3

Performance Analysis of the Modified-Hybrid Optical Neural Network Object Recognition System Within Cluttered Scenes

Ioannis Kypraios
Department of Engineering & Computing, ICTM,
London,
UK

1. Introduction

In literature, we could categorise two broad main approaches for pattern recognition systems. The first category consists of linear combinatorial-type filters (LCFs) (Stamos, 2001) where commonly image analysis is done in the frequency domain with the help of Fourier Transformation (FT) (Lynn & Fuerst, 1998; Proakis & Manolakis, 1998). The second category consists of pure neural modelling methods. (Wood, 1996) has given a brief but clear review of invariant pattern recognition methods. His survey has divided the methods into two further sub-categories of solving the invariant pattern recognition problem. The first sub-category has two distinct stages of separately calculating the features of the training set pattern to be invariant to certain distortions and then classifying the extracted features. The second sub-category, instead of having two separate stages, has a single stage which parameterises the desired invariances and then adapts them. (Wood, 1996) has also described the integral transforms, which fall under the first sub-category of feature extractors. They are based on Fourier analysis, such as the multidimensional Fourier transform, Fourier-Mellin transform, triple correlation (Delopoulos et al., 1994) and others. Part of the first sub-category is also the group of algebraic invariants, such as Zernike moments (Khotanzad & Hong, 1990; Perantonis & Lisboa, 1992), generalised moments (Shvedov et al., 1979) and others. Wood has given examples of the second sub-category, the main representative being based on artificial neural network (NNET) architectures. He has presented the weight-sharing neural networks (LeCun, 1989; LeCun et al. 1990), the high-order neural networks (Giles & Maxwell, 1987; Kanaoka et al. 1992; Perantonis & Lisboa, 1992; Spirkovska & Reid, 1992), the time-delay neural networks (TDNN) (Bottou et al., 1990; Simard & LeCun, 1992; Waibel et al., 1989) and others. Finally, he has included an additional third sub-category with all the methods which cannot be placed under either the feature-extraction feature-classification approach or the parameterised approach. Such methods are image normalisation pre-processing (Yuceer & Oflazer, 1993) methods for achieving invariance to certain distortions. (Dobnikar et al., 1992) have compared the invariant pattern classification (IPC) neural network architecture versus the Fourier Transform method. They used for their comparison black-and-white images. They have proven the generalisation

properties and fault-tolerant abilities to input patterns of the artificial neural network architectures.

An alternative approach for a pattern recognition system has been well demonstrated previously with the Generalised Hybrid Optical Neural Network (G-HONN) filter (object recognition system) (Kypraios, 2010; Kypraios et al., 2004a). G-HONN system combines the digital design of a filter by artificial neural network techniques with an optical correlator-type implementation of the resulting non-linear combinatorial correlator type filter (Jamal-Aldin et al., 1998). The motivation for the design and implementation of the G-HONN object recognition system was to achieve the performance advantages of both artificial neural networks (Looney, 1997; Haykin, 1999; Beale & Jackson, 1990) and the optically implemented correlators (Kumar, 1992). Thus, NNETs exhibit non-linear superposition abilities (Kypraios et al., 2002) of the training set pattern images, learning and generalisation abilities (Kypraios et al., 2004a; Kypraios et al., 2003) over the whole set of the input images. Also, optical correlators allow high speed implementation of the algorithms described.

There are two main design blocks in the G-HONN system, the NNET and a non-linear combinatorial-type correlator (filter) block (Jamal-Aldin, 1998; Casasent, 1984; Caulfield, 1980; Caulfield & Maloney, 1969). Briefly, the original input images pass first through the NNET block and, then, the extracted images from the NNET block's output are used to form a non-linear combinatorial-type correlator filter. Thus the output of the correlator block is a composite image of the G-HONN system's output. To test the system, we correlate it with an input image. Before proceeding to analytical descriptions of the general architecture of the G-HONN system and in an effort to keep consistency between the different mathematical symbolism of artificial neural networks and optical correlators we need to unify their representation. We denote the variables names and functions by non-italic letters (except the vector elements written within the vector, which are written in italic, too), the names of the vectors by italic lower case letters and the matrices by italic upper case. The frequency domain vectors, matrices, variable names and functions are represented by bold letters and the space domain vectors, matrices, variables and functions by plain letters.

Let $h(k,l)$ denote the composite image of the correlator block and $x_i(k,l)$ denote the training set images, where $i = 1,2,\cdots,N$ and N is the number of the training images used in the synthesis of a combinatorial-type filter. The basic filter's transfer function, from the weighed linear combination of x_i, is given by:

$$h(k,l) = \sum_{i=1}^{N} a_i \, x_i(k,l) \tag{1}$$

where the coefficients $\mathbf{a_i}$ $(\mathbf{i = 1,2,...,N})$ are to set the constraints on the peak given by ***c.*** The a_i values are determined from:

$$\boldsymbol{a = R^{-1} c} \tag{2}$$

where ***a*** is the vector of the coefficients $\mathbf{a_i}$ $(\mathbf{i = 1,2,...,N})$, ***R*** is the correlation matrix of t_i and ***c*** is the peak constraint vector. The elements of this are usually set to zeros for false-class objects and to ones for true class objects.

Now, let an image s be the input vector to an artificial neural network's hidden neuron (node), $t_{p\kappa}$ represent the target output for pattern p on node κ and $o_{p\kappa}$ represent the calculated output at that node. The weight from node ι to node κ is represented by $w_{\iota\kappa}$. The activation of each node κ, for pattern p, can be written as:

$$net_{p\kappa} = \sum_{i} \left(w_{\iota\kappa} o_{p\kappa} + b_{\iota} \right) \tag{3}$$

i.e. it is the weighted sum of the calculated output from the node ι to node κ. b_{ι} represents the bias vector of unit ι.

We train a novel-designed NNET with N training set images. The network has N neurons in the hidden layer, i.e. equal to the number of training images. There is a single neuron at the output layer to separate two different object classes. (In a multi-class object recognition problem, the increase of the different classes of objects would require more than one neuron at the output layer to correctly separate all the training images.) From Eq. (3) the net input of each of the neurons in the hidden layer is now given by:

$$net_{x_i} = \sum_{\iota=1}^{m \times n} w_{\iota}^{x_i} s_{\iota}^{x_i} \tag{4}$$

where *net* is the net input of each of the hidden neurons. $w_{\iota}^{x_i}$ are the input weights from the input layer to the hidden neurons for the training image x_i of size $[m \times n]$ in matrix form or of size $[1 \times (m \times n)]$ in vector form. Similarly, for the training image x_N of size $[m \times n]$ in matrix form ($[1 \times (m \times n)]$ in vector form) the net input, net_{x_N} is given by:

$$net_{x_N} = \sum_{\iota=1}^{m \times n} w_{\iota}^{x_N} s_{\iota}^{x_N} \tag{5}$$

From Eqs.(1) and (3) and (5) there is a direct analogy between the combinatorial-type filter synthesis procedure and the combination of all the layers' weighted input vectors.

There are two possible and equivalent custom designs (The Mathworks, 2008) of NNET architectures which could be used to form the basis of the combinatorial-type filter synthesis. In both of the designs each neuron of the hidden layer is trained with only one of the training set images. In effect, $neuron_1$ with the training image x_1, $neuron_2$ with the training image x_2 and so on, ending with $neuron_N$ with the training image x_N. In the first design the number of the input sources is kept constant whereas in the second design the number of the input sources is equal to the number of the training images. In both designs each hidden neuron learns one of the training images. In effect the number of the input weights increases proportionally to the size of the training set:

$$N_{iw} = N \times [m \times n] \tag{6}$$

where N_{iw} is the number of the input weights, N, is the size of the training set equal to the number of the training images and $[m \times n]$ is the size of the image of the training set. The latter design would allow parallel implementation, since all the training images could be

input through the NNET in parallel due to the parallel input sources. However, to allow easier implementation, we chose the former design of the NNET.

Let assume there are three training images of a car, size $[100\times100]$ ($[1\times(100\times100)]$ in vector form), of different angle of view, to pass through the NNET. The chosen first design (see Fig. 1) consists of one input source used for all the training images. The input source consists of 10,000 i.e. $[1\times(100\times100)]$ input neurons equal to the size of each training image (in vector form). Each layer needs, by definition, to have the same input connections to each of its hidden neurons. However, Fig. 1 is referred to as of the fourth layer since there are three hidden layers (shown here aligned under each other) and one output layer. The input layer does not contain neurons with activation functions and so is omitted in the numbering of the layers. Each of the hidden layers has only one hidden neuron. Though the network initially is fully connected to the input layer during the training stage, only one hidden layer is connected for each training image presented through the NNET. Fig. 1 is thus not a contiguous three (hidden) layer network during training, which is why the distinction is made.

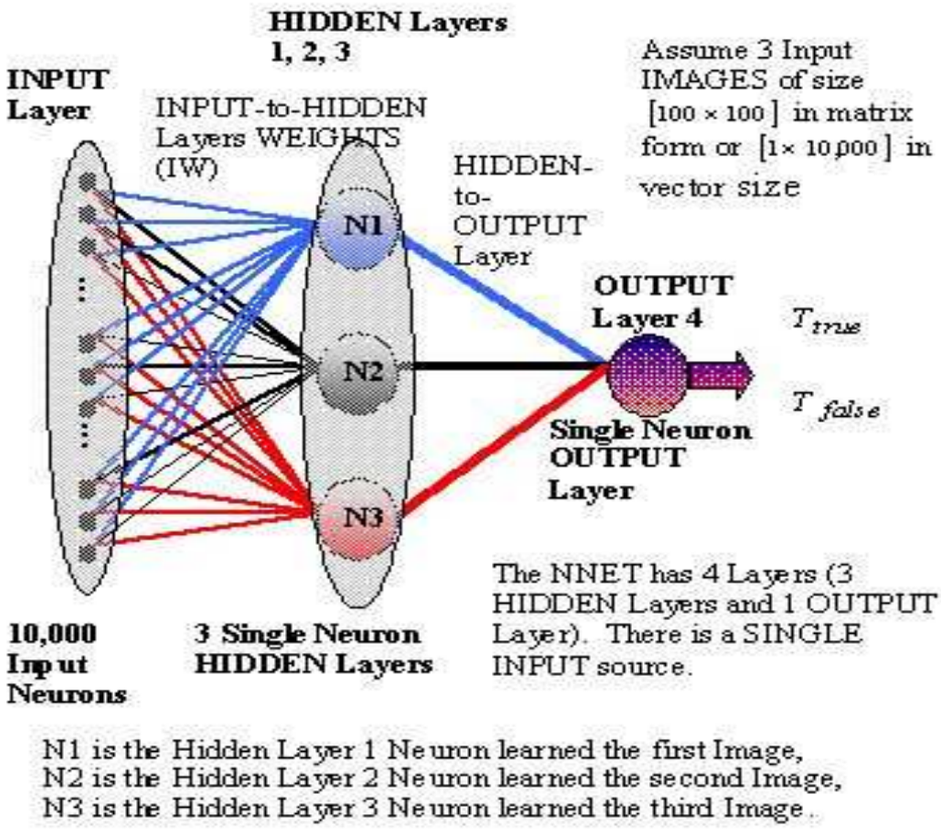


Fig. 1. Architecture of the selected artificial NNET block of the HONN filter.

Next, in section 2 we will give a brief description of the G-HONN system's design and implementation already described with details in the literature. Section 3 describes the M-HONN system. Section 4 focuses on multiple objects recognition and the M-HONN system's

design. It describes the augmented design of the NNET block for accommodating multiple objects recognition of different classes. Section 5 discusses about the performance of M-HONN system with respect to peak sharpness and detectability, distortion range and discrimination ability. We discuss about the M-HONN system and biologically-inspired knowledge learning and representation. Finally, we record the series of tests we conducted with M-HONN system for multiple objects recognition of the same class and of different classes within clutter. Section 6 concludes and suggests future work.

