Study of Particle Swarm Optimisation as Surveillance Object Classifier

Virginia Fernandez Arguedas, Krishna Chandramouli and Ebroul Izquierdo
Multimedia and Vision Research Group,
School of Electronic Engineering and Computer Science
Queen Mary, University of London, Mile End Road, E1 4NS, London, UK
{virginia.fernandez,krishna.chandramouli,ebroul.izquierdo}@eecs.qmul.ac.uk

***Abstract*—Following the recent exponential interest that has been shown in biologically inspired techniques for solving various optimisation challenges in computer vision, in this paper we present a study of *Particle Swarm Optimisation* as a surveillance object classifier. Though inspired by fundamental biological organisms, the memory inherently embedded in particle swarms enables solutions, which are computationally less complex, adept to multi-dimensional problems, and attains global minima. The performance of the proposed technique has been thoroughly evaluated with AVSS 2007 surveillance dataset containing objects such as car and person against well-known kernel machines.**

***Keywords*-Object classification, biologically inspired classifier, particle swarm optimisation, surveillance videos, MPEG-7 features, machine learning**

I. Introduction

The high-level visual semantic information extraction process is conventionally tackled by image clustering and classification algorithms based on machine learning approaches. The problem of image clustering and classification based on low-level features is being addressed by many researchers. As a result, large number of clustering and classification algorithms for multitude of applications have been developed. However, the performance of almost all clustering and classification algorithms is highly dependent on the performance of the optimisation techniques.

Evolutionary computation (EC) uses computational models of evolutionary processes as key elements in the design and implementation of computer - based problem solving systems. In [1] a variety of evolutionary computational models were introduced. Such algorithms share a common conceptual base of simulating the evolution of individual structures via processes of selection and reproduction. These processes depend on the perceived performance (fitness) of the individual structures as defined by an environment. More precisely, evolutionary algorithms maintain a population of structures that evolve according to rules of selection and other operators, such as recombination and mutation. Each individual in the population receives a measure of its fitness in the environment. Selection focuses attention on high fitness individuals, thus exploiting the available fitness information. Although simplistic from a biologist's viewpoint, these algorithms are sufficiently complex to provide robust and powerful adaptive search mechanisms.

Recently an exponential increase of interest has been shown towards the developments in optimisation techniques that have been inspired by problem solving abilities of biological organisms such as bird flocking and fish schooling. One such technique developed by Eberhart and Kennedy is called Particle Swarm Optimisation (PSO). The PSO algorithm has two main assertions as listed below [2].

- Mind is Social: Learning from experience and emulating the successful behaviours of others, people are able to adapt to complex environments through discovery of relatively optimal patterns of attitudes, beliefs and behaviours.
- Particle swarms are a useful computational intelligence methodology: Central to the concept of computational intelligence is system adaptation that enables or facilitates intelligent behaviour in complex and changing environments. Swarm intelligence comprises of three steps namely evaluate, compare and imitate. Each particle goes through these stages by performing simple mathematical operations in solving a more complex optimisation problem.

Object classification in surveillance datasets has attracted researchers attention due to the ever-increasing safety concerns of the public. Norris and McCahill [3] estimated that UK accommodates more than 4.2 million surveillance cameras. The CCTV ubiquitousness generates everyday a large amount of video data, which requires constant human supervision. In an effort to mitigate the dependency of constant supervision of surveillance footage, there exists critical need for automatic and intelligent indexing and classification schemes for objects and events to enable efficient media access, navigation and retrieval. Addressing the challenges related to object indexing and classification, several approaches has been presented based on probabilistic, statistical and biologically inspired classifiers [4]. Many techniques show satisfactory results for general datasets such as movies, sports and news, however, the challenge of retrieving surveillance objects remains largely an open issue.

In this paper, we present a study of Particle Swarm Optimisation as object classifier on Surveillance dataset. The proposed classification approach, *Particle Swarm Classifier*, exploits the recent developments in evolutionary computational algorithms based on biologically inspired optimisation techniques. The performance evaluation and surveillance footage-processing framework is therefore aimed at establishing an evaluation criteria for image clustering/classification techniques against which next generation algorithms could be evaluated. In order to satisfy the necessary constraints of being flexible and robust, the framework is implemented in a plug-in manner with different visual feature plug-in(s) and algorithm plug-in(s) (refer to Figure.1). The proposed approach has been evaluated against three kernel machines for objects extracted from the surveillance dataset. From the study of the evaluation results, we note the improved performance of the proposed approach across all concepts as opposed to the state of the art Support Vector Machine (SVM) performance. The dataset has been specifically designed to be noisy in order to measure the robustness of the evolutionary computational algorithms.

