# Body Language Based Individual Identification in Video Using Gait and Actions

Y. Pratheepan[1], P.H.S Torr[2], JV Condell[1], and G. Prasad[1]

[1] School of Computing and Intelligent Systems, Faculty of Computing and
Engineering, University of Ulster at Magee, Northland Rd, Londonderry, N.I.
{p.yogarajah, j.condell, g.prasad}@ulster.ac.uk
[2] Department of Computing, Oxford Brookes University, Wheatley, Oxford, UK
philiptorr@brookes.ac.uk

**Abstract.** In intelligent surveillance systems, recognition of humans
and their activities is generally the most important task. Two forms
of human recognition can be useful: the determination that an object is
from the class of humans (which is called human detection), and deter-
mination that an object is a particular individual from this class (this is
called individual recognition). This paper focuses on the latter problem.
For individual recognition, this report considers two different categories.
First, individual recognition using "style of walk" i.e. gait and second
"style of doing similar actions" in video sequences. The "style of walk"
and "style of actions" are proposed as a cue to discriminate between two
individuals. The "style of walk" and "style of actions" for each individual
is called their "body language" information.

## 1  Introduction

There is a strong need for smart surveillance systems from security-sensitive
areas e.g. banks to alert security officers to a suspicious individual or unknown
individual wandering about the premises. Visual systems can deal with and
monitor *"familiar"* and *"unfamiliar"* human movements. In order to properly
interact with people an intelligent system has to detect people and identify them
using their body language. Examples of such systems are intelligent security
systems that detect unknown people entering restricted areas or interface robots
that act as an interface for taking known users commands and presenting results.

The aim of this intelligent surveillance system project is a general framework
that groups a number of different computer vision tasks aiming to identify indi-
viduals using their *"style of walk"*, and on the next level to identify individuals
through their *"style of action"*. Gait based individual recognition is explained in
Section 2, action based individual recognition is explained in Section 3. Results
and conclusions are discussed in Sections 4 and 5 respectively.

## 2  Identifying Individuals through Their "Style of Walk"

Humans can identify people known to them over long distances, just by their gait.
This implies that the motion pattern is characteristic for each person. Motion

information is one of the good cues which can be used to recognize individuals. The most recognized and earliest psychophysical study of human perception of gait was carried out by Johansson [1] using moving light displays (MLD). The initial experiments showed that human observers are remarkably good at perceiving the human motion that generated the MLD stimuli. Cutting et al. [2] studied human perception of gait and their ability to identify individuals using MLD. There has been an explosion of research on individual recognition in recent years but most of them are based on gait [3,4,5]. The above psychological evidence and computational approaches justify that individual recognition can be done using people's walking sequences. i.e. human gait has enough information to discriminate individuals.

Most of the research has been done using full-cyclic period human walk sequences. The first question is, if it is not possible to obtain full-cyclic walking sequences then how do we recognise individuals?. The number of frames in individuals walk sequences, even from a particular individual's sequence, may vary because of the speed. Therefore the second question is, how do we calculate a similarity measure between the different number of frames based on the "Body-language" information. We attempt to answer these questions in this paper.

## 2.1   Feature Extraction

This project considers the fronto-parallel view of a walking person to be that which is perpendicular to the direction of walk. The image sequences of four individuals (A, B, C and D) are collected using a stationery camera i.e. walk image sequences with static background. The aim of this work is to recognize individuals using the "style of walk" information alone. That is, the same person can act in walking sequences with different types and colour of clothing, even though the algorithm needs to recognize that individual properly.

To remove the effect of changing clothing colour, only the silhouettes of the walking subjects are used in the representation. In addition, the silhouettes are scaled to remove the effect of changing depth of the walking subject in the view of the camera. For the foreground segmentation the background subtraction is applied and then binarized using a suitable threshold. Morphological operators such as erosion and dilation [6] are first used to further filter spurious pixels. Small holes inside the extracted silhouettes are then filled. A connected component extraction is finally applied to extract a connected region with the largest size motion area of the walking human. The motion area included by the bounding box (Figure 1) is cropped, then re-scaled using a bilinear method [6] into a fixed size $S \times S$ image ($S = 64$, number of pixels). Figure 2 shows normalized silhouettes of two individual's walking sequences.

