G

Gait Biometrics, Overview

Rama Chellappa, Ashok Veeraraghavan, and Narayanan Ramanathan
University of Maryland, College Park, MD, USA

Synonyms

Gait recognition

Definition

Gait is defined as the style or manner of walking. Studies in pyschophysics suggest that people can identify familiar individuals using just their gait. This has led to a number of automated vision-based algorithms that use gait as a biometrics. Such a system usually consists of a video camera capturing images of a person walking within its field of view. Appropriate features such as joint angles or silhouettes are extracted from this video and are then used to compare with the stored gait signatures of known individuals. As with any other biometric system, the system can operate in both the identification and the verification mode. Gait as a biometrics has several advantages compared to traditional biometrics such as fingerprint in that gait is nonintrusive, does not require cooperation from the individual, and can function at moderate distances from the subject.

Introduction

The study of human gait has gathered pace in recent years driven primarily by its potential as a biometrics. Gait-based person authentication has several significant advantages compared to traditional biometrics such as fingerprint or iris. Firstly, gait-based biometric systems do not require the individuals to be cooperative since the input of these systems is the video feed captured by passive cameras. Secondly, gait is a nonintrusive biometrics – it does not require the individuals to wear any special equipment in order to be recognized. Thirdly, gait-based biometric systems have an extended range compared to traditional biometrics – they can operate reliably even when the subjects are tens of meters away from the camera. Finally, such a system harnesses the potential of thousands of surveillance video cameras installed in public locations into a biometric authentication system.

Operation of a Gait-Based Biometric System

The sensor for a gait-based biometric system is a video camera capturing videos of human subjects walking within its field of view. The raw sensor video is then processed to extract relevant features which can then be used for recognition. If the acquisition conditions are expected to be controlled and favorable, then the quality of the

S.Z. Li, A.K. Jain (eds.), *Encyclopedia of Biometrics*, DOI 10.1007/978-1-4899-7488-4,

video will enable the extraction of features such as joint angles from the individual video frames. In more typical uncontrolled settings, the features extracted could either be background subtracted binary images, silhouettes, shapes, or width vectors – all examples of features capturing the extent of the human body to differing amounts of detail. During the training phase, several such sequences of each individual in the gallery are collected and the appropriate features are then stored in the database. During the test phase, each test sequence is compared with the training sequences available in the database, and the similarity is used to perform person authentication.

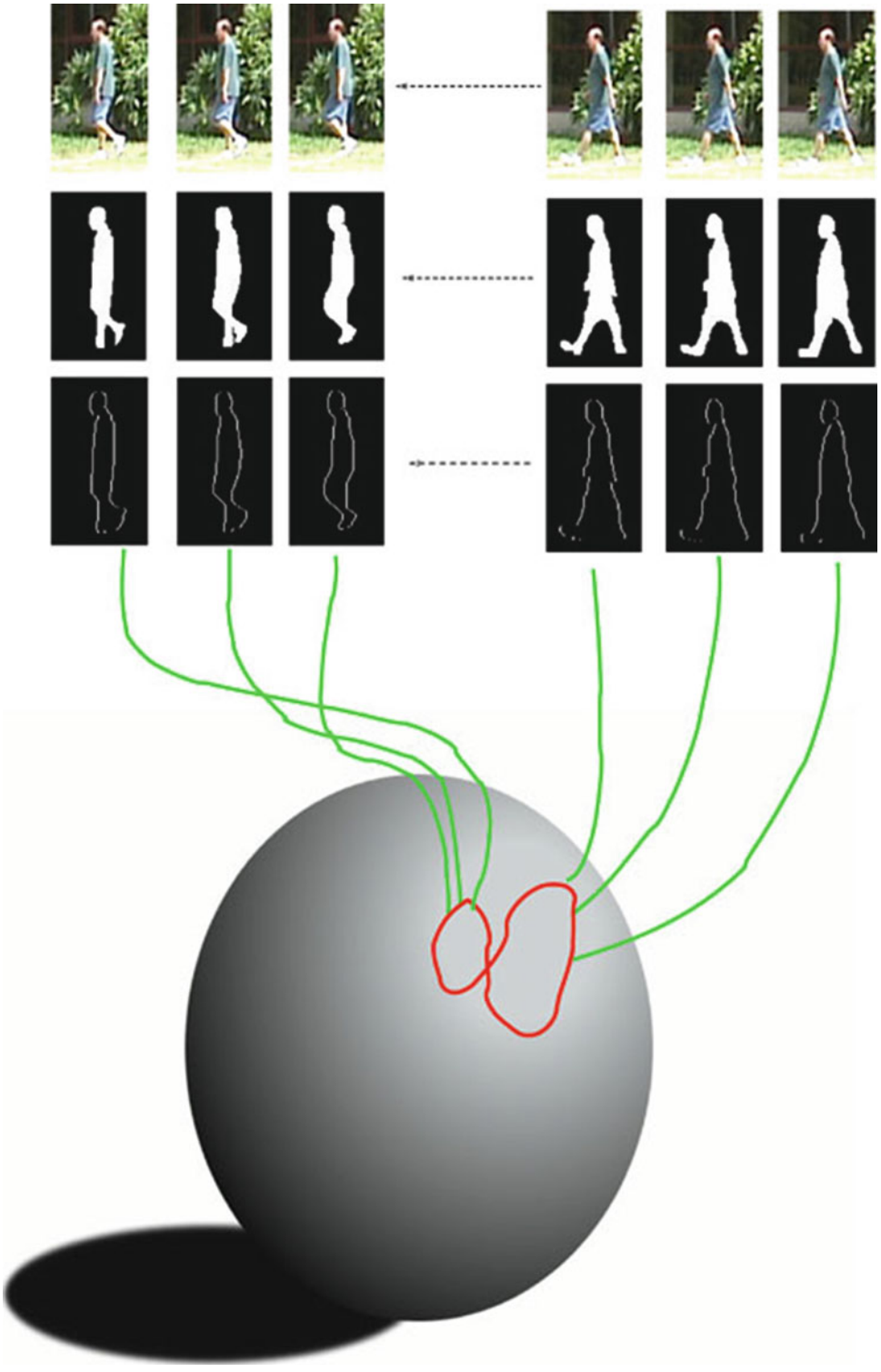
Challenges for Gait-Based Biometric Systems

The discriminative information in gait is present in both the shape of the individual and also in the manner of his/her gait. This means that gait-based biometric systems must be able to model gait as a time series of features or as a dynamical model in order to perform accurate recognition. Static template-based methods which have been used for most other biometric systems need to be adapted to a temporal sequence in order to achieve robust performance. In this regard, another challenge is time alignment of two sequences so that critical events during gait like "mid-stance," "toe-off," etc., are time aligned accurately so that recognition performance is not affected by inaccurate time alignment between postures that occur during gait. Since, gait-based person identification often occurs without any particular viewpoint, view invariance of the feature extracted from the video is another important challenge. This will ensure that recognition performance is robust to changes in the viewpoint of the camera. In scenarios with moderate amounts of acquisition control, one can set up multiple video cameras so as to ensure that the best possible viewpoint which happens to be the fronto-parallel gait is captured on atleast one of the cameras. Another challenge for automated gait-based biometrics is that of changing illumination conditions in the scene. In order to be robust to changing illumination conditions, background subtraction is typically performed on the raw videos before the video data is used in a recognition algorithm. Finally, another important challenge is the variability in the clothing, shoe type, and the surface on which the individuals walk. Obviously, the clothing of the subject especially their type of footwear has significant impact on the gait features observed, and it is important to bear this in mind while developing gait-based biometric systems.

Features for Gait-Based Biometrics

Silhouette: In most gait-based biometric systems, the cameras can be assumed to be static during the short duration of time that they capture the gait of a single individual for verification. This allows simple background models to be built for each of these cameras. Background subtraction then identifies the set of all pixels in the image that belong to the moving individual. Figure 1 shows a sequence of color images captured by a video camera as a person walks through its field of view. Shown below are the binary background subtracted images in which all pixels belonging to the individual are white, while the background is black. This binary image is then scaled to a uniform size so that the feature extracted is independent of the distance of the camera from the subject. Several algorithms for gait-based person identification use this binary silhouette as a feature [1–5].

Shape: "Shape is all the geometric information that remains when location, scale, and rotational effects are filtered out from the object" [6]. Kendall's statistical shape is a sparse descriptor of the shape that describes the shape configuration of k landmark points in an m-dimensional space as a $k \times m$ matrix containing the coordinates of the landmarks. Image space is 2-dimensional, and therefore it is convenient to describe the shape vector as a k-dimensional complex vector. First, a binarized silhouette denoting the extent of the object in an image is obtained. A shape feature is then extracted from this binarized silhouette. This feature vector must be invariant to



Gait Biometrics, Overview, Fig. 1 Graphical illustration of a video sequence obtained during a walking cycle and the corresponding features – silhouette and shape (Courtesy [7])

translation and scaling since the object's identity should not depend on the distance of the object from the camera. So any feature vector that we obtain must be invariant to translation and scale. This yields the pre-shape of the object in each frame. Pre-shape is the geometric information that remains when location and scale effects are filtered out. Let the configuration of a set of k landmark points be given by a k-dimensional complex vector containing the positions of landmarks. Let us denote this configuration as X. Centered pre-shape is obtained by subtracting the mean from the configuration and then scaling to norm one. The centered pre-shape is given by

$$Z_c = \frac{CX}{\|CX\|}, \text{ where } C = I_k - \frac{1}{k} 1_k 1_k^T \quad (1)$$

where I_k is a $k \times k$ identity matrix and 1_k is a k-dimensional vector of ones.

The advantage of using shape feature is that the differential geometric properties of the spherical manifold in which the shapes lie are very well understood, and therefore, appropriate distance measures that can account for translational, rotational, and scale invariances are well defined. For example, consider two complex configurations X and Y with corresponding pre-shapes α and β. The full Procrustes distance between the configurations X and Y is defined as the Euclidean distance between the full Procrustes fit of α and β and is chosen so as to minimize

$$d(Y, X) = \left\| \beta - \alpha s e^{j\theta} - (\alpha + jb) 1_k \right\|, \quad (2)$$

where s is a scale, θ is the rotation, and $(a + jb)$ is the translation. The full Procrustes distance is the minimum full Procrustes fit, i.e.,

$$d_F(Y, X) = \inf_{s,\theta,a,b} d(Y, X). \quad (3)$$

The extracted shape sequence is shown in the bottom row of Fig. 1 with a graphical illustration of the spherical manifold in which shapes lie. Shape is a very popular feature for gait-based biometrics, and several state-of-the-art algorithms perform gait matching as a matching of a sequence of shapes [7–10].

Joint Angles: A very popular feature for gait analysis in the medical and the psychophysics community is the joint angles – i.e., the angles made at each of the limb joints such as the knee, elbow ankle, wrist, etc. There have been a few gait-based biometric algorithms that use joint angles as the feature for matching [11, 12]. The advantage of using joint angles as a feature is the fact that view invariance is automatically achieved while using joint angles as a feature. Nevertheless, the essential problem with using joint angles is the fact that it is very challenging to robustly estimate them from uncontrolled monocular video sequences.

Algorithms for Matching

Most of the features described above have incorporated modest forms of view invariance (at least scale and translational invariance) as a part of the feature. Therefore, the essential task of the algorithm for matching would be to model the dynamics of the feature during gait and use this to perform matching in a manner that is fairly insensitive to the speed of walking.

Dynamic Time Warping (DTW): Dynamic time warping is an algorithm for estimating the nonlinear time synchronization between two sequences of features. The two sequences could be of differing lengths. Experiments indicate that the intrapersonal variations in gait of a single individual can be better captured by nonlinear warping rather than by linear warping [13]. The DTW algorithm which is based on dynamic programming computes the best nonlinear time normalization of the test sequence in order to match the template sequence, by performing a search over the space of all allowed time normalizations. The space of all time normalizations allowed is cleverly constructed using certain temporal consistency constraints. Several gait-based biometric algorithms have used the dynamic time warping algorithm in order to time synchronize and match gait sequences [7, 8]. Recently, the DTW algorithm has also been extended so as to learn the

warping constraints in a class-specific manner in order to improve discrimination between individuals [9].

Hidden Markov Model (HMM): The hidden Markov model (HMM) is a statistical state space model in which the observed shape sequence is modeled as outputs of hidden states whose transitions are assumed to be Markovian. The model parameters of the HMM encode both the transition probabilities between the hidden states and the outputs of hidden states. The advantage of using an HMM is that there exists a wealth of literature on learning the parameters of the HMM and to perform inference using the HMM. Typically, the model parameters for each individual in the gallery is learnt and stored during the training phase. During the test phase, the probability of the observation sequence conditioned on the model parameters is maximized in order to perform recognition. The HMM [2, 3] and its many variants [14] have been successfully used for gait-based person identification.

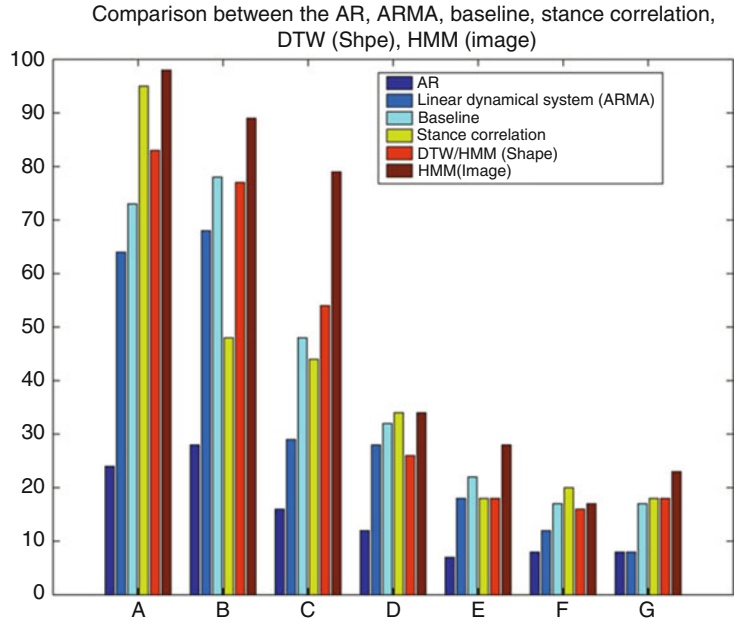
Autoregressive Moving Average Model (ARMA): Matching gait biometrics essentially is a problem of matching time-series data where the feature at each time instant is a silhouette or shape or joint angles. Therefore, traditional time series modeling approaches such as the autoregressive model (AR) and the autoregressive moving average (ARMA) model have also been successfully used for gait-based person identification. The model parameters of the ARMA model are learnt from the training sequences and stored. Given a test sequence, the model parameters for the test sequence are learnt, and the distance between the model parameters is used in order to perform recognition [7].

Model-Based Approaches

Typical feature-based approaches first compute a sequence of features from each video and then match the sequence of features obtained in the test video to those stored in the gallery. Model-based approaches are different in the sense that they fit the sequence of features to a physical model of the human body and its inherent dynamics. For example, a model-based feature extraction process guided principally by biomechanical analysis for gait-based person identification is proposed [15]. The shape model for human subjects is composed of an ellipse to describe the head and the torso, quadrilaterals to describe the limbs, and rectangles to describe the feet. Anatomical data is first used in order to derive shape and motion models that are consistent with normal human body proportions. Prototype gait motion models are then adapted to individuals using the specific characteristics of the extracted features. These individual-specific shape and motion models are then used for gait recognition. A systematic analysis of the model-based approach also showed that cadence and static shape parameters of the human body account for most of the recognition performance.

Experiments on the USF Gait Data

In order to quantitatively test the performance and the viability of gait-based biometrics, a challenging gait database of 122 individuals was collected at the University of South Florida [4] as part of the DARPA Human Identification at a Distance (HID) program. The entire dataset containing over 1,200 videos was separated into 12 different experiments with varying levels of difficulty. The different challenge experiments amounted to varying different covariates during gait, like viewpoint, clothing, surface type, shoe type, time, etc. A bar plot of the recognition performance of various algorithms on the USF dataset (Experiments A–G) is shown in Fig. 2. Experiments A, B, and C correspond to changes in "view," "shoe type," and "view + shoe type," respectively, without y change in the surface of walking, while challenge experiments D, E, F, and G correspond to changes in the surface type from grass to concrete. The experiments indicate that changes in the surface type has significant impact on the recognition performance while view, shoe type affects recognition performance to a much lesser degree.



Gait Biometrics, Overview, Fig. 2 Comparison of various algorithms on the USF gait database (Courtesy [1])

Summary

Gait is thus a novel biometrics that provides significant operational advantages over several other biometrics such as face, fingerprint, iris, etc. Unlike traditional biometrics like fingerprint, gait does not require the active cooperation of the subjects. Moreover, gait is a medium range biometrics in the sense that acquisition distances can be as large as tens of meters. Moreover, in most operational scenarios, it is nonintrusive and does not require the subject to wear any special clothing. Preliminary experiments into gait as a biometrics seem to indicate that the discriminative power of gait is not as strong as that of traditional biometrics such as fingerprints or iris. Therefore, several successful investigations for fusing the gait biometrics with other traditional biometrics in order to boost the identification performance have been performed, and this seems to be an area of immense potential [16].

Related Entries

- ▸ Biometrics, Overview
- ▸ Multibiometrics
- ▸ Surveillance

References

1. G. Veres, L. Gordon, J. Carter, M. Nixon, What image information is important in silhouette-based gait recognition? in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision Pattern Recognition (CVPR 2004)*, Washington, DC, vol. 2, 27 June 2004
2. A. Kale, A. Sundaresan, A. Rajagopalan, N. Cuntoor, A. Roy-Chowdhury, V. Kruger, R. Chellappa, Identification of humans using gait. IEEE Trans. Image Process. **13**(9), 1163–1173 (2004)
3. A. Sundaresan, A. RoyChowdhury, R. Chellappa, A hidden markov model based framework for recognition of humans from gait sequences,in *Proceedings of 2003 International Conference on Image Processing*, Barcelona, vol. 2, 2004, pp. 93–96
4. S. Sarkar, P. Phillips, Z. Liu, I. Vega, P. Grother, K. Bowyer, The human ID gait challenge problem: data sets, performance, and analysis. Pattern Anal. Mach. Intell. IEEE Trans. **27**(2), 162–177 (2005)
5. J. Man, B. Bhanu, Individual recognition using gait energy image. IEEE Trans. Pattern Anal. Mach. Intell. **28**(2), 316–322 (2006)
6. I. Dryden, K. Mardia, *Statistical Shape Analysis* (Wiley, Chichester, 1998)
7. A. Veeraraghavan, A. Roy-Chowdhury, R. Chellappa, Matching shape sequences in video with applications in human movement analysis. IEEE Trans. Pattern Anal. Mach. Intell. **27**(12), 1896–1909 (2005)
8. A. Veeraraghavan, A. Chowdhury, R. Chellappa, Role of shape and kinematics in human movement analysis, in *Proceedings of the 2004 IEEE Computer*

Society Conference on Computer Vision and Pattern Recognition, (CVPR 2004), Washington, DC, vol. 1, 27 June–2 July 2004, pp. I-730–I-737
9. A. Veeraraghavan, R. Chellappa, A. Roy-Chowdhury, The function space of an activity, in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, vol. 1, 2006, pp. 959–968
10. L. Wang, H. Ning, W. Hu, T. Tan, Gait recognition based on procrustes shape analysis, in *Proceedings of the Nineth IEEE International Conference on Image Processing (ICIP, Poster)*, Rochester, vol. III, 2002
11. D. Cunado, J. Nash, M. Nixon, J. Carter, Gait extraction and description by evidence-gathering, in *Proceedings of the International Conference on Audio and Video Based Biometric Person Authentication*, Hilton Rye Town, vol. 48, 1995
12. A. Bissacco, A. Chiuso, Y. Ma, S. Soatto, Recognition of human gaits, in *Conference on Computer Vision and Pattern Recognition*, Hawaii, vol. 2, 2001, pp. 52–57
13. A. Forner-Cordero, H. Koopman, F. van der Helm, Describing gait as a sequence of states. J. Biomech. **39**(5), 948–957 (2006)
14. Z. Liu, S. Sarkar, Improved gait recognition by gait dynamics normalization. IEEE Trans. Pattern Anal. Mach. Intell. **28**(6), 863–876 (2006)
15. D. Wagg, M. Nixon, On automated model-based extraction and analysis of gait, in *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004*, Seoul, 2004, pp. 11–16
16. A. Kale, A. Roychowdhury, R. Chellappa, Fusion of gait and face for human identification, in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004, (ICASSP'04)*, Montreal, vol. 5, 2004

Gait Recognition, Evaluation

Sudeep Sarkar[1], Ravichandran Subramanian[1], and Zongyi Liu[2]
[1]Computer Science and Engineering, University of South Florida, Tampa, FL, USA
[2]Personal device, Device and Services, Microsoft Corporation, Redmond, WA, USA

Synonyms

Gait recognition; Progress in gait recognition

Definition

Gait recognition refers to automated methods that use video or other sensory data of human gait to recognize or to identify a person. Evaluation of gait recognition refers to the benchmarking of progress in the design of gait recognition algorithms on standard, common, datasets.

Introduction

Gait recognition refers to the use of human gait to recognize or to identify a person based on their walking styles. It is a manifestation of overall body geometry, i.e., proportions of the limbs, torso, etc., and body physiology, i.e., bones and musculature. Given this diverse range of body attributes involved in its production, there is a strong possibility for large source of variation in gait among individuals and hence in its potential uniqueness. Given the myriad of factors that determine a person's gait, theoretical modeling of gait is very complex. Thus, the design of gait algorithms is necessarily an iterative process, which involves the intertwined processes of specification of a reasonable algorithmic approach and the experimental evaluation of its performance. In fact, this is true of any biometric algorithm. It is important to constantly evaluate and analyze progress being made at various levels of biometric design. This evaluation can be of three types: at algorithm level, at scenario level, and at operational level, roughly corresponding to the maturity of the biometrics. Given the young nature of gait as a biometric source, relative to mature biometrics such as fingerprints, current evaluations are necessarily at algorithm level. The motivation behind algorithm-level evaluations is to explore possibilities, to understand limitations, and to push algorithmic research toward hard problems. Some of the relevant questions are:

1. Is progress being made in gait recognition of humans?
2. To what extent does gait offer potential as an identifying biometrics?

