International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:11, 2008

A Unified Robust Algorithm for Detection of Human and Non-human Object in Intelligent Safety Application

M A Hannan, A. Hussain, Member, IEEE, S. A. Samad, Member, IEEE, K. A. Ishak and A. Mohamed, Senior Member, IEEE

***Abstract*—**This paper presents a general trainable framework for fast and robust upright human face and non-human object detection and verification in static images. To enhance the performance of the detection process, the technique we develop is based on the combination of fast neural network (FNN) and classical neural network (CNN). In FNN, a useful correlation is exploited to sustain high level of detection accuracy between input image and the weight of the hidden neurons. This is to enable the use of Fourier transform that significantly speed up the time detection. The combination of CNN is responsible to verify the face region. A bootstrap algorithm is used to collect non human object, which adds the false detection to the training process of the human and non-human object. Experimental results on test images with both simple and complex background demonstrate that the proposed method has obtained high detection rate and low false positive rate in detecting both human face and non-human object.

***Keywords*—**Algorithm, detection of human and non-human object, FNN, CNN, Image training.

I. Introduction

With improvement in vehicle safety, security and comfort, a significant progress has been made in the area of detection system for the application of intelligent vehicle. The detection of occupant, non-human object and empty are the fundamental importance research for intelligent vehicle safety system. The general problem of detection and distinguishing a particular class of static objects from all others is a difficult task. However, most of the previous works are limited to either face only or object detection individually. Over the past twenty years, numerous face and object detection systems have been published in the computer vision community. Despite the success of some of these systems in constrained scenarios, the general task of face and object detection still poses a number of challenges with respect to changes in illumination, image scale, image quality, expression and pose.

Many approaches have been proposed for face and object detection in still images that are based on texture, depth, shape and color information, or a combination of them. A comprehensive survey of dynamic scenes for faces and object detection methods are described as follows.

Researches based on object detection approach can be categorized as exemplar and non-exemplar. In non-exemplar based approach, each object of interest generally requires different modeling assumptions [1]. Exemplar-based approach on the other hand avoids making assumptions about the objects of interest in the training set under various conditions and illumination [2]. An automated object detection procedure is developed in [3], to extract samples of the object class containing highest features information. However, the computational demands are high and as a result a portion of manual intervention is needed to keep the computational costs reasonable [4]. Weber *et al.* 2002 described an automated selection technique, to collect distinctive parts using probabilistic model [5]. However, the probabilistic model relies on a small number potentially sensitive to large variations across images. On the object class of interest, Roth *et al.* 2002 and Agarwal *et al.*2004 used feature-efficient learning algorithm and discriminative classifier to learn robust expressive model based on pixel and sparse representation [6, 7].

The approaches applied to face detection are categorized as feature based, view based and shape model based [8]. The basis of feature based approach is the knowledge of human faces, extracted facial features and verification of the potential faces. However, the false detection rate for this approach is large. Alternatively, the view based approach is based on statistical model. It has high verification rate, but the approach is extremely slow due to exhaustive search over the whole image [2]. Colmenarez and Huang presented a hierarchical knowledge-based and information-based maximum discrimination system for face detection in complex backgrounds [9]. Moghaddam and Pentland reported a face detection system based on maximum likelihood on feature vector eigenspace decomposition [10]. However, principal component analysis does not maximize

Manuscript received May 8, 2007. This work is supported by the Malaysian Ministry of Science, Technology and Innovation under the IRPA grant scheme 03-02-02-0017-SR0003/07-03.

M A Hannan is with the University Kebangsaan Malaysia, Malaysia, 43600, Bangi, Selangor, Malaysia, phone:+6-03-89216035;fax:+6-03-89216035;(e-mail: eehannan@vlsi.eng.ukm.my).

A. Hussain is with the University Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysia, (e-mail: aini@vlsi.eng.ukm.my)

S. A. Samad is with the University Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysia. (e-mail: salina@vlsi.eng.ukm.my).

K. A. Ishak is with the University Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysia. (e-mail: ishak@vlsi.eng.ukm.my)

A. Mohamed is with the University Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysia. (e-mail: azah@vlsi.eng.ukm.my).

discrimination and the computation requirement is high on projected eigenspace. Many face detection researcher have reported face detection systems based on neural networks [11-12]. Although general and complex network architectures allow the discriminant function to take advantage of most of the underlying joint distribution of the training patterns, the structure itself is not optimized for that purpose. El-Bakry introduced simple design of modular neural network (MNN) classifier to reduce computational complexity over non-modular alternatives [13]. Huang, et al. proposed polynomial neural network (PNN) method for face detection from cluster image using PCA to achieve higher detection and low positive false rate [14].