The remainder of the paper is structured as follows. In Section II, an overview of the proposed surveillance object classification framework is presented, followed by a brief description of the *Motion Analysis Component* (refer to Section III). The proposed biologically inspired object classifier for surveillance videos, *Particle Swarm Classifier*, is further detailed in Section IV. The experimental results are discussed in Section V, followed by conclusions and future work in Section VI.

II. Biologically inspired object classifier for surveillance videos

The proposed biologically inspired surveillance object classifier framework is presented in Figure 1. The framework integrates three intermediate modules named *motion analysis component*, *appearance features extraction* and *particle swarm classifier*. First, the *motion analysis component* analyses raw surveillance videos to extract the moving objects, the procedure is based on Stauffer and Grimson approach [5] (refer to Section III). Second, the *appearance features extraction* component creates a visual model for each moving object extracted in the motion analysis component. Visual models are constructed from a set of state of the art appearance features selected for their robustness, compact representation and significance for human perception. A description of the selected appearance features is presented in Section V-B. Finally, the *Particle Swarm Classifier*, based on evolutionary computation models, performs surveillance object classification mimicking the effects of bird flocks. A more detailed description of the *Particle Swarm Classifier* is presented in Section IV. The outcome of the classifier is a ranked list of objects retrieved from the image database, which are further evaluated against ground truth.

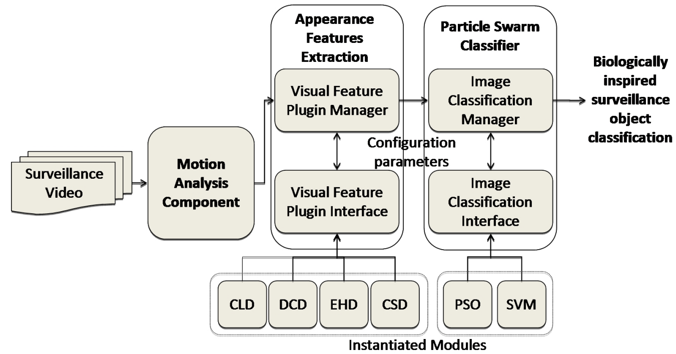

Figure 1. Biologically inspired surveillance object classifier framework

III. Motion Analysis Component

Due to surveillance videos nature, a huge amount of information must be processed, where most of it belongs to their quasi-static background proving no useful data. *Motion analysis component*'s objective is to improve the computational efficiency of the system and to provide movement information about the surveillance video objects. A three-step real-time *Motion Analysis Component* is presented to procure individual blobs to the *Appearance Features Extraction* component.

First, an adaptive background subtraction technique based on Stauffer and Grimson algorithm [5] is performed to remove all the redundant information of the surveillance videos, allowing a faster analysis and providing robustness against external factors, such as changes in illumination or camouflage. Adaptive background subtraction algorithm is a two-step process: (i) modelling a background as a mixture of Gaussians and (ii) modelling each pixel of an image as a weighted mixture of Gaussians to classify it into foreground or background according to the persistence and variance of each of the Gaussians of the mixture. Thus, pixels are classified as foreground if their values do not fit the background distributions formerly calculated. Second, object spatial segmentation is performed grouping the resulting Gaussian mixtures. Consequently, a two-pass connected component algorithm assuming an 8-connection is applied. As a result, foreground moving objects are isolated. Third, temporal segmentation is performed establishing the correspondence of the spatially segmented objects between frames using a linearly predictive multiple hypothesis tracking algorithm based on a set of Kalman filters. Moreover, Kalman filters are used to predict the tracks related to each frame as well as the assignment between the available tracks and the detected blobs in each frame. Despite the many advantages provided by motion analysis component, object detection from surveillance video is affected by (i) the low quality of the image, (ii) lack of contrast, (iii) the image blurring due to the camera motion or (iv) noise and shadows, as highlighted in Figure 2.