3. What factors affect gait recognition and to what extent?
4. What are the critical vision components affecting gait recognition?
5. What are the strengths and weaknesses of different gait recognition algorithms?

An overview of the current evaluation of gait video as a potential biometrics is discussed here, with particular emphasis on the progress with respect to the HumanID Gait Challenge problem that has become the de facto benchmark. A synthesis of gait recognition performances reported on this dataset and other major ones is provided here, along with some suggestions for future evaluations. We also briefly summarize the progress in gait recognition based on inertial or accelerometer data that can be easily collected using modern smartphones.

A Panoramic View of Performance

To take stock of the progress made in gait recognition, we consider a summary of the identification rates reported in literature on different kinds of publicly available experimental protocols and datasets. Since larger datasets have more predictive power than smaller ones, our focus is limited to datasets with >100 subjects. We consider the Southampton Large dataset [1] (indoor and outdoor, 110+ subjects), the HumanID Gait Challenge dataset [2] (outdoor, 122 subjects), the CASIA-B Gait Dataset [3] (indoor, 124 subjects), the CASIA-C Gait Dataset [4] (outdoor, 153 subjects, infrared), the OU-ISIR Large population dataset [5] (indoor, 1,035 subjects), its extension OULP-C1V1-A [6] (indoor, 3,835 subjects), and the very recent TUM GAID dataset [7] (indoor, 305 subjects).

Figure 1 lists the average of the identification rates reported in the literature for matching across different conditions (dataset parameters and **covariates**). Of course, the caveat is that the conclusions are conditioned on the kinds of variations of each **covariate** observed in the respective datasets. Hence, a definitive conclusion is hard to make. However, this kind of summary has some conclusive weight, since it encompasses the findings of multiple research groups. It should provide some directions for focusing future research.

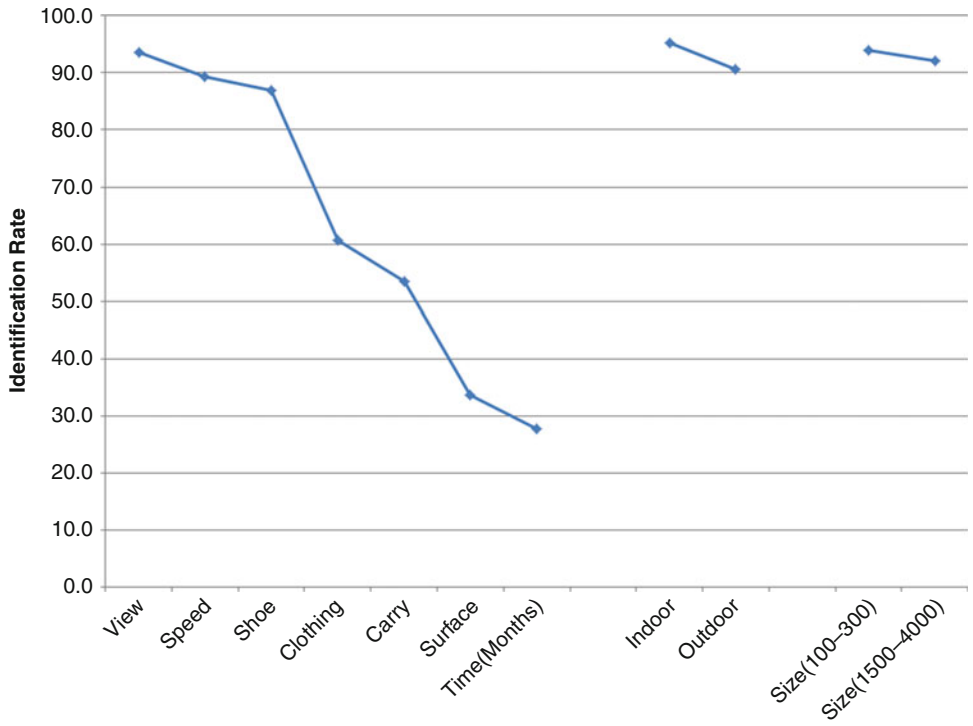
The data shows that the outdoor recognition rates are only a little behind indoor recognition rates. This could be due to the fact that the outdoor datasets considered here have been around for more than 7 years and there are several large indoor datasets that are less than 4 years old. Thus, the outdoor datasets, though inherently more difficult, might have been run through algorithms with specific tuning to get better performance on these datasets.

Recognition rates across the view, shoe, and speed changes are in the reasonable range. Recognition across walking surface-type change, carry condition, clothing, and elapsed time appear to be difficult problems.

Gait recognition rates seem to hold as the dataset size is increased from hundreds to thousands.

The HumanID Gait Challenge Problem

The development of gait biometrics is following a path that is somewhat different from other biometrics, for which serious evaluation benchmarks appeared after years of algorithmic development. It was more than 20 years for face recognition, whereas evaluation framework for gait recognition appeared in less than 10 years after the first publication of vision algorithms for gait recognition. Bulk of the research in gait recognition was spurred by the US DARPA HumanID at a Distance program. The HumanID Gait Challenge problem was formulated in this program to facilitate objective, quantitative measurement of gait research progress on a large dataset [2]. Many gait recognition research papers report performance on this dataset, along with other datasets such as the CASIA and SOTON datasets. One

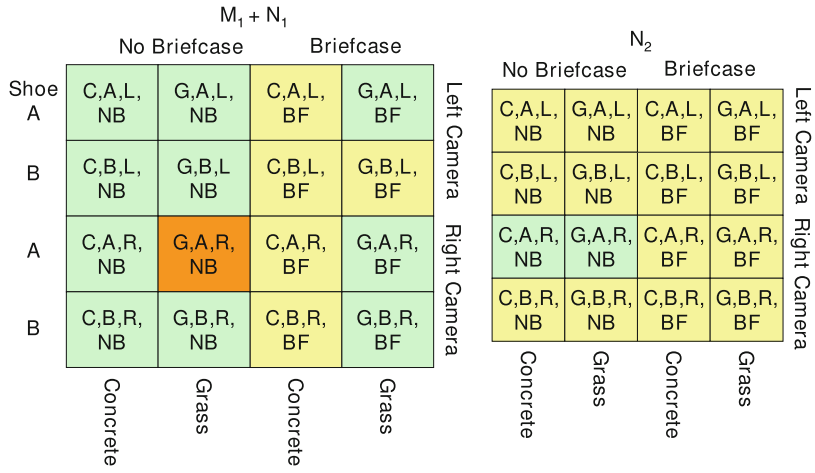


Gait Recognition, Evaluation, Fig. 1 Summary (average) of gait identification rates on datasets with more than 100 subjects, as reported in the literature for different conditions. The first seven performance points are for matching gait templates across different covariates. For example, "carry" refers to matching gait sequences where the hands of the subjects were free to sequences where the subjects were carrying a briefcase/bag. The "time" condition refers to matching gait templates collected at different times with the time gap of 3–6 months. The next two performance points are average of reported rates on datasets collected indoors and outdoors, respectively. The size condition refers to the number of subjects used in the experiments

advantage of the HumanID Gait Challenge dataset is that it also provides a well-defined experimental protocol to report performance, which makes it useful for meta-analysis of reported performances. It is still the only dataset in outdoor conditions with a large number of **covariates**.

The Dataset

The data was collected outdoors. For each person in the dataset, there are combinations of as many as five conditions or **covariates**. The conditions are (i) two camera angles (L and R), (ii) two shoe types (A and B), (iii) two surfaces (grass and concrete), (iv) with and without carrying a brief case (B or NB), and (v) two different dates 6 months apart, May and November. The **covariates** were chosen based on consultation with gait recognition researchers in the HumanID program. These are, of course, not the only variables that can impact gait but were logistically feasible and likely to impact gait the most. Attempt was made to acquire a person's gait in all possible combinations, and there are up to 32 sequences for some persons. Hence, the full dataset can be partitioned into 32 subsets, one for each combination of the 5 **covariates**. The partitioning of the data is visualized in Fig. 2. Each cell refers to a unique combination of view, shoe type, and surface **covariates**. The smaller arrangement of cells represents the data from repeat subjects. Comparisons between these subsets are used to set up challenge experiments; more on this later. The full dataset consists of 1,870 sequences from 122 individuals. This dataset is unique in the



Gait Recognition, Evaluation, Fig. 2 Partitioning of the HumanID gait challenge dataset in terms of its covariates, which are coded as follows: *C* concrete surface, *G* grass surface, *A* first shoe type, *B* second shoe type, *BF* carrying a briefcase, *NB* no briefcase, *M* data collected in May, N_1 new subjects in November data, and N_2 repeat subjects in November. The *light green shaded cells* are used to design the challenge experiments

Gait Recognition, Evaluation, Fig. 3 Frames from (**a**) the left camera for concrete surface, (**b**) the right camera for concrete surface, (**c**) the left camera for grass surface, (**d**) the right camera for grass surface

number of **covariates** exercised. It is the only dataset to include walking on a grass surface. Figure 3 shows some sample frames from this dataset.

In addition to the raw data sequence, there is ancillary information associated with the data. First, for each sequence, there is metadata information about the subject's age, sex, reported

Gait Recognition, Evaluation, Fig. 4 *Top row* shows the color images, cropped around the person, for one sequence. The *bottom row* shows the corresponding, part level, manually specified silhouettes

height, self-reported weight, foot dominance, and shoe information. Second, for a subset of this dataset, manually created **silhouettes** (see Fig. 4) are available. These manual **silhouettes** should not be used to test any recognition algorithm, but they could be used to build models or to study segmentation errors. More details about the process of creating these manual **silhouettes** and the quality checks performed can be found in [8]; here are some salient aspects. Seventy-one subjects from one of the two collection periods (May collection) were chosen for manual **silhouette** specification. The sequences corresponding to these subjects were chosen from the (i) **gallery** set (sequences taken on grass, with shoe type A, right camera view), (ii) **probe** B (on grass, with shoe type B, right camera view), (iii) **probe** D (on concrete, with shoe type A, right camera view), (iv) **probe** H (on grass, with shoe A, right camera view, carrying briefcase), and **probe** K (on grass, elapsed time). The **silhouette** in each frame over one walking cycle, of approximately 30–40 image frames was manually specified. This cycle was chosen to begin at the right heel strike phase of the walking cycle through to the next right heel strike. Whenever possible, this gait cycle was selected from the same 3D location in each sequence. In addition to marking a pixel as being from the background or subject, more detailed specifications in terms of body parts were marked. The head, torso, left arm, right arm, left upper leg, left lower leg, right upper leg, and right lower leg were explicitly labeled using different colors.

The Challenge Experiments

Along with the dataset, the gait challenge problem includes a definition of 12 challenge experiments (A through L), spanning different levels of difficulty. This provides a common benchmark to compare performance with other algorithms. The experiments are designed to investigate the effect on the performance of the five factors, i.e., change in viewing angle, change in shoe type, change in walking surfaces (concrete and grass), carrying or not carrying a briefcase, and temporal differences. The **gallery** set is common for all the experiments and corresponds to the dark colored cell in Fig. 2. The **gallery** consists of sequences with the following **covariates**: grass, shoe type A, right camera, no briefcase, and collected in May along with those from the *new* subjects from November. This set was selected as the **gallery** because it was one of the largest for a given set of **covariates**. The experiments differ in terms of the **probe** sets, which are denoted by the light green shaded cells. The structure of the 12 **probe** sets is listed in Table 1. The signatures are the video sequences of gait. The last two experiments study the impact of elapsed time. The elapsed time **covariate** implicitly includes a change of shoe and clothing because the subjects were not required to wear the same clothes or shoes in both data collections. Because of the implicit change of shoe, it can be safely assumed that a different set of shoes were used in the May and November data collections. This is noted in Table 1 by A/B for shoe type in experiments K and L. The key experiments

Gait Recognition, Evaluation, Table 1 The gallery and probe set specifications for each of gait challenge experiments. The gallery for all of the experiments is (G, A, R, NB, M + N_1) and consists of 122 individuals

Exp.	Probe (Surface, shoe, view, carry, elapsed time) (C/G, A/B, L/R, NB/BF, time)	Number of subjects	Difference
A[a]	(G, A, L, NB, M + N_1)	122	V[b]
B[a]	(G, B, R, NB, M + N_1)	54	S[c]
C	(G, B, L, NB, M + N_1)	54	S + V
D[a]	(C, A, R, NB, M + N_1)	121	F[d]
E	(C, B, R, NB, M + N_1)	60	F + S
F	(C, A, L, NB, M + N_1)	121	F + V
G	(C, B, L, NB, M + N_1)	60	F + S + V
H[a]	(G, A, R, BF, M + N_1)	120	B[e]
I	(G, B, R, BF, M + N_1)	60	S + B
J	(G, A, L, BF, M + N_1)	120	V + B
K[a]	(G, A/B, R, NB, N_2)	33	T[f] + S + C[g]
L	(C, A/B, R, NB, N_2)	33	F + T + S + C

[a]Key experiments
[b]View
[c]Shoe
[d]Surface
[e]carry
[f]Elapsed time
[g]Clothing

are those that involve controlled change in just one **covariate** and are marked with an asterisk in the table. The results from the 12 experiments provide an ordering of the difficulty of the experiments.

Baseline Gait Algorithm

The third aspect of the gait challenge problem is a simple but effective **baseline algorithm** to provide performance benchmarks for the experiments. Ideally, this should be a combination of "standard" vision modules that accomplishes the task. Drawing from the success of template-based recognition strategies in computer vision, a four-part algorithm that relies on **silhouette** template matching was designed. The first part semiautomatically defines bounding boxes around the moving person in each frame of a sequence. The second part extracts **silhouettes** from the bounding boxes using expectation maximization based on the Mahalanobis distance between foreground and background color model at each pixel. Each **silhouette** is scaled to a height of 128 pixels and centered (automatically) in each frame along the horizontal direction so that the centerline of the torso is at the middle of the frame. The third part computes the gait period from the **silhouettes**. The gait period is used to partition the sequences for spatiotemporal correlation. The fourth part performs spatiotemporal correlation to compute the similarity between two gait sequences.

Let $\mathbf{S_P} = \{\mathbf{S_P}(1), \cdots, \mathbf{S_P}(M)\}$ and $\mathbf{S_G} = \{\mathbf{S_G}(1), \cdots, \mathbf{S_G}(N)\}$ be the **probe** and the **gallery silhouette** sequences, respectively. First, the **probe** (input) sequence is partitioned into subsequences, each roughly over one gait period, N_{Gait}. Gait periodicity is estimated based on periodic variation of the count the number of foreground pixels in the lower part of the **silhouette** in each frame over time. This number will reach a maximum when the two legs are farthest apart (full stride stance) and drop to a minimum when the legs overlap (heels together stance).

Second, each of these **probe** subsequences, $\mathbf{S_{Pk}} = \{\mathbf{S_P}(k), \cdots, \mathbf{S_P}(k + N_{\text{Gait}})\}$, is cross-correlated with the given **gallery** sequence, $\mathbf{S_G}$.

$$\mathrm{Corr}(\mathbf{S_{Pk}}, \mathbf{S_G})(l) = \sum_{j=1}^{N_{\mathrm{Gait}}} S\left(\mathbf{S_P}(k+j), \mathbf{S_G}(l+j)\right) \quad (1)$$

where, the similarity between two image frames, $S(\mathbf{S_P}(i), \mathbf{S_G}(j))$, is defined to be the Tanimoto similarity between the **silhouettes**, i.e., the ratio of the number of common pixels to the number of pixels in their union. The overall similarity measure is chosen to be the median value of the maximum correlation of the **gallery** sequence with each of these **probe** subsequences. The strategy for breaking up the **probe** sequence into subsequences allows the algorithm to overcome segmentation errors in some contiguous sets of frames due to some background subtraction artifact or due to localized motion in the background.

$$\mathrm{Sim}(\mathbf{S_P}, \mathbf{S_G}) = \mathrm{Median}_k \left(\max_l \mathrm{Corr}(\mathbf{S_{Pk}}, \mathbf{S_G})(l) \right) \quad (2)$$

The baseline algorithm is parameter free. The algorithm, although straightforward, performs quite well on some of the experiments and is quite competitive with the first generation of gait recognition algorithms.

Performance on the Gait Challenge Problem

We track the best up-to-the-date reported performance of algorithms published after 2004 on the key experiments from the full USF HumanID dataset in Fig. 5. The progress is evident, but so is the stagnation of performance in the some of the key experiments. There is no one method that performs the best on all experiments. It is evident that experiments A (view), B (shoe), and H (carry) have near perfect identification rates. The improvement of performance was not due to "continued engineering" of existing approaches, but involved the redesign of the recognition approaches. For instance, the greatest gains came from approaches that also analyzed **silhouette** shapes [9–11]. Gait dynamics is important, but by itself is not sufficient. All approaches with good performance rely on the **silhouette**.

Experiments D (surface) and K (time) are below 60 % identification rate; however, they have improved over the baseline. Part of the reason for the lack of further improvement on these experiments could be the lack of focus of current algorithm to address these hard issues; effort was expended on the other experiments. Perhaps it is time to focus on these hard **covariates**. How does walking surface affect gait, and how can we compensate for it in the matching process? Are there walking surface gait invariants? Maybe future algorithms will emerge to answer these questions.

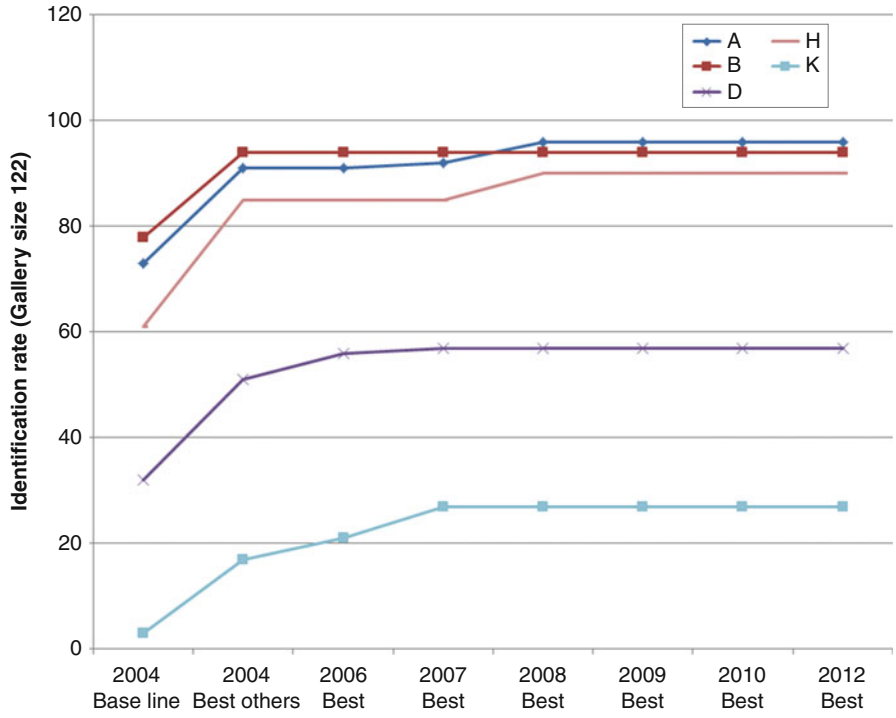
For the study of the time **covariate**, in addition to the HumanID dataset, the SOTON-Temporal dataset [12] can be particularly useful. It has been specifically designed to eliminate the problem of variation in clothing and shoes, which is inherent in data collections done after a period of time. They asked the subjects to wear overalls and walk without shoes.

Other Large Datasets

Table 2 categorizes the current large (>100 subjects) gait datasets in terms of the **covariates** present in them.

The SOTON HID Gait Dataset [13] with 115 subjects was collected mostly indoors and some under outdoor conditions. The indoor SOTON dataset was collected to examine the premise that gait is unique so the background is controlled so as to allow easy segmentation. The same subjects were also filmed walking outdoors to determine whether gait biometrics could be perceived with complex backgrounds. Performances well over 90 % have been reported using a variety of approaches.

The CASIA-B Gait Dataset [3] has 124 subjects with normal walking (6 sequences), coat (2 sequences), and carry bag (2 sequences). The first 4 normal walks are used as the **gallery**. There are 3 **probe** sets, ExpA2 (remaining 2 normal



Gait Recognition, Evaluation, Fig. 5 Improvement in gait recognition algorithms over time with respect to the baseline performance. Results on the full dataset with 122 subjects for the key experiments listed in Table 1. From 2004 to 2012, the best reported performances are better than baseline, on all the experiments

Gait Recognition, Evaluation, Table 2 Statistics and covariates of some gait datasets with more than 100 subjects

	SOTON large	USF HumanId	CASIA-B	CASIA-C	OU-ISIR large	OULP -C1V1-A	TUM GAID
Year	Summer 2001	May, Nov 2001	Jan 2005	Jul 2005	Mar 2009–Aug 2011	Mar 2009–Aug 2011	Jan, Apr 2012
Environment	In & out	Out	In	Out	In	In	In
# subjects	115	122	124	153	1,035	3,835	305
# sequences	2,163	1,870	13,640	1,530	4,140	7,670+	3,370
Covariates							
# views	2	2	11	1	2	2	1
Carry		x	x	x			x
Clothing			x				
Shoe		x					x
Speed				x			
Surface	x	x					
Time		x					x

walks), ExpB (2 coat sequences), and ExpC (2 bag sequences). The galleries and **probes** all contain 11 views (0–180° in 10° increments). Most papers use only the 90° (fronto-parallel) view for evaluation. The average identification rate across papers reported for this dataset is as follows: for ExpA2, it is 97.4 %, ExpB is 60.7 %, and ExpC is 75.2 %.

The CASIA-C Gait Dataset [4] has 153 subjects, captured at night with infrared cameras. It has normal walking (4 sequences), carry bag (2 sequences), fast walk (2 sequences), and slow walk (2 sequences). It is based on only one camera view. The first 3 normal walks are used as the **gallery**. There are 4 **probe** sets, probeN4 (remaining normal walk), probeSlow, probeFast, and probeBag. Among them, the average CCR for probeN4 is 94.9 %, probeSlow is 89.2 %, probeFast is 89.6 %, and probeBag is 49.4 %.

The OU-ISIR Large population dataset and its extension OULP-C1V1-A [6] contain few thousands of subjects. This is the largest dataset to date in terms of number of subjects. However, like many other datasets, these are indoor datasets, collected under controlled background conditions. It is an excellent dataset to study gender and age effects gait and to study dataset scaling issues of gait recognition. The dataset contains subjects with a wide variety of ages and a good mix of genders. Since this is a new dataset, there are very few papers published using this data.

The TUM Gait from Audio, Image, and Depth [7] is a relatively new multimodal dataset with 305 subjects. The **covariates** are carrying a backpack, wearing protective cover over shoes, and elapsed time of 3 months.

Other Types of Gait Sensors

Apart from video, gait can also be captured using floor force sensors and using accelerometers and gyroscopes. Accelerometry data can be used to identify gait parameters such as heel strike, gait cycle frequency, stride symmetry, regularity, cadence, step length, and gait symmetry [14]. The use of accelerometer for gait identification is fairly recent starting around 2004. Most of the papers are from two research groups: Gjovik University, Norway [15, 16], and Hochschule Darmstadt [17, 18]. There is no commonly used dataset to benchmark performance. Even the same research group has often used different datasets for different papers, which makes it hard to judge progress. However, to get a rough idea of the biometric potential for these kinds of gait measurements, one can consider the commonly reported measure of performance, which is in terms of the EER (equal error rate). EER of 15 % is representative of most experiments. Best performances are around 5 % EER. Typical datasets have 40–50 subjects. Very recently, Osaka University has created the largest dataset, with 736 subjects [19].

Future Evaluations

The review of published literature shows that researchers frequently do not conduct standard experiments on a given dataset. This is worse for datasets that do not come with a predefined experimental protocol. Some examples include using only a subset of the data (analyzing only one view of a multiview dataset), creating ad hoc **galleries** and **probes**, and using different metrics to measure performance. While these kinds of creative repartitioning of datasets can be useful to tease out many important factors not brought to fore in the original experiments, they make it hard to compare across algorithms. It would be of great service to the research community if performances on the original experiments were also reported, along with any new ones. Also when new datasets are created, experiments should be precisely defined and metrics carefully constructed.

It is to be expected that each gait research group would collect their own dataset to develop ideas. This is an important process. Given the data-driven nature of biometric research, the key to future progress are such datasets collected to explore issues not considered or raised by existing ones. For instance, the SOTON-Temporal dataset [12] aids in disentangling the time and clothing **covariates**. As of today, there is a need for the better understanding of the variation of gait due to surface conditions and across elapsed time. Also, currently, there is no dataset to explore the matching across time **covariate** for a large number of subjects.

Ideally, the new datasets should consist of gait data from around 1,000 subjects, an order of

magnitude larger than most current large datasets. OU-ISIR dataset is a big step in this direction. It is important to increase the number of subjects so that it is possible to empirically study the scaling of performance with number of subjects.

Some guidance about the required sizes can be found in [20, 21], where statistical reasoning is employed to relate the number of subjects with target error confidences. The data collection should include gait data repeated at regular time intervals of weeks spanning about a year. The dataset should be collected in outdoor conditions, preferably collected at a distance of 300 m to reflect real-world conditions. The dataset should come with a set of well-defined experiments in terms of **gallery** and **probe** sets. These experiments will influence the types of algorithms. For the experiments to be effective at influencing the direction of gait research, the design of the experiments needs to solve the *three bears problem*; the experiments must be neither too hard nor too easy, but just right. If performance on the experiments is easily saturated, then the gait recognition community will not be challenged. If experiments are too hard, then it will not be possible to make progress on gait recognition. Ideally, the set of experiments should vary in difficulty, characterize where the gait recognition problem is solvable, and explore the factors that affect performance. A set of experiments cannot meet this ideal unless the appropriate set of data is collected. It is important to view biometric research as a data-driven algorithm development process rather than algorithm-driven data collection process.

As datasets grow in size, procurement of such datasets and resources for development and evaluation of biometric algorithms will become expensive and time-consuming and will require expertise in computing grids. One intriguing possibility of the future is a cloud-based evaluation system, which can host common dataset(s) and allow the submission of algorithms either as source code or Linux x-86 executable, to enforce a standard experimental protocol and to provide results in a standard format. Over time, the detailed performance data collected on the cloud will help in the deep statistical analysis of different algorithms than is currently possible with just overall recognition performance scores. We will essentially have pairwise similarity scores over entire dataset(s) for many different algorithms. The cloud setup will also lower the computational setup barrier for engaging in biometric research. One such cloud platform, hosted on the Amazon Elastic Cloud, is detailed in [22] for the gait challenge problem.

Related Entries

- ▸ Biometric Verification/Identification/Authentication/Recognition: The Terminology
- ▸ Hidden Markov Models
- ▸ Human Detection and Tracking
- ▸ Gait Recognition, Silhouette-Based
- ▸ Performance Evaluation, Overview
- ▸ Performance Testing Methodology Standardization
- ▸ Psychology of Gait and Action Recognition

References

1. D.K. Wagg, M.S. Nixon, On automated model-based extraction and analysis of gait, in *International Conference on Automatic Face and Gesture Recognition*, Seoul, 2004, pp. 11–16
2. S. Sarkar, P. Jonathon Phillips, Z. Liu, I. Robledo-Vega, P. Grother, K.W. Bowyer, The human ID gait challenge problem: data sets, performance, and analysis. IEEE Trans. Pattern Anal. Mach. Intell. **27**(2), 162–177 (2005)
3. S. Yu, D. Tan, T. Tan, A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition, in *International Conference on Pattern Recognition*, Hong Kong, vol. 4, 2006, pp. 441–444
4. D. Tan, K. Huang, S. Yu, T. Tan, Efficient night gait recognition based on template matching, in *18th International Conference on Pattern Recognition, ICPR 2006*, Hong Kong, vol. 3, 2006, pp. 1000–1003
5. M. Okumura, H. Iwama, Y. Makihara, Y. Yagi, Performance evaluation of vision-based gait recognition using a very large-scale gait database, in *2010 Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*, Washington, DC, 2010, pp. 1–6
6. H. Iwama, M. Okumura, Y. Makihara, Y. Yagi, The ou-isir gait database comprising the large population dataset and performance evaluation of gait

recognition. IEEE Trans. Inf. Forensics Secur. **7**(5), 1511–1521 (2012)
7. M. Hofmann, J. Geiger, S. Bachmann, B. Schuller, G. Rigoll, The {TUM} gait from audio, image and depth (gaid) database: multimodal recognition of subjects and traits. J. Vis. Commun. Image Represent. 25(1), 95–206 (2014)
8. Z. Liu, L. Malave, A. Osuntogun, P. Sudhakar, S. Sarkar, Toward understanding the limits of gait recognition, in *Proceedings of SPIE Defense and Security Symposium: Biometric Technology for Human Identification*, Orlando, 2004, pp. 195–205
9. Z. Liu, S. Sarkar, Improved gait recognition by gait dynamics normalization. IEEE Trans. Pattern Anal. Mach. Intell. **28**(6), 863–876 (2006)
10. A. Kale, A. Sundaresan, A.N. Rajagopalan, N.P. Cuntoor, A.K. Roy-Chowdhury, V. Krüger, R. Chellappa, Identification of humans using gait. IEEE Trans. Image Process. **13**(9), 1163–1173 (2004)
11. X. Yang, T. Zhang, Y. Zhou, J. Yang, Gabor phase embedding of gait energy image for identity recognition, in *8th IEEE International Conference on Computer and Information Technology, CIT 2008*, Sydney, 2008, pp. 361–366
12. D. Matovski, M. Nixon, S. Mahmoodi, J. Carter, The effect of time on the performance of gait biometrics, in *2010 Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*, Washington, DC, 2010, pp. 1–6
13. J.D. Shutler, M.G. Grant, M.S. Nixon, J.N. Carter, On a large sequence-based human gait database, in *International Conference on Recent Advances in Soft Computing*, Nottingham, 2002, pp. 66–71
14. C.C. Yang, Y.L. Hsu, A review of accelerometry-based wearable motion detectors for physical activity monitoring. Sensors **10**(8), 7772–7788 (2010)
15. D. Gafurov, E. Snekkenes, P. Bours, Gait authentication and identification using wearable accelerometer sensor, in *2007 IEEE Workshop on Automatic Identification Advanced Technologies*, Alghero, 2007, pp. 220–225
16. D. Gafurov, J. Hagen, E. Snekkenes, Temporal characteristics of gait biometrics, in *2010 Second International Conference on Computer Engineering and Applications (ICCEA)*, Bali Island, vol. 2, 2010, pp. 557–561
17. C. Nickel, C. Busch, Classifying accelerometer data via hidden Markov models to authenticate people by the way they walk, in *2011 IEEE International Carnahan Conference on Security Technology (ICCST)*, Barcelona, 2011, pp. 1–6
18. C. Nickel, T. Wirtl, C. Busch, Authentication of smartphone users based on the way they walk using k-nn algorithm, in *2012 Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, Piraeus-Athens, 2012, pp. 16–20
19. N.T. Trung, Y. Makihara, H. Nagahara, Y. Mukaigawa, Y. Yagi, Performance evaluation of gait recognition using the largest inertial sensor-based gait database, in *2012-5th IAPR International Conference on Biometrics (ICB)*, New Delhi, 2012, pp. 360–366
20. G. Doddington, M. Przybocki, A. Martin, D. Reynolds, The NIST speaker recognition evaluation–overview, methodology, systems, results, perspective. Speech Commun. **31**(2–3), 225–254 (2000)
21. G. Veres, M. Nixon, J. Carter, Is enough enough? What is sufficiency in biometric data? Lect. Notes Comput. Sci. **4142**, 262 (2006)
22. R. Panchumarthy, R. Subramanian, S. Sarkar, Biometric evaluation on the cloud: a case study with HumanID gait challenge, in *2012 IEEE Fifth International Conference on Utility and Cloud Computing (UCC)*, Chicago, 2012, pp. 219–222

Gait Recognition, Model-Based

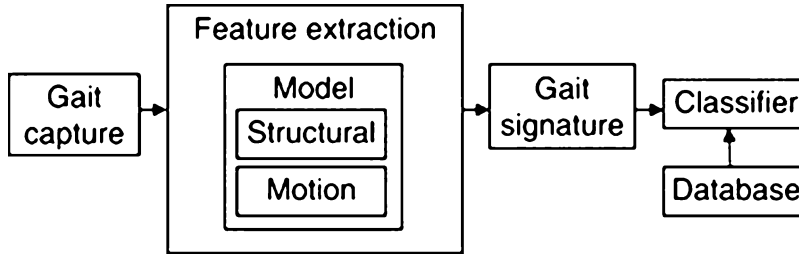
Chew-Yean Yam[1] and Mark Nixon[2]
[1]University of Southampton, Southampton, UK
[2]School of Electronics and Computer Science, University of Southampton, Southampton, UK

Synonyms

Gait models for biometrics; Knowledge-based gait recognition

Definition

Model-based gait recognition relates to the identification using an underlying mathematical construct(s) representing the discriminatory gait characteristics (be they static or dynamic), with a set of parameters and a set of logical and quantitative relationships between them. These models are often simplified based on justifiable assumptions, e.g., a system may assume a pathologically normal gait. Such a system normally consists of gait capture, a model(s), a feature extraction scheme, a gait signature, and a classifier (Fig. 1). The model can be a 2- or 3-dimensional structural (or shape) model and/or motion model that lays the foundation for the extraction and tracking of a moving person. An alternative to a model-based approach is



Gait Recognition, Model-Based, Fig. 1 Components of a typical model-based gait recognition system

to analyze the motion of the human silhouette deriving recognition from the body's shape and motion. A gait signature that is unique to each person in the database is then derived from the extracted gait characteristics. In the classification stage, many pattern classification techniques can be used, such as the k-nearest neighbor approach.

The main advantage of the model-based approach is that it can reliably handle occlusion (especially self-occlusion), noise, scale, and rotation well, as opposed to silhouette-based approaches.

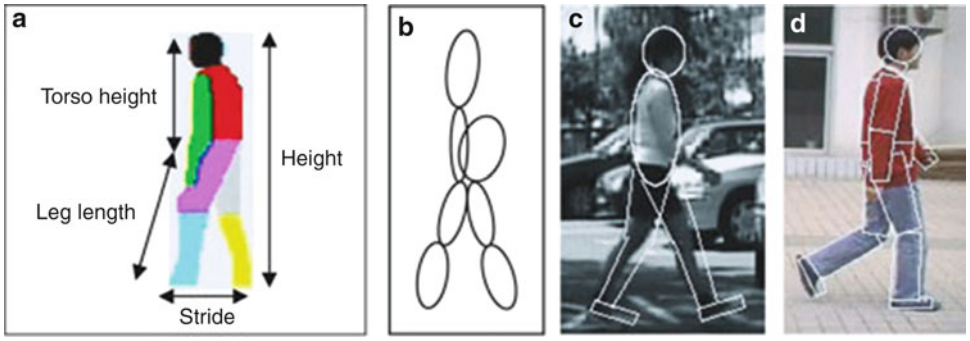
Practical issues that challenge the model-based approach can be divided into two categories, which relate to the *system* and to the *person*. One of the systems-related challenges is viewpoint invariance, while person-related challenges include the effects of *physiological* changes (such as aging, the consistency of gait taken/enrolled at different times, whether our walking pattern changes over a longer period of time), *psychological* changes (mood), and *external factors* (load,footwear, and the physical environment).

The first model-based approach to gait biometrics was by Cunado et al. in 1997 [1, 2], featuring the ability to reliably accommodate self-occlusion and occlusion by other objects, noise, and low resolution. Also, most of the time, the parameters used within the model and their relationship to the gait are obvious, i.e., the mathematical construct may itself contain implicit/explicit meaning of the gait pattern characteristics. Though, it often suffers from high computational cost, this can be mitigated by optimization tools or increased computing power. Gait sequences are usually acquired when the subject is walking in a plane normal to the image capture device since the side view of a moving person reveals most information, though it is possible to use other views.

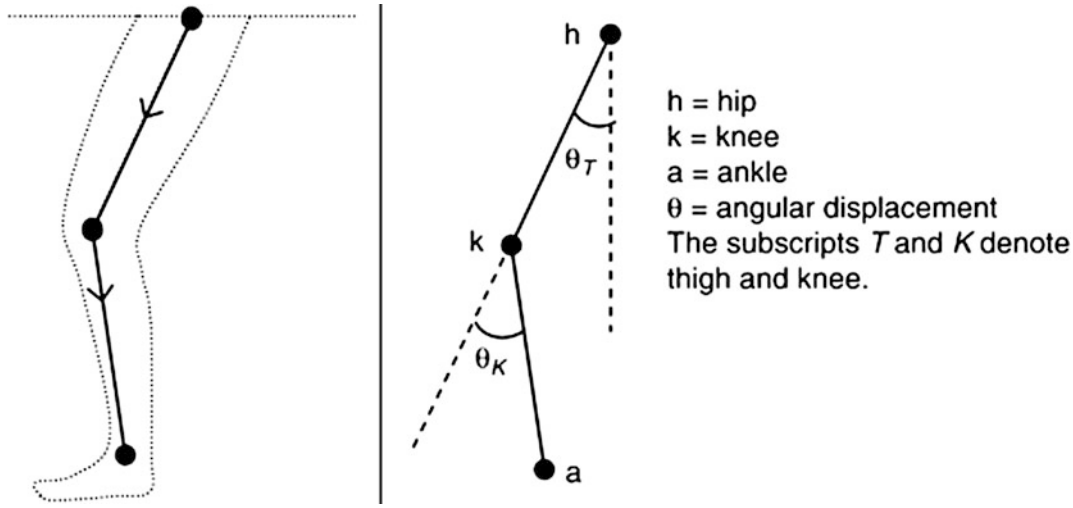
Models

In a typical model-based approach, often, a structural model and a motion model are required to serve as the basis for tracking and feature (moving human) extraction. These models can be 2- or 3-dimensional, though most of the current approaches are 2-dimensional and have shown the capability to achieve promising recognition results on large databases ($>$100 subjects). A structural model describes the topology or the shape of human body parts such as head, torso, hip, thigh, knee, and ankle by measurements such as the length, width, and position. This model can be made up of primitive shapes (cylinders, cones, and blobs), stick figures, or arbitrary shapes describing the edge of these body parts. On the other hand, a motion model describes the kinematics or the dynamics of the motion of each body part. Kinematics generally describe how the subject changes position with time without considering the effect of masses and forces, whereas dynamics account for the forces that act upon these body masses and the resulting motion. When developing a motion model, the constraints of gait such as the dependency of neighboring joints and the limit of motion in terms of range and direction have to be understood.

Bobick et al. used a structural model to recover static body and stride parameters (Fig. 2a) determined by the body geometry and the gait of a person [3]. Lee et al. fit ellipses to seven regions



Gait Recognition, Model-Based, Fig. 2 Example body parameters that are used in structural models. (**a**) Bobick, (**b**) Lee, (**c**) Wagg, (**d**) Wang



Gait Recognition, Model-Based, Fig. 3 Structural model of a lower limb: upper and lower pendulum represents the thigh and the lower leg, respectively, connected at the knee joint

representing the human body (Fig. 2b) and then derived two types of features across time: mean and standard deviation, and magnitude and phase of these moment-based region features [4].

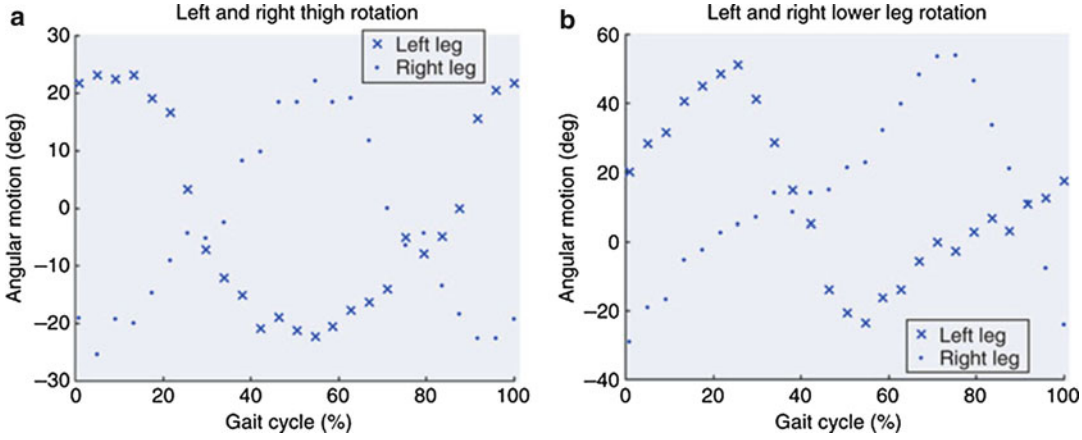
Cunado et al. proposed an early motion model-based approach, based on the angular motion of the hip and thigh [1,2], where the angular motion of the hip and the thigh is described by a Fourier series. For this method, a simple structural model was used and the angular rotation as defined in Fig. 3. Although the motion model is for one leg, assuming that gait is symmetrical, the other leg can be modeled similarly, with a phase lock of $1/2$-period shift (Fig. 4).

Cunado et al. modeled the angular motion of the thigh by

$$\theta_T = a_0 + 2\sum_{1}^{N}[b_k \cos k\omega_0 t - c_k \sin k\omega_0 t],$$

where N is the number of harmonics, ω_0 is the fundamental frequency, and a_0 is the offset. In application, the frequency data was accumulated from a series of edge-detected versions of the image sequence of the walking subject. The gait signature was derived by the multiplication of the phase and magnitude component of the Fourier description.

Later, Yam et al. [5] extended the approach to describe the hip, thigh, and knee angular motion of both walking and running gaits first by an empirical motion model, then by an analytical



Gait Recognition, Model-Based, Fig. 4 Thigh and lower leg rotation of the left and right leg. (**a**) Left and right thigh rotation. (**b**) Left and right lower leg rotation

model motivated by coupled pendulum motion. Similarly, the gait signature is the phase-weighted magnitude of the Fourier description of both the thigh and knee rotation.

Bouchrika et al. [6] have proposed one of the latest motion model-based gait feature extraction using a parametric form of elliptic Fourier descriptors to describe joint displacement:

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} X(t) * S_x \\ Y(t) * S_y \end{bmatrix},$$

where α is the angle, S_x and S_y are the scaling factors, and $X(t)$ and $Y(t)$ are Fourier summations. The joint trajectory is then fitted to the image sequence by optimizing a_0, b_0, α, S_x, and S_y; the motion model fit is implemented by the Hough transform.

Wagg et al. (Fig. 2c) and Wang et al. (Fig. 2d) used a combination of both structural and motion models to track and extract walking human figures [7, 8]. Wagg introduced a self-occlusion model while Wang used the conditional density propagation framework [9] to aid feature extraction.

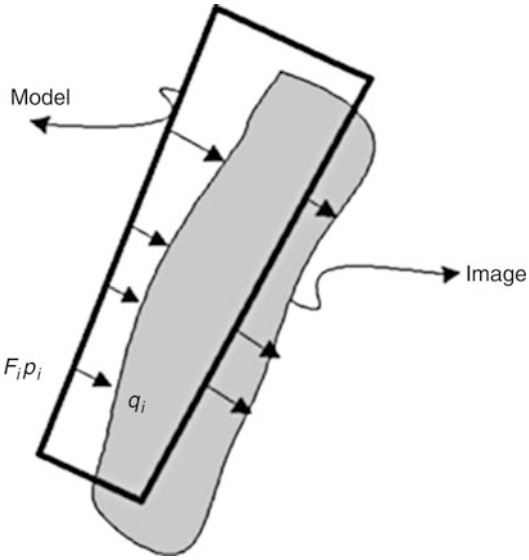
Beyond the 2D models, Urtasun et al. developed a 3D gait motion model derived from a small group of subjects [10]. The joint motion is approximated by a weighted sum of the mean motion and the eigenvectors of sample angular motion vectors. This approach also shows that it is capable of approximating running motion as well.

Feature Extraction

Feature extraction segments interesting body parts for a moving human and extracts static and/or dynamic gait characteristics. The process normally involves model initialization, segmentation, and tracking (estimation) of the moving human from one image to the next. This is a significant step that extracts important spatial, temporal, or spatial-temporal signals from gait. Feature extraction can then be carried out in a concurrent [1, 2, 5, 8] or iterative/hierarchical [7] manner.

A conventional starting point of a gait cycle is the heel strike at the stance phase, although any other stage within a gait cycle can be used. Earlier techniques determine the gait cycle manually; later, many have employed automatic gait cycle detection. A gait cycle can be detected by simply identifying the stance phase; if using a bounding box method, the width of the box has the highest value during the stance phase. Other alternatives are counting the pixels of the human figure, using binary mask (Fig. 5) by approximating the outer region of the leg swing [7].

Gait Recognition, Model-Based, Fig. 5 Binary mask to detect gait cycle. The sum edge strength within the mask varies periodically during the subject's gait and the heel strike being the greatest



Gait Recognition, Model-Based, Fig. 6 Measuring the boundary matching error

Quality of Feature Extraction

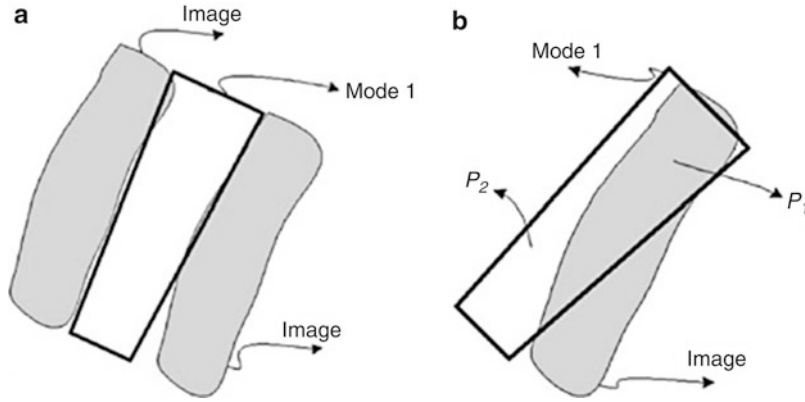
A good model configuration is defined as one that yields a high correlation between the model and the subject's image. Useful measures for computing model and image data correlation include *edge correspondence* and *region correspondence* [8]. Edge correspondence is a measure of how closely model edges coincide with image edges, while region correspondence is a measure of similarity between the image region enclosed by the model and that corresponding to the image of the subject. These two measures are used together. A high edge correspondence indicates that the model is closely aligned with image edges; however, it does not guarantee that the model matches the correct edges. If the initial model configuration is poor, or the subject is occluded, the match may be coincidental. For this reason, region correspondence is also required.

Another measure is a pose evaluation function (PEF) which combines the boundary (edge) matching error and the region matching error to achieve both accuracy and robustness. For each pixel, p_i, in the boundary of the projected human model, the corresponding pixel in the edge image along the gradient direction at p_i (Fig. 6) is searched. In other words, the pixel nearest to p_i and along that direction is desired. Given that q_i is the corresponding pixel and that $\boldsymbol{F}_i$ stands for the vector $\overline{p_i q_i}$, the matching error of pixel p_i to q_i can be measured as the norm $\|\mathbf{F}_i\|$. Then, the average of the matching errors of all pixels in the boundary of the projected human model is defined as the boundary matching error

$$E_b = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{F}_i\|,$$

where N is the number of the pixels in the boundary.

In general, the boundary matching error measures the similarity between the human model and image data, but it is insufficient under certain circumstances, as illustrated in Fig. 7a, where a model part falls into the gap between two body parts in the edge image. Although it is obviously badly fitted, the model part may have a small boundary matching error. To avoid such ambiguities, region information is further considered.



Gait Recognition, Model-Based, Fig. 7 Illustrating the necessity of simultaneous boundary and region matching. (**a**) A typical ambiguity: a model part falls into the gap between two body parts. (**b**) Measuring region matching error

Figure 7b illustrates the region matching. Here, the region of the projected human model that is fitted into the image data is divided into two parts: P_1 is the model region overlapped with the image data and P_2 is the rest of the model region. Then, the matching error with respect to the region information is defined by

$$E_r = |P_2| / (|P_1| + |P_2|)$$

where $|P_i|$, $(i = 1, 2)$ is the area, i.e., the number of pixels in the corresponding region.

Recognition

A gait signature is a discriminatory feature vector that can distinguish an individual. These signatures have invariant properties embedded in a person such as stride length, person's height/width, gait cycle, and self-occlusion and that related to the imaging system such as translation, rotation, scale, noise, and occlusion by other objects. These signatures can be of static [3], dynamic [2, 5], or a fusion of static and dynamic [7, 8] characteristics of gait or with other biometrics [11, 12]. The fusion can happen either at the feature extraction stage or at the classification stage. On the Southampton datasets of 115 subjects filmed indoors (in controlled conditions) and outdoors (with effects of shadows, background objects, and changing illumination), Wagg's approach achieved an overall CCR of 98.6 % on the indoor data and 87.1 % on the outdoor data.

In the case of 3D approach [10], experiments show that the first six coefficients of that motion model can characterize 90 % gait patterns of the database used. This resulted in a very compact gait signature, which requires only the first three coefficients to form separate clusters for each subject. It is interesting that this study found that the first few coefficients could represent physiological characteristics like weight, height, gender, or age, while the remaining ones can be used to distinguish individual characteristics. Another interesting finding is that the nature of the gait signature for running derived from this 3D motion model is similar to that of Yam et al. that is, signature clusters are more dispersed within subject and span more widely within the signature space, as compared to that of walking. Both studies were based on data collected by having subjects running on the treadmill.

Conclusions and Outlook

Using a model is an appealing way to handle known difficulty in subject acquisition and description for gait biometrics. There is a selection of models and approaches which can handle walking and running. Clearly, the use of a model introduces specificity into the feature extraction and description process, though this is generally at the cost of increased computation. Given their advantages, it is then likely that model-based approaches will continue to play a part in the evolution of systems which deploy gait as a biometric. Currently, practical advantages of

three-dimensional (3D) approaches have yet to be explored and investigated. Given that human motion occurs in space and time, it is likely that much information is embedded within the 3D space. Further, 3D approaches may provide a more effective way to handle issues like occlusion, pose, and view point. Therefore, 3D model-based gait recognition may be a good way to move forward.

Related Entries

- ▸ Gait Recognition, Model-Based
- ▸ Gait Recognition, Silhouette-Based
- ▸ Human Detection and Tracking
- ▸ Markerless 3D Human Motion Capture from Images
- ▸ Multibiometrics

References

1. D. Cunado, M.S. Nixon, J.N. Carter, Using gait as a biometric, via phase-weighted magnitude spectra, in *First International Conference on Audio- and Video-Based Biometric Person Authentication*, Crans-Montana, 1997, pp. 95–102
2. D. Cunado, M.S. Nixon, J.N. Carter, Automatic extraction and description of human gait models for recognition purposes. Comput. Vis. Image Underst. **90**(1), 1–41 (2003)
3. A.F. Bobick, A.Y. Johnson, Gait recognition using static, activity-specific parameters, in *Proceedings of IEEE Computer Vision and Pattern Recognition Conference (CVPR'01)*, Kauai, 2001, pp. 423–430
4. L. Lee, W.E.L. Grimson, Gait analysis for recognition and classification, in *Proceedings of Automatic Face and Gesture Recognition*, Washington, DC, 2002, pp. 148–155
5. C.Y. Yam, M.S. Nixon, J.N. Carter, Automated person recognition by walking and running via model-based approaches. Pattern Recognit. **37**, 1057–1072 (2004)
6. I. Bouchrika, M.S. Nixon, Model-based feature extraction for gait analysis and recognition, in *MIRAGE: Computer Vision/Computer Graphics Collaboration Techniques and Applications, INRIA*, Rocquencourt, March 2007
7. D.K. Wagg, M.S. Nixon, On automated model-based extraction and analysis of gait, in *Proceedings of Sixth International Conference on Automatic Face and Gesture Recognition*, Seoul, 2004, pp. 11–16
8. L. Wang, T. Tan, H. Ning, W. Hu, Fusion of static and dynamic body biometrics for gait recognition. IEEE Trans. Circuits Syst. Video Technol. **14**(2), 149–158 (2004). (Special Issue on Image- and Video-Based Biometrics)
9. M. Isard, A. Blake, CONDENSATION – conditional density propagation for visual tracking. Int. J. Comput. Vis. **29**(1), 5–28 (1998)
10. R. Urtasun, P. Fua, 3D tracking for gait characterization and recognition, in *Proceedings of Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, Seoul, 2004, pp. 17–22
11. A. Kale, A.K. RoyChowdhury, R. Chellappa, Fusion of gait and face for human identification, in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, vol. 5, May 2004, pp. 901–904
12. G. Shakhnarovich, L. Lee, T. Darrell, Integrated face and gait recognition from multiple views, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, vol. 1, 2001, pp. 439–446

Gait Recognition, Motion Analysis for

Ahmed Elgammal
Department of Computer Science, Rutgers University, Piscataway, NJ, USA

Synonyms

Appearance-based gait analysis; Silhouette analysis for gait recognition

Definition

The appearance of gait in an image sequence is a spatiotemporal process that characterizes the walker. The spatiotemporal characteristics of gait contain rich perceptual information about the body configuration, the person's gender, the person's identity, and even the emotional states of the person. Motion analysis for gait recognition is a computer vision task that aims to capture discriminative spatiotemporal features (signature) from image sequences in order to achieve human identification. Such a signature ought to be invariant

to the presence of various viewing conditions, such as viewpoint, people clothing, etc. In contrast to model-based gait analysis systems, which is another article, the goal here is to capture gait characteristics without fitting a body model or locating the body limbs, rather by analyzing the feature distribution over the space and time extent of the motion.

Human Gait as a Biometric

Human gait is a valuable biometric cue that has the potential to be used for human identification similar to other biometric features, such as faces and fingerprints. Gait has significant advantages compared to other biometric features since it is easily observable in an unintrusive way, it does not require collaborative subjects, and it is difficult to disguise [1]. Therefore, using gait as a biometric feature has a great potential for human identification in public places for surveillance and for security. A fundamental challenge in gait recognition is to develop robust algorithms that can extract visual gait features invariant to the presence of various conditions that affect people's appearance, as well as conditions that affect people's gait. That includes viewpoint, clothing, walking surface, shoe type, object carried, etc. [2].

Johansson's seminal psychophysical experiments [3] showed that humans can recognize biological motion, such as gait, from Moving Light Displays (MLD). Cutting and Kozlowski [4] showed that humans can also identify friends from their gait using MLD. Motivated by these results, many researchers in different disciplines, have shown that the spatiotemporal characteristics of gait contain rich perceptual information about the body configuration, the person's gender, the person's identity, and even the emotional states of the person. That motivated extensive recent computer vision research on extracting features from gait.

Vision-based human motion tracking and analysis systems have promising potentials for many applications, such as visual surveillance in public area, activity recognition, sport analysis, video retrieval, and human-computer interaction. Extensive research has been done in this area in the last two decades with lots of promising results. For excellent literature surveys in the subject, the reader can refer to [5, 6]. The human body is an articulated object with a large number of degrees of freedom. This fact makes the problems of tracking the body configuration and extracting biometrics very challenging. Besides the articulation nature of the body, the variability in people's appearance adds to the problems. Human gait is a special case of the general problem of human motion analysis, and to some extent, is easier. This is because of the physical constraints on such a motion as well as the periodic nature of it.

The appearance of gait in an image sequence is a spatiotemporal process that characterizes the walker. Gait recognition algorithms, generally, aim to capture discriminative spatiotemporal features (signature) from image sequences in order to achieve human identification. Gait analysis approaches can be categorized according to the way the gait features are extracted for classification. There are two broad categories of approaches: model-based approaches and appearance-based approaches. Model-based approaches, e.g., [1], fit 3D body models or intermediate body representations to body limbs in order to extract proper features (parameters) that describe the dynamics of the gait (see the related entry on "Model-based Gait Recognition" for details). Model-based approaches typically require a large number of pixels on the tracked target to fit their model, i.e., high resolution zoomed-in images are required on the tracked person. In contrast, appearance-based approaches aim to capture a spatiotemporal gait characteristic directly from input sequences without fitting a body model. The appearance-based approaches are mainly motivated by the psychophysical experiments, mentioned earlier, e.g., [3, 4], which showed that spatiotemporal patterns such as moving light displays could capture important gait information without the need of finding limbs. Appearance-based approaches do not require high resolution on subjects, which makes them more applicable in outdoor surveillance

Gait Recognition, Motion Analysis for, Fig. 1 Twenty sample frames from a walking cycle from a side view. Each row represents half a cycle. Notice the similarity between the two half cycles. The right part shows the similarity plot: each row and column of the plot corresponds to one sample. *Darker* means closer distance and *brighter* means larger distances. The *two dark lines* parallel to the diagonal show the similarity between the two half cycles

applications where the subjects can be at a large distance from the camera.

Characteristics and Challenges of Gait Motion

Gait is a 3D articulated periodic motion that is projected into 2D image sequences. Therefore, the appearance of a gait motion in an image sequence is a spatiotemporal pattern, i.e., a spatial distribution of features that changes over time. Researchers have developed several algorithms for capturing gait signature from such spatiotemporal patterns by looking at the space-time volume of features. The observed shapes of the human body, in terms of the occluding contours of the body (silhouettes), are examples of such spatiotemporal patterns, which contain rich perceptual information about the body configuration, the motion performed, the person's gender, the person's identity, and even the emotional states of the person. Objects occluding contours, in general, have a great role in perception [7] and have been traditionally used in computational vision, besides other appearance cues, to determine object category and pose.

The objective of any gait tracking and analysis system is to track the global deformations of contours over time and to capture invariant gait signature from such contours. There are several challenges to achieve this goal.An observed person's contour in a given image is a function of many factors, such as the person's body build (tall, short, big, small, etc.), the body configuration, the person's clothing, and the viewpoint. Such factors can be relevant or irrelevant depending on the application. Modeling these sources of variabilities is essential to achieve successful trackers and to extract gait biometric features. Modeling the human body dynamic shape space is hard, since both the dynamics of shape (different postures) and the static variability in different people's shapes have to be considered. Such shape space lies on a nonlinear manifold.

Figure 1 shows an example of a walking cycle from a side view where each row shows half a walking cycle. The shapes during a gait cycle temporally undergo deformations and self-occlusion. The viewpoint from which the gait is captured imposes self-similarity on the observed shapes over time. This similarity can be noticed by comparing the corresponding shapes at the two rows in Fig. 1. This right part of the figure shows the correlation between these shapes. The similarity between the corresponding shapes in the two half cycles is exhibited by the dark diagonally parallel bands in the correlation plot. The similarity in the observed shapes indicates a nonlinear relation between the observed gait and the kinematics of the gait. This can be noticed by closely inspecting the two shapes in the middle of the two rows in Fig. 1. These two shapes correspond to the farthest points in the walking cycle kinematically (the top has the right leg in front while the bottom has the left leg in front). In the Euclidean visual input space (observed shapes), these two points are very close to each other as can be noticed from the distance plot on the right of Fig. 1. This nonlinear relation between the observed shapes and the kinematics poses a problem to

gait tracking and analysis systems. However, such similarity can be useful in extracting gait features. For example, the temporal self-similarity characteristic has been exploited in the work of BenAbdelkader et al. [8] for gait recognition.

Extracting Gait Signature from Motion

There have been extensive research on appearance-based extraction of gait signatures. Typical preprocessing steps for gait analysis include detecting and tracking the human subject in order to locate a bounding box containing the motion and/or extracting the body silhouette (see the related entry on human detection and tracking).

One of the early papers on gait analysis using spatiotemporal features is the work of Niyogi and Adelson [9] where a spatiotemporal pattern (corresponding to leg motion) was used to detect gait motion in an image sequence represented as an XYT volume. Gait was then parameterized with four angles for recognition. Murase and Sakai [10] used a parametric eigenspace representation to represent a moving object using principle component analysis (PCA). In their work, the extracted silhouettes were projected into an eigenspace where a walking cycle forms a closed trajectory in that space. Spatiotemporal correlations between a given trajectory and a database of trajectories were used to perform the recognition. Huang et al. [11] extended the method using canonical space transformation (CST) based on canonical analysis (CA), with eigenspace transformation for feature extraction.

Little and Boyd [12] exploited the spatial distribution of optical flow to extract spatiotemporal features. From dense optical flow, they extracted scale-independent features capturing the spatial distribution of the flow using moments. This facilitates capturing the spatial layout of the motion or as they call it "the shape of the motion." Periodicity analysis was then done on these features to capture gait signatures for recognition. BenAbdelkader et al. [8] used image self-similarity plots (similar to Fig. 1) to capture the spatiotemporal characteristics of gait. Given bounding boxes around a tracked subject, correlation is used to measure self-similarity between different time frames in the form of similarity plots. PCA analysis was used to reduce the dimensionality of such similarity plots for recognition. Hayfron-Acquah et al. [13] used spatial symmetry information to capture gait characteristics from silhouettes. Given a walking cycle, a symmetry operator was used to extract a symmetry map for each silhouette instance in the cycle. Fourier transform was used to extract descriptors from such symmetry maps for recognition.

Since gait is a temporal sequence, researchers have investigated the use of hidden Markov models (HMM) to represent and capture gait motion characteristics. HMMs have been successfully used in many speech recognition systems, as well as gesture recognition applications. Typically a left-right HMM with a small number of states (three to five) is sufficient to model the gait of each subject in the database, where the HMMs are trained from features extracted from silhouettes. In [14], HMM was used to capture gait dynamics from quantized Hu moments of silhouettes. HMM was also used in [15] with features representing silhouette width distribution.

Lee and Elgammal [16] used bilinear and multilinear models to factorize the spatiotemporal gait process into gait style and gait content factors. A nonlinear mapping was learned from a unit circle (representing a gait cycle) to the silhouettes' shape space. The unit circle represents a unified model for the gait manifold of different people; therefore, any spatiotemporal characteristics of the gait of a specific person should exist on the mapping space. Bilinear and multilinear models were used to factorize such mapping to extract gait signatures.

Manifold-Based Representation for Gait Analysis

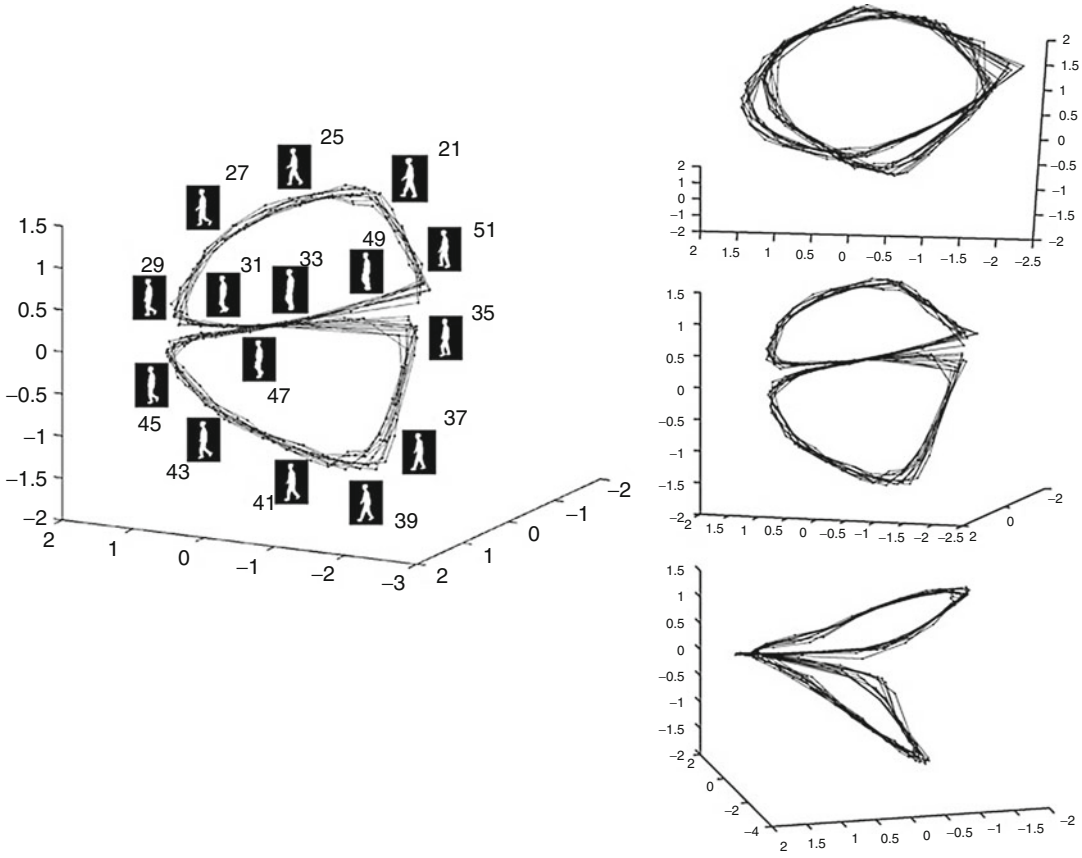
Despite the high dimensionality of the human body configuration space, any body motion is constrained by the physical dynamics, body

constraints, and the motion type. Therefore, many human activities lie intrinsically on low-dimensional manifolds. This is true for the body kinematics as well as for the observed motion through image sequences. For certain classes of motion like gait, facial expression, and simple gestures, considering a single person and factoring out other sources of variability, the deformations will lie on a one-dimensional manifold. Recently many researchers have developed techniques and representations for gait analysis that exploit such manifold structure, whether in the visual space or in the kinematic space, e.g. [17, 18]. Modeling the gait manifold was earlier used for gait recognition in [10].

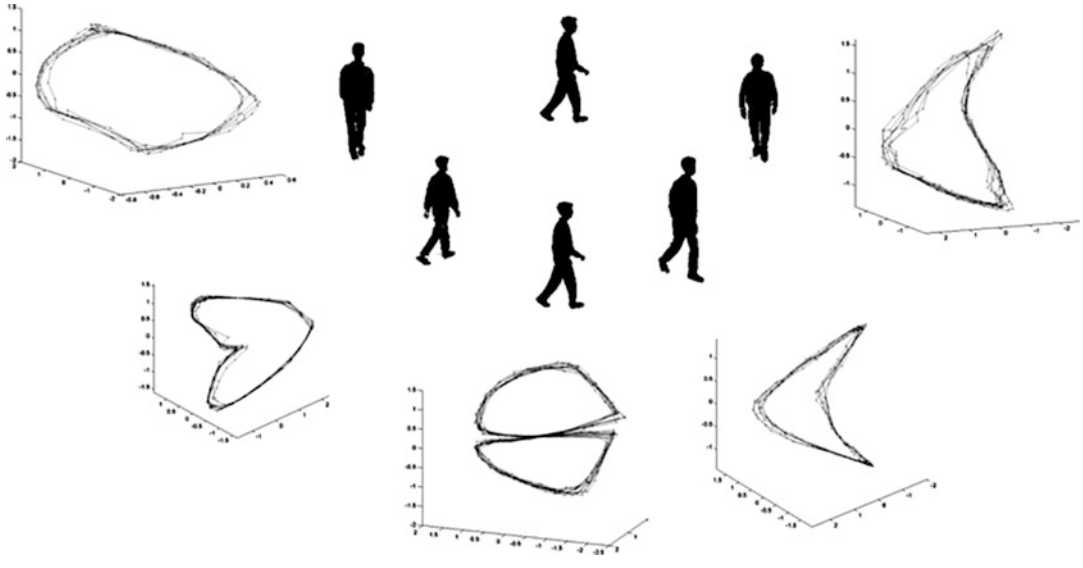
Intuitively, the gait is a one-dimensional closed manifold that is embedded in a high-dimensional visual space. Such a manifold can twist and self-intersect in such high-dimensional visual space. This can be noticed by considering the human silhouette through the walking cycle (as shown in Fig. 1) as points in a high-dimensional visual input space. Given the spatial and the temporal constraints, it is expected that these points will lay on a closed trajectory. In order to achieve a low-dimensional embedding of the gait manifold (manifold embedding), dimensionality reduction techniques can be used. Linear dimensionality reduction can be used to achieve an embedding, as in [10]. However, in such a case the two half cycles would be collapsed to each other because of the similarity in the shape space. Nonlinear dimensionality reduction techniques such as LLE [19], Isomap [20], GPLVM [21], and others can successfully embed the gait manifold in a way that separates the two half cycles. As a result of nonlinear dimensionality reduction, an embedding (and a visualization) of the gait manifold can be obtained in a low-dimensional Euclidean space [17]. Figure 2 illustrates an example embedded manifold for a side view of the walker. The data used are from the CMU Mobo gait data set which contains 25 people from six different view points. Data sets of walking people from multiple views are used in this experiment. Each data set consists of 300 frames and each containing about 8–11 walking cycles of the same person from a certain view points. The walkers were using treadmill which might result in different dynamics from the natural walking. Figure 3 illustrates the embedded manifolds for five different view points of the walker. For a given view point, the walking cycle evolves along a closed curve in the embedded space; i.e., only one degree of freedom controls the walking cycle, which corresponds to the constrained body pose as a function of the time. Such a conclusion conforms to the intuition that the gait manifold is one dimensional.

As can be noticed in Fig. 3, The manifold twists in the embedding space given the different viewpoints, which impose different self-occlusions. The least twisted manifold is the manifold for the back view as this is the least self-occluding view (left most manifold in Fig. 3. In this case the manifold can be embedded in a two-dimensional space. For other views, the curve starts to twist to be a three-dimensional space curve. This is primarily because of the similarity imposed by the view point which attracts far away points on the manifold closer. The ultimate twist happens in the side view manifold where the curve twists to get the shape of the numeral 8 where each cycle of the eight (half eight) lies in a different plane. Each half of the "eight" figure corresponds to half a walking cycle. The cross point represents the body pose where it is totally ambiguous from the side view to determine from the shape of the contour which leg is in front, as can be noticed in Fig. 2. Therefore, in a side view, a three-dimensional embedding space is the least that can be used to discriminate the different body poses. Embedding a side view cycle in a two-dimensional embedding space results in an embedding similar to that shown in top right of Fig. 2 where the two half cycles lie over each other. Interestingly, despite that the side view is the most problematic view of the gait, most gait recognition systems seem to favor such view for recognition! Different people are expected to have different manifolds. However, such manifolds are all topologically equivalent.

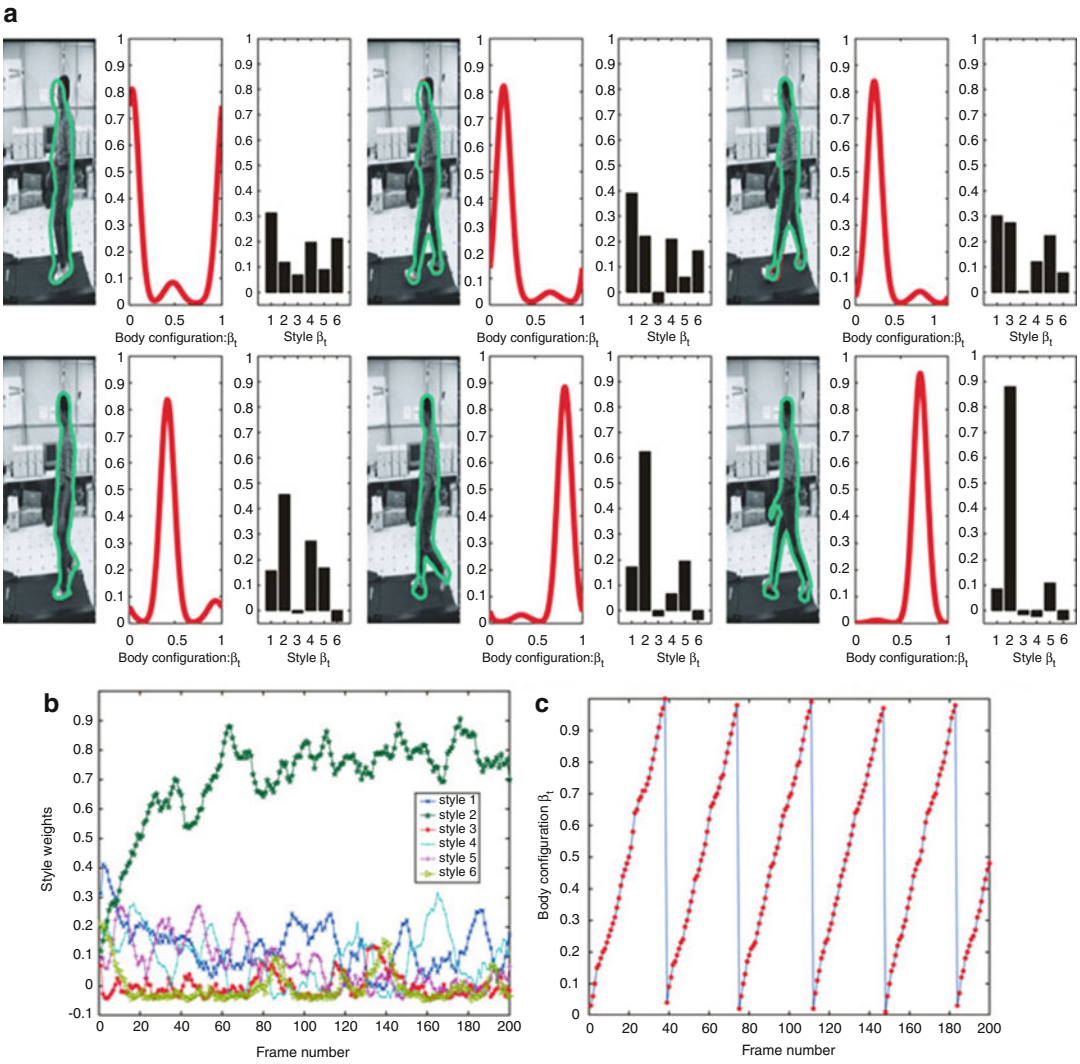
The example embeddings shown here are for silhouette data, i.e., the *visual manifold* of the gait is embedded. Similar embedding can be



Gait Recognition, Motion Analysis for, Fig. 2 Embedded gait manifold for a side view of the walker. *Left*: sample frames from a walking cycle along the manifold with the frame numbers shown to indicate the order. Ten walking cycles are shown. *Right*: three different views of the manifold (© IEEE)



Gait Recognition, Motion Analysis for, Fig. 3 Embedded manifolds for five different views of the walkers. Frontal view manifold is the rightmost one and back view manifold is the leftmost one. The view of the manifold that best illustrates its shape in the 3D embedding space is visualized (© IEEE)



Gait Recognition, Motion Analysis for, Fig. 4 Adaptive Contour Tracking of Gait: (**a**) tracking through sample frames. (**b**) Adapting to the target style. (**c**) The tracked body configuration showing a constant speed dynamic system (From [22])

obtained for kinematic data, in such a case the *kinematic manifold* of the gait is embedded. In such a case PCA would be sufficient to achieve an embedding. The importance of such embedded representations is that they provide a low-dimensional representation for tracking the gait motion. Only a one-dimensional parameter is needed to control and track the gait motion. This leads to a simple constant speed dynamic model for the gait. Figure 4 shows an example of gait contour tracking system [22] that uses an embedded representation of the gait manifold. As a result, a constant speed linear dynamics is achieved (Fig. 4b). The tracker can also adapt to the tracked person shape style and identify that style from a database of styles (Fig. 4c).

Explicit manifold representation for gait is not only useful for tracking and pose estimation, but also can be used in gait recognition systems. Different people are expected to have different manifolds for the appearance of their gait. However, such manifolds are all topologically

equivalent to a unit circle. A person's gait manifold can be thought of as a twisted circle in the input space. The spatiotemporal process of gait is captured in the twist of a given person's manifold. Therefore, a person's gait signature can be captured by modeling how a unit circle (an ideal manifold) can deform to fit that person's gait manifold. This can be achieved by fitting a nonlinear warping function between a unit circle and a given person's silhouette sequence. In [23] this approach was used to capture gait signatures by factorizing the warping functions' coefficient space to obtain a low-dimensional gait signature space for recognition.

Summary

Appearance-based analysis of gait is motivated and justified by psychophysical experiments. Appearance-based approaches for gait recognition aim to extract a gait signature from the spatial and temporal distribution of the features on a tracked subject without the need to fit a body model or to locate limbs. Such approaches have proved very successful in gait recognition and are applicable in scenarios where the gait biometric features can only be extracted from a distance. There are many limitations to the current gait recognition systems including achieving invariant to viewing conditions, such as viewpoint invariant. Recent progress in manifold-based representation of gait, as well as factorized models, such as multilinear tensor models provides potential solutions to such problems.

Related Entries

▸ Gait Recognition, Model-Based
▸ Human Detection and Tracking

References

1. D. Cunado, M.S. Nixon, J. Carter, Automatic extraction and description of human gait models for recognition purposes. Compt. Vis. Image Underst. **90**, 1–41 (2003)
2. S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, K.W. Bowyer, The humanid gait challenge problem: data sets, performance, and analysis. IEEE Trans. PAMI **27**(2), 162–177 (2005)
3. G. Johansson, Visual motion perception. Sci. Am. **232**, 76–88 (1975)
4. J.E. Cutting, L.T. Kozlowski, Recognizing friends by their walk: gait perception without familiarity cues. Bull. Psychon. Soc. **9**(5), 353–356 (1977)
5. D.M. Gavrila, The visual analysis of human movement: a survey. Comput. Vis. Image Underst. **73**(1), 82–98 (1999). doi:http://dx.doi.org/10.1006/cviu.1998.0716
6. T.B. Moeslund, A. Hilton, V. Krüger, A survey of advances in vision-based human motion capture and analysis. Comput. Vis. Image Underst. **104**(2), 90–126 (2006). doi:http://dx.doi.org/10.1016/j.cviu.2006.08.002
7. S.E. Palmer, *Vision Science, Photons to Phenomenology* (MIT, Cambridge, 1999)
8. C. BenAdbelkader, R. Cutler, L. Davis, Motion-based recognition of people using image self-similarity, in *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, Washington, DC, 2002, pp. 254–259
9. S. Niyogi, E. Adelson, Analyzing and recognition walking figures in xyt, in *Proceedings of IEEE CVPR*, Seattle, 1994, pp. 469–474
10. H. Murase, R. Sakai Moving object recognition in eigenspace representation: gait analysisand lip reading. Pattern Recogn. Lett. **17**, 155–162 (1996)
11. P. Huang, C. Haris, M. Nixon, Recogising humans by gait via parametric canonical space. Artif. Intell. Eng. **13**, 359–366 (1999)
12. J.J. Little, J.E. Boyd, Recognizing people by their gait: the shape of motion. Videre J. Comput. Vis. Res. **1**(2), 2–32 (1998)
13. J.B. Hayfron-Aquah, M.S. Nixon, J.N. Carter, Automatic gait recognition by symmetry analysis. Pattern Recogn. Lett. **24**, 2175–2183 (2003)
14. Q. He, C. Debrunner, Individual recognition from periodic activity using hidden markov models, in *Proceedings of the Workshop on Human Motion*. (Humo) (Dec 2007), HUMO, Rio de Janeiro. IEEE Computer Society, Washington, DC
15. A. Kale, A. Sundaresan, A.N. Rajagopalan, N.P. Cuntoor, A.K. Roy-Chowdhury, V. Kruger, R. Chellappa, Identification of human using gait. IEEE Trans. Image Process. **13**(9), 1163–1173 (2004);
16. C.S. Lee, Elgammal, A.: Gait style and gait content: bilinear model for gait recogntion using gait re-sampling, in *Proceedings of FGR*, Seoul, 2004, pp. 147–152
17. A. Elgammal, C.S. Lee, Inferring 3d body pose from silhouettes using activity manifold learning, in *Proceeedings of CVPR*, Washington, DC, vol. 2, 2004, pp. 681–688
18. R. Urtasun, D.J. Fleet, A. Hertzmann, Fua, P.: Priors for people tracking from small training sets, in *ICCV*, Beijing, 2005, pp. 403–410

19. S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linar embedding. Science **290**(5500), 2323–2326 (2000)
20. J. Tenenbaum, Mapping a manifold of perceptual observations, in *Proceedings of Advances in Neural Information Processing (NIPS)*, Denver, vol. 10, 1998, pp. 682–688
21. N.D. Lawrence, Gaussian process models for visualisation of high dimensional data, in *Proceedings of NIPS*. BMVC, Oxford, 2004
22. C.S. Lee, A. Elgammal, Style adaptive bayesian tracking using explicit manifold learning, in *Proceedings of British Machine Vision Conference*, Oxford, 2005
23. C.S. Lee, A. Elgammal, *Audio- and Video-Based Biometric Person Authentication Conference AVBPA*, Terrytown, 2005

Gait Recognition, Silhouette-Based

Jeffrey E. Boyd[1] and James J. Little[2]
[1]University of Calgary, Calgary, AB, Canada
[2]University of British Columbia, Vancouver, BC, Canada

Synonyms

Appearance-based gait analysis; Silhouette Analysis-Based Gait Recognition for Human Identification

Definition

Silhouette-based gait recognition is the analysis of walking human figures for the purpose of biometric recognition. Gait biometrics offers the advantage of covertness; acquisition is possible without the awareness or cooperation of the subject. The analysis may apply to a single static image or to a temporal sequence of images, i.e., video.

Introduction

The phenomenon of gait is the "coordinated, cyclic combination of movements that result in human locomotion" [1]. Gait is necessary for human mobility and is therefore ubiquitous and easy to observe.

The common experience of recognizing a friend from a distance by the way they walk has inspired the use of gait as a biometric feature. In fact, Cutting and Kozlowski [2], using moving light displays to isolate the motion stimulus, demonstrated that humans can indeed identify familiar people from gait. In their experiments, seven subjects identified the gaits of a subset of six subjects correctly at a rate of 38 %. While this rate is less than adequate for biometrics, it is significantly better than random (17 % in for their sample size) and validates the human source of inspiration.

To convert a gait into a feature vector suitable for biometrics, one can *characterize the motion* in the gait, e.g., by analyzing joint angles and limb trajectories or by measuring the overall pattern of motion. Alternatively, one can *measure critical body dimensions* such as height or limb lengths. In the later approach, biometric features can be measured statically, but the motion in the gait provides a convenient mechanism to reveal joint positions and, consequently, limb lengths. McGeer's work on *passive dynamic walkers* [3, 4] reveals the extent to which gait motion relates to body mass and limb lengths: in the passive dynamic model of a human gait, the motion is a stable limit cycle that is a direct result of body mass and limb length. Factors not accounted for in McGeer's original model are muscle activation (gravity powers a passive dynamic walker), walking surface, injury, and fatigue. Intuitively, the motion in a gait is a reflection of the mass and skeletal dimensions of the walker. McGeer's passive dynamic model leads to more sophisticated models that account for some of these other factors. For example, see the work of Kuo [5, 6].

Confounding factors in gait biometrics include clothing and footwear. Clothing can change the observed pattern of motion and make it difficult to accurately locate joint positions. The effect of footwear is more complex. Some variation in footwear causes changes in muscle activation but

causes no outwardly visible change in the pattern of motion [7], whereas other footwear changes will alter gait.

Silhouette-based gait recognition extracts the form of a walking subject and then computes a feature vector that describes either the pattern of motion in the gait or the physical dimensions of the subject. A classifier then matches the feature vector against previously acquired examples for identification or verification.

Silhouettes

Definitions of silhouette are often ambiguous: some definitions refer to the region covered by a figure, whereas other definitions refer to the boundary between a figure and its background. In the context of silhouette-based gait biometrics, we assume that the silhouette refers to the region, rather than the border. Nevertheless, there are related examples that use the boundary, e.g., see Baumberg and Hogg [8].

To form a silhouette of a walking figure requires the segmentation of image pixels into foreground (the moving figure) and background (everything else) sets of pixels. The silhouette is the set of foreground pixels. The easiest way to acquire a reliable silhouette is *chroma-keying* [9], which relies on color disparities between a backdrop and the foreground subject. The background color (usually green or blue) is chosen to make the color discrimination robust. Figure 2d shows an example of chroma keying in gait analysis. The unusual color of the backdrop makes the subject aware that they are under surveillance, negating the covertness of gait biometrics.

Background subtraction obviates the need for a colored backdrop by measuring the naturally occurring scene behind the subject. This entails estimating the statistical properties (usually in the luminance and color) of every pixel over one or more frames of video. By comparing the background estimate with subsequent frames of video, one can classify foreground pixels as those that do not match the background. The classifier can be as simple as thresholding of the absolute difference between the background and video frames. In most cases, the background estimation and subtraction are merged into an online system that continuously computes pixel differences and then updates the background for each frame of video. Background subtraction requires that the background and camera be stationary. Stauffer and Grimson [10] describe a widely used background subtraction method that uses a multimodal estimate of background statistics to produce reliable silhouettes of moving objects. Their method is robust in the presence of some background motions (e.g., rustling leaves or swaying tree branches).

The projection of motion in a scene onto a camera image plane is called a motion field. When a human figure is walking, segmenting moving from slow or stationary pixels in the motion field will extract a silhouette of the figure. Additionally, a motion field provides richer information than a simple silhouette because it indicates not only where the subject is moving but also how fast the various body parts are moving. In general, it is not possible to measure a motion field, but one can measure optical flow, an approximation to the motion field that is sufficient for biometric gait recognition. If one imagines the luminance of pixels to be a fluid that can flow around an image, the optical flow estimates the movement of that fluid. It is, in part, related to the motion field, but is not necessarily equal to the motion field in all cases. Barron et al. [11] provide a comparative survey of some well-known optical flow algorithms. For example, see Fig. 2a.

Most silhouette-based biometric gait analysis focuses on a view of the subject orthogonal to the sagittal plane of the subject, i.e., the subject walks across the field of view rather than toward or away from the camera. We believe that this preference exists because front or rear views of the subject show mostly side-to-side motion and do not reveal either joint location or the complex patterns of limb motion.

Marker-based motion capture, e.g., Johansson's moving light displays, offers a counterpoint to silhouettes that are less practical

for biometrics but are useful for gaining insight into the perceptual issues surrounding gait [12, 13].

Duration of Observation

In general, it is desirable to observe the gait as long as possible. One way to extend the duration of an observation indefinitely is to have a subject walk on a treadmill in front of a stationary camera, e.g., see Fig. 2b. However, this requires the cooperation and awareness of the subject.

Alternatively, allowing the camera to pan with the motion of the subject can extend the observation time without the subject walking on a special apparatus. However, when the camera moves, the images acquired contain both the movement of the subject and the background. The changing background makes accurate background subtraction difficult.

Using a static camera simplifies both the apparatus and the processing to extract the silhouette, but the duration of observation is limited by the time it takes the subject to cross the field of view of the camera. The actual duration will vary with the angular width of the field of view, the distance between the subject and the camera, and the speed of the subject. The practical limit on distance to subject depends on the resolution of the camera. Higher resolutions allow the subject to be further away while maintaining enough pixel coverage to measure biometric feature vectors accurately. In examples reported in the literature that use a static cameras and subjects walking on the ground, the typical duration of observation is approximately three to six strides.

Periodicity and Synchronization

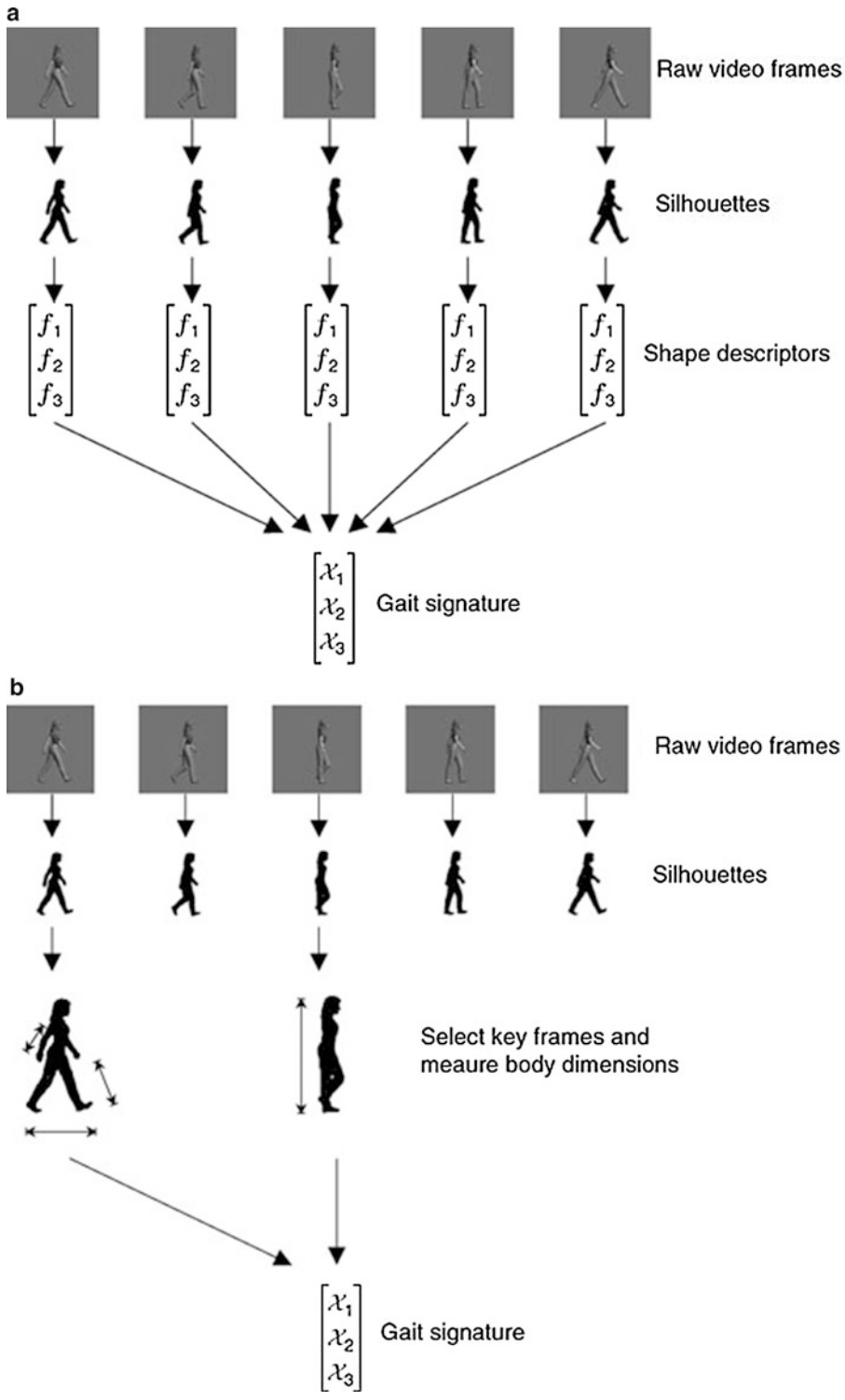
Gait is a periodic phenomenon, so the silhouette of a walker varies with position in the gait cycle. Consequently, it is necessary to synchronize measurements of the silhouette to positions in the gait cycle. In turn, this requires measurement of the frequency of the gait and establishment of a phase reference within the gait cycle.

The method used to perform the synchronization depends on the particular measurements acquired and can serve to differentiate gait analysis methods. For example, Little and Boyd [14] measure the frequency from the oscillations of the centroid of the figure. To establish a phase reference, they use the phase of an oscillating measurement. In methods that measure height, e.g., Ben-Abdelkader et al. [15], the frequency of oscillations of the figure height gives the frequency of the gait. Positions of maxima in the height correspond to the positions in the gait where the swinging leg is vertical, thus defining a phase reference.

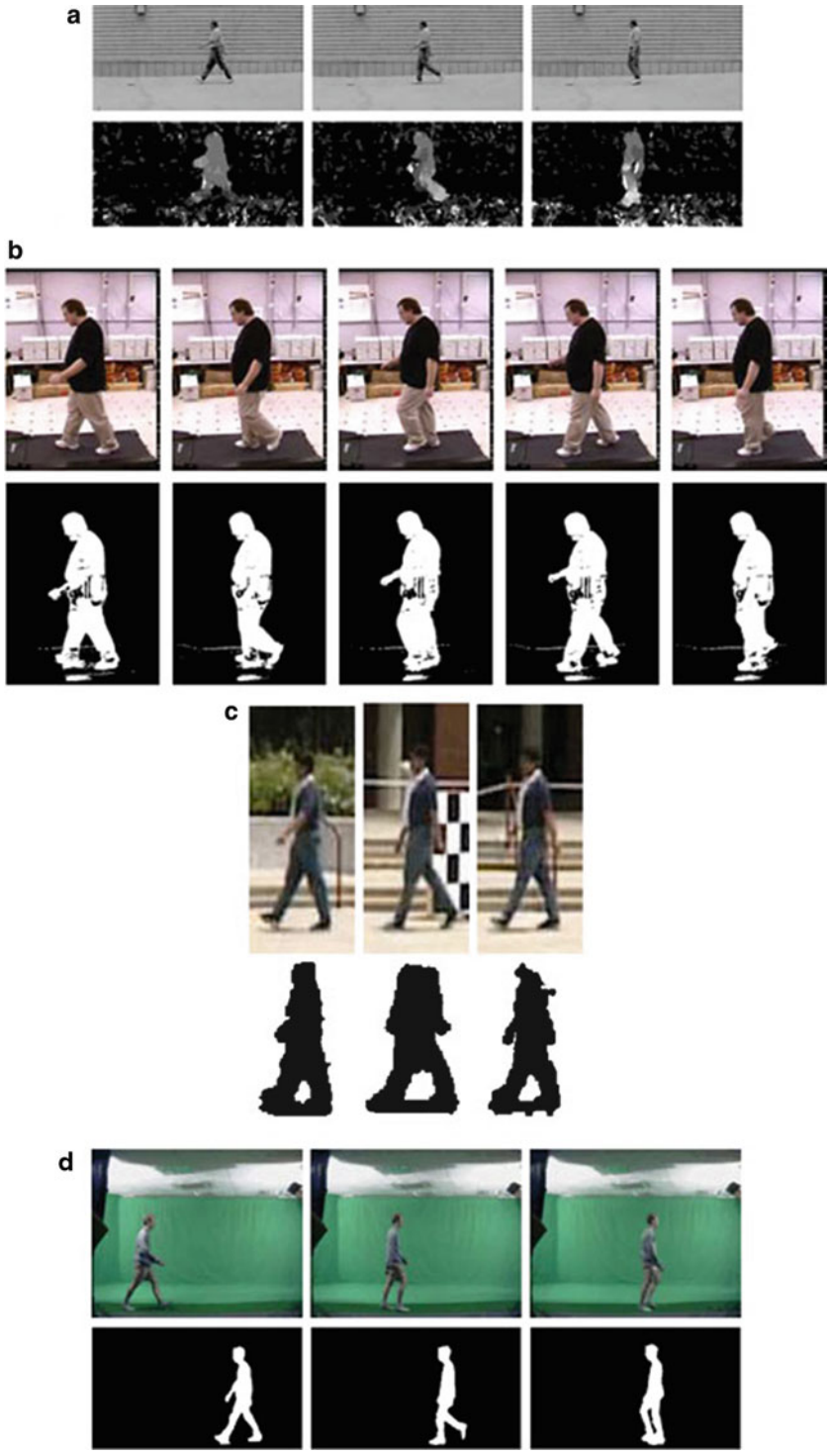
Conversion of Silhouettes to Features

A necessary step in silhouette-based gait recognition is conversion of a temporal sequence of silhouettes into a *gait signature*, i.e., a feature vector suitable for classification. One approach is to extract features that characterize the silhouette shapes and their variation over time, as illustrated schematically in Fig. 1a.

As an example, Little and Boyd [14] use geometric moments to describe a silhouette within a single frame of video. The moments include geometric centers, i.e., the average position of pixels in the silhouette, sometimes called the center of mass. Weighting the pixel positions by corresponding optical flow values gives geometric moments sensitive to rapid limb movement. Little and Boyd also use eccentricity [16], based on higher-order geometrical moments. A further step is necessary to combine the shape description for silhouettes in individual frames to a feature vector representative of the entire gait. Cyclic oscillations in the silhouette shape moments result naturally from a gait, so Little and Boyd exploit this to collect the individual shape descriptions into a single feature vector of the relative phases of the moment oscillations. Shutler and Nixon [17] describe a variation on this approach that uses Zernike moments to represent an accumulated shape over the duration of a gait cycle.



Gait Recognition, Silhouette-Based, Fig. 1 Themes in silhouette-based gait recognition: (**a**) shape descriptors of the silhouette combine to form a gait signature from the motion of the gait, or (**b**) critical body dimensions are measured from key frames within the gait cycle. Existing methods use variations on both of these themes and can even combine them.



Gait Recognition, Silhouette-Based, Fig. 2 Example images from gait databases suitable for testing silhouette-based gait recognition: (**a**) Little and Boyd [14], (**b**) MOBO [21], (**c**) HumanID Gait Challenge [22], and (**d**) Shutler et al. [23]. All examples show raw video images in the top row and silhouettes or magnitude of the optical flow (Little and Boyd only) in the bottom row. The silhouettes shown for the Shutler et al. do not correspond to the images above.

Ben-Abdelkader et al. [18] also exploit the periodic nature of a gait to form feature vectors. Periodicity and symmetry in a gait mean that similar shapes occur throughout the cycle of a gait. A feature vector built from measures of the silhouette self-similarity over period forms the basis for gait recognition. Periodicity in the self-similarity measures establishes the frequency of the gait. Hayfron-Acquah et al. [19] characterize the silhouette shape in a single frame by measuring symmetries in the outline of the silhouette to produce a *symmetry map*. The average of these symmetry maps over a gait cycle gives the gait signature used for recognition. Boyd [20] uses an array of phase-locked loops to measure the frequency, amplitude, and phase of pixel intensity oscillations due to a gait. The amplitudes and relative phases form a vector of complex phasors that acts as gait signature for recognition.

Rather than relying on the connection between gait and body structure to form a gait signature, one can use feature vectors that relate directly to body dimensions as shown in Fig. 1b. For example, Bobick and Johnson [21] measure stride and torso lengths, and Ben-Abdelkader et al. [15] measure height and stride characteristics. Collins et al. [22] identify key frames in a gait sequence for both the double-support (two feet on the ground) and mid-stride phase of a gait. From these key frames they measure cues related to height, width, and other body proportions and movement-related characteristics such as stride length and amount of arm swing.

Data Sets

A database of sample gaits is essential for developing a silhouette-based gait recognition system. Little and Boyd [14] provided one of the earliest databases featuring seven sample gaits for each of six subjects, for a total of 42 gait sequences (Fig. 2a).

Gross and Shi [23] created the Motion of Body (MOBO) database (Fig. 2b). It features gait samples for 25 subjects. Each subject walks on a treadmill under four different conditions (slow, fast, on an incline, and carrying a ball) and from a variety of viewing angles. Segmented silhouettes are part of the database.

Sarkar et al. [24] present a large (1.2 GB) gait database as part of the *HumanID Gait Challenge Problem* associated with the Defense Advanced Research Projects Agency (DARPA) HumanID project (Fig. 2c). The database contains samples for 122 subjects acquired in multiple sessions and under variable conditions. The *challenge problem* specifies a series of tests using the database as well as a reference algorithm to facilitate comparative testing by researchers.

Shutler et al. [25] created a database featuring over 100 subjects (Fig. 2d). The database contains sequences acquired over multiple sessions and features subjects walking from both left to right and right to left. Subjects walk on the ground or on treadmills and in front of green screens (for chroma keying) or in outdoor scenes.

Examples

Bhanu and Han [26] estimate upper bounds on the performance of gait recognition by equating gait with body dimensions, presented as plots of recognition rate versus gallery size for varying assumptions of accuracy. As one might expect with upper bounds, these rates are optimistic. Random guessing is a good lower bound on performance, but any practical biometric system must be much better. One can reasonably expect that a gait biometric system should perform at least as well as humans on moving light displays [2], i.e., 38 % from a gallery of six.

Within these broad bounds, there are numerous examples of existing silhouette-based gait recognition systems. Most of these have been tested with one or more of the databases mentioned earlier. Examples include the work of Hayfron-Acquah et al. [19], Shutler and Nixon [17], Collins et al. [22], Bobick and Johnson [21], Ben-Abdelkader et al. [15, 18], Liu and Sarkar [27], Robledo and Sarkar [28], Lee and Grimson [29], Little and Boyd [14, 20], and Wang et al. [30]. The best reported correct classification rates

(CCR) are better than 90 % from a gallery of approximately 100 people.

Summary

Human experience supported by psychological observation suggests that humans can be recognized by their gaits, which inspires gait biometric systems. Silhouette-based gait recognition systems convert images from a video gait sequence to silhouettes of the walker. Dynamic shape or body dimensions are measured from the silhouettes and combined to form a gait signature used for recognition. There are several databases available for testing silhouette-based gait recognition and numerous published examples of successful recognition using these databases.

Related Entries

- ▸ Gait Recognition, Motion Analysis for
- ▸ Gait Recognition, Model-Based
- ▸ Human Detection and Tracking

References

1. J.E. Boyd, J.J. Little, Biometric gait recognition, in *Advanced Studies in Biometrics: Summer School on Biometrics*, Alghero, June 2–6, 2003. Revised Selected Lectures and Papers (Lecture Notes on Computer Science), vol. 3161/2005 (Springer, 2005), pp. 19–42
2. J.E. Cutting, L.T. Kozlowski, Recognizing friends by their walk: gait perception without familiarity cues. Bull. Psychon. Soc. **9**(5), 353–356 (1977)
3. T. McGeer, Passive dynamic walking. Int. J. Robot. Res. **9**(2), 62–82 (1990)
4. T. McGeer, Passive walking with knees, in *IEEE International Conference on Robotics and Automation*, Cincinnati, pp. 1640–1645, 1990
5. A.D. Kuo, A simple model of bipedal walking predicts the preferred speed-step length relationship. J. Biomech. Eng. **123**, 264–269 (2001)
6. A.D. Kuo, Energetics of actively powered locomotion using the simplest walking model. J. Biomech. Eng. **124**, 113–120 (2002)
7. V. von Tscharner, B. Goepfert, Gender dependent emgs of runners resolved by time/frequency and principal pattern analysis. J. Electromyogr. Kines. **13**, 253–272 (2003)
8. A.M. Baumberg, D.C. Hogg, Generating spatiotemporal models from examples, in *Sixth British Conference on Machine Vision*, Birmingham, vol. 2, pp. 413–422, 1995
9. R. Brinkmann, *The Art and Science of Digital Compositing* (Morgan-Kaufmann, San Diego, 1999)
10. C. Stauffer, W.E.L. Grimson, Adaptive background mixture models for real-time tracking, in *Computer Vision and Pattern Recognition*, Fort Collins, vol. II, pp. 246–252, 1998
11. J.L. Barron, D.J. Fleet, S. Beauchemin, Performance of optical flow techniques. Int. J. Comput. Vis. **12**(1), 43–77 (1994)
12. N.F. Troje, Decomposing biological motion: a framwork for analysis and synthesis of human gait patterns. J. Vis. **2**, 371–387 (2002)
13. L. Omlor, M.A. Giese, Extraction of spatio-temporal primitives of emotional body expressions. Neurocomputing **70**(10–12), 1938–1942 (2007)
14. J.J. Little, J.E. Boyd, Recognizing people by their gait: the shape of motion. Videre **1**(2), 1–32 (1998)
15. C. Ben-Abdelkader, R. Cutler, L. Davis, Person identification using automatic height and stride estimation, in *16th International Conference on Pattern Recognition*, Quebec, 2002, pp. 377–380
16. D.H. Ballard, C.M. Brown, *Computer Vision* (Prentice-Hall, Englewood Cliffs, 1982)
17. J.D. Shulter, M.S. Nixon, Zernike velocity moments for sequence-based descriptiion of moving features. Image Vis. Comput. **24**, 343–356 (2006)
18. C. Ben-Abdelkader, R. Cutler, L. Davis, Motion-based recognition of people in eigengait space, in *5th IEEE International Conference on Automatic Face and Gesture Recognition*, Washington, DC, pp. 267–272, 2002
19. J.B. Hayfron-Acquah, M.S. Nixon, J.N. Carter, Automatic gait recognition by symmetry analysis. Pattern Recogn. Lett. **24**, 2175–2183 (2003)
20. J.E. Boyd, Synchronization of oscillations for machine perception of gaits. Comput. Vis. Image Underst. **96**, 35–59 (2004)
21. A. Bobick, A. Johnson, Gait recognitin using static activity-specific parameters, in *Computer Vision and Pattern Recognition 2001*, Kauai, vol. I, pp. 423–430, 2001
22. R.T. Collins, R. Gross, J. Shi, Silhouette-based human identification from body shape and gait, in *Automatic Face and Gesture Recognition*, Washington, DC, pp. 351–356, 2002
23. R. Gross, J. Shi, The cmu motion of body (mobo) database. Technical report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, 2001
24. S. Sarkar, J. Phillips, Z. Liu, I. Robledo, P. Prother, K.W. Bowyer, The humanid gait challenge problem: data sets, performance, and analysis. IEEE Trans. Pattern Anal. Mach. Intell. **27**(2), 162–177 (2005)

25. J. Shutler, M. Grant, M. Nixon, J. Carter, On a large sequence-based human gait database, in *Fourth International Conference on Recent Advances in Soft Computing*, Nottingham, pp. 66–71, 2002
26. B. Bhanu, J. Han, Bayesian-based performance prediction for gait recognition, in *IEEE Workshop on Motion and Video Computing*, Orlando, pp. 145–150,2002
27. Z. Liu, S. Sarkar, Improved gait recognition by gait dynamics normalization. IEEE Trans. Pattern Anal. Mach. Intell. **28**(6), 863–876 (2006)
28. I. Robledo, S. Sarkar, Statistical motion model based on the change of feature relationships: human gait-based recognition. IEEE Trans. Pattern Anal. Mach. Intell. **25**(10), 1323–1328 (2003)
29. L. Lee, W. Grimson, Gait analysis for recognition and classification, in *Automatic Face and Gesture Recognition*, Washington, DC, pp. 148–155, 2002
30. L. Wang, T. Tan, H. Ning, W. Hu, Silhouette analysis-based gait recognition for human identification. IEEE Trans. Pattern Anal. Mach. Intell. **25**(12), 1505–1518 (2003)

Gait, Forensic Evidence of

Niels Lynnerup and Peter K. Larsen
Laboratory of Biological Anthropology, Institute of Forensic Medicine, University of Copenhagen, Copenhagen, Denmark

Synonyms

Gait analysis; Perpetrator identification

Definition

Forensic evidence of gait, or forensic gait analysis, may be defined as the analyses of gait performed in the service of the law. Usually, this involves the analyses of criminal cases with the aim to characterize the gait of a perpetrator and often to compare the gait of a perpetrator with the gait of a suspect. The results of the analyses may furthermore need to be presented in court. The methods involved in forensic gait analyses comprise morphological assessment of single gait features and kinematic assessment of body movement, often combined with photogrammetrics. The latter means that body segment lengths, stride length, etc. may be quantified and used in direct comparisons.

Introduction

Forensic analysis of gait has a lot of common ground with biometric gait recognition, but there are also some major differences. In terms of image capture, the imagery used in forensic gait analysis is mostly always acquired from CCTV, with the perpetrator specifically trying to conceal identity. In biometric systems, image capture of a person, or registration, takes place under specific circumstances, designed to maximize data quality, and obviously a person will willingly follow a set of guidelines in order to ensure proper registration. On the other hand, a bank robber might try to hide his or her face to avoid facial recognition or wear baggy clothes to blur body morphology.

Biometric gait recognition systems may operate with various false accept or reject rates, which govern how exclusive the system is and reflect the number of "wrong" registrations that can be tolerated. For example, a relatively high false reject rate (i.e., rejecting a person who otherwise should be cleared) is not a problem if the system is meant for a screening function, where rejection simply leads to an additional identity check. It is possible to generate computer models which can identify people by their gait with more than 90 % success [1, 2], but these models are still based on a small number of people and require optimal conditions seldom found outside the laboratory [3]. Alternative biometric approaches use a description of a subject's silhouette, often with reportedly improved recognition performance [4]. In forensic gait analysis, the analysis is often specifically carried out to match a perpetrator with a suspect. If the case is made that there is a match, then the suspect may be sentenced. This places a certain onus on the gait analysis and the scientists carrying

out the analyses, and the prosecution and the defense may well challenge the findings of the gait analysis. This also means that when presenting the results of a forensic gait analysis, one has to be familiar with the legal prerequisites for the legal statements and how expert evidence is adjudicated.

Technical Issues

Bank CCTV systems are often set up not to capture gait specifics, but rather to give fields of view covering office spaces, teller machines, etc. and also often to supervise the bank employees. It is a not uncommon experience when perusing CCTV footage after a bank robbery that the perpetrator is seen moving behind desks and tellers, so that only the upper part of his body is filmed. Also, the CCTV system may vary quite a lot in terms of technical quality, e.g., image capture frequencies, digital versus analog data storage, color versus b/w cameras (the latter often of sharper quality), and numerous supplier-dependent video and computer systems (e.g., in terms of data compression of video images).

The recording frequency should ideally be about 15 Hz allowing the examination of dynamic features such as lateral instability in the knee at heel strike. Others have found a similar frequency sufficient for obtaining joint angles [5] and for automatic recognition of gait [2]. Lower recording frequencies may also be sufficient to examine features that are more static, although the gait will have a "jerky" appearance. Even at a low 5 Hz recording frequency, it has proved possible to examine gait parameters such as dorsal/plantar flexion at heel strike, degree of "push-off" at toe-off, and knee flexion during stance. At even lower recording frequencies, where the images really are still image series, specific gait-related characteristics may be noticed, e.g., a perpetrator with a bowlegged left knee. This means that even just one single image of the gait can sometimes be useful, if the gait feature captured can be deemed characteristic.

Gait

The ability to recognize other individuals is fundamental to human life. Identification by gait is a part of this process. Shakespeare made use of this in his play "The Tempest" where Ceres said: "High'st queen of state, Great Juno, comes; I know her by her gait." Psychophysiological studies have proved that the human being can recognize the sex of a walker [6] and friends and colleagues [7, 8] with a success rate up to 70–80 %.

The authors derive from the institution that has conducted what are, so far, the only scientific approaches to gait analysis for evidential procedures. The entry describes how evidential analysis was derived and presented in two forensic investigations [9, 10].

Gait analyses is performed by first gaining a purely morphological, anthroposcopic impression of the gait of a perpetrator. We then combine the basic ability to recognize people with biomechanical knowledge in order to give statements as to whether a suspect could have the same identity as a perpetrator in a given case by comparing the suspect's posture and joint angles during gait with the perpetrator's. A checklist has been developed for forensic gait analysis (Table 1). First described are the general characteristics of the perpetrator's gait following which are analyzed each of the joint rotations and segment movements found relevant for forensic gait analysis (by trial and error). When a profile of the perpetrator has been completed, each item of the list is compared to the recording of the suspect and stated if agreement (A), no agreement (N), or comparison not possible (-) is found. An item can be incomparable because either the joint rotation/movement cannot be analyzed due to the poor quality of the surveillance recordings or the recording of the suspect differs too much in some way from the recording of the crime such as differences in shoulder angles between suspect and perpetrator because of elevated shoulders in one of the recordings.

The middle column is used for notes and specific observations.

Gait, Forensic Evidence of, Table 1 IFM Copenhagen gait description form/checklist. The rightmost column is marked up either with "A" for agreement, "N" for no agreement, and "-" for incomparable (see text)

General	Notes on gait of perpetrator/suspect
Long/short steps, stiff/relaxed gait with narrow/wide distance between the feet	
Signs of pathologic gait	
Feet/ankle joint	
Outward rotation	
Inversion/eversion	
Dorsal/plantar flexion at heel strike	
Degree of "push-off" at toe-off	
Knee	
Varus/valgus	
Knee flexion during stance	
Hip/pelvis	
Pelvis abduction/adduction	
Pelvis rotation	
Pelvis tilt	
Upper body	
Lateral flexion of spinal column	
Forward/backward leaning	
Rotation of the upper body during walk	
Shoulders	
Angle in frontal plane	
Forward/backward rotation	
Neck/head	
Posture in sagittal plane	
Head movements in frontal plane	
Quality of recordings/other precautions	

There have been several automated assessments of feature analysis for forensic and biometric purposes which show that there is a natural match between technique and observed performance [5]. Their features include foot angle (degree of outward rotation), the step length, and the mean hip joint angle, among others. Several other characteristic features have also been identified: inversion/eversion in the ankle during stance, lateral flexion in the dorsal column of the spine, and the knee angle in the frontal plane that would show lateral instability of the knee and signs of a person being bowlegged/knock-kneed. Furthermore, some of the characteristic features found were so special, such as limping, that it was not necessarily expected to be found in the 11 randomly selected subjects.

It should be stressed that a rather wide definition of "gait analysis" is used, so that basically all bodily movements may be studied. Posture and stance may be quite specific. For example, when standing, one leg is more often weight-bearing than the other, there may be marked lordosis, the neck and shoulders may be more or less slouched, and so on. These stance-related characteristics have a bearing on how a person initiates or stops walking and should thus also be involved in the analysis.

All the above features may be judged purely morphologically, but it may be of great evidentiary value to attach numbers to these features. Thus, the morphological approach is combined with photogrammetry in order to acquire specific measurements of body segment lengths and heights.

Gait, Forensic Evidence of, Fig. 1 Measuring screens put up in a department store, in order to calibrate the CCTVs [10]

Photogrammetry in Association with Gait Analysis

Photogrammetry literally means measuring by photography. Photogrammetry enables the measurement of unknown values in two-dimensional space (2D) using known values within a single image [9, 10]. Another basic application of photogrammetry is measuring objects in three-dimensional space (3D) using photographs taken from different sides and angles. Zhao et al. [11] have also worked with video sequences in this respect. Jensen and Rudin [9] used a 2D method to measure the stature and several segment lengths in two different cases and found excellent agreement between perpetrator and suspect. Lynnerup and Vedel [10, 12] used a 3D method in the investigation of a bank robbery where the perpetrator was recorded simultaneously from two different cameras and found good agreement in bodily measurements when comparing the perpetrator to the suspect.

A first step in photogrammetry is calibration of the CCTV cameras. This is done by placing frames with targets on the locations (Fig. 1). The frames are photographed with both the surveillance video cameras and a calibrated digital camera. Using the digital camera images and special software (PhotoModeler Pro®), the points are measured and subsequently imported as control points ("fiduciary points"). A feature in PhotoModeler Pro® allows determination of the internal parameters of the surveillance video cameras, e.g., focal length, and subsequent calculation of the exact placement of the cameras. After calibration, still images from the surveillance cameras are input in PhotoModeler Pro®. The photogrammetrical method described here has the advantage that there is no need to ascertain the position of the perpetrator in relation to a measuring device. After calibration by fiduciary points, the photogrammetrical analysis produces points in a 3D space, and an evaluation of the goodness of fit may be made directly in the software. This then allows measuring body segment lengths, stature, etc. of a perpetrator in various locations and with various body stances (Fig. 2). The selection of anatomical points is done by choosing specific points such as the top of the head, eyes, and joint center points on an image. This selection is made by judging anatomical landmarks, clothing displacement, comparison with images just before and after the chosen photo, etc. When then focusing on the other images of the same situation, but from other cameras, the program will indicate the epi-lines (the "line of sight") from the first image, as well as a line connecting the two joints. After selecting the identical anatomical points in this image, it is immediately

Gait, Forensic Evidence of, Fig. 2 Screen shots of PhotoModeler Pro® interface, showing selection of points. Simultaneous images from different CCTV cameras are used to pinpoint concurrent anatomical points (and markers) seen from different POV. The lines between the points are to scale and thus hold accurate measures of distance [10]

apparent how good the fit is and whether the points selected in the first image are adequate. Thus, the 3D coordinates are calculated not only by a simple averaging of points chosen from the two images but reflect a dynamic process where the tightness of the intersections of the epi-lines is minimized. The absolute error associated with measuring using photogrammetry as described is small. For instance, the height of a desk (bolted to the floor and not moved between the incident and the analysis) was measured by photogrammetry (result: 89.3 cm) and compared to an actual physical measurement (result: 90.0 cm), thus the error was 7 mm or less than 1 %. Intra- and interobserver tests of photogrammetric measurements of bodily segments seem to indicate that the error associated with clearly identifiable body points, such as top of the head, eyes, earlobes, and among others, is small. On the other hand, if the body points are hidden or obscured by clothing, such as joint center points, then there is some variation, which needs to be taken into account.

Currently, research focuses on implementing the possibility of performing accurate measurements of a perpetrator even though images are from only one camera. To do this, a measuring screen is used, the contours of which can be accurately measured by the software, which is physically placed near to where the perpetrator was standing (it needs to place the perpetrator on a specific point on the floor). If the screen is oriented perpendicular to the camera, then the screen can be imported as a virtual screen overlaid the crime video footage (Fig. 3). The perpetrator can then be measured against this screen, akin to seeing a person standing in front of a light source and whose shadow is cast of a screen or wall behind him.

Gait, Forensic Evidence of, Fig. 3 Using the back-projection screen method (see text)

Comparing Gait and Photogrammetry

As the forensic analysis mostly pertains to comparisons of perpetrators and suspects, then gait analyses and measuring of the suspect also have to be carried out. Owing to legal exigencies, this may be performed under very different settings and conditions, comprising hidden and overt image capture for gait, and hidden and overt photogrammetry. In some cases, legal circumstances have ruled out hidden image capture; in other cases, the defense counsel was invited to be present (but without the knowledge of the suspect); and finally, the suspect has sometimes been filmed completely overt. Ideally, it is felt that gait image capture should be performed hidden, so as the suspect does not know he is being filmed. This is to ensure that the gait is not "changed." Preferably, the setting for performing the image capture should to some extent mimic the crime scene. For example, if at the crime scene there was a step at the entrance, which the suspect engaged in a distinct fashion, then filming the suspect engaging a somewhat likewise step would be obvious for comparison. If the crime scene images show a perpetrator walking down a corridor, either against or away from the CCTV camera, then a setting at police offices with a long corridor may be suitable. The filming usually takes place with ordinary DV cameras and is done by forensic technicians, but the setting would have been discussed in advance. For instance, a policeman can be instructed to accompany the suspect, but walking at a speed that matches the velocity of the perpetrator, because the gait speed may influence some of the features. For example, a lateral instability in the knee will be more pronounced at a higher gait speed.

The photogrammetric measurement of the suspect is most easily performed overtly. Usually a corner in an office is identified with points fixed on the wall, and the suspect is asked to stand in the corner. Using two or three digital cameras, coupled to a computer, several sets of images of high quality for subsequent photogrammetry can be rapidly acquired. While height could be just as easily acquired using a stadiometer, it is found that the same measuring method (photogrammetry) should be used for comparing perpetrator and suspect. While at first glance stadiometer-measured stature might seem as a "gold standard," it is also found that people almost automatically straighten themselves when asked to stand against a stadiometer, meaning in fact that a better agreement between subsequent measurements of stature by photogrammetry has been found than between photogrammetry and a stadiometer. Of course, measuring the suspect by photogrammetry also makes it easier to measure other heights, such as floor to eye, floor to shoulder, and floor to earlobe.

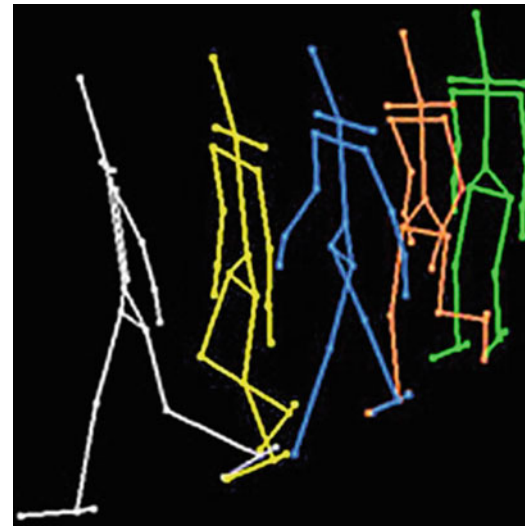
Schöllhorn et al. [13] concluded that "identification of individuality seems to be impossible with single variables or specific parameters of single variables," so the more gait characteristics and bodily measurements of the perpetrator that can be extracted and compared to the suspect, the better.

The Nature of Forensic Statements

In statements to the police, it is noted what image material has been available and what manner of image-enhancing techniques had been used. The results of the above analyses are then presented, each followed by a separate conclusion, and each conclusion always summing up what features were found to indicate concordance between the suspect and the perpetrator, as well as features which seemed to indicate incongruity. Each item may therefore be seen as constituting single pieces of evidence. This renders a statistical approach, for instance, the calculation of likelihood ratios for identity, based on the prevalence of certain facial and bodily traits, problematical [14].

Using the data sheets for gait analyses and photogrammetry fulfills three of the four guidelines in the Daubert standard, a legal precedent set by the Supreme Court of the United States [15], for determining whether expert witnesses' testimony is admissible as evidence: (1) the testimony in court is based on an empirically used technique, (2) the technique has been published in peer-reviewed literature, and (3) it is generally accepted for use in forensic medicine. The last Daubert guideline states that the reliability of the technique has been tested and potential error rates known.

Image-based comparison will probably never achieve specific identification such as that which is associated with DNA typing and fingerprinting. However, analyzing gait and measuring stature and segment lengths of a perpetrator from surveillance video have the possibility of becoming valuable forensic tools because the gait and the measures are an integrated part of the offender. At present, the methods can be used effectively to exclude a suspect if the gait and anthropometrical measures of the suspect and perpetrator are entirely different from each other. On the other hand, if the perpetrator and suspect do have a similar gait and similar measures, it can only be stated in court that the suspect cannot be excluded as the perpetrator. To give a more specific statement of the value of evidence, a database with gait characteristics and measures for a population of which the perpetrator and suspect could be referenced against can be constructed. In theory, this might mean that if a perpetrator and a suspect are measured to have an unusual height, i.e., either very tall or very small, then this might in itself increase the likelihood of concordance between them, whereas very average heights would lower the likelihood (because then it might be almost anybody). In actuality, such databases are rather restricted, with often only specific subsamples of the entire population represented; populations also change in terms of, e.g., immigration; and finally the perpetrator might well be from an entirely different part of the world. If comparing with such databases, it is important to stress that "given the perpetrator/suspect are drawn



Gait, Forensic Evidence of, Fig. 4 Line models produced by the PhotoModeler® based on the selected anatomical points, showing the gait

from the same population as the database," then their stature is more or less common, and the likelihood of concordance between them is more or less likely.

Future work will probably focus on a better integration of gait characteristics and photogrammetry in order to perform dynamic measurements of gait (basically "animating" the line models, cf. Fig. 4). This has the potential of calculating angles of flexion – extension in the major joints, step length, degree of side-to-side movement of the torso during walking, etc. These parameters may then further assist in discriminating between suspects and more specifically in identifying individual traits of gait.

Related Entries

- ▸ Gait Biometrics, Overview
- ▸ Gait Recognition, Model-Based
- ▸ Gait Recognition, Motion Analysis for
- ▸ Gait Recognition, Silhouette-Based

References

1. D. Cunado, M.S. Nixon, J.N. Carter, Automatic extraction and description of human gait models for recognition purposes. Comput. Vis. Image Underst. **90**, 1–41 (2003)
2. R. Urtasun, P. Fua, 3D tracking for Gait Characterization and Recognition. *Computer Vision Laboratory*. Swiss Federal Institute of Technology, Lausanne, Switzerland. Report No. IC/2004/04 (2004)
3. M.S. Nixon, T.N. Tan, R. Chellappa, *Human Identification Based on Gait* (Springer, New York, 2006)
4. S. Rahati, R. Moravejian, F.M. Kazemi, Gait recognition using wavelet transform, in *Fifth International Conference on Information Technology: New Generations, 2008*, Las Vegas (Universitat Triet, Germany, 7–9 Apr 2008), pp. 932–936
5. D.K. Wagg, M.S. Nixon, On automated model-based extraction and analysis of gait, in *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, Seoul (IEEE, Seoul, 17–19 May 2004), pp. 11–19
6. L.T. Kozlowski, J.E. Cutting, Recognizing the sex of a walker from a dynamic point-light display. Percept. Psychophys. **21**, 575–580 (1977)
7. J.E. Cutting, L.T. Kozlowski, Recognizing friends by their walk: gait perception without familiarity cues. Bull. Psychon. Soc. **9**, 353–356 (1977)
8. D. Jokisch, I. Daum, N.F. Troje, Self recognition versus recognition of others by biological motion: viewpoint-dependent effects. Perception **35**(7), 911–920 (2006)
9. S.C. Jensen, L.I. Rudin, Measure: an interactive tool for accurate forensic photo/videogrammetry, in *Investigative and Trial Image Processing 1995 July 13*, ed. by L.I. Rudin, S.K. Bramble (SPIE, San Diego, 1995), pp. 73–83
10. N. Lynnerup, J. Vedel, Person identification by gait analysis and photogrammetry. J. Foren. Sci. **50**(1), 112–118 (2005)
11. G. Zhao, G. Liu, L. Hua, M. Pietikainen, 3D gait recognition using multiple cameras, in *Seventh International Conference on Automatic Face and Gesture Recognition, 2006, FGR 2006*, 10–12 Apr 2006, Southampton (2006), pp. 529–534
12. N. Lynnerup, B. Sejrsen, J. Vedel, Identification by facial recognition, gait analysis and photogrammetry: the Anna Lindh murder, in *Forensic Anthropology: Case Studies from Europe*, ed. by M. Brickley, R. Ferllini (Charles C. Thomas, Springfield, 2007), pp.232–244
13. C.G.G. Aitken, *Statistics and the Evaluation of Evidence for Forensic Scientists* (Wiley, Chichester, 1995)
14. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993)
15. W.I. Schollhorn, B.M. Nigg, D.J. Stefanyshyn, W. Liu, Identification of individual walking patterns using time discrete and time continuous data sets. Gait Posture **15**(2), 180–186 (2002)

Gaussian Mixture Models

Douglas Reynolds
Lincoln Laboratory, MIT, Lexington, MA, USA

Synonyms

Gaussian mixture density; GMM

Definition

A Gaussian mixture model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component

densities. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as vocal tract-related spectral features in a speaker recognition system. GMM parameters are estimated from training data using the iterative expectation-maximization (EM) algorithm or maximum a posteriori (MAP) estimation from a well-trained prior model.

Introduction

A Gaussian mixture model is a weighted sum of M component Gaussian densities as given by the equation

$$p(\mathbf{x}|\lambda) = \sum_{i=1}^{M} w_i g(\mathbf{x}|\mu_i, \mathbf{\Sigma}_i), \qquad (1)$$

where $\mathbf{x}$ is a D-dimensional continuous-valued data vector (i.e., measurement or features); $w_i, i=1,\ldots,M$, are the mixture weights; and $g(\mathbf{x}|\mu_i, \Sigma_i), i = 1,\ldots,M$ are the component Gaussian densities. Each component density is a D-variate Gaussian function of the form

$$g(\mathbf{x}|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)' \Sigma_i^{-1}(\mathbf{x}-\mu_i)\right\}, \qquad (2)$$

with mean vector $\boldsymbol{\mu}_i$ and covariance matrix Σ_i. The mixture weights satisfy the constraint that $\sum_{i=1}^{M} w_i = 1$.

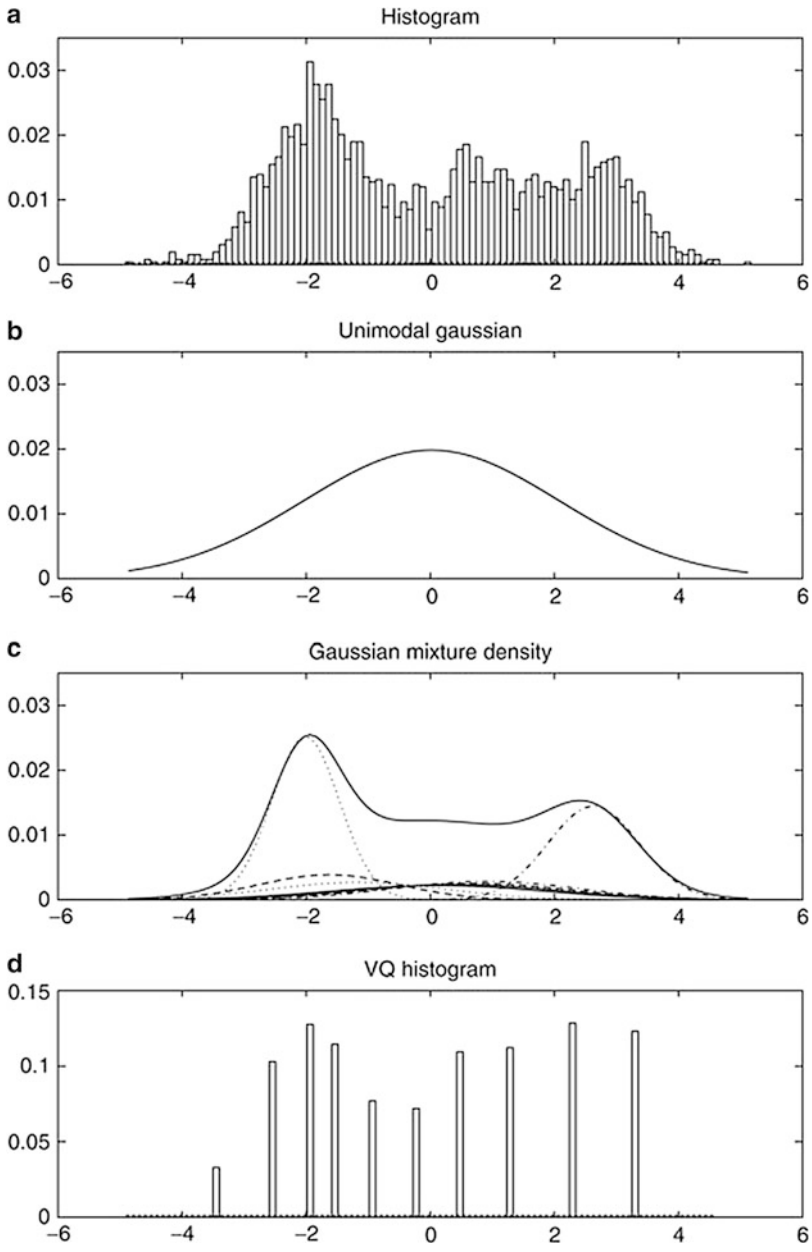
The complete Gaussian mixture model is parameterized by the mean vectors, covariance matrices, and mixture weights from all component densities. These parameters are collectively represented by the notation

$$\lambda = \{w_i, \boldsymbol{\mu}_i, \Sigma_i\} \quad i = 1,\ldots,M. \qquad (3)$$

There are several variants on the GMM shown in Eq. (3). The covariance matrices, Σ_i, can be full rank or constrained to be diagonal. Additionally, parameters can be shared, or tied, among the Gaussian components, such as having a common covariance matrix for all components. The choice of model configuration (number of components, full or diagonal covariance matrices, and parameter tying) is often determined by the amount of data available for estimating the GMM parameters and how the GMM is used in a particular biometric application.

It is also important to note that since the component Gaussian are acting together to model the overall feature density, full covariance matrices are not necessary even if the features are not statistically independent. The linear combination of diagonal covariance basis Gaussians is capable of modeling the correlations between feature vector elements. The effect of using a set of M full covariance matrix Gaussians can be equally obtained by using a larger set of diagonal covariance Gaussians.

GMMs are often used in biometric systems, most notably in speaker recognition systems [1, 2], due to their capability of representing a large class of sample distributions. One of the powerful attributes of the GMM is its ability to form smooth approximations to arbitrarily shaped densities. The classical unimodal Gaussian model represents feature distributions by a position (mean vector) and a elliptic shape (covariance matrix) and a vector quantizer (VQ), or nearest neighbor model represents a distribution by a discrete set of characteristic templates [3]. A GMM acts as a hybrid between these two models by using a discrete set of Gaussian functions, each with its own mean and covariance matrix, to allow a better modeling capability. Figure 1 compares the densities obtained using a unimodal Gaussian model, a GMM, and a VQ model. Plot (a) shows the histogram of a single feature from a speaker recognition system (a single cepstral value from a 25-s utterance by a male speaker); plot (b) shows a unimodal Gaussian model of this feature distribution; plot (c) shows a GMM and its ten underlying component densities; and plot (d) shows a histogram of the data assigned to the VQ centroid locations of a ten-element codebook. The GMM not only provides a smooth overall distribution fit, its components



Gaussian Mixture Models, Fig. 1 Comparison of distribution modeling. (**a**) Histogram of a single cepstral coefficient from a 25-s utterance by a male speaker, (**b**) maximum likelihood unimodal Gaussian model, (**c**) GMM and its ten underlying component densities, and (**d**) histogram of the data assigned to the VQ centroid locations of a ten-element codebook

also clearly detail the multimodal nature of the density.

The use of a GMM for representing feature distributions in a biometric system may also be motivated by the intuitive notion that the individual component densities may model some underlying set of *hidden* classes. For example, in speaker recognition, it is reasonable to assume the acoustic space of spectral-related features corresponding to a speaker's broad phonetic events, such as vowels, nasals, or fricatives. These acoustic classes reflect some general speaker-dependent vocal tract configurations that are useful for characterizing speaker identity. The spectral shape of the ith acoustic class can in turn be represented by the mean $\boldsymbol{\mu}_i$ of

the ith component density, and variations of the average spectral shape can be represented by the covariance matrix Σ_i. Since all the features used to train the GMM are unlabeled, the acoustic classes are hidden in that the class of an observation is unknown. A GMM can also be viewed as a single-state HMM with a Gaussian mixture observation density or an ergodic Gaussian observation HMM with fixed, equal transition probabilities. Assuming independent feature vectors, the observation density of feature vectors drawn from these hidden acoustic classes is a Gaussian mixture [4, 5].

Maximum Likelihood Parameter Estimation

Given training vectors and a GMM configuration, the parameters, λ, are estimated which, in some sense, best match the distribution of the training feature vectors. There are several techniques available for estimating the parameters of a GMM [6]. By far the most popular and well-established method is maximum likelihood (ML) estimation.

The aim of ML estimation is to find the model parameters which maximize the likelihood of the GMM given the training data. For a sequence of T training vectors $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$, the GMM likelihood, assuming independence between the vectors (the independence assumption is often incorrect but is needed to make the problem tractable.), can be written as

$$p(X|\lambda) = \prod_{t=1}^{T} p(\mathbf{x}_t|\lambda). \tag{4}$$

Unfortunately, this expression is a nonlinear function of the parameters λ and direct maximization is not possible. However, ML parameter estimates can be obtained iteratively using a special case of the expectation-maximization (EM) algorithm [7].

The basic idea of the EM algorithm is, beginning with an initial model λ, to estimate a new model $\bar{\lambda}$, such that $p\left(X|\bar{\lambda}\right) \geq p(X|\lambda)$. The new model then becomes the initial model for the next iteration and the process is repeated until some convergence threshold is reached. The initial model is typically derived by using some form of binary VQ estimation.

On each EM iteration, the following re-estimation formulas are used which guarantee a monotonic increase in the model's likelihood value:

Mixture weights

$$\bar{w}_i = \frac{1}{T}\sum_{t=1}^{T} \Pr(i|\mathbf{x}_t, \lambda). \tag{5}$$

Means

$$\bar{\boldsymbol{\mu}}_i = \frac{\sum_{t=1}^{T} \Pr(i|\mathbf{x}_t, \lambda)\,\mathbf{x}_t}{\sum_{t=1}^{T} \Pr(i|\mathbf{x}_t, \lambda)}. \tag{6}$$

Variances (diagonal covariance)

$$\bar{\sigma}_i^2 = \frac{\sum_{t=1}^{T} \Pr(i|\mathbf{x}_t, \lambda)\,\mathbf{x}_t^2}{\sum_{t=1}^{T} \Pr(i|\mathbf{x}_t, \lambda)} - \bar{\mu}_i^2, \tag{7}$$

where ${\sigma_i}^2$, x_t , and μ_i refer to arbitrary elements of the vectors σ_i^2, $\mathbf{x}_t$, and $\boldsymbol{\mu}_i$, respectively.

The a posteriori probability for component i is given by

$$\Pr(i|\mathbf{x}_t, \lambda) = \frac{w_i\, g(\mathbf{x}_t|\mu_i, \boldsymbol{\Sigma}_i)}{\sum_{k=1}^{M} w_k\, g(\mathbf{x}_t|\mu_k, \boldsymbol{\Sigma}_k)}. \tag{8}$$

Maximum A Posteriori (MAP) Parameter Estimation

In addition to estimating GMM parameters via the EM algorithm, the parameters may also be estimated using maximum a posteriori (MAP) estimation. MAP estimation is used, for example, in speaker recognition applications to derive speaker model by adapting from a universal background model (UBM) [8]. It is also used

in other pattern recognition tasks where limited labeled training data is used to adapt a prior, general model.

Like the EM algorithm, the MAP estimation is a two-step estimation process. The first step is identical to the "Expectation" step of the EM algorithm, where estimates of the sufficient statistics (these are the basic statistics needed to be estimated to compute the desired parameters. For a GMM mixture, these are the count and the first and second moments required to compute the mixture weight, mean, and variance) of the training data are computed for each mixture in the prior model. Unlike the second step of the EM algorithm, for adaptation these "new" sufficient statistic estimates are then combined with the "old" sufficient statistics from the prior mixture parameters using a data-dependent mixing coefficient. The data-dependent mixing coefficient is designed such that mixtures with high counts of new data rely more on the new sufficient statistics for final parameter estimation and mixtures with low counts of new data rely more on the old sufficient statistics for final parameter estimation.

The specifics of the adaptation are as follows. Given a prior model and training vectors from the desired class, $X = \{\mathbf{x}_1 \ldots, \mathbf{x}_T\}$, the probabilistic alignment of the training vectors into the prior mixture components is determined (Fig. 2a). That is, for mixture i in the prior model, $\Pr\left(i \,|\, \mathbf{x}_t, \lambda_{\text{prior}}\right)$ is computed as in Eq. (8).

Then compute the sufficient statistics for the weight, mean, and variance parameters ($\mathbf{x}^2$ is shorthand for $\text{diag}(\mathbf{x}\mathbf{x}')$):

$$n_i = \sum_{t=1}^{T} \Pr\left(i \,|\, \mathbf{x}_t, \lambda_{\text{prior}}\right) \text{ weight}, \tag{9}$$

$$E_i(x) = \frac{1}{n_i} \sum_{t=1}^{T} \Pr\left(i \,|\, \mathbf{x}_t, \lambda_{\text{prior}}\right) \mathbf{x}_t \text{ mean}, \tag{10}$$

$$E_i(\mathbf{x}^2) = \frac{1}{n_i} \sum_{t=1}^{T} \Pr\left(i \,|\, \mathbf{x}_t, \lambda_{\text{prior}}\right) \mathbf{x}_t^2 \text{ variance}. \tag{11}$$

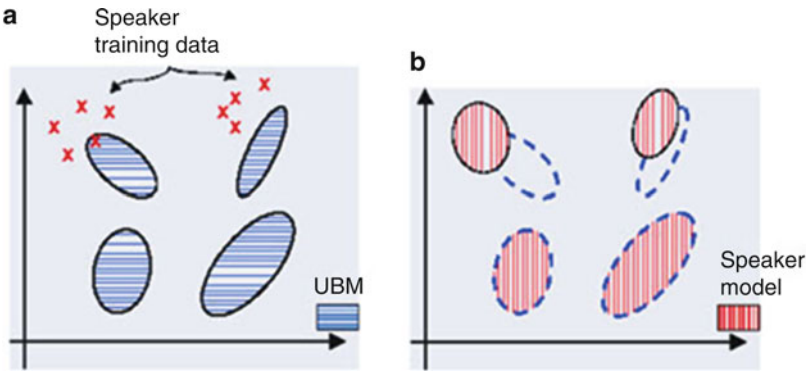
This is the same as the "Expectation" step in the EM algorithm.

Lastly, these new sufficient statistics from the training data are used to update the prior sufficient statistics for mixture i to create the adapted parameters for mixture i (Fig. 2b) with the equations:

$$\hat{w}_i = \left[\alpha_i^w n_i / T + \left(1 - \alpha_i^w\right) w_i\right] \gamma \quad \text{adapted mixture weight}, \tag{12}$$

$$\hat{\mu}_i = \alpha_i^m E_i(\mathrm{x}) + \left(1 - \alpha_i^m\right) \mu_i \quad \text{adapted mixture mean}, \tag{13}$$

$$\hat{\sigma}_i^2 = \alpha_i^v E_i(\mathbf{x}^2) + \left(1 - \alpha_i^v\right) \left(\sigma_i^2 + \hat{\mu}_i^2\right) \quad \text{adapted mixture variance}. \tag{14}$$



Gaussian Mixture Models, Fig. 2 Pictorial example of two steps in adapting a hypothesized speaker model. (**a**) The training vectors (x's) are probabilistically mapped into the UBM (prior) mixtures. (**b**) The adapted mixture parameters are derived using the statistics of the new data and the UBM (prior) mixture parameters. The adaptation is data dependent, so UBM (prior) mixture parameters are adapted by different amounts

The adaptation coefficients controlling the balance between old and new estimates are $\{\alpha_i{}^w, \alpha_i{}^m, \alpha_i{}^v\}$ for the weights, means, and variances, respectively. The scale factor, γ, is computed over all adapted mixture weights to ensure that they sum to unity. Note that the sufficient statistics, not the derived parameters, such as the variance, are being adapted.

For each mixture and each parameter, a data-dependent adaptation coefficient $\alpha_i{}^\rho, \rho \in \{w, m, v\}$, is used in the equations mentioned earlier. This is defined as

$$\alpha_i^\rho = \frac{n_i}{n_i + r^\rho}, \tag{15}$$

where r^ρ is a fixed "relevance" factor for parameter ρ. It is common in speaker recognition applications to use one adaptation coefficient for all parameters ($\alpha_i{}^w = \alpha_i{}^m = \alpha_i{}^v = n_i/(n_i + r)$) and further to only adapt certain GMM parameters, such as only the mean vectors.

Using a data-dependent adaptation coefficient allows mixture-dependent adaptation of parameters. If a mixture component has a low probabilistic count, n_i, of new data, then $\alpha_i{}^\rho \to 0$ causing the de-emphasis of the new (potentially undertrained) parameters and the emphasis of the old (better trained) parameters. For mixture components with high probabilistic counts, $\alpha_i{}^\rho \to 1$, causing the use of the new class-dependent parameters. The relevance factor is a way of controlling how much new data should be observed in a mixture before the new parameters begin replacing the old parameters. This approach should thus be robust to limited training data.

Related Entries

- ▸ Session Effects on Speaker Modeling
- ▸ Speaker Matching
- ▸ Speaker Recognition, Overview
- ▸ Universal Background Models

Acknowledgments This work was sponsored by the Department of Defense under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the US Government.

References

1. J. Benesty, M. Sondhi, Y. Huang (eds.), *Springer Handbook of Speech Processing*, vol. XXXVI (Springer, Berlin, 2008)
2. C. Müler (ed.), *Speaker Classification I: Fundamentals, Features, and Methods*. Lecture Notes in Computer Science, vol. 4343/2007 (Springer, Berlin, 2007)
3. R. Gray, Vector quantization. IEEE ASSP Mag. **1**(2), 4–29 (1984)
4. D.A. Reynolds, A Gaussian mixture modeling approach to text-independent speaker identification. PhD thesis, Georgia Institute of Technology, 1992
5. D.A. Reynolds, R.C. Rose, Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Trans. Acoust. Speech Signal Process. **3**(1), 72–83 (1995)
6. G. McLachlan (ed.), *Mixture Models* (Marcel Dekker, New York, 1988)
7. A. Dempster, N. Laird, D. Rubin, Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. **39**(1), 1–38 (1977)
8. D.A. Reynolds, T.F. Quatieri, R.B. Dunn, Speaker verification using adapted Gaussian mixture models. Digit. Signal Process. **10**(1), 19–41 (2000)

Gender Classification

Guodong Guo
Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, USA

Synonyms

Gender recognition; Sex classification; Gender prediction

Definition

Gender classification is to determine a person's gender, e.g., male or female, based on his or her biometric cues. Usually facial images are used to extract features and then a classifier

is applied to the extracted features to learn a gender recognizer. It is an active research topic in Computer Vision and Biometrics fields. The gender classification result is often a binary value, e.g., 1 or 0, representing either male or female. Gender recognition is essentially a two-class classification problem. Although other biometric traits could also be used for gender classification, such as gait, face-based approaches are still the most popular for gender discrimination.

Introduction

A sex difference is a distinction of biological and/or physiological characteristics associated with either males or females of a species. These can be of several types, including direct and indirect. Direct is the direct result of differences prescribed by the Y-chromosome, and indirect is a characteristic influenced indirectly (e.g., hormonally) by the Y-chromosome [14].

The most obvious differences between males and females include all the features related to reproductive role, notably the endocrine (hormonal) systems and their physiological and behavioral effects, including gonadal differentiation, internal and external genital and breast differentiation, and differentiation of muscle mass, height, and hair distribution [14].

In recognizing the difference between males and females of humans, one can call it sex or gender classification. Historically, "sex" is used to characterize the biological and physical properties associated with males and females, while "gender" is more related to "social roles" or "gender roles." Nowadays, however, gender is commonly used even to refer to the physiology of non-human animals, without any implication of social identity or gender roles [10]. So in this description, either sex classification or gender classification can be used without any real difference. Actually, in computational gender classification, "gender"

Gender Classification, Fig. 1 Illustration of gender difference in face images: female (*top*) and male (*bottom*). Each row has face images of the same person at different ages (Sources are from the FG-NET database [3].)

is used very often without involving any social identity.

For human sex classification, usually the face images are used, although many other biometric cues could be used too. So the study is mainly about how to extract features in face images so that the male and female can be discriminated by using a classifier.

As shown in Fig. 1, one can see the difference between a female and a male, even though the facial appearance changes a lot with aging (in each row).

Applications of Gender Classification

Automated human gender classification has a number of applications. Some of the typical applications will be presented in the following:

1. *Business intelligence*. Gender classification can be useful for business intelligence. For advertizing or better customer service, it is needed to know more about the customers, in terms of their age, gender, ethnicity, etc. So gender attribute can be combined with age and ethnicity to group customers into different groups [7]. The challenge is how to improve the gender classification accuracy when the image quality is not good and the image resolution is small.
2. *Access control*. A security entrance might use sex information for surveillance and control of access to certain areas. For example, some areas may only be accessed by a man or woman. In some cases, the gender information can be combined with age for access control, e.g., only allowing female adults to access a certain place.
3. *Image filtering*. Gender classification can be useful for online image filtering or image database organization. To better organize the huge amount of image and video data, gender could play an important role by dividing the images and videos into two groups, male-centered or female-centered.
4. *Soft biometrics*. Gender can be a useful soft biometric trait, as other soft biometric traits, e.g., age, ethnicity, eye color, weight, height, etc. Soft biometrics can be useful to enhance a traditional biometric system such as face recognition.

Gender Classification Approaches

Computational gender classification is a standard pattern recognition problem. A typical framework is shown in Fig. 2. For an input face image, the first step is to perform face detection. Then the detected face will be aligned with other faces based on the eye coordinates. The purpose of alignment is to correct some possible transformations such as translation, rotation, and scaling and make all faces to have the same size. Facial features will then be extracted from the aligned face images. Classifiers can be trained using the extracted facial features and the corresponding gender label. It is a supervised learning process. Finally, the learned classification function can be used to recognize the gender for a test face image.

Most existing sex classification approaches use frontal view face images [7]. Earlier works applied the artificial neural networks for sex classification. For example, Golomb et al. [5] built the SEXNET for sex classification and reported an accuracy of 91.9 % obtained from the experiments on 90 photos of 45 males and 45 females. This is probably the first computational approach to sex classification. Brunelli and Poggio [2] used the HyperBF networks for gender classification with an accuracy of 79 % experimented

Gender Classification, Fig. 2 A typical framework for gender classification

on 168 images of 21 males and 21 females; Gutta et al. [9] used a hybrid classifier combining the artificial neural networks and decision trees and reported an accuracy of 96 % on the FERET database, a popular face recognition database.

Later on, the SVM classifier [16] became a dominating method in gender recognition. For example, Moghaddam and Yang [13] used the nonlinear SVM for gender recognition on the FERET database with an accuracy of 96.6 % on 1,755 images of 1,044 males and 711 females. The feature is just the raw face images with reduced sizes. Graf and Wichmann [6] used dimensionality reduction methods (e.g., PCA and LLE) with the SVM classifier. Jain and Huang [11] used independent component analysis (ICA) and linear discriminant analysis (LDA) with the SVM for gender classification.

AdaBoost [4] is another effective classifier for gender recognition. For example, Shakhnarovich et al. [15] reported a gender recognition system using web-collected face images. Their sex classification framework is about the same as their face detector with rectangle filters on integral images for feature extraction and the AdaBoost for feature selection and classification. Baluja and Rowley [1] used simple features (intensity comparisons on neighborhood pixels) with the AdaBoost method and reported an accuracy of 94.4 % on 2,409 face images of 1,495 males and 914 females from FERET. Wu et al. [17] proposed to use a Look Up Table (LUT) based AdaBoost for gender classification. Xu and Huang [18] described a SODA-Boosting method (a variant of the AdaBoost) for gender classification. They reported an accuracy of 92.82 % on 4,109 face images (703 males and 498 females) from FERET database. Yang and Ai [19] used the local binary patterns (LBP) feature with the AdaBoost classifier for age, gender, and ethnicity classification. An accuracy of 93.3 % was reported on 3,540 face images from FERET.

Guo et al. [8] empirically studied the influence of human aging on gender classification. Based on the experimental results on a large database with 8,000 face images, they found that the gender classification accuracies on children and elderly are lower than the adults significantly. This study result shows that gender classification is affected by age.

Makinen and Raisamo [12] performed an evaluation of various gender classification methods using two databases. Based on extensive experiments, they found that the SVM with 36×36 face image size as input achieved the best gender classification rate. Further, facial alignment is important for sex classification. A good alignment of facial features could improve the sex classification accuracy.

Performance Measure

The gender classification results are usually measured by the classification accuracy on test images. In the literature, one can find the reported accuracies with different experimental setup, such as the number of face images in databases, the validation scheme (e.g., five-fold cross validation), etc. As a pattern recognition problem, the performance of gender classification can also be measured by the confusion matrix, which contains true positive, true negative, false positive, and false negative rates, especially when the number of females and males is very unbalanced. Another measure is the ROC (receiver operating characteristic) curve.

Challenges

In a constrained environment with cooperative users (e.g., looking at the camera, displaying a frontal view, and showing a neutral expression), the gender classification accuracies can be pretty high, e.g., around 90 % or even higher. However, in an unstrained environment with noncooperative users, the captured face images can have many different variations, and the image resolutions can be very low. In those cases, the accuracies

of gender recognition can be much lower. The difficulties are about the same as those encountered by unconstrained face recognition. These difficulties need to be dealt with especially for business intelligence applications where the working environment can vary significantly and the users cannot be expected to be cooperative. New research efforts are needed to extract invariant or more robust features to characterize the face images with respect to different variations. In addition, the gender recognition performance is also influenced by human aging, as demonstrated in [8].

Discussion

Gender classification is still a very challenging problem, especially in an unconstrained environment with noncooperative users. Continuous research efforts are needed to advance the performance of gender recognition. In addition to static images, it might be interesting to investigate how motion or dynamic information of the humans (i.e., moving faces) could be utilized to improve gender classification accuracies.

Gender classification itself is an interesting problem, and it can be helpful for biometrics-based human recognition by serving as a useful soft biometric trait. Gender recognition can be combined with other soft biometric cues, such as age and ethnicity to improve gender classification accuracy, or serve better as soft biometric cues for human recognition.

Summary

Gender classification is an interesting problem with many applications. Significant progress has been made by researchers and high accuracies can be obtained for gender recognition with cooperative users in a constrained environment. However, it is still challenging for gender classification in an unconstrained environment with noncooperative users, which prevails in many real applications including business intelligence or Internet image analysis. As a soft biometric trait, gender recognition can help to improve biometrics-based human recognition.

Related Entries

▸ Soft Biometrics

References

1. S. Baluja, H.A. Rowley, Boosting sex identification performance. Int. J. Comput. Vis. **71**(1), 111–119 (2007)
2. R. Brunelli, T. Poggio, Hyperbf networks for gender classification, in *Proceedings of the DARPA Image Understanding Workshop*, San Diego, 1992, pp. 311–314
3. FGNET, The FG-NET aging database (2002), http://www.fgnet.rsunit.com/
4. Y. Freund, R. Schapire, Experiments with a new boosting algorithm, in *Proceedings of the Thirteenth International Conference on Machine Learning*, Bari, 1996, pp. 148–156
5. B. Golomb, D. Lawrence, T. Sejnowski, Sexnet: a neural network identifies sex from human faces, in *Advances in Neural Information Processing Systems*, ed. by R.P. Lippmann, J.E. Moody, D.S. Touretzky, vol. 3 (Morgan Kaufmann Publishers, San Mateo, 1991), pp. 572–577
6. A. Graf, F. Wichmann, Gender classification of human faces, in *International Workshop on Biologically Motivated Computer Vision*, Tübingen, 2002, pp. 491–500
7. G. Guo, Human age estimation and sex classification, in *Video Analytics for Business Intelligence*, ed. by C. Shan et al. SCI 409 (Springer, Berlin/New York, 2012), pp. 101–131
8. G. Guo, C. Dyer, Y. Fu, T.S. Huang, Is gender recognition influenced by age? in *IEEE International Workshop on Human-Computer Interaction, in Conjunction with ICCV'09*, Kyoto, Japan, 2009, pp. 2032–2039
9. S. Gutta, H. Wechsler, P. Phillips, Gender and ethnic classification, in *Proceedings of the IEEE International Workshop on Automatic Face and Gesture Recognition*, Nara, 1998, pp. 194–199
10. D. Haig, The inexorable rise of gender and the decline of sex: social change in academic titles, 1945–2001. Arch. Sex. Behav. **33**(2), 87–96 (2004)
11. A. Jain, J. Huang, Integrating independent component analysis and linear discriminant analysis for gender

classification, in *IEEE International Conference on Automatic Face and Gesture Recognition*, Seoul, 2004
12. E. Makinen, R. Raisamo, Evaluation of gender classification methods with automatically detected and aligned faces. IEEE Trans. Pattern Anal. Mach. Intell. **30**(3), 541–547 (2008)
13. B. Moghaddam, M.-H. Yang, Learning gender with support faces. IEEE Trans. Pattern Anal. Mach. Intell. **24**(5), 707–711 (2002)
14. Sex differences in humans, in http://en.wikipedia.org/wiki/Sex_differences_in_humans
15. G. Shakhnarovich, P. Viola, B. Moghaddam, A unified learning framework for real time face detection and classification, in *International Conference on Automatic Face and Gesture Recognition*, Washington, DC, 2002
16. V.N. Vapnik, *Statistical Learning Theory* (Wiley, New York, 1998)
17. B. Wu, H. Ai, C. Huang, S. Lao, Lut-based adaboost for gender classification, in *International Conference on Audio and Video-Based Person Authentication*, Guildford, 2003
18. X. Xu, T. Huang, SODA-boosting and its application to gender recognition, in *AMFG 2007*, Rio de Janeiro, ed. by S.K. Zhou, W. Zhao, X. Tang, S. Gong. LNCS, vol. 4778 (Springer, Heidelberg, 2007), pp. 193–204
19. Z. Yang, H. Ai, Demographic classification with local binary patterns, in *International Conference on Biometrics*, Seoul, 2007, pp. 464–473