The size of these cropped images is considerable. Therefore a dimensionality reduction needs to be applied before applying the classification algorithm. This is done using Principal Components Analysis (PCA) on those images.
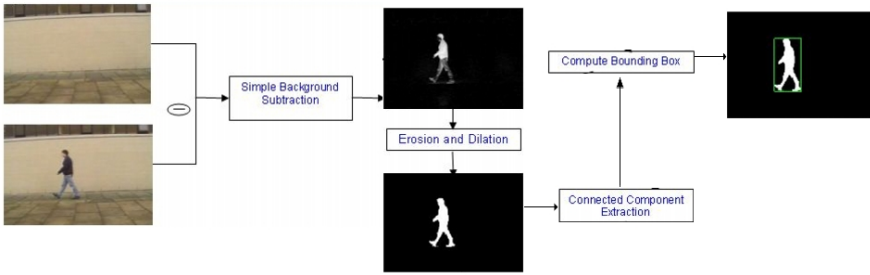
**Fig. 1.** Silhouette extraction method



**Fig. 2.** The silhouettes of individuals A and B's walking sequences

**Dimensionality Reduction Using PCA** is applied successfully in many applications. Sirovich et al [7] has applied this technique for face recognition. In this work also the PCA technique is used for individual recognition.

The image set is defined as $\{y(i)|i = 1, 2, \ldots, M\}$, where $M$ is the number of images in the set. Next the average image $\hat{y}$ is calculated, which is the mean image of all images in the set. An image matrix $P$ is constructed by subtracting $\hat{y}$ from each image and stacking the resulting vectors column-wise. Let, $P$ be $N \times M$, where $N$ is the number of pixels in each image.

In this experiment 114 images are used for training, i.e. $M = 114$. Each image size is $64 \times 64$, i.e. $N = 4096$ and $N > M$. If we consider the covariance matrix $Q$, where $Q = PP^T$, then $Q$ is $N \times N$ and $N > M$. Calculation of the eigenvectors of a matrix as large as $Q$ is computationally intensive. For improvement the implicit covariance matrix $\hat{Q}$ is used, where: $\hat{Q} = P^T P$.

Note that $\hat{Q}$ is an $M \times M$ matrix and therefore much smaller than $Q$. Here $M$ eigenvectors of $\hat{Q}$ can be computed. These can be computed much faster than the first $M$ eigenvectors of $Q$ due to the disparity in the size of the two matrices. It can be shown that the $M$ largest eigenvalues and corresponding eigenvectors of $Q$ can be determined from the $M$ eigenvalues and eigenvectors of $\hat{Q}$ as: $\lambda_i = \hat{\lambda}_i$ and $e_i = \hat{\lambda}^{-1/2} P \hat{e}_i$ [8]. Here, $\lambda_i$ and $e_i$ are the $i^{th}$ eigenvalue and eigenvector of $Q$, while $\hat{\lambda}_i$ and $\hat{e}_i$ are the $i^{th}$ eigenvalue and eigenvector of $\hat{Q}$. Previous research has proved that only a few eigenvectors are necessary for visual recognition. Therefore, we use the first $k$ eigenvectors calculated corresponding to the largest

$k$ eigenvalues. The $k$-dimensional subspace spanned by these eigenvectors is called the *eigenspace*. Singular Value Decomposition (SVD) is applied to the data set as $N$ is much larger than $M$ [8]. It is not viable, however, when more than $M$ eigenvectors are needed.
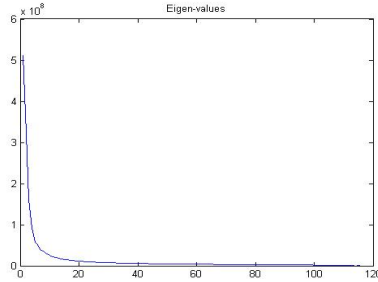


**Fig. 3.** Eigenvalues obtained from applying SVD algorithm on the entire training sequence

Figure 3 shows the eigenvalues calculated by the SVD algorithm for the training image sequence. The first 40 eigenvalues are greater than zero and then the eigenvalues tend to zero. Therefore, we use the first 40 eigenvectors corresponding to these 40 eigenvalues to reconstruct silhouette sequences. Figure 4 shows the reconstruction result of a silhouette image. Using this dimensionality reduction algorithm, each image in the sequence can be represented by a $k$ dimensional vector and this vector is represented as a **feature vector** for each image.
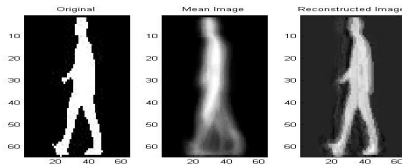


**Fig. 4.** Left image: Silhouette obtained from test sequence. Middle image: Mean image. Right image: Reconstructed image using first 40 eigenvectors.