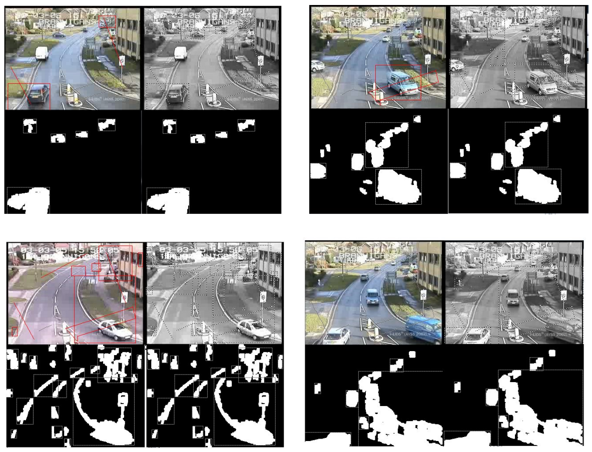

Figure 2. Background subtraction and spatial segmentation techniques results can be observed for four different problematic situations as low quality image (top-left),videos with inaccurate background substration (top-right), videos with camera movement (bottom-left) and objects merged due to noise and shadow (bottom-right)

IV. Particle Swarm Classifier

In this section, the proposed biologically inspired object classifier is further detailed. The *Particle Swarm Classifier (PSC)* is an evolutionary computational algorithm based on Self Organising Maps (SOM). PSC proposes to follow SOM structure and connectivity properties, however, in order to provide a more adaptive and robust approach its in-built trainer is replaced for a biologically inspired trainer based *Particle Swarm Optimisation (PSO)*. The PSO algorithm proposes an optimisation technique following birds flock behaviour to search for an optimal solution to train the input feature vectors of each object (refer to Section IV-A). After the training stage, classification is performed comparing the input feature vector with each neuron's feature vector from a competitive neural network (refer to Section IV-B).

A. Particle Swarm Optimisation(PSO)

A critical need of indexing, classification and retrieval algorithms has been reported for the analysis of surveillance videos, in order to mitigate the dependency on human supervision. Although the performance of the machine learning techniques has largely been improved, the results are still far away from results generated by human cognition. Addressing this problem, recent developments in optimisation techniques have been inspired by problem solving abilities of biological organisms such as bird flocking and fish schooling. One such technique developed by Eberhart and Kennedy is called Particle Swarm Optimisation (PSO).

In the PSO algorithm [6], the birds in a flock are symbolically represented as particles. These particles are considered to be "flying" through the problem space searching for the optimal solution [7]. A particle's location in the multidimensional problem space represents one solution for the problem. When a particle moves to a new location, a different solution to the problem is generated. This solution is evaluated by a fitness function that provides a quantitative value of the solution's utility. The velocity and position of each particle moving along each dimension of the problem space will be altered with each generation of movement. The particles at each time step are considered to be moving towards particle's personal best ($pbest$) and swarm's global best ($gbest$). The motion is attributed to the velocity and position of each particle. Acceleration (or velocity) is weighted with individual parameters governing the acceleration being generated for c_1 and c_2. The equations governing the velocity and position of each particle are presented in Equations 1 and 2.

$$\begin{aligned} v_{it}(t+1) &= v_{id}(t) + c_1(pbest_i(t) - x_{id}(t)) \\ &+ c_2(gbest_d(t) - x_{id}(t)) \end{aligned} \tag{1}$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \tag{2}$$

- $v_{id(t)}$ represents the velocity of particle in dimension at time t
- $pbest_i(t)$ represents the personal best solution of particle i at time t
- $gbest_d(t)$ represents the global best solution for $d-$dimension at time t
- $x_{id}(t)$ represents the position of the particle x in $d-$dimension at time t
- c_1, c_2 constant parameters

The trajectory of each individual in the search space is adjusted by dynamically altering the velocity of each particle; according to the particle's own problem solving experience and the problem solving experience of other particles in the search space. The first part of Equation 1 represents the velocity at time t, which provides the necessary momentum for particles to move in the search space. During the initialization process, the term is set to '0' to symbolize that the particles begin the search process from rest. The second part is known as the "cognitive component" and represents the personal memory of the individual particle. The third term in the equation is the "social component" of the swarm, which represents the collaborative effort of the particles in achieving the globally best solution. The social component always clusters the particles toward the global best solution determined at time t.

B. Particle Swarm Classifier

The network architectures and signal processes used to model nervous systems can be categorised as feedforward, feedback and competitive. Feedforward networks [8] transform a set of input signals into a set of output signals. The desired input-output transformation is usually determined by external, supervised adjustment of the system parameters. In feedback networks [9], the input information defines the initial activity state of the feedback system, and after state transitions the asymptotic final state is identified as the outcome of the computation. In competitive learning networks, neighbouring cells in a neural network compete in their activities by means of mutual lateral interactions and develop adaptively into specific detectors of different signal patterns.

In competitive neural networks, active neurons reinforce their neighbourhood within certain regions, while suppressing the activities of other neurons [10]. This is called on-centre/off-surround competition. The objective of *Self Organising Maps (SOM)* is to represent high-dimensional input patterns with prototype vectors that can be visualized in a usually two-dimensional lattice structure [11]. Each unit in the lattice is called a neuron, and adjacent neurons are connected to each other which results in a clear topology of how the network fits itself to the input space. Input patterns are fully connected to all neurons via adaptable weights, and during the training process, neighbouring input patterns are projected into the lattice, corresponding to the adjacent neurons. SOM enjoys the merit of input space density approximation and independence of the order to input patterns.

In the basic SOM training algorithm the input training vectors are trained with Equation 3

$$m_n(t+1) = m_n(t) + g_{cn}(t)[x - m_n(t)] \tag{3}$$

where m is the weight of the neurons in the SOM network, $g_{cn}(t)$ is the neighbourhood function that is defined as in Equation 4,

$$g_{cn}(t) = \alpha(t)exp(\frac{||r_c - r_i||^2}{2\alpha^2(t)}) \tag{4}$$

where, $\alpha(t)$ is the monotonically decreasing learning rate and r represents the position of the corresponding neuron.

The proposed *Particle Swarm Classifier (PSC)* appeared to further improve the performance of the SOM classifier, considering biologically inspired optimisation techniques to approach the classifier performance to human cognition. In the PSC the weight of the neurons m_d is optimized with PSO. The optimisation is achieved by evaluating the $L1$ norm between the input feature vector and the feature vector of the winner node. The global best solution obtained after the termination of the PSO algorithm is assigned as the feature vector of the winner node. The training process is repeated until all the input training patterns are exhausted. In the testing phase, the distance between the input feature vector is compared against the trained nodes of the network. The label associated with the node is assigned to the input feature vector.

V. EXPERIMENTAL RESULTS

In this section, we present an outline of the experimental methodology adopted to evaluate the performance of the proposed biologically inspired surveillance object classifier. First, the surveillance dataset and ground truth are explained in Section V-A. Moreover, for the creation of visual models to represent each object for further classification, four MPEG-7 visual descriptors were selected and further discussed in Section V-B. Finally, in order to make a fair evaluation of the proposed classification approach, we compare the retrieval results of the PSC in individual descriptor spaces for two concepts against three different support vector kernels which are further discussed in Section V-C.

A. Dataset

AVSS 2007 dataset [1] was used to evaluate the presented biologically inspired surveillance video object classifier, providing indoor and outdoor videos summing a total of 35000 images. For evaluation purposes, three outdoor videos, with a total of 13400 images, were analysed with variable lighting conditions as well as different levels of difficulty. The surveillance footage includes several challenges such as noise, low quality image, camera movement or blurring increasing the difficulty of its analysis. To investigate the performance of our surveillance video object classifier a ground truth was developed selecting a relatively small sized set of blobs extracted from the dataset and manually annotated with two predefined concepts, *Car* and *Person*. A total of 1377 objects were included and annotated in the ground truth, of which 50% of objects were annotated as “Cars” against 10% annotated as “Person” and the remaining 40% were annotated as “Unknown”. Largely, these unknown blobs are composed of noise and those blobs which could not be assigned to either one of the above concepts. Instead of ignoring the blobs labelled as “Unknown”, our dataset included these blobs to explicitly study the effect of noise on the performance of the retrieval models.

B. MPEG-7 Visual Feature Extraction

In this section, the set of selected MPEG-7 descriptors, chosen by their robustness, compact representation and significance for human perception are further explained as well as their similarity measurements [12].

Colour Layout Descriptor (CLD) is a very compact and resolution-invariant representation of the spatial distribution of colour in an arbitrarily-shaped region [13]. CLD’s similarity measure *CLD* uses a weighted Euclidean distance function for each colour component (Y, Cb, Cr).

Colour Structure Descriptor (CSD) describes spatial distribution of colour in an image, but unlike colour histograms, *CSD* also describes local colour spatial distribution.

Dominant Colour Descriptor (DCD) describes global as well as local spatial colour distribution in images for fast search and retrieval. *DCD* provides a description on the distribution of the colour within an analysed image by

[1] http://www.eecs.qmul.ac.uk/ãndrea/avss2007_d.html

storing only a small number of representative colours or *dominant colours*.

Edge Histogram Descriptor (EHD) An edge histogram in the image space represents the frequency and the directionality of the brightness changes in the image. *EHD* provides a description for non-homogeneous texture images and captures the spatial distribution of edges whilst providing ease of extraction, scale invariance and support for rotation-sensitive and rotation-invariant matching.

C. Experimental evaluation of Particle Swarm Classifier

The *Particle Swarm Classifier* is built upon a SOM network. Each neuron of the SOM network is trained using *Particle Swarm Optimisation* techniques. The PSO model implemented is a combination of cognitive and social behaviour. The structure of the PSO is fully connected in which a change in a particle affects the velocity and position of other particles in the group as opposed to partial connectivity, where a change in a particle affects the limited number of neighbourhood in the group. Each dimension of the feature set is optimized with 50 particles. The size of the SOM network is pre-fixed with the maximum number of training samples to be used in the network. The stopping criteria threshold is experimentally determined for different individual feature space. The value of the threshold indicated the closeness in solving the $L1$ optimisation problem between the neuron weights and the input features.

In Figure 3, performance comparison of PSC based retrieval is evaluated against different kernels namely polynomial, radial basis function and sigmoid function for Support Vector Machines (SVM). From the results, we can note the average retrieval performance for the concept "Car" across all feature spaces is 50% at recall 1, while for the concept "Person", the average performance of all classifiers drops significantly at recall 0.5. Generally, the performance of the proposed Particle Swarm Classifier based on Particle Swarm Optimisation (named as PSO in the graphs) overcomes the performance of the different SVM kernels with exception of EHD feature. As it is noted from the results, the performance of the classifier varies according to the feature space. This could be largely attributed to the extraction of different features and the matching functions involved in these distinct feature spaces.

VI. Conclusion

In this paper, a biologically inspired object classifier for surveillance videos was presented providing a more adaptive and robust classifier able to analyse real-world surveillance scenarios. The proposed *Particle Swarm Classifier* was evaluated against individual descriptor space using three support vector machines kernels. A detailed study of the results was carried out, noting that the evaluation retrieval framework achieved an average 10% improvement as opposed to the results obtained using SVMs. The improvement provided by the presented biologically inspired optimisation is limited due to the range of selected appearance features and their use as individuals.

For the future work we will extend the study to include more concepts and novel non-MPEG-7 visual features, as well as, studying the fusion of different visual features and their combined training using PSO. Similarly, a relevance feedback module will be included for online training of the system.

Acknowledgment

The research was partially supported by the European Commission under contract FP7-SEC 261743 VideoSense.

References

[1] R. Kicinger, T. Arciszewski and K.D. Jong. *Evolutionary computation and structural design: A survey of the state-of-the-art*. In: Computers & Structures, vol. 83, no. 23-24, pp. 1943-1978, 2005, Elsevier.

[2] R.C. Eberhart, Y. Shi and J. Kennedy. *Swarm intelligence*. San Mateo, CA: Morgan Kaufmann, 2001.

[3] M. McCahill and C. Norris, *Estimating the extent, sophistication and legality of CCTV in London*. In: CCTV, pp. 51-66, 2003.

[4] K. Chandramouli, T. Piatrik and E. Izquierdo. *Biological Inspired Methods for Media Classification and Retrieval*. In: Multimedia Analysis, Processing and Communications, pp.81-109,2011, Springer.

[5] C. Stauffer and W.E.L. Grimson, *Learning patterns of activity using real-time tracking*. In: IEEE Trans. Pattern Anal. Mach. Intell., vol.2, pp.747-757, 2000, IEEE.

[6] R.C. Eberhart and Y. Shi. *Tracking and optimizing dynamic systems with particle swarms*. In: Proceedings of the 2001 Congress on Evolutionary Computation, vol.1, 2001.

[7] C.W. Reynolds. *Flocks, herds and schools: a distributed behavioural model*. In: Computer Graphics, pp. 25-34, 1987.

[8] D.E. Rumelhart, G.E. Hinton and R.J. Williams. *Learning internal representations by error propagation*. In: Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol.1, pp.318-362, 1986.

[9] J.J. Hopfield. *Neural Networks and physical systems with emergent collective computational abilities*. In: Proc. Natl. Acad. Sci., vol.79, pp.2554-2558, 1982.

[10] R. Xu and D. Wunch II. *Survey of Clustering Algorithms*. In: IEEE Trans. Neural Network, vol.6, no.3, pp.645-678, 2005.

[11] T. Kohonen. *The Self Organizing Map*. In: Proceedings of IEEE, vol.78, no.4, pp.1464-1480, 1990.

[12] B.S. Manjunath, P. Salembier and T. Sikora. *Introduction to MPEG-7: multimedia content description interface*. John Wiley & Sons Inc, 2002.

[13] T. Sikora. *The MPEG-7 visual standard for content descriptor - An overview*. In: IEEE Trans. Circuits Syst. Video Technol., vol.11, no.6, pp.696–702, 2002.

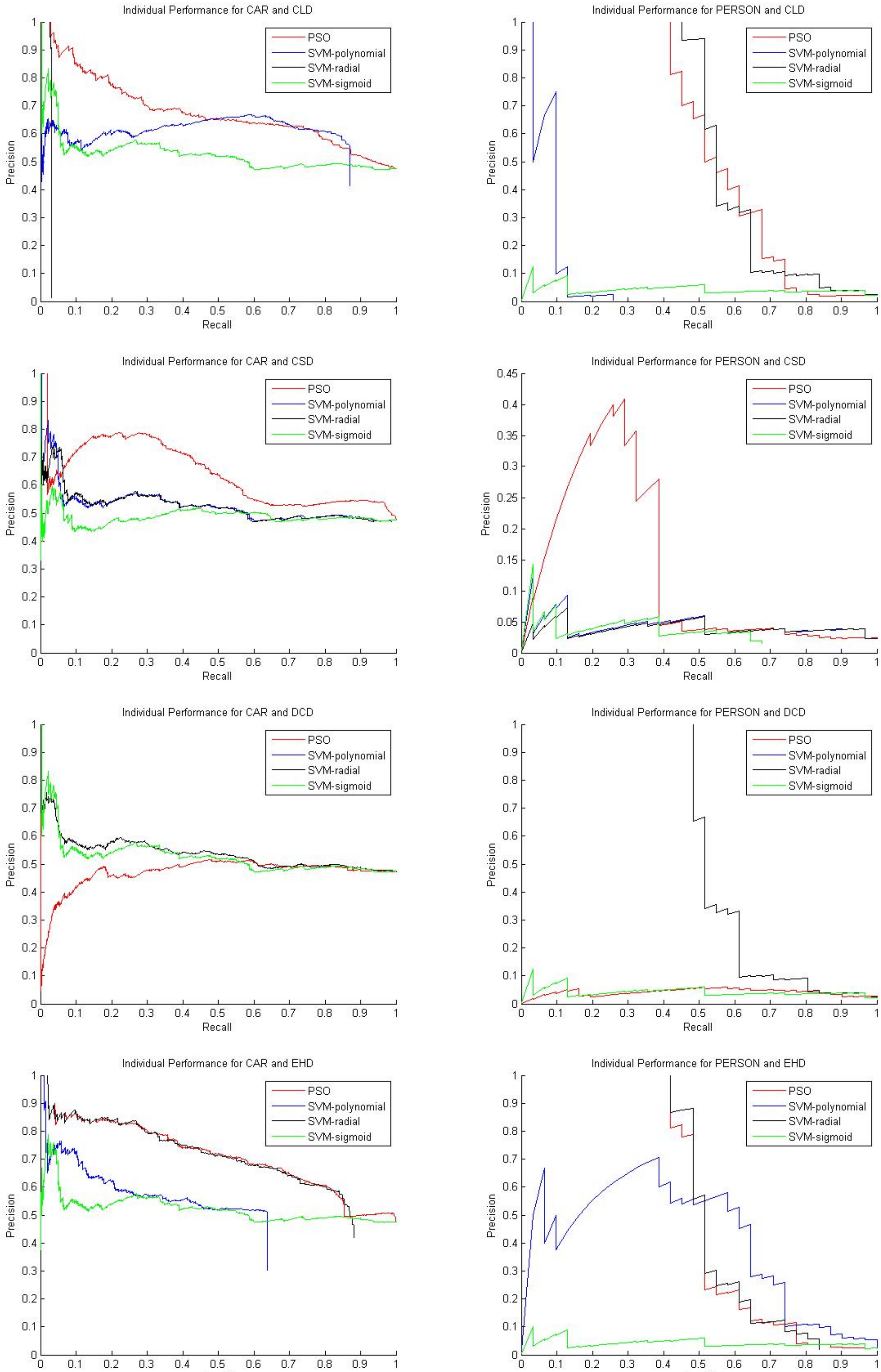

Figure 3. Performance of the proposed biologically inspired Particle Swarm Classifier and Support Vector Machines in Individual Feature Spaces for Concepts 'Car' and 'Person'