MySQL *task_queue* table, with information on each entry's status. Clicking the right-mouse button on one of the entries displays a sub menu, allowing you to view the corresponding sample or task code, reprocess the entry, delete it or set its priority. If the task has completed or raised an error, the sub menu also allows you to view the task's standard output and exception data.


File Edit View History Bookmarks Tools Help
Tunnel Toolchain v5
The Image Module
precede - Definition from th...
Queue Tasks Datasets Subjects Samplesets Analysis Fusers Add Help
Cleanup (keep errors)Cleanup (all)Refresh

Task	Sample_ID	Priority	Status	Progress	Hold
sample downloader	828	0			
sample downloader	829	0			
sample downloader	830	0			
sample downloader	831	0			
sample downloader	832	0			
sample downloader	895	0	error		
sample downloader	896	0	error		
sample downloader	897	0	error		
sample downloader	898	0	error		
sample downloader	940	0	error		
sample downloader	943	0	success		
sample downloader	944	0	Importing files		
sample downloader	945	0			
sample downloader	946	0			
sample downloader	947	0			
sample downloader	948	0			
sample downloader	949	0			
sample downloader	950	0			
sample downloader	951	0			
sample downloader	952	0			

Host ID	Task	Sample ID	Status	Progress
richard-desktop:2KuEq2	sample downloader	944	Importing files	
richard-desktop:CA9g1V	sample downloader	1664	Downloading sample from Tunnel server	
richard-desktop:mUDTAZ	sample downloader	1220	Downloading sample from Tunnel server	

Figure G.1: The web-interface's processing queue

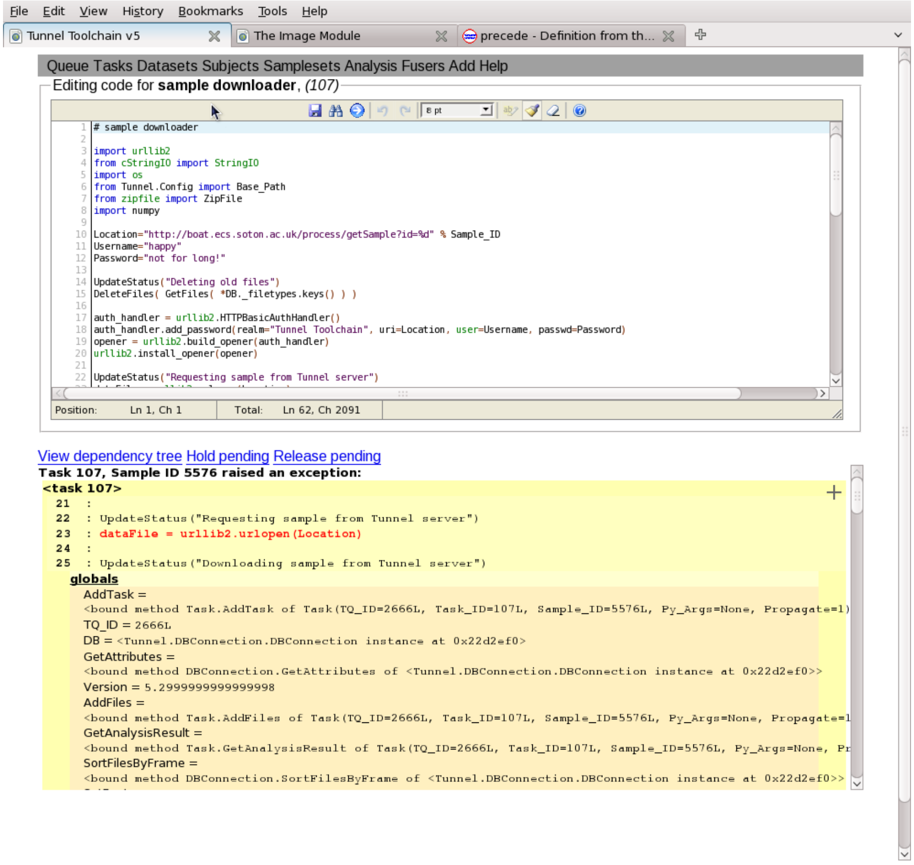
Below the queue is is set of options for clearing all completed tasks and refreshing the queue. The table below the queue show the status of each connected processing client.

G.3 Tasks

Clicking on **Tasks** at the top of the page will present a menu with all the tasks, along with options to add a new task or view the execution dependency tree. Clicking on one of the tasks will load a new page displaying the task's code in a powerful javascript code editor, featuring syntax highlighting and line numbering. There is also a link underneath the editor, which loads a version of the task queue that only shows instances of the current task. Figure G.2 shows the task editor, with the task queue displaying the traceback of a failed execution.

G.4 Datasets

The datasets page can be accessed from the top menu bar of the web-interface by clicking **Datasets**. The datasets page display information on all the datasets contained within the toolchain's database. Clicking the "+" link at the left of an entry will open it up, displaying further information on the dataset, such as it's sessions. Opening a session will show the subjects that feature in that session. Each subject's entry contains their corresponding samples. The samples are coloured according to the "quality" of the extracted gait cycle (where green is good, and red is bad). This can prove helpful when trying to identify problems in a dataset. Clicking the right mouse button on a dataset, session, subject or sample will present a pop-up menu, allowing


File Edit View History Bookmarks Tools Help
Tunnel Toolchain v5
The Image Module
precede - Definition from th...
Queue Tasks Datasets Subjects Samplesets Analysis Fusers Add Help
8 pt
View dependency tree Hold pending Release pending