2. General HONN filter's design and implementation

The novel design of NNET's architecture of the G-HONN system is implemented as a feedforward multi-layer architecture trained with a backpropagation algorithm. It has a single input source (as explained in the previous section) of input neurons equal to the size of the training image in vector form. In effect, for the training image $x_{i=1\ldots N}$ of size $[m \times n]$, there are $[m \times n]$ input neurons in the single input source. The input weights are fully-connected from the input layer to the hidden layers. There are N_{iw} input weights proportional to the size of the training set. The number of the hidden layers, N_l, is equal to the number of the images of the training set, N:

$$N = 1,2,3,\cdots,i \tag{7}$$

$$N_l = N \tag{8}$$

Each hidden layer consists of a single neuron. The layer weights are fully connected to the output layer. If we set the output layer to have a single output neuron, then the number of the layer's weights, N_{lw} equals the number of the training images N:

$$N_{lw} = N \times N_{opn} \tag{9}$$

where $N_{opn} = 1$ is the number of the output neurons. There are bias connections to each one of the hidden layers:

$$N_b = N_l \tag{10}$$

where N_b is the number of the bias connections. But from Eq. (8), Eq. (10) becomes:

$$N_b = N \tag{11}$$

Assuming there is only a single output neuron in the output layer, then there is only one target connection for that output neuron.

We apply Nguyen-Widrow (Nguyen & Widrow, 1989; Nguyen & Widrow, 1990) initialisation algorithm for setting the initial values of the input weights, the layer weights and the biases. The transfer function of the hidden layers is set as the Log-Sigmoidal function. When a new training image is presented to the NNET we leave connected the input weights of only one of the hidden neurons. In order not to upset any previous learning of the rest of the hidden layer neurons we do not alter their weights when the new image is input to the NNET. It is emphasised that there is no separate feature extraction

stage (The Mathworks, 2008; Talukder & Casasent, 1999; Casasent et al., 1998) applied to the training set images. To achieve faster learning we used a modified steepest descent (Looney, 1997; The Mathworks, 2008) back propagation algorithm based on heuristic techniques. This adaptive training algorithm updates the weights and bias values according to the gradient descent momentum and an adaptive learning rate:

$$\Delta w(i,i+1) = \mu \times \Delta w(i-1,i) + \alpha \times \mu \times \frac{\Delta P_f}{\Delta w(i+1,i)} \tag{12}$$

$$\Delta b(i,i+1) = \mu \times \Delta b(i-1,i) + \alpha \times \mu \times \frac{\Delta P_f}{\Delta b(i+1,i)} \tag{13}$$

$$\alpha = \left\{ \begin{array}{l} \alpha = \alpha + \varepsilon \quad \text{if } \Delta P_f < 0 \\ \alpha = \text{no change} \quad \text{if } 0 < \Delta P_f \;\&\&\; \Delta P_f > \max(P_f) \\ \alpha = \alpha - \varepsilon \quad \text{if } \Delta P_f > \max(P_f) \end{array} \right\} \tag{14}$$

where now variable i is the iteration index of the network and is updated every time all the training set images pass through the NNET. Δw is the update function of the input and layer weights, Δb is the update function of the biases of the layers and μ is the momentum constant. The momentum (Looney, 1997; Haykin, 1999; Beale & Jackson, 1990; The Mathworks, 2008) allows the network to respond not only to the local gradient, but also to recent trends in the error surface. Thus, it acts like a low-pass filter by removing the small features in the error surface of the NNET. The employment of momentum in the training algorithm allows the network not to get stuck in a shallow local minimum, but to slide through such a minimum. P_f is the performance function, usually set as being the mean square error (mse) (Looney, 1997; Haykin, 1999) and ΔP_f is the derivative of the performance function. The learning rate is indicated with the letter α. It adapts iteratively based on the derivative of the performance function ΔP_f. In effect, if there is a decrease in the ΔP_f, then the learning rate is increased by the constant ε. If ΔP_f increases but the derivative does not take a value higher than the maximum allowed value of the performance function, $\max(P_f)$, then the learning rate does not change. If ΔP_f increases more than $\max(P_f)$, then the learning rate decreases by the constant ε. The layer weights remain connected with all the hidden layers for all the training set and throughout all the training session.

Hence, now that we have described the design and implementation of the G-HONN filter (object recognition system) we can proceed with a detailed description of the modified-HONN filter.

3. Modified-HONN system implementation

We can make the following qualitatively observations for the G-HONN system. Though the combinatorial-type filters (Samos, 2001) contain no information on non-reference objects in the training set used during their synthesis, the NNET includes information for reference and non-reference images of the true-class object. That can be explained due to the NNET interpolating non-linearly (Kypraios et al.,2002) between the reference images included in

the training set and forcing all the non-reference images to follow the activation graph. Moreover the NNET generalizes between all the reference and non-reference images. Quantitatively, we could demonstrate the above observations as follows. The average training set image $\bar{x}$ in the space domain of the combinatorial-type filters is given by:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{15}$$

In the frequency domain Eq. (15) is written as:

$$\bar{\boldsymbol{x}} = \frac{1}{\mathbf{N}} \sum_{i=1}^{N} \boldsymbol{x}_i \tag{16}$$

The non-linear activation function of each hidden neuron of an artificial neural network with a non-linear activation function such as the sigmoidal function $f_s(\ )$ can take the form:

$$f_s(x) = \alpha \frac{1 - \exp(\beta x)}{1 + \exp(\beta x)} \tag{17}$$

where α and β shift the graph of the function with respect the x-axis and y-axis and are called the saturation level and slope. It can be shown (Kypraios et al., 2009) that the output y_N of an artificial neural network with a non-linear activation function corresponding to an input s_i for $i=1,\ldots,N$ (where N is the number of the training set images) is written as:

$$\begin{aligned} &f\left(\sum_{i=1}^{N} s_i g_i - \theta \right) = \\ &= \alpha \frac{1 - k\exp\left(\beta s_1 g_1\right)\exp\left(\beta s_2 g_2\right)\exp\left(\beta s_3 g_3\right)\cdots\exp\left(\beta s_{N-1} g_{N-1}\right)\exp\left(\beta s_N g_N\right)}{1 + k\exp\left(\beta s_1 g_1\right)\exp\left(\beta s_2 g_2\right)\exp\left(\beta s_3 g_3\right)\cdots\exp\left(\beta s_{N-1} g_{N-1}\right)\exp\left(\beta s_N g_N\right)} \end{aligned} \tag{18}$$

where $k = \exp(-\beta\theta)$ takes a constant value (and g_i the neural network node' weights). Therefore from Eq. (16) and Eq. (18) it is shown that any artificial neural network with a non-linear activation function can non-linearly interpolate through the different training set views of the true-class object. Thus, the average training set image $\bar{x}$ in the space domain of the NNET is given by:

$$\bar{x} = \frac{1}{N} f_\kappa(x_i) \tag{19}$$

where $f_\kappa(\ )$ is the activation function of node κ in the space domain. Eq. (19) is written in the frequency domain as:

$$\bar{\boldsymbol{x}} = \frac{1}{\mathbf{N}} \mathbf{f}_\kappa(\mathbf{x}_i) \tag{20}$$

The activation function $f_\kappa(\ )$ of node κ against the training set images x_i is plotted in Fig. 2. On the activation function graph the true-class object values (which is similar for the false-

class object) are marked with + . Now, if we mark on the plot the activation function values for the training image at 30º and 40º degrees object poses, then the activation function for the training image at 35º degrees will be located on the graph between the values of the activation function for the 30º and 40º degree inputs. The actual activation function values for the training set images of x_{30}, x_{40} and x_{35} are located in the area included under the graph for activation function values greater or equal to the pre-specified true-class object classification level, in this case shown we assume it is set at +40.

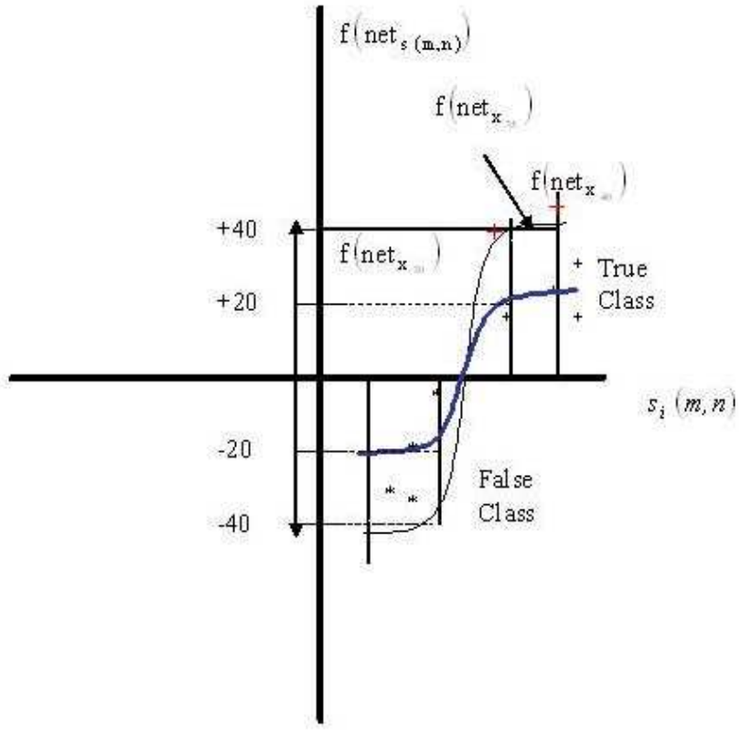


Fig. 2. It shows the activation function graph of node κ against the training set images x_i .

Motivated by these observations, we apply an optical mask to the filter's input (see Fig. 3). The mask is constructed by the weight connections of the reference images of the true-class object and is applied to all the tested images. Modified-HONN (M-HONN) system is described as follows:

$$\Gamma_c = W^{x_c} \times L^{x_c} = \begin{bmatrix} w_{11}^{x_c} & w_{12}^{x_c} & L & w_{1n-1}^{x_c} & w_{1n}^{x_c} \\ w_{21}^{x_c} & w_{22}^{x_c} & L & w_{2n-1}^{x_c} & w_{2n}^{x_c} \\ & & & & \\ & & & & \\ w_{m1}^{x_c} & w_{m2}^{x_c} & L & w_{mn-1}^{x_c} & w_{mn}^{x_c} \end{bmatrix} \times \begin{bmatrix} l_{11}^{x_c} L\, l_{1q}^{x_c} \\ l_{21}^{x_c} L\, l_{2q}^{x_c} \\ \\ \\ l_{n1}^{x_c} L\, l_{nq}^{x_c} \end{bmatrix} \quad (21)$$

where W^{x_c} and L^{x_c} are the input and layer weights from the input neuron of the input vector element at row m and column n to the associated hidden layer for the training image $x_c(m,n)$. $l_{mn}^{x_c}$ are the input and layer weights from the hidden neuron of the layer vector element at row m and column n to the associated output neuron q. This time, instead of multiplying each training image with the corresponding weight connections as for the G-HONN system's implementation, we keep constant the weight connection values, setting