On-road vehicle face detection is a difficult task due to variability in scale, location, orientation, pose, race, facial expression, and occlusion. Rowley et al. 1998 [11] proposed a NN-based face detection method using pre-processed image to train a multilayer NN to learn the face and non face examples. Sung and Poggio developed a system for face and non face detection based on distribution and multilayer NN [12]. Osuna et al. 1997 presented a support vector machine (SVM) based approach for frontal view face detection, ensemble of feed-forward neural networks trained by the back-propagation algorithm [15]. Papageorgiou et al. 2000 also proposed SVM to detect faces using wavelets for feature extraction and classification [16]. Viola and Jones [17] is developed AdaBoost learning algorithm for very fast face detection using simple features.

In this paper, we intend to focus on the detection of human face and non-human object in a frame that could be used in the vehicle for application such as driver assistance systems and airbag deployment decision. The detection of both classes of object in one frame is challenging than that of individual detection. To meet the challenge, detection of human face and non-human object is done by FNN with correlation between input image and hidden units. The post-processing step is done to reduce the false generation from FNN. CNN performs a verification procedure to provide the decision based on human face and non-human object using post-processing dataset which involved lighting correction and histogram equalization of linear function.

II. System Structure and Algorithm Description

This section explains the overall system structure and the algorithm for detection and classification of occupant, non-human object and non-object as shown in Fig. 1. The system detection employed two different hierarchical classification architecture of neural networks namely fast neural network (FNN) and classical neural network (CNN). The proposed system is a combination of FNN and CNN, in which the FNN extracts any positive detection including false detection. The output of FNN is then fed to the CNN to verify which region is indeed the system detection. This proposed combined network is quite robust in terms of its detection accuracy and computational efficiency when compared to a single network, which is unable to fully eliminate false detection problem.

In the proposed system architecture as shown in Fig. 1, the FNN first extracts a sub-image from the test image to detect object and false detection. Next, post-processing strategies are applied to convert normalized outputs back into the same units that were used for the original targets using 2D-multiple detection, 3D-multiple detection and elimination of overlapping detection [20]**.** There is some assumption that FNN may introduce some false detection due to variation in the lighting conditions, for example lighting from the side of the object, which changes its overall appearance. To solve this problem, a process of adjusting intensity values is done automatically using histogram equalization or lighting correction function. Then we take one step further using classical neural network as an object verification procedure in order to reduce the number of false detection.

At this time, lighting normalization is performed by mapping the features to some fixed locations in an N x M image. The mapping is assumed to be an affine transformation, computed iteratively as in [11]. Each normalized image is then reprocessed to account for different lighting conditions and contrast using linear fit function to the intensity of the image.

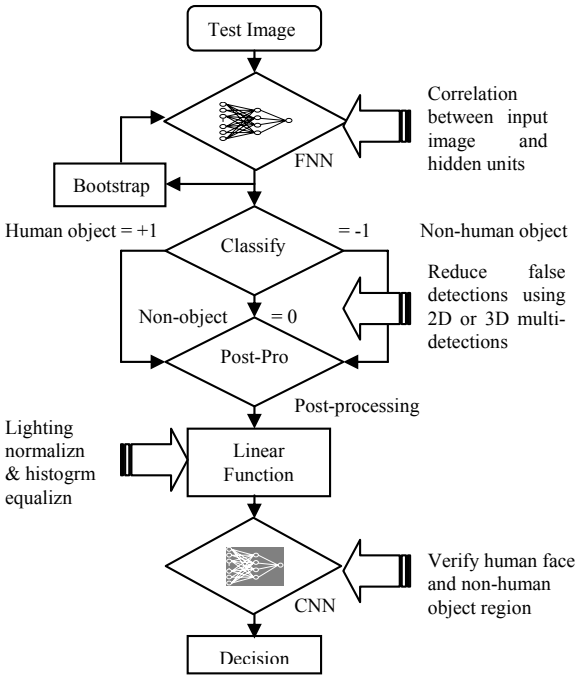

Fig. 1 NN algorithm for the occupancy detection

The result is subtracted out from the original image to correct lighting differences. Then, histogram equalization is performed to correct for different camera gains and to improve contrast [18]. Fig. 2 shows some examples method using linear fit function and histogram equalization on the object regions before it can be fed into the second detector. The 1st row is original, the 2nd row is by linear fit, while the 3rd row is lighting corrected and the 4th row is the histogram

equalization image. In general there are two advantages using this approach: a) detection task is speed-up by using fast neural network method, b) to accommodate the variation in illumination, the lighting normalization is adopted into the system.