Editing code for **sample downloader**, *(107)*

```
# sample downloader

import urllib2
from cStringIO import StringIO
import os
from Tunnel.Config import Base_Path
from zipfile import ZipFile
import numpy

Location="http://boat.ecs.soton.ac.uk/process/getSample?id=%d" % Sample_ID
Username="happy"
Password="not for long!"

UpdateStatus("Deleting old files")
DeleteFiles( GetFiles( *DB._filetypes.keys() ) )

auth_handler = urllib2.HTTPBasicAuthHandler()
auth_handler.add_password(realm="Tunnel Toolchain", uri=Location, user=Username, passwd=Password)
opener = urllib2.build_opener(auth_handler)
urllib2.install_opener(opener)

UpdateStatus("Requesting sample from Tunnel server")
```

Position: Ln 1, Ch 1 Total: Ln 62, Ch 2091

Task 107, Sample ID 5576 raised an exception:

```
<task 107>
 21  :
 22  : UpdateStatus("Requesting sample from Tunnel server")
 23  : dataFile = urllib2.urlopen(Location)
 24  :
 25  : UpdateStatus("Downloading sample from Tunnel server")
   globals
     AddTask =
     <bound method Task.AddTask of Task(TQ_ID=2666L, Task_ID=107L, Sample_ID=5576L, Py_Args=None, Propagate=1)
     TQ_ID = 2666L
     DB = <Tunnel.DBConnection.DBConnection instance at 0x22d2ef0>
     GetAttributes =
     <bound method DBConnection.GetAttributes of <Tunnel.DBConnection.DBConnection instance at 0x22d2ef0>>
     Version = 5.2999999999999998
     AddFiles =
     <bound method Task.AddFiles of Task(TQ_ID=2666L, Task_ID=107L, Sample_ID=5576L, Py_Args=None, Propagate=1
     GetAnalysisResult =
     <bound method Task.GetAnalysisResult of Task(TQ_ID=2666L, Task_ID=107L, Sample_ID=5576L, Py_Args=None, Pr
     SortFilesByFrame =
     <bound method DBConnection.SortFilesByFrame of <Tunnel.DBConnection.DBConnection instance at 0x22d2ef0>>
```

FIGURE G.2: The web-interface's task editor

you to add a task, or hold or continue processing for all the samples in the corresponding item. The popup menu also provides options to display the demographics for the item or add all its corresponding samples to a chosen sampleset.

Clicking on a sample will take you to an overview of the chosen sample. The overview features basic information such as the sample's subject, session, dataset, along with much more advanced information including the gait cycle finder's diagnostics, the average silhouettes and a video player for the sample. The sample overview also provides links to check the camera calibration and display a processing queue filtered by only that sample.

G.5 Subjects

The subjects page provides a convenient mechanism for viewing small thumbnails of every subject's face. This makes it easier to locate subjects who are returning to provide data, and do not know their tunnel ID. Clicking on a subject's thumbnail will display a filtered version of the datasets page, only showing the specified subject's samples.

G.6 Samplesets

The samplesets page shows a summary of the samplesets in the tunnel system. Samplesets provide a convenient way of grouping samples in sets, which is useful for creating gallery and probe sets for analysis. Clicking the right-mouse button in the left table will present a pop-up menu with options for deleting, splitting or editing the chosen sampleset, along with creating a new sampleset. The samplesets are populated using the datasets page, by clicking the right mouse button on the dataset/session/subject/sample of choice and selecting the *Add to sampleset* option.

G.7 Analysis

The analysis page provides an interface for viewing and editing analysis experiments. Clicking the right mouse button on the left list of analysis experiments will present a popup-menu allowing you to create a new analysis task. The right hand box allows you to configure the task. When *Start Analysis* is clicked, the system adds an instance of the chosen root task to the `task_queue` table, with the given configuration options placed as a pythonic dictionary in the `py_args` field. The processing task can access these parameters through the *Py_Args* variable, or by directly accessing the options by their names. The user configurable options are specified as `analysis property` items in the tunnel's MySQL `attributes` table. When a completed analysis experiment is selected, the results are displayed below the two tables.

Appendix H

Publications

Sina Samangooei, John D. Bustard, Richard D. Seely, Mark S. Nixon and John N. Carter. On Acquisition and Analysis of a Dataset Comprising of Gait, Ear and Semantic Data. In *Multibiometrics for Human Identification*; upcoming book. B. Bhanu and V. Govindaraju, eds. Cambridge University Press.

- Contributed towards areas discussing the technical aspects of the Biometric Tunnel.

Richard D. Seely, Michela Goffredo, John N. Carter and Mark S. Nixon. View Invariant Gait Recognition. In *Handbook of Remote Biometrics: for Surveillance and Security*, Springer. ISBN 978-1-84882-384-6

- Contributed towards a comprehensive literature review of modern gait analysis techniques and an overview of the Biometric Tunnel.

Richard D. Seely, Sina Samangooei, Lee Middleton, John N. Carter and Mark S. Nixon. The University of Southampton Multi-Biometric Tunnel and introducing a novel 3D gait dataset. In *Proceedings of IEEE Conference on Biometrics: Theory, Applications and Systems, BTAS 08*, September 2008.

- This paper was given as an oral presentation, describing the collection of the large multi-biometric dataset and presenting early results using the first one-thousand samples.

Michela Goffredo, Richard D. Seely, John N. Carter and Mark S. Nixon. Markerless view independent gait analysis with self-camera calibration. In *Proceedings of the Eighth IEEE International Conference on Automatic Face and Gesture Recognition*, September 2008.

- Made technical contributions towards project; setting up a capture system and assisting in the capture of a small dataset.

Richard D. Seely, John N. Carter and Mark S. Nixon. Spatio-temporal 3D Gait Recognition. In *3D Video Analysis, Display and Applications*, February 2008.

- Presented a poster outlining the Biometric Tunnel and results from the small evaluation dataset, collected to further develop the computer-vision algorithms within the system.