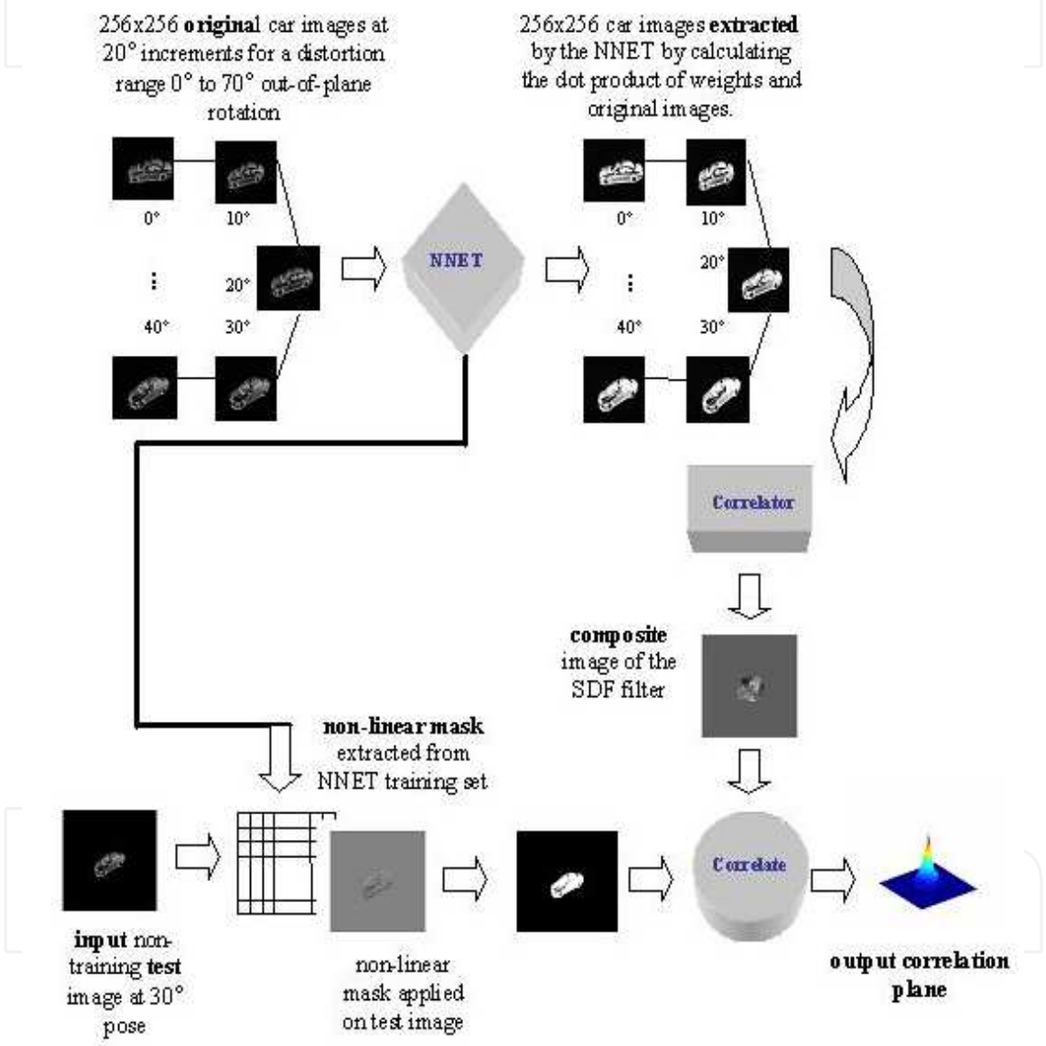


Fig. 3. M-HONN system block diagram.

them to be equal with a randomly chosen image included in the training set $x_c(m,n)$. The matrix Γ_c is used for creating the optical mask for the M-HONN system's implementation. The transformed image $S_{i=1\cdots N}(m,n)$ calculated from the dot product of the matrix elements of Γ_c with the corresponding training image matrix elements of $X_{i=1\cdots N}(m,n)$ is given by:

$$S_{i=1\cdots N} = \Gamma_c \cdot X_{i=1\cdots N}(m,n) \tag{22}$$

Thus, the M-HONN system's transfer function is formulated as follows:

$$M-HONN = \sum_{i=1\cdots N}^{N} a_i \cdot S_i(m,n) \tag{23}$$

In Eq. (23) we have chosen to constrain the correlation peak height values as we did with the constrained-HONN (C-HONN) system's implementation, but we can also easily re-write the system's transfer equation for the case of the unconstrained peak height values as with the unconstrained-HONN (U-HONN) system's implementation (Mahalanobis, 1994; Kypraios et al., 2004b).

4. Multiple objects recognition

Multiple objects of the same class can be accommodated by the G-HONN type filters to be recognised within an input cluttered image due to the shift invariance properties inherited by its correlator unit. Thus, in the M-HONN system all the training set images pass through the NNET unit. This time, instead of multiplying each training image with the corresponding weight connections (mask) as for the C-HONN filter, we keep constant the weight connection values, setting them to be equal with a randomly chosen image included in the training set. All the test set images are multiplied with the same randomly chosen image's weight connection values. Then, the training set images, after being transformed (masked) through the NNET unit by being multiplied with the mask, pass through the correlator unit where they are correlated with the masked test set images. In effect, the cross-correlation of each masked test set image with the transformed training set images (reference kernel) returns an output correlation plane peak value for each cross-correlation step. Hence, the maximum peak height values of the output correlation plane correspond to the recognised true-class objects.

4.1 Modified NNET block architecture for multiple objects of different classes recognition

As for all the HONN-type systems (Kypraios et al., 2004; Kypraios et al., 2003; Kypraios et al., 2009), in the M-HONN system's NNET block (unit) there is a single input source used for all the input data. Assuming we have N = 3 input still images or video frames of size 256×256 in pixels, then the input source consists of 65.536 i.e. [1(256×256)] input neurons equal to the size of each training image or frame (in vector form). Each layer needs, by definition (Hagan et al., 1996), to have the same input connections to each of its hidden neurons. Therefore, the shown NNET architecture is referred to as N+1 = 3+1 = 4, four-layered since there are, N = 3, three hidden neurons (though shown here aligned under each other, they do not belong in the same hidden layer but rather create three separate hidden layers each of a single hidden neuron) and one output layer. Each of the hidden layers consist of only one hidden neuron. The input layer does not contain neurons with activation functions and so is omitted in the numbering of the layers. Though the network initially is fully connected to the input layer during the training stage, only one hidden layer is connected for each training image presented through the NNET. NNET is thus not a

contiguous three layer network during training, which is why the distinction is made. In effect, neuron_1 is trained with the training still image or video frame x_1, neuron_2 is trained with the training still image or video frame x_2 and so on, ending with neuron_N being trained with the training still image or video frame x_N. Thus, the number of the input weights increases proportionally to the size of the training set:

$$N_{iw} = N \times [m \times n] \tag{24}$$

where N_{iw} is the number of the input weights, N, is the size of the training set equal to the number of the training images and [m×n] is the size of the image of the training set.

Fig. 4 shows the modified NNET block architecture for accommodating multiple objects for more than one class recognition. As for all the family of G-HONN filters, NNET is implemented as a feedforward multi-layer architecture trained with a backpropagation algorithm. It has a single input source of input neurons equal to the size of the training image or video frame in vector form. In effect, for the training still image or video frame $x_{i=1\ldots N}$ of size [m×n], there are [m×n] input neurons in the single input source. The input weight are fully connected from the input layer to the hidden layers. There are N_{iw} input weights proportional to the size of the training set. The number of the hidden layers, N_l is equal to the number of the images or video frames of the training set N:

$$N = 1,2,3,\cdots,i \quad \text{and} \quad N_l = N \tag{25}$$

We have set to each hidden layer to contain a single neuron. The layer weights are fully connected to the output layer. Now, the number of the layer weights N_{lw} is equal to:

$$N_{lw} = N \times N_{opn} \quad \text{and} \quad N_{opn} = N_{classes} \tag{26}$$

where N_{opn} is the number of the output neurons and $N_{classes}$ is the number of the different classes. In effect, we have augmented the output layer by adding more output neurons, one for each different class. On Fig. 4 we assume $N_{classes} = 2$. Thus:

$$N_{opn} = N_{classes} = 2 \quad \text{so, there are} \quad N_{lw} = N \times 2 \tag{27}$$

and

$$N_{class1\ lw} = N \ , \ N_{class2\ lw} = N \tag{28}$$

where $N_{class1\ lw}$ and $N_{class2\ lw}$ are the layer weights corresponding to object class 1 and object class 2, respectively. There are bias connections to each one of the hidden layers:

$$N_b = N \tag{29}$$

where N_b is the number of the bias connections. There are $N_{target\ w}$ target connections from the N_{opn} output neurons of the output layer:

$$N_{target\ w} = N_{opn} \tag{30}$$

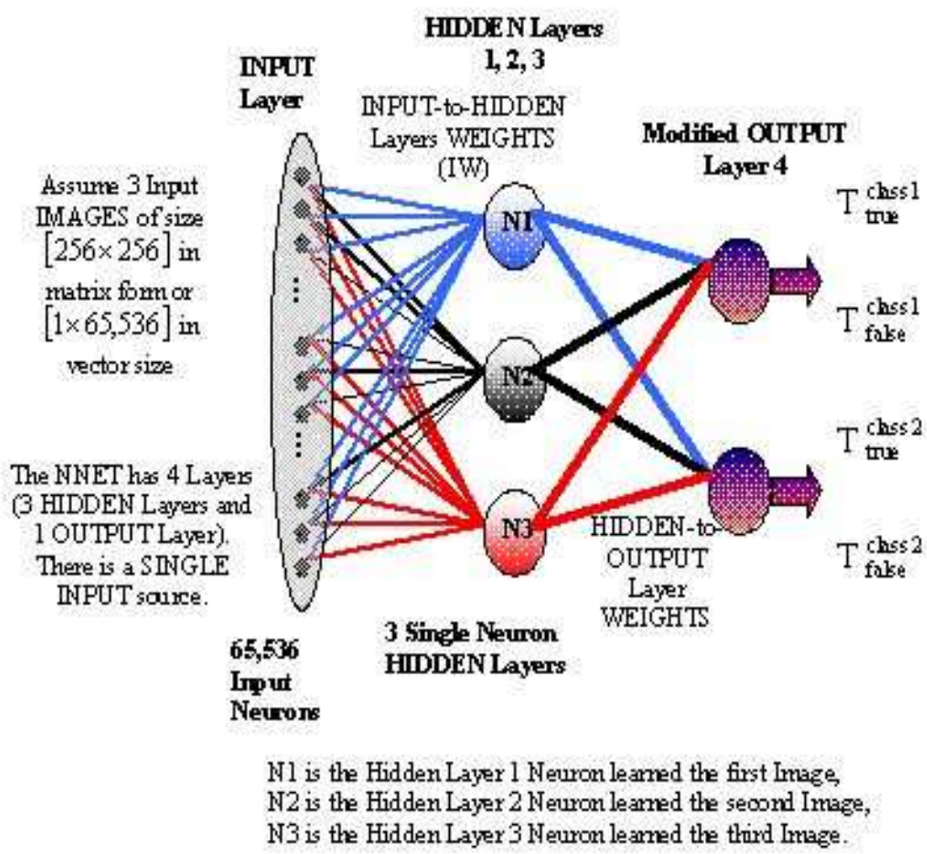


Fig. 4. Modified NNET block of the M-HONN system for multiple objects recognition of different classes

Thus, now for $N_{classes} = 2$ there will be N transformed images being created for class 1 and N transformed images being created for class 2. Then, both sets of transformed images are used for the synthesis of the system's composite image. M-HONN system for multiple objects recognition of different class objects is written as follows:

$$M\text{-}HONN = \sum_{i=1LN_{classes}\times N}^{N_{classes}\times N} a_i \cdot S_i^{class}(m,n)$$
$$= a_1 \cdot \left(\Gamma_c^{class} \times X_1(m,n)\right) + a_2 \cdot \left(\Gamma_c^{class} \times X_2(m,n)\right) + L + \quad (31)$$
$$+ \cdots a_N \cdot \left(\Gamma_c^{class} \times X_N(m,n)\right)$$

or in the frequency domain the above equation is re-written as:

$$\boldsymbol{M\text{-}HONN} = \sum_{i=1LN_{class}\times N}^{N_{class}\times N} \mathbf{a}_i \cdot \boldsymbol{S}_i^{class}(\boldsymbol{m},\boldsymbol{n}) \quad (32)$$

The above Eq. (31) in spatial domain, and Eq. (32) in frequency domain describe the M-HONN system's transfer function for multiple objects recognition (where the upper script *class* is used for the class index, i.e. for Fig. 4 we have *class* = *class1, class2*). Thus, the M-HONN filter (robust object recognition system) is composed of a non-linear space domain superposition of the training set images or from the video frames of the training set video sequences. As for all the HONN-type systems, the multiplying coefficient now becomes a non-linear function of the input weights and the layer weights, rather than a simple linear multiplying constant as used in a constrained linear combinatorial-type filter synthesis procedure. The non-linear M-HONN system is inherently shift invariant and it may be employed in an optical correlator as would a linear superposition constrained-type filter, such as the synthetic discriminant function (SDF) -type (Bahri & Kumar, 1988) filters. It may be used as a space domain function in a joint transform correlator architecture or be Fourier transformed and used as Fourier domain filter in a 4-f Vander Lugt (Vander Lugt, 1964) type optical correlator.