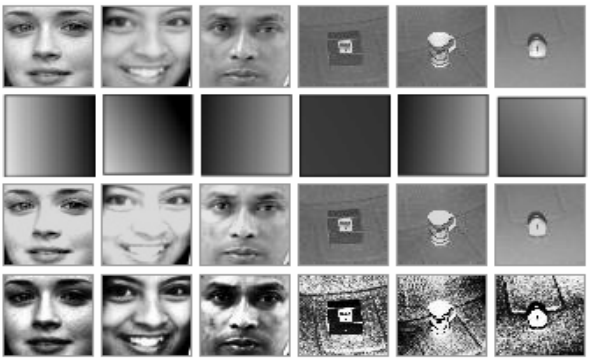

Fig. 2 Linear fit function and histogram equalization

III. Training Network Principles

The key idea to train neural network is based on network properties as shown in Table 1. The first step of the training procedure is to collect the training data, which are human indicating by face, non-human object and no objects images.

TABLE 1
NEURAL NETWORK PROPERTIES

Network Features	**FNN**	**CNN**
Number of Input Unit	30	25
Number of Hidden Unit	15	10
Activation Function	Sigmoid	Sigmoid
Pre-processing Type	-	Lighting Normalization
Number of Bootstrap Iteration	3	3
Trained Image Size	25x25	25x25

For that, it is easy to get a representative sample of images which contain human faces and non-human object but much more difficult to get a representative sample of non-object images as shown in Fig. 3. The collected training data is then labeled to fit to training system. A number of techniques were applied to training neural network. Among them back-propagation algorithm is used for successful design of multilayer feed forward networks. Each image produce three images, whose must have some invariance to position, rotation and scale like randomly mirrored, rotated up to 5^0 from 0 axes. This operation is done by a best-fit of lighting correction and histogram equalization.

Fig. 3 Representative training sample images

The histogram equalization of a sub-image is defined by,

$$H_I^{(\alpha)}(z) = \frac{1}{I}\sum_{v}\delta(z - I^{(\alpha)}(v)) \quad (1)$$

where $H^{(\alpha)}_I(z)$ is the histogram equalization of I sub-image with α normalizing constant. The input vector of I sub-image is z, δ is the coefficient of I and $I^{(\alpha)}(v)$ is the normal sub-band image through linear convolution.

A bootstrap algorithm is used to add non-object images to the training database to improve the neural network performance during training period and automatically clipping false detection to inserting these into current training set. In bootstrap algorithm, weights are adjusted according to their classification errors. The equation used in the bootstrap training procedure for adjusting weights is achieved by,

$$w_j = c\exp\left[-y_j\sum_{t=1}^{i}\phi_t(x_j)\right] \quad (2)$$

where w_j is weighting of sample x_j, y_j is the label of sample x_j, c is the initial weight for negative samples, i is the current node index and Φ_t is the ith boosting classifier in cascade.

A. Training Human Images

In order to classify human as face, non-human object and non-object, we need training examples for each set. The positive training images indicated as human faces are collected from various sources. A face cropping program was employed to manually label the position of the eyes and center of the mouth so that the face images are aligned and both scale and position invariant. All face images are scaled to a uniform size of 25x25 pixels. A set of 7344 face images generated from 1836 face samples was collected by randomly scaling down the input image to 1.2 for each step in the pyramid, translating up to half a pixel and rotating by random variation from -5^o to 5^o.

Fig. 4 show the sum square error (SSE) of human face FNN and CNN training in epochs. Initially, it can be seen that the error recovering is faster, however, it is slower later on. The individual FNN and CNN training of human face leads to a similar adaptation on the training data as suggested by the asymptotic convergences of both curves.

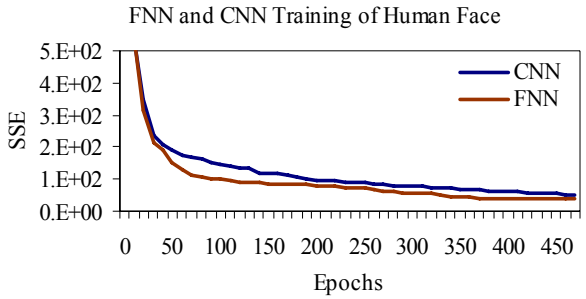

Fig. 4 SSE of FNN and CNN of human face training

B. Training Non-human Object Images

Non-human object images refer to images without human faces but instead contain non-human objects such as grocery bag or others. The negative training images that represent the non-human object images are collected from our own databases. The negative training images are also cropped, aligned and scaled similar to the positive training image data set. The non-human objects image data set contains various objects such as computer mouse, porcelain cup etc. A set of 6000 non-human object images are generated from various non-human objects samples.

Fig. 5 shows the SSE of non- human object FNN and CNN training in epochs. Again, the results agree with previous training of CNN and FNN using the face image dataset.

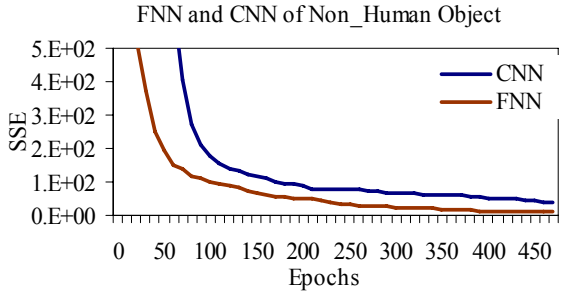

Fig. 5 SSE of FNN and CNN of non-human object

C. Training Non-objects Images

It was mentioned that non-object images are more complex since the definition of a non-object is much broader and richer. The images containing neither a human face nor specific objects are used to train the non-object class. We adopted the bootstrapping step, which was successfully used by Sung and Poggio [12]. In this work, some 8000 non-object images in our collection were randomly cropped from the original non-object images. Then, we trained the network with the initial set of human face, non-human object and non-object images to produce a neural network output of 1, -1 and 0, respectively. After training, the misclassified sub-images with output 0 are collected. Next, we selected those misclassified images randomly to add to the training set of the non-object class. The training iteration is continued adding misclassified sub-images to the non-object examples until the training error is large enough. The trained model is then obtained and used to verify the human face and non-human object candidates.

IV. MLP and FFT Algorithm for Detection

The human and non-human object detection algorithm based on two dimensional cross correlations between test image and 25x25 sliding window is adopted as in [20]. This proposed detection system uses the multilayer perceptron (MLP) and Fast Fourier transformation. The sliding window is represented by the neural network weights situated between the input unit and hidden layer. In the convolution theorem, convolution of x with y is identical to the result of Fourier transformation X and Y in the frequency domain. Therefore, multiplying X and Y in the frequency domain point by point and then the cross-product is transformed back into spatial domain via the inverse Fourier transform yields the same results. As cross correlation is in frequency domain, detection process can be speed up.

During detection, a sliding sub image I of size mxn is extracted from the tested image of SxT and fed to the neural network. Fig. 6 shows the MLP neural network structure used for the detection of human, non-human object and non-object.

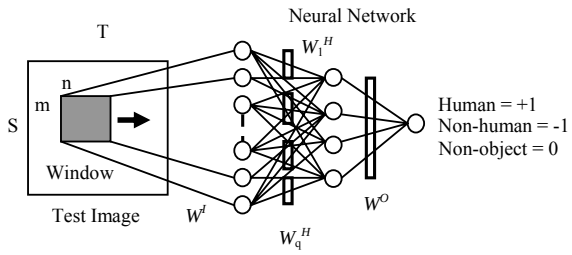

Fig. 6 The MLP neural network structure used

Let w_i be the vector of weights between the input sub image and the hidden layer. This vector has a size of mn, can be represented as mxn matrix. The output of hidden neurons h_i can be calculated in a $2D$ space as follows:

$$h_i = g\left[\sum_{j=1}^{m}\sum_{k=1}^{n} w_i(x,y)I(x,y) + b_i\right] \qquad (3)$$

where, g is the sigmoid function representing as, $g(x) = 1/(1+e^{-x})$ for neural network output 1, 0 and -1 and b_i is the bias of each hidden neuron ith. Equation (3) represents the output of each hidden neuron for a particular sub-image I. It can be computed for the whole image, z as follows:

$$h_i(u,v) = g\left[\sum_{j=-m/2}^{m/2}\sum_{k=-n/2}^{n/2} w_i(x,y)z(v+x, u+y) + b_i\right] \qquad (4)$$

Equation (4) represents a cross correlation operation. Given any two functions $f(x, y)$ and $d(x, y)$, their cross correlation can be obtained by:

$$f(x,y)\otimes d(x,y)=\sum_{m=-\infty}^{m=\infty}\sum_{n=-\infty}^{n=\infty} f(m,n)d(x+m,y+n) \quad (5)$$

Thus, we can write equation (4) as follows:

$$h_i(u,v) = g\left[w_i \otimes z + b_i\right] \quad (6)$$

Where, h_i is the output of the hidden neuron *ith* and h_i *(u,v)* is the activity of the *ith* hidden unit for whole test image. Now the above cross correlation can be expressed in terms of the Fourier Transform as follows,

$$z \otimes w_i = F^{-1}(F(z) \bullet \overline{F}(w_i)) \quad (7)$$

Where, F bar is the conjugated complex of Fourier transform. Hence, by evaluating this cross correlation, a speed up ratio can be obtained comparable to conventional neural networks. