Human Recognition by Gait Analysis Using Neural Networks

J. Elías Herrero-Jaraba, Carlos Orrite-Uruñuela,
David Buldain, and Armando Roy-Yarza

Dept. of Electronic and Communication Engineering
Maria de Luna, 1, Zaragoza, 50015 SPAIN

{jelias, corrite, buldain, armanroy}posta.unizar.es

Abstract. This paper presents a new method to recognize people by their gait, which forms part of a major project to detect and recognize automatically different human behaviours. The project is comprised of several stages, but this paper is focused on the last one, i.e. recognition. This stage is based on the Self-Organizing Map, assuming that the previous stage of feature extraction has already been solved. Although this previous stage is solved with manual extraction of human model points, the obtained results demonstrate the viability of the neural approach to the recognition of these kind of temporal sequences.

1 Introduction

This paper shows a new approach in the area of biometrics, i.e., the automatic identification of an individual based on their physical or behavioural characteristics. In the present approach we describe a new method for human recognition based on automatic gait analysis. Gait is a new biometric aimed at recognising someone by the way they walk with several notable advantages over other biometrics systems. It allows recognition from a great distance, where other biometrics systems might fail and, on the other hand, it is a non-invasive technique in the sense that the subject need not even know they are being recognised. This kind of recognition has been studied in others works [8], but in our case we implemented the solution by using neural networks, on account of which the system is more versatile and flexible. This system begins with a pre-processing stage in which the suitable data are extracted, defined as link points of several limbs of the human body.

The work forms part of a major project to detect and recognize automatically different human behaviours. It is comprised of four stages: segmentation of moving objects in the image, tracking, sub-segmentation, and recognition. Sub-segmentation stage deals with the detection of the body parts over the whole segmentation of the human shape. Nonetheless, this paper is focused on the last stage, and we want to demonstrate the viability of the approach before the project is fully finished.

The recognition stage deals with the processing of temporal sequences (humans walking in a scene). Therefore, this kind of data could be conducive to use recurrent neural networks. Recurrent Self-Organizing Map (RSOM) [7] is an example of them.

J.R. Dorronsoro (Ed.): ICANN 2002, LNCS 2415, pp. 364-369, 2002.

RSOM includes a recurrent difference vector in each unit, which allows storing temporal context from consecutive input vectors fed to the map. Another example is SARDNET [3], which extends the Kohonen Map architecture with activation retention and decay to create unique distributed response patterns for different sequences. Finally, the Kohonen map is used in [1], but relies on spaces of functions, instead of the usual spatial inputs.

In our case, we are going to use a standard Self-Organizing Map [6]. The network activates only one cell (the "winner" cell) when a new input data vector is presented, as if it were static pattern recognition. The sequence of these cells, in response to a sequence of input vectors taken over some period of time draws a trajectory over the map. These trajectories define the kind of movement, or the human identity, i.e. the trajectories include the dynamical information of data. In this paper we are going to focus on human identification, but it is very easy to extend the method to other identification tasks.

The organization of the paper is as follows: section 2 shows our approach, explaining the data acquisition and the needed pre-processing. In Section 3, we comment the accomplished experiments. Finally, section 4 concludes this paper by commenting the conclusions about our work and presenting the future lines of research.

2 Our Approach

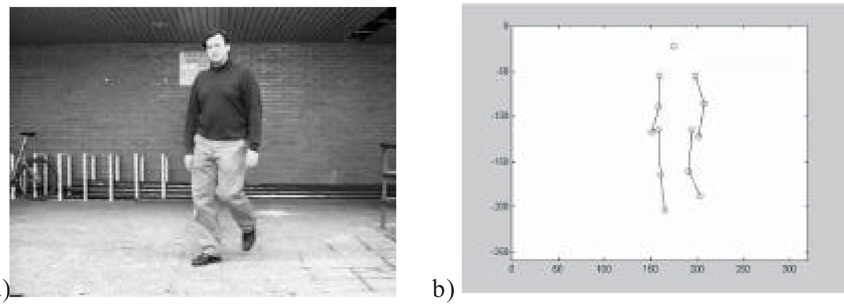
As we have already commented in previous section, this paper is focused on the recognition of human identities by using a neural network (Self-Organizing Map). Nevertheless, the data must be processed before being presented to the map. Likewise, the dimension of the data space is very large, thus we have to reduce it by considering only the most significant projections, i.e. to choose only those principal components with a high variance. Finally, a map must be adequately trained.

2.1 Data Acquisition

Because of the fact that Sub-segmentation task is not developed yet (a preliminary version has been actually finished), we manually worked out the learning data.

At beginning, we grabbed some sequences of 6 people (models) moving in different directions at outdoor scenes. From these sequences, we obtained the model-shots (a piece of sequence in which a particular model is walking) of each one of the models in AVI format. An example frame from these model-shots is shown in Fig. 1a.

After, we obtained the control points (see Fig. 1b) by clicking with the mouse in the body part. These control points (*CP*) are defined as a 2N-dimensional array, thus we obtain a new *CP* in each frame. Let $CP_j = \{y_1 \;\; \cdots \;\; y_N \;\; x_1 \;\; \cdots \;\; x_N\}^T$ be the *CP* of the frame *j*, where $\{x_i, y_i\}_{i=1 \ldots N}$ are the (x, y)-coordinates of the element *i* of CP_j, and *N* is the number of control points (in our case, *N=13*).



a) b)

Fig. 1. a) Frame of a model-shot, and b) Control points obtained manually in the frame.

2.2. Pre-processing

Pre-processing is comprised of three steps: Computing relative coordinates, scaling, and Principal Analysis Components (PCA). Even though the CP coordinates can be directly considered as SOM input data, the SOM performance is much higher if these data are previously pre-processed. Otherwise, the neural network has to decode the whole complexity of the data.

Relative Coordinates: Since the model is moving in the camera vision field, the SOM may interpret the same position of the body in various parts of the image in different way. This is not an advisable characteristic, and we need therefore to design an invariant system to model the position.

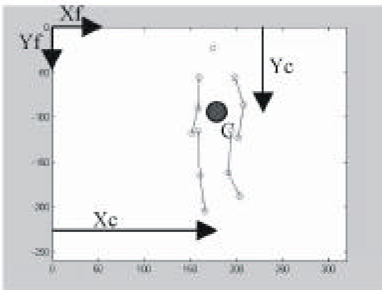


Fig. 2. Relative coordinates and transformations needed to compute them.

Thus, the frame reference is defined at the top left-hand corner of the image, and C is defined as the centre of the CP, whose coordinates in frame reference are *(Xc, Yc)* (see Fig. 2). The data reference is transformed to its centre reference, which is calculated in according to Eq. 1.

$$C = \begin{bmatrix} Xc \\ Yc \end{bmatrix} = \begin{bmatrix} \frac{1}{N} \cdot \sum_{i=N+1}^{2N} CP_j(i) \\ \frac{1}{N} \cdot \sum_{i=1}^{N} CP_j(i) \end{bmatrix} \quad \textbf{(1)}$$

Scaling: The translated data are invariant with respect to the position of *CP* on the frame. Nevertheless, we need to normalize this data. Normalization is comprised of a scaling with respect to the height and width of the model shown in the image.

Principal Component Analysis: Owing to the fact that the data present a high dimensionality (2xN-dimensional), we reduce the input data of the SOM with a Principal Component Analysis (PCA). It takes into account only the maximal variability axes of the input data, and projects all data into these axes.

The number of projections (t < N) depends on the proportion (p), i.e. percentage, which is needed to explain the variance exhibited in the input data space. Let ${}_i$ be the eigenvalues of the covariance matrix of the input data. The total variance V_T is the sum of all the eigenvalues, as we can see in Eq. 2 [2].

$$\sum_{i=1}^{t} {}_i \geq p \cdot V_T \qquad and \qquad V_T = \sum_{i=1}^{N} {}_i \tag{2}$$

Thus, in our experiments we have obtained for a proportion of 95% the next number of projected components: $t=12$.

2.3 Recognition

This process is based on Self-Organizing Map (SOM). Nowadays, this kind of neural network is applied to several engineering applications [5]: process and systems analysis, process analysis and monitoring, statistical pattern recognition, or robotics.

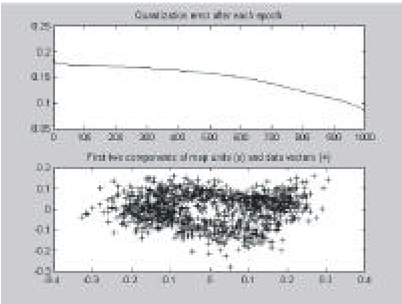
The chosen topology of the SOM is a rectangle map of 20 x 20 cells. Likewise, the number of iterations in the training period has to be large enough to increase the final statistical accuracy; these parameters appear in Table 1.

The process begins with an initialisation period in which the initial values for the weights are chosen randomly, with the only restriction being that they should be different.

During the training period, there exist two phases: ordering and tuning [4]. The first phase takes charge of adapting the weights with a broad neighbourhood to generate a global ordering of the map over the data space, and the tuning phase provides the fine adjustment of the weights. The parameters used in both phases are shown in Table 1. By way of example, the ordering phase appears in Figure 3, where the quantization error and the ordering of two first components are drawn.

Table 1. Parameters used in training period. R_INI and R_FIN are the initial and final radius of neighborhood, ALPHA_INI is the initial value of rate learning, ALPHA_TYPE is the decreasing function of rate learning, and TRAINLEN is the number of iterations.

	R_INI	R_FIN	ALPHA_INI	ALPHA_TYPE	TRAINLEN
Ordering	15	1	0.9	Inv.	1000
Tuning	2	2	0.01	Const.	4000

a) 

b) 

Fig. 3. a) Graphical representation of the ordering phase. b) "Winner" cells labelled with the number of model.

The whole input data used for learning is presented to the trained net, and the "winner" unit of each input pattern is detected. This set of neurons is labelled with the number of the models in a supervised way. Therefore, we obtain a matrix like the one in Figure 3b. As can be seen in this figure the map works as a static pattern recogniser, and the close units on the SOM lattice are mapped to input patterns close in the input space due to the fact that the SOM keeps the property of topology preservation. It exists the possibility that the same units of the map were excited by patterns of various models, but each model shows different positions during the walking cycle. Therefore, this last property is used to exploit our system as recogniser of temporal sequences.

3 Experiments

In this section, we show the results for 6 models and only one walking direction. Nevertheless, if we wanted to identify models in *n* directions, *n* maps would have to be added working in parallel. As can be seen in Figure 4, the trajectories for each model are clearly different. Otherwise, these are drawn over the distance map (which represents the mean distance of a neuron's centroid from their closest neighbours) in which four regions appear. These regions can be recognised as the stages in a human walking cycle, and every model crosses over them but in a different way, determining its particular trajectory. There exist various deviations from the mean trajectory, but it is because of the fact that point extraction was accomplished manually. Nevertheless, this problem does not affect to the system performance, since the percentage of these deviations is much smaller.

4 Conclusions

We have presented a new approach in the use of Self-Organizing Maps with successful results. The SOM is able to discriminate the gait details of humans with a promising results. Besides it generates four clear regions associated to the stages in a human walking cycle. This result promises a good behaviour of the SOM as biometric automatic recogniser.

The future work is focused on adapting this system to be able to recognize behaviour or gestures of people. The development of an application for activities recognition in real-time is considered as the main goal of our work. This work has been supported by the Spanish government, and the European Commission, FEDER, project Nº. 2000/0347 “Tele-observation and moving patterns identification system”.

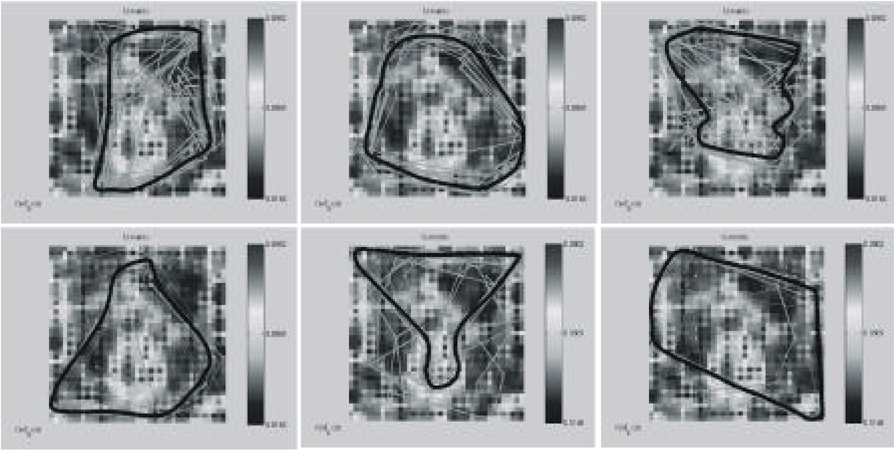


Fig. 4. Trajectories of each one of the models used in our experiments. The mean trajectory is drawn as a wider line than real trajectories.

References

[1] J.C. Chappelier, A. Grumbach, “A Kohonen map for temporal sequences”, NEURAP’95 Conference, Marseille, France, March 1996.

[2] T. F. Cootes, , and C. J. Taylor, “Statistical models of appearance for computer vision”, Technical report, www.isbe.man.ac.uk/bim/refs.html, 2001.

[3] D.L. James, R. Miikkulainen, “ SARDNET: A Self-organizing feature map for sequences”, Neural Processing Systems, 7, 1995.

[4] T. Kohonen, “The Self-organizing map”, Proc. of the IEEE, Vol. 78, nº 9, pp. 1464-1480, September 1990.

[5] T. Kohonen, “Engineering Applications of the Self-organizing map”, Proc. of the IEEE, Vol. 84, nº 10, pp. 1358-1384, October 1996.

[6] T. Kohonen, “Self-Organizing Maps”, 2nd Edition, Springer-Verlag Berlin Heidelberg, New York, 1997

[7] T. Koskela, “Temporal sequence processing using recurrent SOM”, Engineering Systems, Australian, Vol. 1, pp. 290-297, April 1998.

[8] Y. Yacoob, M.J. Black, “Parameterized modelling and recognition of activities”, Sixth Int. Conf. on Computer Vision, India, January 1998.