

Knowledge discovery in panoramic X-rays for postmortem identification

Tom Van de Wiele ^a Vinicius Tragante do Ó ^a Hendrik Blockeel ^a

Ricardo Henrique Alves da Silva ^b

^a *Departement Computerwetenschappen, Katholieke Universiteit Leuven, Belgium*

^b *Faculdade de Odontologia de Ribeirão Preto, Universidade de São Paulo, Brazil*

Abstract

The estimated age and sex of an unknown person are important indicators in post mortem identification. In this study models are built for estimating these indicators automatically using panoramic X-rays. The used filtered dataset contains 303 images where the age group (7 groups) and sex is known. Two different techniques for extracting the features from the images are compared: automatic extraction (AE) and principal component analysis (PCA). To improve the quality of the prediction, feature selection is applied afterwards. This was done using both filter and wrapper techniques. The wrapper technique makes use of evolutionary algorithms. A novel evolutionary technique, adaptive binary continuous particle swarm optimization (ABCPSO) is introduced. The obtained results from the novel technique are comparable to the best studied standard evolutionary techniques, and present superior performance in some cases although not statistically significant. The results show that when the optimal subset of features is used the sex prediction reaches an average accuracy of 63.1% (AE) and 73.9% (PCA). For age prediction the adapted average accuracy is 34.5% and 36.0%, respectively. The Naive Bayes classifier produced the best results among five well known classifiers.

1 Introduction

Identity, by definition, is the ensemble of characteristics that allow the individualization of a person or an object, distinguishing it from any other person or object[1]. Identification is the act by which the identity of an individual is established, and it is unquestionable the value of *postmortem* human identification, since social relations and/or civil, administrative, commercial and legal demands claim this form of recognition.

By knowing that the main external factors that can limit the retrieval of information starting from the remainings of a body, and also restrict the whole process of human identification, are the elements present or associated to decomposition or physical agents (such as mutilations, fire, explosions, among others), the dental structure, capable of being evaluated with radiographic images, comes as an important element in this identification task and in Criminology, due to the high probability that the dental characteristics are never the same in any two persons, and also due to the high degree of physical and chemical resistance of the dental structure[2]–[4].

According to Mincer *et al.* (2008), the panoramic radiograph allows the visualization of all the teeth, jaw, mandible and some face bones, that is, the whole maxilomandibular complex, in a single picture. This technique presents several advantages, such as: symmetry of the images, a complete overview of the lower face, ease of generation of such images, lack of film in the oral cavity, lower dose of radiation and exposition time if compared to the complete intraoral radiographic exam[5].

However, this technique also presents several disadvantages: image distortion, as all structures are put in the same layer; the difficulty of performing such exam *postmortem*, regarding the position of the body and differences in density (due to the absence of some tissues and structure). However, the panoramic technique has been reasonably developed, and it is already possible to know in advance the magnification of each X-Ray equipment.

Manual radiograph comparison is highly time-consuming and requires high levels of ability and accuracy. With this in mind, the comparison of dental records via computer systems emerges as the most adequate method for manipulation of big volumes of information, keeping accuracy, consistency and low cost[6].

We present in this paper a comparison between evolutionary algorithms and also different classifiers in the task of finding relevant attributes for determining gender and age group based on panoramical radiographs, using two different techniques for attribute generation: Automatic Extracion (AE) via Image Processing techniques, based on general guidelines provided by forensic odontology experts, and Principal Component Analysis. The panoramic radiograph samples used in the current study belong to the Service of the Kernel for Culture and Extension Support in Odontological Diagnosis [Serviço do Núcleo de Apoio a Cultura e Extensão em Diagnóstico Odontológico (NACEDO), original name] from the School of Dentistry of Ribeirão Preto of the University of São Paulo (FORP-USP), and is an ensemble of radiographs taken with two different X-Ray equipments: Ortophos-Siemens®(Sirona Dental), analogical, and Veraviewepocs Digital Panoramic®(J. Morita Co.), digital. The analogical radiographs were digitalized with a professional Epson®Scanner.

2 Methodology

This work consists of three main parts. In the first part the images are preprocessed. This transformation ensures a uniform positioning of all images. In this part each input image is also enhanced to facilitate the detection of features. After detection of the features in the second part the found features are finally selected. In this step only the most relevant features for prediction of the sex and age are retained.

2.1 Preprocessing

Preprocessing of the images is twofold. In the first step the input image is transformed to a standard format which makes the automatic extraction of the features applicable. Afterwards the transformed image is enhanced by applying two image filters.