An image in the individual sequence can be mapped to a point $f(i)$, where $f(i) = [e_1, e_2, \ldots, e_k]^T y(i)$, in the eigenspace. Here, $f(i)$ is the $k$th dimensional feature vector for image $y(i)$. Therefore, a sequential movement can be represented as a trajectory in the eigenspace. An example of walking patterns for four different individuals is shown in Figure 5. This is the eigenspace representation of the individual trajectory manifolds.
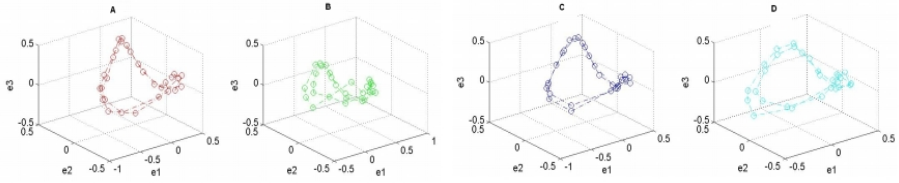
**Fig. 5.** Trajectory of 4 individuals in 3D space made of first 3 eigenvectors

## 2.2 Recognition Stage

Walking sequences are periodic and cyclic. In the walking sequence all the phases which match the known poses are called a *walk-cycle*. An example of a human walk-cycle is shown in Figure 6. If an image sequence contains a walk-cycle then that sequence can be called a full walking sequence.

Murase et al [9] and He et al [10] assumed their database had the full cyclic walking sequence data. If any sequence does not have a walk-cycle then we can say those sequences are *partial* sequences. For the partial case the above two methods did not give any solution for partial data comparison. To solve this problem, we needed to find the longest common subsequence from both partial walking sequences. To find the common subsequence, we used the Longest Common Subsequence (LCS) algorithm [11].

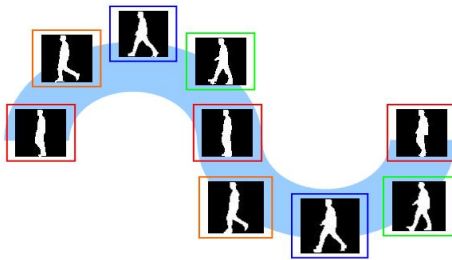In our work we assume walk sequences may consist of full cyclic motion data or partial motion data.



**Fig. 6.** The silhouettes from a walk-cycle

**Longest Common Subsequence (LCS) Algorithm** finds the longest subsequence that two sequences have in common, regardless of the length and number of intermittent non-matching symbols. For example, the sequences **"abcdefg"** and **"axbydezzz"** have a sequenced length of four **"abde"** as their longest common subsequence. Figure 7 given below has more information.
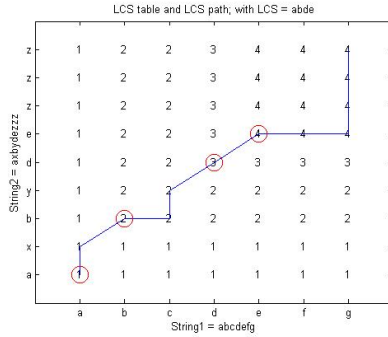
**Fig. 7.** The longest common sequence "abde" found from two strings "abcdefg" and "axbydezzz" using LCS algorithm

Formally, the LCS problem is defined as follows: Given a sequence $X = (x_1, x_2, \ldots, x_m)$, and a sequence $Y = (y_1, y_2, \ldots, y_n)$, find a longest common sequence $Z = (z_1, z_2, \ldots, z_k)$. The solution to the LCS problem involves solving the following recurrence equation, where the cost for the edit operations stored in $C$ is:

$$C(i,j) = \begin{cases} 0 & if \quad (i=0) \;\; or \;\; (j=0) \\ C(i-1, j-1)+1 & if \quad (i,j>0), \; (x_i = y_j) \\ max \; [C(i, j-1), C(i-1, j)] & if \quad (i,j>0), \; (x_i \neq y_j) \end{cases} \qquad (1)$$

Using LCS as a similarity measure between two sequences has the advantage that the two sequences can have different lengths and have intermittent non-matches. In the context of individual recognition, this allows for the use of partial walking sequences with noisy inputs.

Given two image sequences ($S_1$ and $S_2$) the 40-dimensional feature vector is calculated, as described before, for each frame in those sequences. Further we normalize those vectors to unit length to then apply the correlation measure between two frames: $c = x^T y$, where $c$ is the correlation value between 0 and 1, $x$ and $y$ are the normalized vectors of the corresponding frames from $S_1$ and $S_2$ respectively. These correlation values are stored in a matrix $C$. The rows and columns in matrix $C$ represent the frames from sequence $S_1$ and $S_2$ respectively. Each value in matrix $C$ tells the degree of similarity between the frames from $S_1$ and $S_2$. From experimental results the correlation value greater then or equal to 0.7 gives similar frames. The threshold value is defined as 0.7 for good experimental results. Now the most similar frames corresponding to both these sequences can be calculated. To do this, we need to find the maximum value for each row. If that maximum value is greater than a threshold value we can say the frames represented by the rows and columns are similar (or equal). It is important to find the similar frames before applying LCS algorithm as there is a calculation in LCS algorithm that, if $(x_i = y_j)$ then $c(i,j) = c(i-1, j-1) + 1$. Using the Longest Common Sequence algorithm we can find a set of pair of frames, which

are the similar frames from two walk sequences $S_1$ and $S_2$. The values from the matrix $C$ corresponding to each pair of frames from the set are summed and finally we find the mean value. This mean value gives the final measure of the two sequences.

## 3   Individual Identification Using "Style of Actions"

The image sequences of three individuals (A, B, and C) are taken. The local space-time intensity gradient $(S_x, S_y, S_t)$ is calculated at each space-time point $(x, y, t)$ for each frame in the sequence. We find the absolute values of the temporal derivatives in all pixels of the image frame and ignore all space-time points $(x, y, t)$ for which the temporal derivative is below some threshold, thus performing the calculation mostly on the points which participate in the event. Let the thresholded image with the detected object be divided into $10 \times 10$ bins. Each bin has a numeric value of the total number of pixels which have non-zero values. The values in each bin are stacked into a 100-dimensional vector. This 100-dimensional vector is used as a feature to represent each frame (Figure 8).
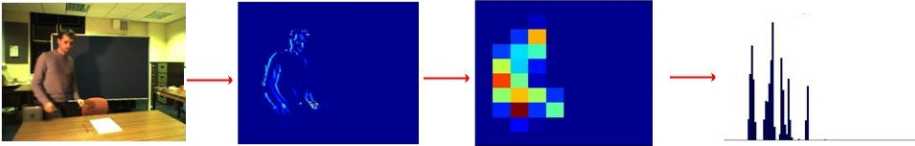


**Fig. 8.** 100D feature vector generation for image frame in sequence

There are many actions available from the individuals image sequences: enter, pick up the pen, sit down, write their name, get up and leave etc. The frames are clustered into groups to represent the actions (here we assume the number of groups = 9). Consider the whole image frames from these three image sequences and apply the K-means algorithm. After clustering the 100-D vectors in each cluster we represent similar actions. We calculate the mean of the vectors from each cluster and this mean vector is the "code-word". Each code-word represents a different action.

For the purpose of individual re-identification a similarity measure is needed to match two image sequences, and find whether or not the same individual appears in those sequences. If two image sequences are captured from a particular person in a different time then the length of these two sequences need not be with the equal number of frames depending on both the speed and movement performance. Therefore image sequence alignment is important for the similarity measure. Here the DTW (Dynamic Time Warping) [11] alignment algorithm is used for sequence alignment with an assumption that start and end frames are correctly aligned.

Method: a) Take an image sequence that needs to be matched to another sequence. b) For each frame in the image sequence, find the most suitable code-word. This operation converts an image frame sequence into a code-word sequence. c) Apply the DTW algorithm for code-word sequences to do alignment and find the distance. Actually this alignment is based on actions.

## 4    Experiments and Results

**Using Gait Sequences:** In Table 1, the already known individual's data is represented in rows. Frames 1 to 8 are taken from this data as a partial sequence. The column data are taken from the unknown individuals and we need to find which individuals appear in those sequences. Frames from 4 to 13 are taken from this data as partial data. We can expect that the common subsequence should contain frames 4 to 8. Due to noise it varies slightly.

**Table 1.** Similarity matrix for known - unknown sequences

| Individuals | 1-Seq | 2-Seq | 3-Seq | 4-Seq |
|:-----------:|:-----:|:-----:|:-----:|:-----:|
| S1 | **0.9458** | 0.8305 | 0.8908 | 0.8542 |
| S2 | 0.7586 | **0.9877** | 0.8036 | 0.8006 |
| S3 | 0.8979 | 0.8748 | **0.9571** | 0.8867 |
| S4 | 0.8735 | 0.7285 | 0.8783 | **0.9031** |

For the maximum value in each column it can find the corresponding row (i.e, column represents the unknown person). This maximum value indicates the high similarity between the row and column. Therefore individuals appearing in corresponding rows and columns are the same person. Here also the threshold value is defined as 0.9. Therefore if the highest value is greater than this threshold value then we accept that the same person is available in both sequences.