5. Performance analysis

We have constructed a data set of input images of an S-type Jaguar car model at 10° increments of out-of-plane rotation at an elevation angle of approximately 45° to be used for the M-HONN system. A second set of images was constructed for the Police car model Mazda Efini RX-7 at the same elevation angle to serve as the out-of-class data for discrimination tests (see Fig. 5). A third data set was created of the background images of typical car parks (see Fig. 6) and the images of the S-type car model and the Mazda RX-7 car model added in the background scene. The size of all the images was $\left[256\times256\right]$ and all the images are in grey-scale bitmap format. All the input training images (and all the input test set images) for M-HONN system are concatenated row-by-row into a vector of size $\left[1\times\left(256\times256\right)\right]$ prior to input to the NNET block. Normally this size of image is impossibly large for

processing by any artificial neural network architecture, since to be implemented by enough input and layer weights:

$$\begin{aligned} N_{iw} &= 10\times\left[256\times256\right] \\ &= 10\times65,536 \\ &= 655,360 \end{aligned} \tag{33}$$

Thus, for a training set of $N=10$ individual vectors of size $\left[256\times256\right]$, there would, in total, be more than half-a-million input weight connections needed. Thus the selective weight connection architecture is employed to overcome this problem. To overcome this problem we developed a novel selective weight connection architecture (see Section 2). Also, applying the heuristic training algorithm with momentum and an adaptive learning rate into the NNET training session (Nguyen & Widrow, 1989; Nguyen & Widrow, 1990), has speeded up the learning phase and reduced the memory size needed to complete fully the training session. Here, it worth mentioning that the NNET block and, in overall, M-HONN system is able to process input still images and video frames for all the test series in few a msec with a Dual Core CPU at 2.4 GHz with 4.0GB RAM. Additionally, due to the

generalization properties exhibited by a NNET architecture, the number of the training images decreases, in comparison to the typical number of images required for the training set of linear combinatorial filters (such as the SDF filter).

Fig. 5. RX-7 Mazda Efini Police patrol car used in the training and test sets

Fig. 6. Car park scene used in the training and test sets

It was proven experimentally that by choosing different values of the classification levels for the true-class Cl_T and false-class Cl_F objects, one can control the M-HONN system's behaviour to suit different application requirements, similarly with all the HONN-type systems. Thus we define:

$$\Delta Cl = \left| Cl_T - Cl_F \right| \tag{34}$$

where ΔCl is the absolute distance of the classification levels between the true-class objects and the false-class objects. When we increase ΔCl, then the resulting M-HONN system behaves more like a high-pass biased filter, which generally gives sharp correlation peaks and good clutter suppression but is more sensitive to intra-class distortions. Now, when we decrease ΔCl, then the resulting M-HONN system behaves more like a minimum variance synthetic discriminant function (MVSDF) (Kumar, 1986) filter with relatively good intra-class distortion invariance but producing broad correlation peaks. In effect, when ΔCl increases, the M-HONN system possesses better discriminatory properties but when ΔCl decreases the M-HONN system has better generalising properties. By plotting the isometric

correlation planes of M-HONN system for different ΔCl values, one could observe that by increasing ΔCl value it leads to an increased emphasis of the high spatial frequency content of the composite images comprising M-HONN system, which in turn leads to a more localised response, sharper peaks, and reduction in the plane's sidelobes. By decreasing ΔCl value it leads to an increased emphasis on peripheral lower spatial frequency content of the composite images comprising M-HONN system, which in turn leads to a broader peaks in the correlation plane.

Next, we summarise the tests series for assessing M-HONN system's peak sharpness and detectability, distortion range, and discrimination ability, which we have all described them in full details in our previous work (Kypraios et al., 2008). We focus afterwards in analysing the performance of the M-HONN object recognition system within cluttered scenes.

5.1 Peak sharpness and detectability

Here we assessed (Jamal-Aldin et al., 1997; Jamal-Aldin et al., 1998; Kumar & Hassebrook, 1990) M-HONN system's ability to detect non-training in-class images that are oriented at the intermediate angle of view between the training images (Refregier, 1990; Refregier, 1991). The training set consisted of still images out-of-plane rotated between $[20\ 70]$ degrees at increments of 20°. We tested the M-HONN system with the true-class object's intermediate car poses over the same range at 10° increments. Two randomly chosen intermediate car poses, at 130° and at 140°, were added in the training set of the M-HONN system to create a false-class. We set the target of the false-class object to be $T_{false} = -40$ and of the true-class object to be $T_{true} = +40$. The M-HONN system had no information on the non-training, intermediate car images in the construction of its composite image. We explicitly constrained the correlation peak in the constraint matrix. Thus, we constrained the correlation peaks in the constraint matrix to be $+1$ for the images of the true-class object and 0 for the images of the false-class object. The randomly chosen mask Γ_c applied on both the training set and the test set was built from the training set image at 60°, i.e. c= 60°:

$$\Gamma_{60^\circ} = W^{x_{60^o}} \times L^{x_{60^o}} = \begin{bmatrix} w_{11}^{x_{60^o}} & w_{12}^{x_{60^o}} & L & w_{1n-1}^{x_{60^o}} & w_{1n}^{x_{60^o}} \\ w_{21}^{x_{60^o}} & w_{22}^{x_{60^o}} & L & w_{2n-1}^{x_{60^o}} & w_{2n}^{x_{60^o}} \\ & & & & \\ w_{m1}^{x_{60^o}} & w_{m2}^{x_{60^o}} & L & w_{mn-1}^{x_{60^o}} & w_{mn}^{x_{60^o}} \end{bmatrix} \times \begin{bmatrix} l_{11}^{x_{60^o}} & L & l_{1q}^{x_{60^o}} \\ l_{21}^{x_{60^o}} & L & l_{2q}^{x_{60^o}} \\ & & \\ l_{n1}^{x_{60^o}} & L & l_{nq}^{x_{60^o}} \end{bmatrix} \quad (35)$$

where $W^{x_{60^o}}$ and $L^{x_{60^o}}$ are the matrices of the input and layer weights. $w_{mn}^{x_{60^o}}$ are the input weights from the input neuron of the input vector element at row m and column n to the associated hidden layer for the training image $x_{60^o}(m,n)$ at 60° angle of view. $l_{mn}^{x_{60^o}}$ are the layer weights from the hidden neuron of the layer vector element at row m and column n to the associated output neuron. We set q = 1 since the output layer had only one neuron for a single class of objects. In M-HONN system, instead of multiplying each training image with the corresponding weight connections as done for the constrained- HONN (C-HONN)

system, we keep constant the weight connection values which are set to be equal to a (randomly) chosen image included in the training set, here to be $x_{60^\circ}(m,n)$.

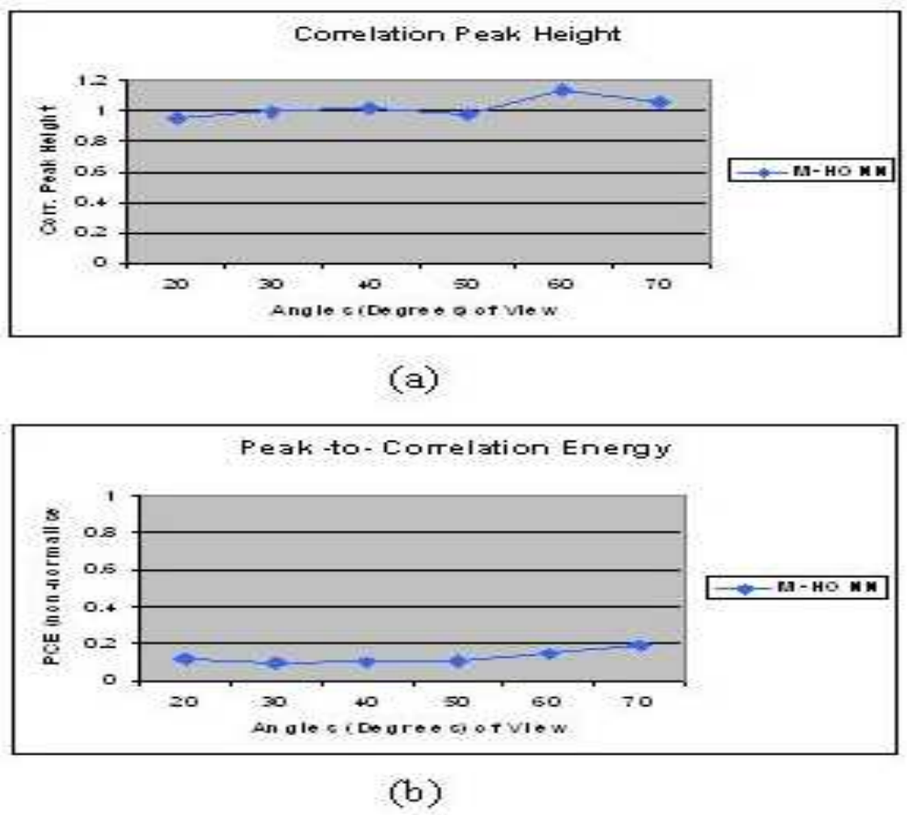


Fig. 7. (Adapted by Kypraios et al., 2008) shows (a) correlation peak-height versus the out-of-plane rotation angles of the object over the range of 20° to 70°. We tested the M-HONN system with the true-class object's intermediate car poses over the same range out-of-plane rotated at 10° increments; (b) the non-normalised PCE values of the test images at 10° increments versus the angles of view over the range of 20° to 70°.

Fig. 7 (a) shows the plot of the correlation-peak height for each input image for the M-HONN system. From the plot it is shown that the M-HONN system is invariant to the out-of-plane rotation, since it has produced consistent correlation peaks for both the in-class training and non-training images around the fixed-correlation peak-height value of unity. The consistency of the correlation peak values that the M-HONN system has exhibited demonstrate the system's ability to interpolate well between the intermediate car poses at 10° increments. Earlier (Kypraios et al., 2008; Kypraios, 2009; Kypraios, 2010), we have shown the NNET includes information for reference and non-reference images of the true-class object. Hence, the NNET interpolates non-linearly between the reference and non-reference images to follow the activation function graph. Moreover, the NNET is able to generalize between all the reference and non-reference images.

Fig. 7 (b) shows the non-normalised peak-to-correlation energy (PCE) (Kumar & Hassebrook, 1990) values for the M-HONN system. From the graph, it can be observed that the M-HONN system produced PCE values for the intermediate non-training images close to those produced by the training car images. In effect, the system maintains correlation peak sharpness for the in-class training and non-training images.

5.2 Distortion range

The second tests series (Jamal-Aldin et al., 1997; Jamal-Aldin et al., 1998; Kumar & Hassebrook, 1990) was carried out to assess the distortion range (Refregier, 1990; Refregier, 1991, Kypraios et al., 2008) of the M-HONN system. The training set consisted of images for a distortion range over 0° to 90°. We used several smaller test sets, which consisted of two in-class training images at a widely separated angle within the range $\left[10^\circ\ 20^\circ\ 30^\circ\ 40^\circ\ 50^\circ\ 60^\circ\ 70^\circ\ 80^\circ\right]$ and a third non-training in-class image lying on the bisector angle of the two in-class training images (see Fig. 8 (a)). The intermediate non-training car pose image was $\hat{\Theta}_3 \in \left[5^\circ, 40^\circ\right]$ i.e. $\Theta_3 = \left[5^\circ\ 10^\circ\ 15^\circ\ 20^\circ\ 25^\circ\ 30^\circ\ 35^\circ\ 40^\circ\right]$. Three randomly chosen training images, out-of-plane rotated at $110^\circ, 130^\circ$ and 140°, were added in the training set of the M-HONN system which fell inside the false-class. The targets of the true-class objects and of the false-class objects were found to be best set for the tests series as $T_{true} = +40$ and $T_{false} = -40$. The M-HONN system has no information built into it on the test images of the intermediate car poses. We constrained the correlation peaks in the constraint matrix to be $+1$ for the images of the true-class object and 0 for the images of the false-class object.

Fig. 8 (b) shows the correlation-peak height for each input image for the M-HONN system. It is found the system has good performance in recognising all the intermediate car poses of the test set. The correlation-peak height of the in-class input images, intermediate between two training images, lie within a band of greater than 76% of the pre-specified peak-height constant in the constraint matrix *C* for the M-HONN system. From the graph it can be observed that the system tolerated orientation over a range of $\hat{\Theta}_3 \in \left[5^\circ, 40^\circ\right]$.

5.3 Discrimination ability

The third tests series (Jamal-Aldin et al., 1997; Jamal-Aldin et al., 1998; Kumar & Hassebrook, 1990) was carried out to assess the discrimination ability (Refregier, 1990; Refregier, 1991, Kypraios et al., 2008) of the M-HONN system. In the tests, M-HONN system tried to discriminate between objects of different classes while retaining invariance to in-class distortions. The training set consisted of images of the Jaguar S-type for adistortion range over 20° to 70° at 10° increments. The test set consisted of one training image out-of-plane rotated at 40° of the Jaguar S-type and a second image of the out-of-class RX-7 Police patrol car at the same angle of out-of-plane rotation. Two different training set configuration of still images were experimented with. Firstly, we added two images of the Jaguar S-type at 130° and 140° for the false-class of the system's training set. We constrained the false-class images of the objects to zero correlation peak-height in the synthesis of the M-HONN system's composite image. Secondly, we conducted experiments

with no inclusion in the system's composite image of any false-class images. For both cases, we aimed in observing if there was any change in the class separation ability of the M-HONN system. We constrained the true-class objects to unity correlation peak-height and we used the same as before Targets for the false- and true- class images of the NNET block. The target of the false-class object is $T_{false} = -40$, and the Target of the true-class object is $T_{true} = +40$ in the training set of the NNET block for the M-HONN system. It had no built-in information on the test images.

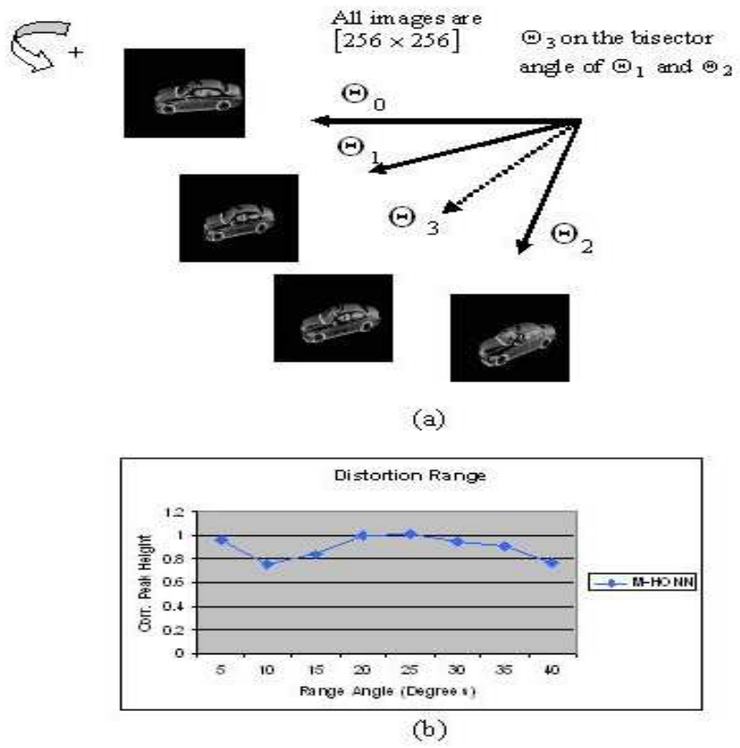


Fig. 8. (Adapted by Kypraios et al., 2008) shows (a) the reference angle, Θ0, and the two in-class training images at the angles Θ1 and Θ2. The test image is on the bisector at angle Θ3; (b) the correlation peak-heights for each input image over a range of Θ3 = [5° 40°] for the M-HONN system.

From the conducted experiments we drew Table 1. The forth column of Table 1 records the values taken for the in-class training image and the fifth column contains all the values taken for the out-of-class training image. It is shown from the third column of Table 1 that M-HONN system gave sufficient discrimination ability between the two objects, the Jaguar S-type car and the RX-7 Police patrol car. It produced 12% class separation (with the false-class images included in the synthesis of the composite image with zero correlation peak-height constraint). By not including any false-class images in the system's composite image, but by setting to unity correlation peak-height constraint the true-class images and keeping constant the target of the false-class object to $T_{false} = -40$ and the target of the true-class object to $T_{true} = +40$, the M-HONN system increased the class separation to 27%. Thus, the

two different training set still images configurations that we had experimented with (in first, false-class images zero peak constrained and, in second, false-class images not included in the system's composite image) helped us make a useful observation about the M-HONN system's ability to distinguish between two different classes. More specifically, the false-class images included in the composite image, and zero peak constrained, were taken from the true-class object in different poses not included in the training set. In effect, when we tested the RX-7 police patrol car images, the system separated the input images from the trained images (unity peak constrained) of the true-class object and the false-class images (of the same true-class but zero peak constrained) as a third class. Apparently, that caused the drop of the M-HONN system discrimination ability by almost half. We have found to be a solution to the problem by including false-class images not belonging in the same the true-class object but from a different one which it could increase further the discrimination ability of M-HONN system.

Correlation Peak Height

Object Recognition System	False Class BG	Discrimination Ability %	Jaguar S-type	RX-7 Mazda Efini
M-HONN	zero – peak constrained	11.9967	1.0878	0.9573
M-HONN	NOT included	26.9758	1.1299	0.8251

Table 1. (adapted by Kypraios et al., 2008) Discrimination Ability of M-HONN system

5.4 Clutter tolerance

5.4.1 Training sets

We have conducted several tests (Kypraios et al., 2009) for evaluating the performance of the M-HONN system in recognising multiple objects of the same class or of different classes. Several training sets were created for testing the system's performance with still images and with video sequences. The first training set consisted of still images of the Jaguar S-type car for a distortion range over 0° to 360° out-of-plane rotated at 10° increments. The second training set consisted of still images of the RX-7 Police patrol car for a distortion range approximately over 0° to 360° out-of-plane rotated at 10° increments. The third training set consisted of video frames of a Ferrari Testarossa car within a background clutter scene. The fourth training set consisted of still images of different car park scenes. A fifth training set consisted of video frames we have taken showing a sequence of a Jaguar S-type car and a Ferrari Testarossa car within a background clutter scene. All the training and test sets of the still images and of the video sequence frames were used in grey-scale bitmap format, and they were sized to 256x256. All the test and train input still images and video frames were concatenated row-by-row into a vector form prior being processed by the NNET block of the M-HONN system.

5.4.2 Biologically-inspired knowledge representation and learning

As S. Haykin in his work on artificial neural network architectures (S. Haykin, 1999) observes, pattern recognition systems need to be re-designed in novel architectures, if they are to be solving more complex problems. He argues that such novel architectures should be designed with separate blocks of a recognition unit and a knowledge learning unit, and that the implementation of such designs can be only possible with the combination of artificial neural networks architectures with other tools as a hybrid. Some of the elements (S. Haykin, 1999) that such biologically-inspired hybrid systems need to exploit are, the non-linearity of the input information, learning and adaptation to the input information, and provide an attentional mechanism for the hybrid system to be able to select certain information to be included in its learning against other input. Therefore, knowledge representation and learning becomes a central issue in the design and implementation of such hybrid biologically-inspired pattern recognition systems (Lee & Portier, 2007).

Aler et al. in their work discuss the knowledge representation and its role in knowledge learning (Aler et al., 2000). Aler et al. argue the effects that altering the knowledge representation can have on the problem knowledge learned and problem solving. They consider any problem solving system to consist of a domain theory which specifies the task to be solved, the initial problem states and the aimed problem goals, and a control knowledge which guides the decision-making process. They were able to demonstrate the effects of knowledge representation to the efficiency of the problem solving process.

Recent work we have conducted (Kypraios, 2010) has demonstrated the problem solving ability of the HONN-type systems, such as the M-HONN system for multiple objects recognition. We have shown the system is able to solve, in particular, different visual tasks. Fig. 9 shows the first problem we have tested M-HONN system for recognising different angles of view of the input object. The training set consisted of still images of the Jaguar S-type car out-of-plane rotated over a range 0° to 170° to belong in the true-class, and still images of the Jaguar S-type car out-of-plane rotated over a range 180° to 360° to belong in the false-class. The true-class images were constrained to unit correlation peak-height in the synthesis of the M-HONN system's composite image, and the false-class images were constrained to zero correlation peak-height in the synthesis of the M-HONN system's composite image (see Fig. 10). We have set the true-class target classification levels (here we assume there is only one *class=1*, so there is no need to set any target connections for a second output neuron) to be $T_{true}^{class\,1} = +40$, and for false-class the target classification levels were set to be $T_{false}^{class\,1} = -40$. The test set consisted of multiple Jaguar S-type car objects inserted in plain background at different non-training out-of-plane rotation angles over a range 0° to 360°. As shown on Fig. 9, M-HONN system was able to correctly recognise the Jaguar S-type car poses over the range 0° to 170° to belong in the true-class, and the Jaguar S-type car poses over the range 180° to 360° to belong in the false-class. We have indicated with the solid line the recognised true-class objects and with the dashed line the recognised false-class objects.

Fig. 11 shows the second test we conducted to demonstrate the system's ability of problem solving where we want the system to recognise only the true-class objects of the Jaguar S-type car over a range 0° to 360°, and reject the false-class objects of the RX-7 Mazda Efini

Fig. 9. It shows the first visual problem for testing the M-HONN object recognition system's ability of problem solving. M-HONN system tries to recognise certain angles of view of the input object while rejecting others. The training set consisted of still images of the Jaguar S-type car out-of-plane rotated over a range 0° to 170° to belong in the true-class, and still images of the Jaguar S-type car out-of-plane rotated over a range 180° to 360° to belong in the false-class. We have indicated with the solid line the recognised true-class objects and with the dashed line the recognised false-class objects.

Police patrol car over approximately the same range. The training set consisted of still images of the Jaguar S-type car out-of-plane rotated over a range 0° to 360° to belong in the true-class, and still images of the RX-7 Mazda Efini Police patrol car out-of-plane rotated over approximately a range 0° to 360° to belong to the false-class. The true-class images were constrained to unit correlation peak-height in the synthesis of the M-HONN system's composite image, and the false-class images were constrained to zero correlation peak-height in the synthesis of the M-HONN system's composite image. We have set the true-class target classification levels (here we assume there is only one *class=1*, so there is no need to set any target connections for a second output neuron) to be $T_{\text{true}}^{\text{class 1}} = +240$, and for false-class the target classification levels were set to be $T_{\text{false}}^{\text{class 1}} = -240$. Here, we have set higher target classification level values for increasing the inter-class discrimination abilities of the M-HONN system. It worth mentioning that we could have set *class=2* and, then, set a target classification level for $T_{\text{true}}^{\text{class 2}}$ with no need to include false-class objects, but adjust the constraint matrix of the system's composite image to class 1 and class 2 different fixed correlation peak-height values. The test set consisted of non-training input still images of Jaguar S-type car objects inserted in plain background at different out-of-plane rotation angles over a range 0° to 360°, and input still images of RX-7 Mazda Efini Police patrol cars inserted in plain background at different out-of-plane rotation angles over a range 0° to 360°. As shown on Fig. 11, M-HONN system was able to successfully recognise the Jaguar S-type car poses over the range 0° to 360° to belong in the true-class, and the RX-7 Mazda Efini car poses over approximately the same range 0° to 360° to belong in the false-class. Again, we have indicated with the solid line the recognised true-class objects and with the dashed line the recognised false-class objects.

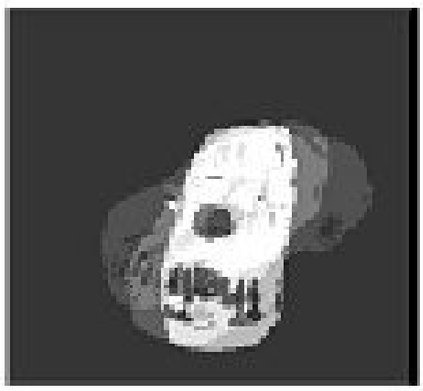


Fig. 10. It shows the composite image the M-HONN system synthesised for a training set consisting of still images of the Jaguar S-type car out-of-plane rotated over a range 0° to 360°.

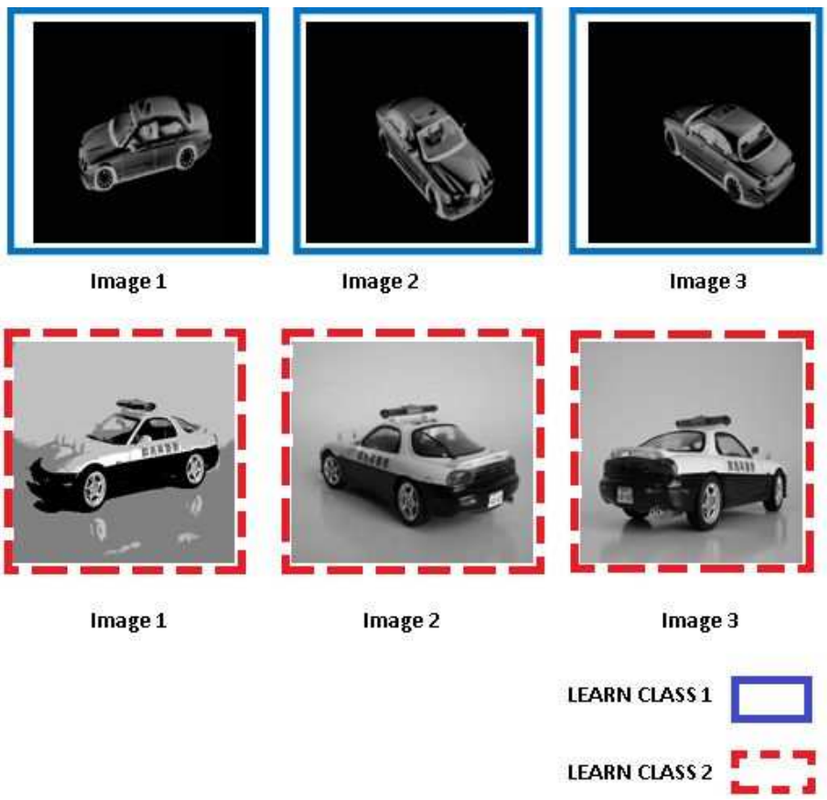


Fig. 11. It shows the second visual problem for testing the M-HONN object recognition system's ability of problem solving. M-HONN system tries to recognise only the true-class objects of the Jaguar S-type car and reject all the false-class objects. The training set consisted of still images of the Jaguar S-type car out-of-plane rotated over a range 0° to 360° to belong in the true-class, and still images of the RX-7 Mazda Efini Police patrol car out-of-plane rotated over approximately a range 0° to 360° to belong in the false-class. We have indicated with the solid line the recognised true-class objects and with the dashed line the recognised false-class objects.

Additional tests we conducted have demonstrated the effect that has the input mask Γ_c in the synthesis of the composite image of the M-HONN system. Thus, in one of the tests we created a training set consisting of Jaguar S-type car objects out-of-plane rotated over a range 20° to 70° at 10° increments to belong in true-class 1, RX-7 Mazda Efini Police patrol car objects out-of-plane rotated over approximately a range 20° to 70°at 10° increments to belong in true-class 2, and a random car park scene to belong in false-class. The true-class 1 images were constrained to unit correlation peak-height, the true-class 2 images were constrained to half-a-unit correlation peak-height, and the false-class images were constrained to zero correlation peak-height in the synthesis of the M-HONN system's composite image. We have set the true-class 1 and true-class 2 target classification levels to be $T_{\text{true}}^{\text{class 1}} = +180$ and $T_{\text{true}}^{\text{class 2}} = +90$, and for false-class 1 and false-class 2 the target classification levels were set to be $T_{\text{false}}^{\text{class 1}} = -10$ and $T_{\text{false}}^{\text{class 2}} = -10$. The test set (see Fig. 12) consisted of input still images of a Jaguar S-type car object and a RX-7 Mazda Efini Police patrol car object both inserted in a car park background scene (not the one we included in the training set), positioned off-the-centre and out-of-plane rotated over a range 20° to 70°. During the process of inserting the objects in to the car park scene some Gaussian noise is added, too. The M-HONN system was able to correctly discriminate between class 1 and class 2. However, in this test the emphasis was to study the effect that knowledge representation in the form of the composite image synthesis has on the problem solving. In effect, as shown in Fig. 13, when we have chosen to build the input mask Γ_c, which we applied it on both the training set and the test set, from the training set image of true-class 1 object of the Jaguar S-type car, then the system synthesised its composite image by non-linearly revealing more features for the true-class 1 object of the Jaguar S-type car, allowing less features for the true-class 2 object of the RX-7 Mazda Efini Police patrol car and completely suppressing any features of the background car park scene. When we have chosen to build the input mask Γ_c from the training set image of true-class 2 object of the RX-7 Mazda Efini Police patrol car, then the system synthesise its composite image (see Fig. 14) by non-linearly revealing more features for the true-class 2 object of the RX-7 Mazda Efini Police patrol car, allowing less features for the true-class 1 object of the Jaguar S-type car and completely suppressing any features of the background car park scene.

Fig. 12. It shows one of the test set input images used for assessing the M-HONN system's performance within clutter

From the above observations and conducted experiments, the M-HONN system, as all the HONN-type systems, combine in their design a knowledge representation unit being the optical correlator block with a knowledge learning unit being the NNET block. Moreover, HONN-type systems, such as M-HONN, have been proven in previous work we have done (Kypraios et al., 2002) to non-linearly combine the weighted, extracted by the NNET block, input training set. In effect, in HONN-type systems the attentional mechanism is provided by the extracted weights of the NNET block to be able to select certain features to be included in its composite image against other ones. Additionally, the M-HONN system, as shown above, can learn and adapt to the input information depending on the created training set itself. Here, the created training set comprises the domain theory of the task to be solved, the initial problem states and the problem goals are given by the true-class and false-class classification levels, and the synthesised composite image provides the control knowledge which guides the decision-making process.

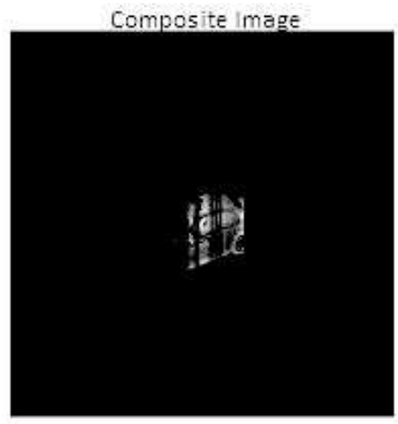


Fig. 13. It shows the synthesised composite image of the M-HONN system. The training set set consisted of Jaguar S-type car objects out-of-plane rotated over a range 20° to 70° at 10° increments to belong in true-class 1, RX-7 Mazda Efini Police patrol car objects out-of-plane rotated over approximately a range 20° to 70°at 10° increments to belong in true-class 2, and a random car park scene to belong in false-class. When we have built the input mask from the training set image of true-class 1 object of the Jaguar S-type car, then the system synthesised its composite image by non-linearly revealing more features for the true-class 1 object of the Jaguar S-type car, allowing less features for the true-class 2 object of the RX-7 Mazda Efini Police patrol car and completely suppressing any features of the background car park scene.

5.4.3 Multiple objects recognition

Here, we summarise several tests we previously conducted for explicitly testing the M-HONN system's ability to recognise multiple objects of different classes (Kypraios et al., 2008). In the first series of conducted tests, the training set consisted of three Jaguar S-type car images out-of-plane rotated at 40° 60° and 80° to belong in class 1, and three Ferrari Testarossa extracted video frames from a recorded video sequence to belong in class 2. For our application purposes it was found to be adequate to set $T_{true}^{class1} = +40$ and $T_{true}^{class2} = +40$ for the true-class target classification levels and $T_{false}^{class1} = -40$ and $T_{false}^{class2} = -40$ for the false-

class classification levels. We constrained true-class 1 of the Jaguar S-type object images to unit correlation peak-height constraint, and true-class 2 of the Ferrari Testarossa car to half-a-unit correlation peak-height constraint in the synthesis of the M-HONN system's composite image. Fig. 15 (a) and Fig. 15 (b) show the output correlation planes response of the M-HONN system for class 1, which are normalised to the overall maximum correlation plane peak-height value for all the input images. Fig. 15 (c) and Fig. 15 (d) show the output correlation plane response of the M-HONN system for class 2, which are normalised to the overall maximum correlation plane peak-height value for all the input images. From the recorded results we have shown that the M-HONN system has accommodated the recognition of class 1 and class 2 objects by output neuron 1 and output neuron 2, respectively.

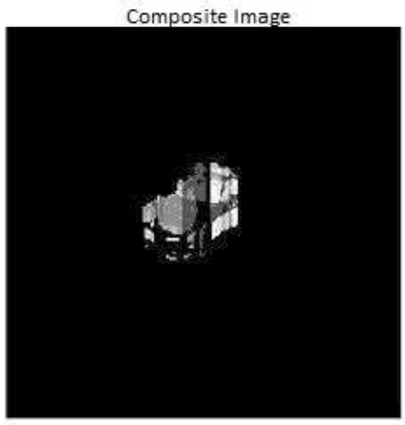


Fig. 14. It shows the synthesised composite image of the M-HONN system. The training set set consisted of Jaguar S-type car objects out-of-plane rotated over a range 20° to 70° at 10° increments to belong in true-class 1, RX-7 Mazda Efini Police patrol car objects out-of-plane rotated over approximately a range 20° to 70°at 10° increments to belong in true-class 2, and a random car park scene to belong in false-class. Now we have built the input mask from the training set image of true-class 2 object of the RX-7 Mazda Efini Police patrol car, then the system synthesised its composite image by non-linearly revealing more features for the true-class 2 object of the RX-7 Mazda Efini Police patrol car, allowing less features for the true-class 1 object of the Jaguar S-type car and completely suppressing any features of the background car park scene.

