Heuristic Method of Feature Selection for Person Re-identification based on Gait Motion Capture Data

Henryk Josiński[2], Agnieszka Michalczuk[1], Daniel Kostrzewa[3], Adam Świtoński[1], and Konrad Wojciechowski[1,2]

[1] Polish-Japanese Institute of Information Technology, Aleja Legionów 2, 41-902 Bytom, Poland
{amichalczuk,aswitonski,kwojciechowski}@pjwstk.edu.pl
[2] Silesian University of Technology, Institute of Computer Science, Akademicka 16, 44-100 Gliwice, Poland
{Henryk.Josinski,Konrad.Wojciechowski}@polsl.pl
[3] Silesian University of Technology, Department of Industrial Informatics, Krasińskiego 8, 40-019 Katowice, Poland
Daniel.Kostrzewa@polsl.pl

Abstract. The authors present a heuristic method of feature selection for gait mocap data, based on the exIWO metaheuristic which is characterized by both the hybrid strategy of the search space exploration and three variants of selection of individuals as candidates for next population. The proposed method was evaluated by the accuracy of person re-identification based on the selected feature subset. Because of the high-dimensional nature of motion data, feature selection was preceded by the data dimensionality reduction.

Keywords: feature selection, exIWO algorithm, gait analysis, person re-identification, dimensionality reduction, MPCA algorithm

1 Introduction

Gait is defined as coordinated, cyclic combination of movements which results in human locomotion [1]. A unique advantage of gait as a biometric is that it offers potential for recognition at a distance or at low resolution or when other biometrics might not be perceivable [2]. Gait can be captured by two-dimensional video cameras of surveillance systems or by much accurate motion capture (*mocap*) systems which acquire motion data as a time sequence of poses. In the latter case the movement of an actor wearing a special suit with attached markers is recorded by NIR (Near Infrared) cameras. Positions of the markers in consecutive time instants constitute basis for reconstruction of their 3D coordinates.

Although the inconvenience of the acquisition process excludes direct application of the mocap system for human identification in a surveillance system, high precision of mocap recordings make them useful in the development stage

of methods for solving the *person re-identification problem* which can be formulated as the question of when person detections in different views or at different time instants can be linked to the same individual [3].

Motion data lie in high-dimensional space, but the components of gait description are correlated, what allows to transform a high-dimensional data into a low-dimensional equivalent representation while retaining most of the information regarding the underlying structure or the actual physical phenomenon [4]. The dimensionality reduction problem can be solved, *inter alia*, by encoding an image object as a general tensor of second or higher order [5].

Feature (attribute, variable) selection is expected to simplify object description, discover most discriminative features and give a chance for more precise classification. Most methods involve searching the space of attributes for the subset that is most likely to predict the class best [6]. In the *filter* approach features are selected on the basis of statistical properties, i.e. by ranking them with correlation coefficients. *Wrappers* assess subsets of variables according to their usefulness to a given predictor. In practice, one needs to define: (i) how to search the space of all possible variable subsets; (ii) how to assess the prediction performance of a learning machine to guide the search and halt it; and (iii) which predictor to use [7]. Making use of the wrapper methodology, the authors applied the exIWO metaheuristic to choose a correct subset of predictive attributes. The exIWO constitutes an expanded version of the Invasive Weed Optimization (IWO) algorithm. The authors of the original method [8] from University of Tehran were inspired by observation of dynamic spreading of weeds and their quick adaptation to environmental conditions. The exIWO proposed by the authors of the present paper is characterized by both the hybrid strategy of the search space exploration and three variants of selection[4] of individuals as candidates for next population. According to the authors' knowledge, the algorithm has never been used for the purpose of feature selection.

Main goal of the research is an evaluation of the influence of a heuristic method of feature selection on accuracy of person re-identification based on gait mocap data.

The organization of this paper is as follows – section 2 contains a brief description of both human motion acquisition procedure and gait data tensor representation. Application of the MPCA algorithm for gait data dimensionality reduction is presented in section 3, whereas adaptation of the exIWO metaheuristic to the feature selection is discussed in section 4. Section 5 deals with procedure of the experimental research along with its results. The conclusions are formulated in section 6.

2 Gait Data Acquisition and Representation

Gait sequences were recorded in the Human Motion Laboratory (HML) of the Polish-Japanese Institute of Information Technology (`http://hm.pjwstk.edu.`

[4] It is necessary to mention that the term "selection" is used in the present paper in the following meanings: (i) feature selection, (ii) selection of individuals.

pl) by means of the Vicon Motion Kinematics Acquisition and Analysis System equipped with 10 NIR cameras with the acquisition speed of 100 to 2000 frames per second at full frame resolution of 4 megapixels and 8-bit grayscale (Fig. 1). The gait route was specified as a 5 meters long straight line. The acquiring process started and ended with a T-letter pose because of requirements of the Vicon calibration process. As a result of the acquisition procedure 353 sequences for 25 men aged 20-35 years were stored in a gait database.

Fig. 1. Recording session in the HML.

Tensor object is a multidimensional object, the elements of which are to be addressed by indices. The number of indices determines the order of the tensor object, whereas each index defines one of the tensor modes. Gait silhouette sequences are naturally represented as third-order tensors with column, row, and time modes [9].

Description of each of the consecutive poses of a gait sequence depends on the assumed skeleton model. For a typical model containing 22 segments and a global skeleton rotation (Fig. 2), description of a single pose comprises values of 69 Euler angles. Three additional values are required for specification of a global translation.

The third-order tensor representation is based on modes indexed, respectively, by numbers of components of Euler angles ("angle mode"), numbers of skeleton components ("pose mode"), and numbers of sequence frames ("time mode"). Gait sequences consist of 128 frames. Euler angles increase the total number of features characterizing a single sequence to 8832.

3 Data Dimensionality Reduction – the MPCA algorithm

Multilinear projection of tensor objects for the purpose of dimensionality reduction is the basis of the multilinear principal component analysis (MPCA) which is the multilinear extension of the PCA method. According to the authors of the MPCA algorithm: "Operating directly on the original tensorial data, the proposed MPCA is a multilinear algorithm performing dimensionality reduction in

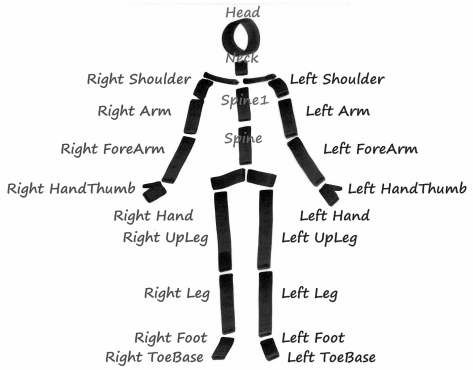

Fig. 2. Segments of the skeleton model.

all tensor modes seeking those bases in each mode that allow projected tensors to capture most of the variation present in the original tensors" [9]. Input data are required to be tensor objects and the algorithm operates directly on tensors producing as output low-dimensional *feature tensors*.

In the MPCA an elementary matrix algebra is extended by two operations: tensor unfolding and the product of a tensor by a matrix (Fig. 3). The unfolding (matricization) transforms a tensor into a matrix along a specified mode. In other words, the tensor is decomposed into column vectors. The n-mode multiplication of a tensor x by a matrix U ($x \times_n U$) is realized by the product of the tensor x unfolded along the n-mode ($x_{(n)}$) by the matrix U, followed by the folding operation which creates a resultant tensor: $x \times_n U = \text{fold}_n(U\ \text{unfold}_n(x)) = \text{fold}_n(Ux_{(n)})$. To put it another way, tensor x is projected in the n-mode vector subspace by the projection matrix U.

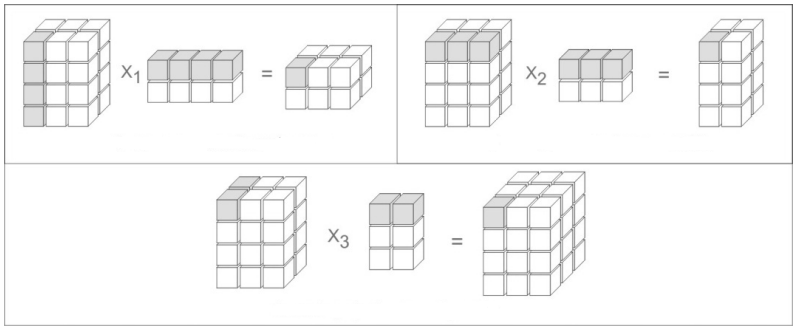

Fig. 3. Multiplication of a tensor by a matrix.

The MPCA algorithm consists of the following steps:

1. Preprocessing – normalization of all M N-order input tensor samples $x_m \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ $(m = 1..M)$ to zero mean value $(\tilde{x}_m)$. I_n denotes the n-mode dimension of the tensor $(n = 1..N)$.
2. Learning phase which has an iterative character:
   (a) Initialization – for each mode n:
      - Computation of the matrix $\Phi^{(n)*}$: $\Phi^{(n)*} = \sum_{m=1}^{M} \tilde{X}_{m(n)} \tilde{X}_{m(n)}^T$, where $\tilde{X}_{m(n)}$ is the n-mode unfolded matrix of $\tilde{x}_m$.
      - Eigendecomposition of the matrix $\Phi^{(n)*}$.
      - Selection of P_n most significant eigenvectors which form a projection matrix $\tilde{U}^{(n)}, \tilde{U}^{(n)} \in \mathbb{R}^{I_n \times P_n}, P_n \leq I_n$. Eigenvectors are evaluated on the basis of the corresponding eigenvalues. Variation coverage $Q^{(n)}$ in the n-mode after the truncation of the eigenvectors beyond the P_n-th could be calculated as follows: $Q^{(n)} = \sum_{i_n=1}^{P_n} \lambda_{i_n}^{(n)} / \sum_{i_n=1}^{I_n} \lambda_{i_n}^{(n)}$ where $\lambda_{i_n}^{(n)}$ is the i-th eigenvalue of the matrix $\Phi^{(n)*}$. In practice, only one user-defined value of variation coverage Q determines the number of selected eigenvectors for each mode separately.
   (b) Iterative local optimization of the projection matrices $\tilde{U}^{(n)}$ $(n = 1..N)$:
      - Computation of the matrix $\Phi^{(n)}$: $\Phi^{(n)} = \sum_{m=1}^{M} \left(X_{m(n)} - \bar{X}_{(n)}\right) \cdot \tilde{U}_{\Phi^{(n)}} \cdot \tilde{U}_{\Phi^{(n)}}^T \cdot \left(X_{m(n)} - \bar{X}_{(n)}\right)^T$, where $\bar{X}_{(n)} = \frac{1}{M} \sum_{m=1}^{M} X_{m(n)}$ and $\tilde{U}_{\Phi^{(n)}} = \left(\tilde{U}^{(n+1)} \otimes \tilde{U}^{(n+2)} \otimes \ldots \otimes \tilde{U}^{(N)} \otimes \tilde{U}^{(1)} \otimes \ldots \otimes \tilde{U}^{(n-1)}\right)$; $\otimes$ denotes the Kronecker product.
      - Eigendecomposition of the matrix $\Phi^{(n)}$.
      - Construction of the matrix $\tilde{U}^{(n)}$ which consists of the P_n eigenvectors corresponding to the largest P_n eigenvalues of the matrix $\Phi^{(n)}$.
      - In our version of the MPCA algorithm the local optimization stage is repeated until a user-specified number of iterations is reached.
3. Reduction phase – projection of the input samples using the matrices $\tilde{U}^{(n)}$ $(n = 1..N)$ results in construction of the low-dimensional representation of the input samples with Q % variation captured, in the form of feature tensors $y_m \in \mathbb{R}^{P_1 \times P_2 \times \ldots \times P_N}$ $(m = 1..M)$: $y_m = x_m \times_1 \tilde{U}^{(1)^T} \times_2 \tilde{U}^{(2)^T} \times_3 \ldots \times_N \tilde{U}^{(N)^T}$.

Our MPCA implementation is based on the Jama-1.0.2 library (`http://math.nist.gov/javanumerics/jama/`).

4 Feature Selection – the exIWO algorithm

The simplified pseudocode describes the exIWO using terminological convention consistent with the "natural" inspiration of the authors of the original IWO version. Consequently, the words "*individual*", "*plant*", and "*weed*" are treated as synonyms.

```
Create the first population.
For each individual from the population:
  Compute the value of the fitness function.
While the stop criterion is not satisfied:
  For each individual from the population:
    Compute the number of seeds.
    For each seed:
      Draw the dissemination method.
      Create a new individual.
      Compute the value of its fitness function.
  Select individuals for a new population.
Return the best individual.
```

The optimization process starts with a random initialization of the first population. An individual is represented by a binary vector of a length equal to the entire number of features. "1" at position i in the vector denotes that the i-th feature belongs to the considered subset. Each weed, i.e. each subset of features constructed by the exIWO is evaluated by means of the 1NN classifier using five-fold cross-validation on a training set. Thus, the fitness function is equivalent to the classification accuracy expressed by means of the Correct Classification Rate (CCR) which indicates the percentage of correctly classified cases.

The number of seeds S_w produced by a single weed depends on the value of its fitness function f_w in the following way:

$$S_w = S_{min} + \left\lfloor (f_w - f_{min}) \left(\frac{S_{max} - S_{min}}{f_{max} - f_{min}} \right) \right\rfloor \tag{1}$$

where S_{max}, S_{min} denote maximum and minimum admissible number of seeds generated, respectively, by the best population member (fitness f_{max}) and by the worst one (fitness f_{min}).

According to the terminological convention the hybrid strategy of the search space exploration proposed by the authors of the paper can be called "dissemination of seeds". It consists of 3 methods randomly chosen for each seed: spreading, dispersing and rolling down. The draw procedure is based on the pseudorandom number generator of the uniform distribution on the interval $[0, 1)$. The sum of probabilities assigned to the particular methods should be equal to 1.

The *spreading* consists in random disseminating seeds over the whole of the search space, i.e. independently of the location of a parent plant.

The *dispersing* is based on the idea proposed in the original IWO version. The degree of difference between the individual and his offspring can be interpreted as the distance between the parent plant and the place where the seed falls on the ground. The distance is described by normal distribution with a mean equal to 0 and a standard deviation truncated to nonnegative values. The standard deviation is decreased in each algorithm iteration as follows:

$$\sigma_{iter} = \left(\frac{iter_{max} - iter}{iter_{max}} \right)^m (\sigma_{init} - \sigma_{fin}) + \sigma_{fin} \tag{2}$$

where *iter* denotes the current iteration ($iter \in [1, iter_{max}]$). Consequently, the distance is gradually reduced. The number of iterations $iter_{max}$ is used as stop criterion. The symbols σ_{init}, σ_{fin} represent, respectively, initial and final values of the standard deviation, whereas m is a nonlinear modulation factor. From the practical point of view, the distance between plants, rounded to the nearest integer value, is interpreted as the number of transformations of the parent individual. A single transformation is a simple binary mutation of a randomly chosen element of the feature vector.

The *rolling down* can be interpreted as a movement of a seed towards a "better" location with respect to the fitness function. The term "neighbours" stands for individuals located at the distance equal to 1 (transformation) from the current plant. The best adapted individual is chosen from among k neighbours (k is a parameter of the method), whereupon its neighbourhood is analyzed in search of the next best adapted individual. This procedure is repeated $k-1$ times giving the opportunity to select the best adapted individual found in the last iteration as a new one. Thus, the method enables exploration of the vicinity of the parent individual's location in the search space.

The term "selection" refers also to competitive exclusion of individuals. Candidates for next population are selected in a deterministic manner according to one of the following methods: global, offspring-based and family-based. Set of candidates for the *global* selection consists of all parent plants and all their newly created descendants. By contrast, the *offspring-based* selection is limited solely to the descendants and thus should decrease the risk of stagnation at non-optimal points in the search space [10]. If the best individual so far was grown in the current population, then despite the fact that it cannot be retained in the next population it will be stored with an eye to the final optimization result. According to the rules of the *family-based* selection [11], each plant from the first population is a protoplast of a separate family. A family consists of a parent weed and its direct descendants. Only the best individual of each family survives and becomes member of the next population. For all 3 aforementioned methods cardinality of a population remains constant in all algorithm iterations. The aspect of individuals' selection was also taken into account in the experimental research.

5 Experimental Research

The goal of the experiments was to evaluate the influence of feature selection for accuracy of person re-identification based on gait mocap data. Dimensionality of the recordings was previously reduced by means of the MPCA. Subsequently, gait sequences of each of the 25 actors were divided evenly into training set containing 180 samples and test set composed of 173 samples. The training set was subject to feature selection performed by the exIWO. Finally, samples from the test set were classified using the 1NN method in the space defined by the selected features. The findings of the initial experiments determined the most appropriate values of the exIWO parameters for the considered problem. They were collected in Table 1.

Table 1. Basic parameters of the exIWO used for feature selection

Description	Value
Population cardinality	$\{10, 50\}$
Number of iterations (stop criterion)	$\{20, 50, 100\}$
Number of seeds sowed by a weed ($S_{max} = S_{min}$)	2
Initial value of standard deviation σ_{init} (dispersing)	8.735
Final value of standard deviation σ_{fin} (dispersing)	0.01
Nonlinear modulation factor m (dispersing)	2.59
Number k of examined neighbours and neighbourhoods (rolling down)	3

The parameters from Table 1 were supplemented by different combinations of probabilities (p_{spr}, p_{disp}, p_{roll}) assigned to the particular methods of dissemination of seeds: spreading, dispersing and rolling down. The probabilities were taken from the range of $[0, 1]$ with a step value of 0.1 taking into account the fact that $p_{spr} + p_{disp} + p_{roll} = 1$. Only in case of the family-based selection values of p_{spr} were limited to $\{0, 0.1, 0.2, 0.3\}$, because this method is assumed to give a chance for the preservation of characteristic features of the plants family, what implicates a marginal importance or even absence of the "random oriented" spreading. Hence, in total, 170 combinations of probabilities were tested for the given set of tensor samples (i.e. for the given variation coverage Q). The number of trial runs in the presence of a single configuration of the probabilities was equal to 5. The workstation used for experiments is described by the following parameters: 2×Intel Xeon X5650 2.66GHz (×6) processor, RAM 12GB 1600MHz. Table 2 includes averaged results of the experiments conducted on several data sets previously reduced by the MCPA according to the given value of Q. The results comprise the classification accuracy (CCR) of the test set: (i) in the space defined by the selected features, (ii) in the space of all features, and total execution time of both selection and final classification. Mean values of the accuracy of the final classification test are illustrated in Fig. 4. The x-axis values denote entire number of features for the given Q (see Table 2).

Analysis of the results leads to the following remarks: (i) if number of features is too low (such as for example for $Q = 85$), the classification accuracy is not satisfying, (ii) subsets of high quality features are already constructed by the exIWO in early populations – extending the computational process through increase of numbers of individuals and iterations yielded very modest accuracy profit at best and turned out to be unrewarding ($Q \in \{91, 96\}$), (iii) increase of the number of features does not guarantee accuracy improvement ($Q = 89$ vs $Q \in \{91, 93, 99\}$), (iv) selection on the smaller feature subset gives accuracy similar to the larger subset in time which is several times shorter than that related to the larger subset ($Q = 89$ using 210 weeds vs $Q = 96$ using 510 plants), (v) feature selection leads to the improvement of accuracy, but not in all cases ($Q = 99$). Cardinality of the final selected subset of features oscillated in particular experiments from 42 to 64% of cardinality of the entire set.

Table 2. Results of the experiments

Variation coverage Q [%]	Entire number of features	Entire number of individuals	Mean CCR with feature selection [%]	Mean CCR without feature selection [%]	Mean execution time [s]
85	16	5050	97.68	97.17	584.61
89	30	510	99.74	99.15	52.29
89	30	210	99.69	99.15	27.87
91	33	5050	99.40	98.58	554.30
91	33	510	99.41	98.58	52.72
93	36	510	99.58	99.15	64.52
93	36	210	99.63	99.15	35.96
96	84	5050	99.84	99.72	948.01
96	84	510	99.71	99.72	97.03
99	270	510	97.85	99.15	298.60
99	270	210	97.51	99.15	109.68

Influence of the method of individuals' selection on the classification accuracy turned out to be not significant. Hence, the summary of this aspect of the research will be limited only to description of each method by means of a triple $a/b/c$, where a denotes number of cases (i.e. rows in Table 2) in which the mean CCR of the given method (not included in Table 2) was greater than the mean CCR for all 3 methods ($\mathrm{CCR}_m > \mathrm{CCR}_{all}$), b represents number of cases in which both mean values were equal, while c – number of cases in which $\mathrm{CCR}_m < \mathrm{CCR}_{all}$. Thus, descriptions of particular methods are as follows: global – 3/4/4, offspring-based – 8/2/1, family-based – 2/3/6. The outcome of the family-based method is supposed to result from the random initialization of the first population, instead of using any "well-considered" method.

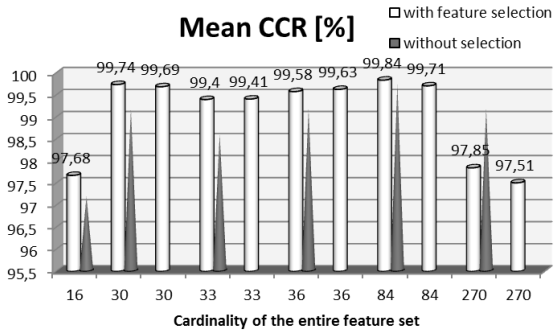

Fig. 4. Dependence between entire number of features and re-identification accuracy.

6 Conclusion

The research revealed the usefulness of the exIWO for solving feature selection problem. However, the adaptation of the metaheuristic requires determination of the following components: a representation of a single solution, a method of initialization of the first population, admissible transformations of an individual, a formula of a fitness function, a stop criterion, and a thorough choice of appropriate values of algorithm parameters. Besides, evaluation of quality of the selected feature subset based on classification accuracy is classifier-dependent and rather time-consuming. With regard to this last aspect, the selection was preceded by the data dimensionality reduction performed by the MPCA algorithm. Future research will focus on alignment of parameters of both algorithms and feature selection for video data.

Acknowledgments. This work was supported by projects NN 516475740 and DEC-2011/01/B/ST6/06988 from the Polish National Science Centre.

References

1. Boyd, J.E., Little, J.J.: Biometric Gait Recognition. LNCS, Vol. 3161, pp. 19–42. Springer (2005)
2. Nixon, M.S., Tan, T.N., Chellappa, R.: Human Identification Based on Gait. Springer (2006)
3. Doretto, G., Sebastian, T., Tu, P., Rittscher, J.: Appearance-based person reidentification in camera networks: problem overview and current approaches. Journal of Ambient Intelligence and Humanized Computing, Vol. 2, pp. 127–151. Springer (2011)
4. Law, M.H.C., Jain, A.K.: Incremental nonlinear dimensionality reduction by manifold learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28(3), pp. 377–391. (2006)
5. Yan, S., Xu, D., Yang, Q., Zhang, L., Tang, X., Zhang, H.-J.: Discriminant Analysis with Tensor Representation. Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 526–532. (2005)
6. Witten, I.H., Frank, E.: Data Mining. Practical Machine Learning Tools and Techniques. Elsevier (2005)
7. Guyon, I., Elisseeff, A.: An Introduction to Variable and Feature Selection. Journal of Machine Learning Research, Vol. 3, pp. 1157–1182. (2003)
8. Mehrabian, R., Lucas, C.: A novel numerical optimization algorithm inspired from weed colonization. Ecological Informatics, Vol. 1(4), pp. 355–366. Elsevier (2006)
9. Lu, H., Plataniotis, K.N., Venetsanopoulos, A.N.: MPCA: Multilinear Principal Component Analysis of Tensor Objects. IEEE Transactions on Neural Networks, Vol. 19(1), pp. 18–39. (2008)
10. Michalewicz, Z., Fogel, D.B.: How to Solve It: Modern Heuristics. Springer (2004)
11. Tao, G., Michalewicz, Z.: Inver-over Operator for the TSP. LNCS, Vol. 1498, pp. 803–812. Springer (1998)