Significant progress has been made in individual identification using their "style of walk" over the past few years. Compared with "style of walk", not much progress has been made in individuals recognition using their "style of actions".

**Using Action Sequences:** For this experiment we have chosen 3 individuals: A, B and C with similar height and width. Table 2 shows the time normalized accumulated distance (see Figure 9) between two sequences. The diagonal elements represent the smallest value in each row. The smallest value shows that individuals appearing in the two sequences are similar based on their "way of doing actions" i.e. "body language".

We could also apply our LCS based individual recognition technique to this "style of actions" based individual recognition as well.

**Table 2.** Similarity matrix for known - unknown sequences

| Individuals | A1 | B1 | C1 |
|---|---|---|---|
| A2 | **0.0608** | 0.2213 | 0.2976 |
| B2 | 0.2543 | **0.0698** | 0.2588 |
| C2 | 0.2514 | 0.1179 | **0.0917** |
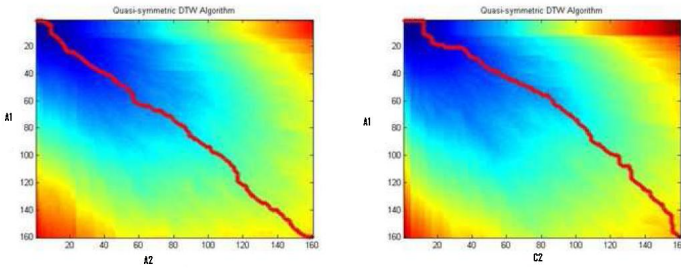


**Fig. 9.** The path with the minimum distance between two individuals sequences. Path towards the diagonal gives the the minimum distance.

## 5   Conclusion

This paper has implemented methods for individual recognition based on *"style of walk"* and *"style of actions"*. Both of these methods can be applied to outdoor surveillance e.g. when an unknown individual enters a restricted area and indoor surveillance e.g. when an unknown individual is in an office. The individual recognition method considered the sequence of particular actions (sit down, write name, stand up etc.). It gave good similarity measures between two different individuals actions sequences. To keep the spatial invariance between individuals the individuals were carefully selected with similar height and width. To improve this system we need a larger training set.

Further, to scale up this system into an automated system, as body-language information fully depends on body parts motion, so the body parts must be detected automatically. Optical flow [12] motion detection methods are shown to be more useful for finding particular moving regions e.g. body parts. The main constraint in the DTW method is that the first and last frames in the video sequence should be aligned properly. Ongoing work applies N-Cut clustering to individuals video sequences.

## References

1. Johansson, G.: Visual motion perception. Science American 232(6), 76–88 (1975)
2. Cutting, J., Prott, D., Kozlowski, L.: A biomechanical invariant for gait perception. Journal of Experimental Psychology: Human Perception and Performance 4(3), 357–372 (1978)

3. Little, J.J., Boyd, J.E.: Recognizing people by their gait: The shape of motion. Computer Vision Research 1(2), 1–32 (1998)
4. Nixon, M., Carter, J., Nash, J., Huang, P., Cunado, D., Stevenage, S.: Automatic gait recognition. IEE Colloquium Motion Analysis and Tracking, 31-36 (1999)
5. Johnson, A., Bobick, A.: Gait recognition using static, activity specific parameters. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 423–430 (2001)
6. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 2nd edn. Pearson Education, London
7. Sirovich,, Kirby, M.: Low dimensional procedure for the charecterization of human faces. Journal of Optical Society of America 4(3), 519–524 (1987)
8. Nayar, S.K., Murase, H., Nene, S.A.: Parametric appearance representation in early visual learning, ch. 6. Oxford University Press, Oxford (1996)
9. Murase, H., Sakai, R.: Moving object recognition in eigenspace representation: gait analysis and lip reading. Pattern Recognition Letters 17(2), 155–162 (1996)
10. He, Q., Debrunner, C.H.: Individual recognition from periodic activity using hidden markov models. In: Proceedings IEEE Workshop on Human Motion, pp. 47–52 (2000)
11. Guo, A., Siegelmann, H.: Time-warped longest common subsequence algorithm for music retrieval. In: Proc. Fifth International Conference on Music Information Retrieval, pp. 10–14 (2004)
12. Condell, J.V., Scotney, B.W., Morrow, P.J.: Adaptive Grid Refinement Procedures for Efficient Optical Flow Computation. International Journal of Computer Vision (1), 31–54 (2005)