Automatic conversion to the standard format consists of four actions. All non-grayscale images are transformed to grayscale values in the first step. Then the vertically oriented images are rotated to obtain a uniform orientation. The last two steps consist of centralising the image around its vertical axis of symmetry and resizing the image to the standard pixel dimensions(500×945 pixels).

In order to make detections easier, an image enhancing step was performed. This comes down to the improvement of two characteristics: luminance and contrast. Two filters are used to obtain this: the INDANE filter[7] and the image sharpening filter[8]. First the INDANE filter is applied to the original image. The image sharpening filter is then applied to the obtained image.

2.2 Feature extraction

Two different techniques for automatic feature extraction are applied. The first technique, Automatic Extracion (AE), extracts the features by parsing the panoramical X-ray for the regions of interest. Based on the pixels representing these regions of interest, the features are calculated afterwards. The second technique calculates the features immediately based on the image itself using Principal Component Analysis (PCA). PCA is a well known transformation whereas the used AE is a novel approach. An overview of the feature extraction is presented in Figure 1.

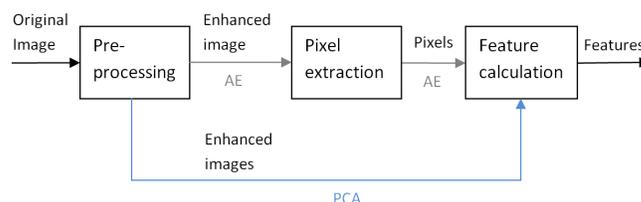


Figure 1: Overview of the feature extraction

2.2.1 Automatic extraction(AE)

In the automatic extraction of the features two regions of interest are considered: the chin line and the teeth. These regions are located by parsing the edges of the enhanced images. The canny filter[9] is chosen for the edge detection because of its good properties for detecting true weak edges. Features are then calculated for both the chin and the teeth using the location of the regions of interest (ROI).

2.2.2 Principal Component Analysis(PCA)

The second used technique for extracting features automatically is the Principal Component Analysis (PCA)[10]. It is a technique that is applied in several domains as it is able to reveal underlying structures in complex datasets using analytical solutions from linear algebra[11]. The idea is to represent a dataset using a limited number of relevant components, the principal components. When the dataset consists of facial images, these principal components are called *eigenfaces*. Based on the information from the AE the principal components are calculated for the 303 selected images by first applying the INDANE - and then the image sharpening filter. The 24 most significant eigenfaces are used as the features from the PCA. This number was chosen to be equal to the number of features in AE in order to be able to compare the results.

2.3 Feature selection

Feature selection is the selection of the relevant features for determination of a certain property. It is a technique that is applied in many fields such as data mining, machine learning, pattern recognition and signal processing[12]. When the number of involved features is huge, feature selection will be necessary. Also when there is a limited number of features it is an appropriate step when not all features are relevant. Previous works[12]–[13] on datasets with a limited number of features (<30) showed that feature selection was able to improve the accuracy of the prediction.

Algorithms for feature selection fall into two broad categories: wrappers, which use the learning algorithm itself to evaluate the usefulness of features; and filters which evaluate features according to heuristics based on general characteristics of the data[13]. Techniques from both categories are applied and results are compared.

2.3.1 Filter feature selection

Filter methods operate independently of any learning algorithm. Undesirable features are filtered out of the data before induction starts. Filters typically make use of all the available training data when selecting a subset of features. This approach has proven to be much faster than wrappers and hence can be applied to large data sets containing many features. Since they are more general, they can be used with any learner, unlike the wrapper, which must be re-run when switching from one learning algorithm to another. The problem with this approach is that there is no guarantee that the selected features will produce good results in combination with the used classifier.

Two well known filter techniques are used: the Fisher criterium[14] and the ReliefF algorithm[15].

2.3.2 Wrapper feature selection

Wrapper feature selection, on the other hand, evaluates the fitness of the features using the classifier itself. For every considered subset, the classifier needs to be retrained and its estimated global accuracy must be recalculated. In general, this approach yields better results than the filter methods because the search for an appropriate subset is guided by the generation process[16]. The main disadvantages of this approach are the high computational cost and the specific selection of the features for one classifier.

Evolutionary techniques The generation of the features is done using a variety of evolutionary and alternative algorithms. Genetic Algorithms(GA)[17]–[19] and Particle Swarm Optimization(PSO) techniques [12][16][18][20] have been used several times for this purpose. These methods search for an optimal subset in the binary search space with 2^N possible subsets, being N the number of features. With 24 features, exhaustive search in this search space would be too time consuming. In this work we use genetic, particle swarm and alternative search methods.

The standard genetic algorithm[21](SGA) is used, as well as some extensions. The first three insert random immigrants[22] (GARIS for the low rate:2% and GARIL for the high rate:30% of random immigrants) into the population. Two of them have a fixed number of random immigrants whereas the third[23] (SSORIGA) makes use of a dynamic rate. The last genetic algorithm makes use of a dynamic population size[24](VPGA).

As the particle swarm algorithms were originally constructed for continuous problems, an adaptation must be made to be able to represent binary values. The first used version of the PSO algorithm makes use of the standard algorithm (SPSO), in which all continuous values are converted to their corresponding binary values when the subsets are evaluated. As with the use of this technique all of the members of the swarm tend to converge to the same local optimum after a few generations, an improved binary continuous PSO was developed: ABCPSO. The last used PSO[26] represents the particles directly in a binary way(BPSO). In this implementation the represented binary representations are determined in each generation using a biased random process.

Two alternative search methods are considered: a hybrid PSO-GA algorithm[24](HGAPSO) and a random search process(RS). The hybrid algorithm makes use of both the standard GA and the ABCPSO. Every 5th generation a hybrid process takes place, and the information from both techniques is exchanged.

Adaptive binary continuous particle swarm optimization(ABCPSO) The proposed algorithm is the standard continuous PSO algorithm with 3 extensions. All particles have a range of [0,1] in all their dimensions and the corresponding binary value is equal to 1 when the continuous value is greater than 0.5, and 0 otherwise. Every dimension represents one feature.

Every particle X_i is represented by a point in the D -dimensional search space: $X_i = (x_{1,i}, x_{2,i}, \dots, x_{N,i})^T$ in the standard PSO algorithm. In the initial generation random vectors are chosen within the search space. The best visited position for each particle X_i is stored in the vector $P_i = (p_{1,i}, p_{2,i}, \dots, p_{N,i})^T$. The best visited position by the entire swarm is also taken into account. This point is represented by $P_g = (p_{1,g}, p_{2,g}, \dots, p_{N,g})^T$. These two vectors are initialized to the random initial vectors in the first generation. The positions of all particles are adapted every generation according to the speed vector V . This vector is also determined randomly in the first generation. The adaption of the speed and position vector for each particle goes as follows:

$$V_i = w * V_i + c_1 * r() * (P_i - X_i) + c_2 * r() * (P_g - X_i)$$

$$X_i = X_i + V_i$$

This means that the new position is the old position added with the new speed. The speed is calculated using three factors:

- **Inertion** The inertion factor w is used to control the influence of the previous speed. The choice of this factor is a trade-off between local and global exploration of the particles. A high inertion factor facilitates global exploration while a lower value for w stimulates local exploration.
- **Personal factor** Every particle is attracted by its personal best position. This, in combination with the inertion, guarantees continuous exploration throughout the algorithm.
- **Social factor** The last part of the speed equation ensures that all particles are attracted by the best visited position so far.

Furthermore, there are two constants in the speed equation: c_1 a c_2 and a random function $r()$, which generates a random number in the interval [0,1]. The constants determine the magnitude of the attraction to the local and global maxima and are chosen as in[27]. In that work both constants are 1.49445 and the inertion factor is $0.5 + \frac{r()}{2}$.

An important advantage of this algorithm is the simplicity of the operations. The population size is kept constant and every particle survives from the first till the last generation. In each generation every particle is adapted with a rather simple equation.

When the evaluation of the function values is slow(which is the case in this work), the efficiency of the algorithm is less important and extensions can be made without a significant loss of time. Three extensions are made:

1. **Random immigrants** – The first extension is based on the random immigrants introduced in the genetic algorithm. When the new positions of the particles are calculated, the worst 10% of the particles are replaced by random particles. The aim of this extension is to increase the diversity in the population while there is no loss of information about the best members of the population.
2. **Repulsion** – This extension aims to increase the diversity in the population and is based on the natural movement of particles in a swarm. In order to avoid collision, the particles undergo a process of mutual repulsion after calculation of the new position. All couples are compared mutually, and if the Manhattan distance between the particles is lower than a certain threshold (here $\frac{\#features}{4}$), both particles are repelled in relation to each other. This happens in the opposite direction with a Manhattan distance of $\min(\frac{\#features}{4}, \frac{\#features}{20 * distance^2})$.
3. **Local exploration** – The addition of the previous two extensions increases the diversity in the population but reduces the capacity of local exploration around the provisional maximum. To enable local exploration the population is split into two independent swarms every X generations (here 5). After these X generations the information of both swarms is combined and a new split of the population is calculated.

The first swarm (which contains 90 % of the particles in this work) executes the previously described standard PSO algorithm with the extensions of random immigrants and repulsion. The remaining randomly chosen part performs local exploration around the provisional best point in the search space. In this second swarm there are no random immigrants or repulsive forces. The previous speeds are randomized and the global and local bests are set to the provisional best point. In this way a randomized exploration around the maximum is guaranteed. The norm of the previous speeds is determined by the distribution of the particles. It is set to the mean distance from the provisional optimum to the best 5% of the population.

Despite the repulsion in the major part of the total swarm, convergence is still possible. Every X generations the two swarms are merged and the global maximum is set to the best value of both swarms. After calculation of the initial speeds for the new second swarm, the particles of the previous second swarm are put back to their previous positions such as their previous local best positions. This way the variation of those particles is restored.

2.3.3 Evaluation of the subsets

Evaluation of the subsets is performed using stratified K-fold cross validation. K is chosen to be 10 which is a rather high value. This ensures that a large part of the dataset can be used to train the classifier. Before the start of each run the images are selected in such a way that all considered classes have equal sizes. This ensures an unbiased prediction. The accuracy of a single prediction is 1 if the predicted class equals the actual class of the image and 0 otherwise. When the age class is predicted an adapted accuracy is used. A wrong prediction of an age by 1 age group is assigned an accuracy of $\frac{1}{4}$ when the actual age group is not the highest or the lowest age group and $\frac{1}{2}$ otherwise.

Five well known classifiers are compared: naive bayes, decision trees, 3 nearest neighbors, neural networks and support vector machines.

3 Results

3.1 Comparison of classifiers and evolutionary algorithms

To compare the classifiers, both the filter and wrapper approaches are used with all the classifiers to select the features of varying search spaces (by changing the selected images). Among all the classifiers, the naive bayes method resulted in the best accuracies in general. A comparison of all evolutionary techniques can be seen in table 1.

This table represents the results of applying all evolutionary techniques on 50 different search spaces for each used classifier. The population size(10) and number of generations(20) are chosen to be small. This way the techniques can be compared before they converge. All search spaces consist of the sex prediction using the features from the AE. Two indicators for the quality of the techniques are used. The first is the average function value of the best member of the population in all 250 simulations. The other indicator

represents the number of victories of each technique for a given search space and classifier. All technique(s) that obtain the best function value are considered victorious. It is very important to note that all considered evolutionary techniques have a comparable number of evaluations. All of them have a fixed number of generations and all but one have a fixed population size. The limits for the initial and maximal population in the genetic algorithm with variable population size are chosen in such a way that the number of evaluations is close to the number of evaluations in the other techniques.

The best five techniques(ranked on the average maximal function value) are retained for further simulations, these are underlined in table 1. All of the selected evolutionary techniques have a comparable average function value and number of victories. The developed ABCPSO improved the performance of the standard binary continuous PSO significantly. The binary PSO on the other hand produced the best results overall.

Table 1: Comparison of evolutionary techniques

Technique	Average maximal evaluation	Victories	Rank
<u>SGA</u>	.619	33	5
GARIS	.611	29	7
GARIL	.616	25	6
<u>SSORIGA</u>	.625	32	4
VPGA	.601	6	8
SPSO	.593	1	9
<u>ABCPSO</u>	.626	39	3
BPSO	.635	52	1
<u>HGAPSO</u>	.632	45	2
RS	.592	3	10

3.2 Used subsets

The results are compared for all the considered feature selection methods. Evaluation of the subsets of each size S where S varies from 1 to 24 features are plotted for all four sex/gender – AE/PCA couples. The subset of size S contains only those S features with the best evaluated function values. In the case of wrapper feature selection this function value is the number of times each feature is selected in the optimal subset of a run. 20 runs are used with a population size of 25 and 50 generations for all of the 5 best evolutionary techniques.

The results of sex prediction can be found in Figure 2. The results of the age prediction are displayed in Figure 3. In both images the results using the features from the AE are displayed on the left part. The results for the predictions using the features from the PCA are shown on the right.

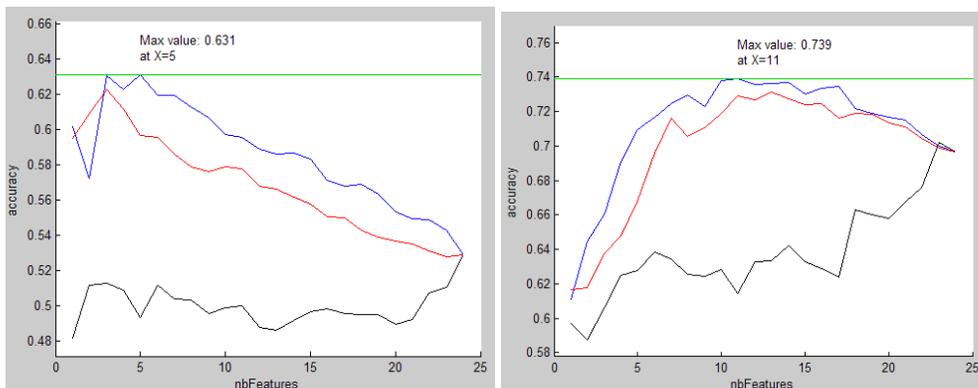


Figure 2: Sex prediction using AE (left) and PCA (right). The red color represents the Fisher method, Black is used for ReliefF and Blue stands for the Wrapper approach. The ReliefF method needs more attributes in order to achieve better performance. The Wrapper method obtains the best results overall.

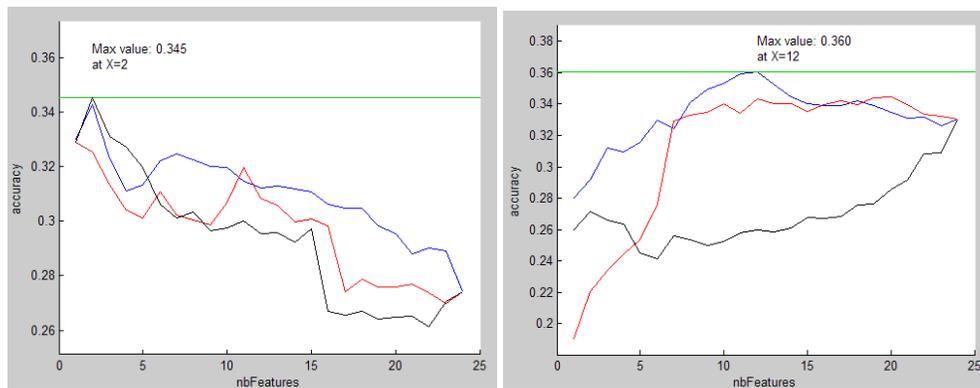


Figure 3: Age prediction using AE (left) and PCA (right). The red color represents the Fisher method, Black is used for ReliefF and Blue stands for the Wrapper approach. In most not all attributes are important for reaching a better accuracy, varying between 1 (Fisher with AE) and 24 (Relief F with PCA) for the optimal number of attributes.

4 Discussion

The obtained results show that the wrapper approach yields better results in 3 of the 4 considered predictions. At least one of the subsets generated by the filter techniques comes close to the predictive accuracy of the optimal subset generated by the wrapper approach in all of the 4 considered predictions. When using the wrapper approach there is only a marginal improvement, but at a very high computational cost, that doesn't compensate the gain in performance.

Another remark is the fact that the optimal subsets contain a lot more features when the eigenfaces are used. The reason for this is the higher number of irrelevant features in the features from the automatic extraction, despite these being suggested by the experts. The explanation for that is the difficulty in extracting exactly the parts that the experts point, since the quality of the radiographs varies, and not all the points suggested were detected by the algorithms.

The optimal subsets have an estimated predictive accuracy of 74% for sex prediction and 36% for the age group prediction. Better results are obtained for both sex- and age prediction when the eigenfaces are used as features, what implies that PCA is a suitable method for extracting useful information from radiographs. In the sequence of this work other types of radiographs will be investigated, since other parts of human faces can also be useful in the task of human identification.

References

- [1] G. Arbenz, *Identidade e identificação: conceitos gerais*, In: Medicina Legal e Antropologia Forense. Rio de Janeiro: Atheneu, 1998. p.105-127.
- [2] R. Oliveira, E. Daruge, L. Galvão, A. Tumang, *Contribuição da odontologia legal para a identificação "post-mortem"*, Rev Bras Odontol 1998; 55(2): 117-22.
- [3] H. Mincer, J. Chaudhry, J. Blankenship, E. Turner, *Postmortem dental radiography*, J Forensic Sci 2008; 56(2):405-7.
- [4] R. Silva, S. Carvalho, A. Sales-Peres, C. Lopes-Junior, *A utilização de imagens na identificação humana em odontologia legal*, Radiol Bras 2009; 42(2):125-30.
- [5] M. Oliveira, M. Correia, J. Barata, *Aspectos relacionados ao emprego da radiografia panorâmica em pacientes infantis*, Rev Fac Odontol Porto Alegre 2006; 47:15-9.
- [6] Nassar DEM, Ammar HH. A neural network system for matching dental radiographs. Pattern Recognition 2007; 40: 65-79.

- [7] L. Tao, V. Asari *An Integrated Neighborhood Dependent Approach for Nonlinear Enhancement of Color Images*, Proceedings of the International Conference on Information Technology: Coding and Computing, 2004.
- [8] *imfilter*, in Matlab Web Server Manual, 27 April 2010, available at <http://www.mathworks.com/access/helpdesk/help/toolbox/images/imfilter.html>.
- [9] J. Canny, *A computational approach to edge detection*, in IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986.
- [10] A. Miranda, Y Le Borgne, G. Bontempi, *New Routes from Minimal Approximation Error to Principal Components*, Neural Processing Letters, Springer, Volume 27, Number 3, 2008.
- [11] Wikipedia, *Principal component analysis*, on Wikipedia 2 May 2010, available at http://en.wikipedia.org/wiki/Principal_component_analysis.
- [12] X. Wang, J. Yang, X. Teng, W. Xia, R. Jensen, *Feature selection based on rough sets and particle swarm optimization*, in Pattern Recognition Letters 28 (2007) 459471.
- [13] M. Hall, *Correlation-Based Feature Selection for Discrete and Numeric Class Machine Learning*, Proceedings of the 17th International Conference on Machine Learning (2000) 359-366.
- [14] R. Duda, P. Hart, *Pattern classification and scene analysis*, Wiley, 1973.
- [15] I. Kononenko, S. Hong, *Attribute Selection for Modeling*, In: Future Generation Computer Systems, ISSN 0167 - 739X, 13 (2-3), 181-195, 1997.
- [16] Y. Liu, Z. Qin, Z. Xu, X. He, *Feature Selection with Particle Swarms*, in Computational and Information, Heidelberg: Springer-Verlag (2004) 425430.
- [17] M. Dash, H. Liu, *Feature Selection for Classification*, in Intelligent Data Analysis, 1 (1997) 131-156.
- [18] Y. Saeys, I. Inza, P. Larrañaga, *A review of feature selection techniques in bioinformatics*, in Bioinformatics, 19 (2007) 25072517.
- [19] H. Vafaie, I. Imam, *Feature selection methods: genetic algorithms vs. greedy-like search*, Proceedings of International Conference on Fuzzy and Intelligent Control Systems (1994).
- [20] L. Chuang, H. Chang, C. Tu, C. Yang, *Improved binary PSO for feature selection using gene expression data*, in Computational Biology and Chemistry, 32 (2008), 2938.
- [21] J. Holland, *Adaptation in natural and artificial systems*, Ann Arbor, MI: University of Michigan Press.
- [22] J. Grefenstette. Genetic algorithms for changing environments. *Parallel Problem Solving from Nature*, Amsterdam, 2 (1992) 137-144.
- [23] V. Tragante, R. Tinós, *Control of the number of random immigrants in genetic algorithms for protein structure prediction*, in Proceedings of the 11th Annual conference on Genetic and evolutionary computation, 2009.
- [24] X. Shi, Y. Liang, H. Leeb, C. Lub, L. Wang, *An improved GA and a novel PSO-GA-based hybrid algorithm*, in Information Processing Letters, 93 (2005) 255-261.
- [25] R. Eberhart, Y. Shi, *Comparison between Genetic Algorithms and Particle Swarm Optimization*, Evolutionary Programming VII (1998), Lecture Notes in Computer Science 1447, 611-616. Springer.
- [26] J. Kennedy, R. Eberhart, *A Discrete Binary Version of the Particle Swarm Optimization*, in Proc. of the conference on Systems, Man, and Cybernetics SMC97, 1997 4104-4109.
- [27] R. Eberhart, Y. Shi, *Tracking and optimizing dynamic systems with particle swarms*, in Proceedings of Congress on Evolutionary Computation. 2001, Seoul.