In the second series of conducted tests, we aimed to assess the ability of the M-HONN system to recognise multiple objects of different classes within a cluttered video sequence. Fig. 16 shows indicatively four of the video frames from the recorded video sequence. The frame rate of the video sequence was 25 frames per second (fps). The training consisted of images of the Jaguar S-type car out-of-plane rotated over 20° to 80° degrees at 20° increments. We added two images of the Jaguar S-type car out-of-plane rotated at 130° and 140° to fall inside the false-class object for increasing the peak sharpness and class discrimination abilities of the M-HONN system. For our conducted tests we found the best values for the true-class 1 and true-class 2 classification values to be $T_{\text{true}}^{\text{class 1}} = +240$ and $T_{\text{true}}^{\text{class 2}} = +240$, respectively, and the best values for the false-class 1 and false-class 2

classification values to be $T_{false}^{class\ 1} = -240$ and $T_{false}^{class\ 2} = -240$. We constrained true-class 1 of the Jaguar S-type object images to unit correlation peak-height constraint, true-class 2 of the Ferrari Testarossa car to half-a-unit correlation peak-height constraint, and false-class 1 and false-class 2 to zero correlation peak-height constraints in the synthesis of the M-HONN system's composite image. Fig. 16 shows the locked window unit of chosen size 70x70 on top of the maximum correlation peak-height values. With the dashed line we have shown the secondary correlation peaks of the output plane and with the solid line we have shown the maximum correlation peak-height value of the output plane. M-HONN system successfully suppressed the unknown background clutter throughout the length of the video sequence and recognised correctly class 1 and class 2 objects. It is emphasised that we have not included any background information in the training set of the system.

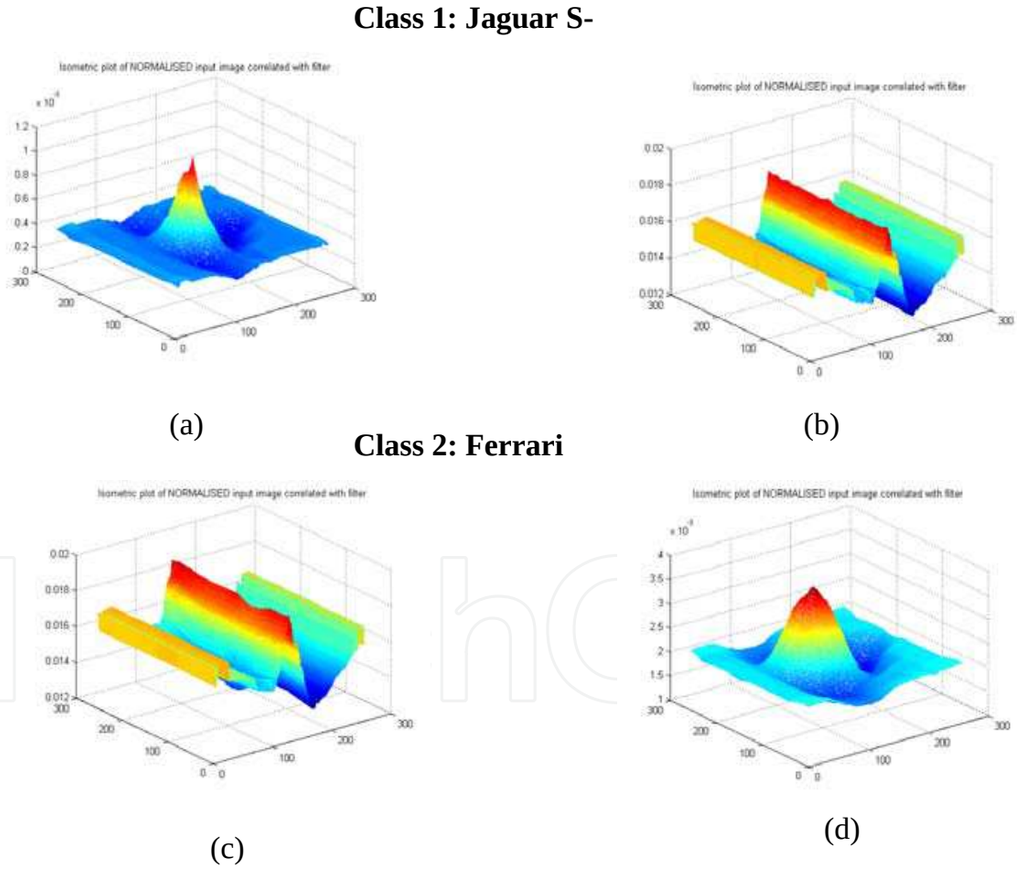


Fig. 15. (Adapted by Kypraios et al., 2009) It shows (a) for the first output layer neuron, and (b) for the second output layer neuron the isometric output correlation plane response of the M-HONN system for Class 1 (normalised to the maximum correlation plane peak-height value), and (c) and (d) the isometric output correlation plane response of the M-HONN system for Class 2 (normalised to the maximum correlation plane peak-height value).

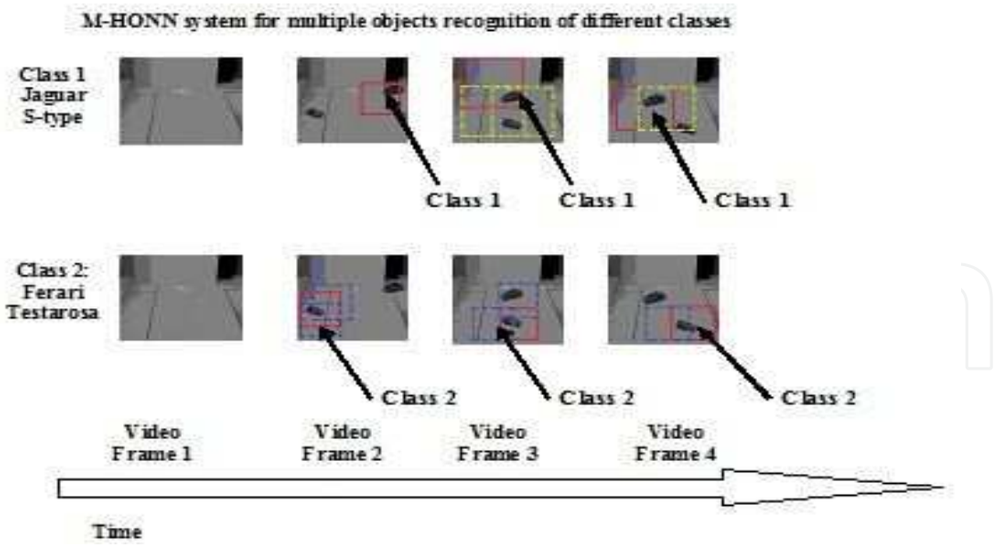


Fig. 16. It shows indicatively four of the video frames from the recorded video sequence. The locked window unit is on top of the maximum correlation peak-height values. With the dashed line we have shown the secondary correlation peaks of the output plane and with the solid line we have shown the maximum correlation peak-height value of the output plane.

6. Conclusion and future work

We have described the design and implementation of the M-HONN system. In particular, we focused in the design and implementation of the M-HONN system for multiple objects recognition of the same and of different classes. The inherited shift invariance properties by the optical correlator block of the system can accommodate for the recognition of multiple objects of the same class. The cross-correlation of each masked test set image with the transformed reference kernel returns an output correlation plane peak value for each cross-correlation step. Thus, the maximum peak height values of the output correlation plane correspond to the recognised true-class objects. By augmenting the output layer of the NNET block of the M-HONN system we can accommodate for the recognition of multiple objects of different classes. In effect, we increase the number of the output layer neurons proportionally with the number of the different object classes. We assign one output neuron to each different class. It was proven experimentally that by choosing different values of the classification levels for the true-class Cl_T and false-class Cl_F objects we can control the M-HONN system's behaviour and it can be varied from more like a high-pass biased filter, which generally gives sharp correlation peaks and good clutter suppression but is more sensitive to intra-class distortions, to more like a MVSDF filter behaviour, which generally gives broader correlation peaks but is more robust to intra-class distortions of the input objects.

We have assessed the performance of the M-HONN system by conducting several series of tests. We assessed the system's ability to detect non-training in-class images that are oriented at the intermediate angle of view between the training images. From the recorded

results, we were able to show the system's ability to interpolate well between the intermediate car poses. The system maintained correlation peak sharpness for the in-class training and non-training images. More specifically, the M-HONN system is able to interpolate non-linearly between the reference and non-reference images to follow the activation function graph. The NNET block is able to generalize between all the reference and non-reference images. Next, we have tested the M-HONN system's distortion range. From the recorded results, we have shown that the system has exhibited a high distortion range recognising all the intermediate car poses of the test set over the range $\hat{\Theta}_3 \in \left[5^\circ, 40^\circ \right]$ (bisector angle). The third series of tests we conducted were for assessing the discrimination ability of the M-HONN system. From the recorded results, we have shown that the system successfully discriminate between objects of different classes while retaining invariance to in-class distortions.

We have analysed the M-HONN system's biologically-inspired hybrid design and we have found to combine a knowledge representation unit being the optical correlator block with a knowledge learning unit being the NNET block, as for the G-HONN type systems. We conducted several experiments for testing the system's problem solving abilities. The M-HONN system was able to solve the visual task of recognising certain Jaguar S-type car poses to belong in the true-class from other Jaguar S-type car poses. Also, the M-HONN system was able to solve the visual task of recognising only the true-class objects of the Jaguar S-type car.

The last series of tests aimed to assess the M-HONN system's performance of recognising multiple objects of different classes within clutter. We have tested the system with a recorded video sequence. The system successfully suppressed the unknown background clutter during the whole length of the video sequence and recognised correctly class 1 and class 2 objects. In overall, the M-HONN system was able to correctly recognise true-class objects out-of-plane rotated, translated off-the-centre and inserted into background scenes. It is emphasised that the system was able to recognise the true-class objects within an unknown background clutter scene since we have not included any background information in its training set. Additionally, all the invariance properties were simultaneously exhibited by the M-HONN system with a single pass over the input data sets. In effect, as we could see from its transfer function, M-HONN system is not either a multiple stages-type of filter or any pre-processing of the input data is required for maintaining its invariance properties. There is no need for a separate background segmentation pre-processing stage prior the system's object tracking as in the case of other motion based segmentation and object tracking techniques. Instead, the M-HONN system is able to successfully suppress the background clutter and track throughout the video sequence the recognised true-class object.

In future, we would like to assess the performance of each output neuron of the M-HONN system's NNET block individually and record separately their performance metrics values for the detectability, distortion range, and discrimination ability. Also, we believe that the M-HONN system's design can be extended to accommodate three-dimensional (3D) object recognition. Similarly to stereo vision systems (Lowe, 1987; Xu & Zhang, 1996; Sumi et al., 2002), the M-HONN system's design can be extended with a second input mask for

processing different angles of the captured data, and incorporating the corresponding transformed images into its composite image synthesis.

7. References

Aler, R., Borrajo, D. & Isasi, P. (2000). Knowledge representation issues in control knowledge learning, *Proceedings of 25th International Conference on Machine Learning*, Morgan Kaufmann, pp. 1-8, ISBN 1558607072

Bahri, Z. & Kumar, B. V. K. (1988). Generalized Synthetic Discriminant Functions, *Journal of Optical Society of America,* Vol.5, No.4, pp. 562-571

Beale, R. & Jackson, T. (1990). *Neural Computing : An Introduction*, Institute of Physics Publishing, Hilger, ISBN 0852742622, 9780852742624, Bristol, Philadelphia

Bottou, L., Fogelman- Soulié, F., Blanchet, P. & Lienard, J. S. (1990). Speaker Independent Isolated Digit Recognition : Multilayer Perceptrons vs Dynamic Time Warping, *Neural Networks,* Vol.3, pp. 453-465

Casasent, D. (1984). Unified synthetic discriminant function computational formulation, *Applied Optics,* Vol.23, No.10, 1620-1627

Casasent, D., Neiberg, L. M. & Sipe, M. A. (1998). Feature Space Trajectory Distorted Object Representation for Classification and Pose Estimation, *Optical Engineering,* Vol.37, No.3, pp. 914-923

Caulfield, H. J. & Maloney, W. T. (1969). Improved discrimination in optical character recognition, *Applied Optics,* Vol.08, No.11, 2354-2356

Caulfield, H. J. (1980). Linear combinations of filters for character recognition: a unified treatment, *Applied Optics,* Vol.19, No.23, 3877-3878

Delopoulos, A., Tirakis, A. & Kollias, S. (1994). Invariant Image Classification Using Triple-Correlation-Based Neural Networks, *IEEE Transactions on Neural Networks*, Vol.5, No.3, pp. 392-408

Dobnikar, A., Ficzko J., Podbregar, D. & Rezar, U. (1991/92). Invariant pattern classification neural network versus FT approach, *Microprocessing and Microprogramming,* Vol. 33, pp. 161-168

Giles, C. L. & Maxwell, T. (1987). Learning, Invariance and Generalisation in Higher-Order Neural Networks, *Applied Optics,* Vol.26, pp. 4972-4978

Hagan, M. T., Demuth, H. B. & Beale, M. H. (1996). *Neural Network Design,* PWS Publishing, ISBN 0-9717321-0-8, Boston, MA

Haykin, S. (1999). *Neural Networks-A Comprehensive Foundations,* 2nd Edition, Prentice Hall International, Inc.

Jamal-Aldin, L. S., Young, R. C. D. & Chatwin, C. R. (1997). Application of non-linearity to wavelet-transformed images to improve correlation filter performance, *Applied Optics*, Vol. 36, No. 35, pp. 9212-9224

Jamal-Aldin, L. S., Young, R. C. D. & Chatwin, C. R. (1998). Synthetic discriminant function filter employing nonlinear space-domain preprocessing on bandpass-filtered images, *Applied Optics*, Vol. 37, No. 11, pp. 2051-2062

Kanaoka, T., Chellapa, R., Yoshitaka, M. & Tomita, S. (1992). A Higher-Order Neural Network for Distortion Invariant Pattern Recognition, *Pattern Recognition Letters,* Vol.13, pp. 837-841

Khotanzad, A. & Hong, H. (1990). Invariant Image Recognition by Zernike Moments, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 12, pp. 489-497

Kumar, B. V. K. & Hassebrook, L. (1990) Performance measures for correlation filters, *Applied Optics*, Vol. 29, No. 20, pp. 2997-3006

Kumar, B. V. K. (1986). Minimum Variance Synthetic Discriminant Functions, *Journal Optics Society America A,* Vol.3, pp. 1579-1584

Kumar, B. V. K. (1992). Tutorial Survey of Composite Filter Designs for Optical Correlators, *Applied Optics,* Vol.31, No.23, 4773-4801

Kypraios, I. (2009). A Comparative Analysis of the Hybrid Optical Neural Network-type Filters Performance within Cluttered Scenes, *51st International Symposium ELMAR, IEEE Region 8/IEEE Croatia/EURASIP,* Vol.1, pp. 71-77

Kypraios, I. (2010). *Hybrid optical neural network-type filters for multiple objects recognition within cluttered scenes,* Object Recognition, InTech, ISBN 978-953-307-222-7

Kypraios, I. (February 2010). A Digital/Optical Neuronal Model for the Cognitive Interaction Between Retina Sensor and the Human Brain Visual Cortex, *Poster Presentation, INCF UK Node Congress Analysing and Modelling the Neural Systems in Health and Disease, Informatics Forum,* Edinburgh

Kypraios, I., Lei, P. W., Birch, P. M., Young, R. C. D. & Chatwin C. R. (2008). Performance Assessment of the Modified-Hybrid Optical Neural Network Filter, *Applied Optics,* Vol.47, No.18, pp. 3378-3389

Kypraios, I., Young, R. C. D. & Chatwin, C. R. (2004b). Performance assessment of Unconstrained Hybrid Optical Neural Network filter for Object Recognition Tasks in Clutter, *Optical Pattern Recognition XV, Proceedings of SPIE,* Vol.5437, pp. 51-62

Kypraios, I., Young, R. C. D. & Chatwin, C. R. (2009). Modified-Hybrid Optical Neural Network Filter for Multiple Objects Recognition within Cluttered Scenes, *Optics and Photonics for Information Processing III, Proceedings SPIE,* Vol.7442, pp. 74420P-74420P-12

Kypraios, I., Young, R. C. D., Birch, P. M. & Chatwin, C. R. (2004a). Object Recognition Within Cluttered Scenes Employing a Hybrid Optical Neural Network Filter, *Optical Engineering Special Issue on Trends in Pattern Recognition,* Vol.43, pp. 1839-1850

Kypraios, I., Young, R. C. D., Birch, P. M. & Chatwin, C. R. (2003). A non-linear training set superposition filter derived by neural network training methods for implementation in a shift invariant optical correlator, *Proceedings SPIE Defense & Security, Optical Pattern Recognition XIV,* Vol. 5106, pp. 84-95, Orlando, Florida, USA

Kypraios, I., Young, R. C. D., Chatwin C. R. (2002). An Investigation of the Non-Linear Properties of Correlation Filter Synthesis and Neural Network Design, *Asian Journal of Physics,* Vol.11, No.3, pp. 313-344

LeCun, Y. (1989). Generalisation and Network Design Strategies, *Connectionism in Perspective,* Pfeirer, R., Schreter, Z., Fogelman-Soulié, F. & Steels, L., Elsevier Science, Amsterdam

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., Jackel, L. D. (1990). Handwritten Digit Recognition with a Backpropagation Network, *Advances in Neural Information Processing Systems,* Touretzky, D., Morgan Kaufmann, Vol.2, pp. 396-404

Lee, I. & Portier, B. (11-13 July 2007). An empirical study of knowledge representation and learning within conceptual spaces for intelligent agents, *6th IEEE/ACIS International Conference Computer and Information Science*, pp. 463-468, Melbourne, Australia

Looney, C. G. (1997). *Pattern Recognition Using Neural Networks-Theory and Algorithms for Engineers and Scientists,* Oxford University Press, New York-Oxford

Lowe, D. G. (1987). Three-dimensional object recognition from single two dimensional images, *Artificial Intelligence, Elsevier,* Vol. 31, No. 3, pp. 355-395

Lynn, P. A. & Fuerst, W. (1998). *Introductory Digital Signal Processing- with Computer Applications,* John Wiley & Sons Ltd.

Mahalanobis, A., Kumar, B. V. K., Song, S., Sims, S. R. F. & Epperson, J. F. (1994). Unconstrained Correlation Filters, *Applied Optics,* Vol.33, No.17, pp. 3751-3759

Nguyen, D. & Widrow, B. (1989). The Truck Backer-Upper: An Example of Self-Learning in Neural Networks, *Proceedings of the IEEE International Joint Conference on Neural Networks,* Vol.2, pp. 357-363

Nguyen, D. & Widrow, B. (1990). Improving the Learning Speed of 2-Layer Neural Networks by Choosing Initial Values of the Adaptive Weights, *Proceedings of the IEEE International Joint Conference on Neural Networks,* Vol.3, pp. 21-26

Perantonis, S. & Lisboa, P. (1992). Translation, Rotation and Scale Invariant Pattern Recognition by High-Order Neural Networks and Moment Classifiers, *IEEE Transactions on Neural Networks,* Vol.3, No.2, pp. 241-251

Proakis, J. G. & Manolakis, D. G. (1988). *Introduction to Digital Signal Processing,* Prentice Hall International Paperback Editions

Refregier, Ph. (1990). Filter design for optical pattern recognition: multicriteria optimisation approach, *Optics Letters*, Vol. 15, No. 15, pp. 854- 856

Refregier, Ph. (1991). Optimal trade-off filters for noise robustness, sharpness of the correlation peak and Horner efficiency, *Optics Letters,* Vol. 16, No. 11, pp. 829-831

Shvedov, A., Schmidt, A. & Yakubovich, V. (1979). Invariant Systems of Features in Pattern Recognition, *Automation Remote Control,* Vol. 40, pp. 131-142

Simard, P. & LeCun, Y. (1992). Reverse TDNN : An Architecture for Trajectory Generation, *Advances in Neural Information Processing Systems,* Moody, J., Hanson, S. & Lipmann, R., Morgan Kauffmann, Vol.4, pp. 579-588

Spirkovska, L. & Reid, M. (1992). Robust Position, Scale and Rotation Invariant Object Recognition Using Higher-Order Neural Networks, *Pattern Recognition,* Vol.25, pp. 975-985

Stamos, E. (2001). *Algorithms for Designing Filters for Optical Pattern Recognition,* D.Phil. Thesis, Department of Electronic and Electrical Engineering, University College London

Sumi, Y., Kawai, Y., Yoshimi, T. & Tomita, F. (2002). 3D Object recognition in cluttered environments by segment-based stereo vision, *International Journal of Computer Vision, Springer,* Vol. 46, No. 1, pp. 5-23

Talukder, A. & Casasent, D. (1999). Non-Linear Features for Product Inspection, *Optical Pattern Recognition X, Proceedings of SPIE,* Vol.3715, pp. 32-43

The Mathworks (August 2008). Neural Network Processing Toolbox 13: User's Guide for version Matlab 6.5, Available from http://www.mathworks.com

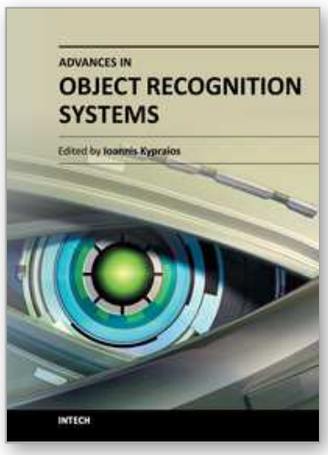
Vander Lugt, A. (1964). Signal Detection By Complex Spatial Filtering, *IEEE Transactions on Information Theory,* Vol.10, pp. 139-145

Waibel, A., Hanazawa, T., Hinton, G , Shikano, K. & Lang, K. (1989). Phoneme Recognition Using Time-Delay Neural Networks, *IEEE Transactions on Acoustics, Speech Signal Processing,* Vol.37, No.3, pp. 328-339

Wood, J. (1996). Invariant Pattern Recognition: A Review. *Pattern Recognition,* Vol.29, No.1, pp. 1-17

Xu, G. & Zhang, Z. (1996). *Epipolar Geometry in Stereo, Motion, and Object Recognition: A Unified Approach,* Springer, Kluwer Academic Publishers, ISBN 0-7923-4199-6, The Netherlands

Yuceer, C. & Oflazer, K. (1993). A Rotation, Scaling and Translation Invariant Pattern Classification System, *Pattern Recognition,* Vol.26, No.5, pp. 687-710



Advances in Object Recognition Systems

Edited by Dr. Ioannis Kypraios

ISBN 978-953-51-0598-5
Hard cover, 170 pages
Publisher InTech
Published online 09, May, 2012
Published in print edition May, 2012

An invariant object recognition system needs to be able to recognise the object under any usual a priori defined distortions such as translation, scaling and in-plane and out-of-plane rotation. Ideally, the system should be able to recognise (detect and classify) any complex scene of objects even within background clutter noise. In this book, we present recent advances towards achieving fully-robust object recognition. The relation and importance of object recognition in the cognitive processes of humans and animals is described as well as how human- and animal-like cognitive processes can be used for the design of biologically-inspired object recognition systems. Colour processing is discussed in the development of fully-robust object recognition systems. Examples of two main categories of object recognition systems, the optical correlators and pure artificial neural network architectures, are given. Finally, two examples of object recognition's applications are described in details. With the recent technological advancements object recognition becomes widely popular with existing applications in medicine for the study of human learning and memory, space science and remote sensing for image analysis, mobile computing and augmented reality, semiconductors industry, robotics and autonomous mobile navigation, public safety and urban management solutions and many more others. This book is a "must-read" for everyone with a core or wider interest in this "hot" area of cutting-edge research.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Ioannis Kypraios (2012). Performance Analysis of the Modified-Hybrid Optical Neural Network Object Recognition System Within Cluttered Scenes, Advances in Object Recognition Systems, Dr. Ioannis Kypraios (Ed.), ISBN: 978-953-51-0598-5, InTech, Available from: http://www.intechopen.com/books/advances-in-object-recognition-systems/performance-analysis-of-the-modified-hybrid-optical-neural-network-filter-for-objects-recognition-wi

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

www.intechopen.com

IntechOpen

IntechOpen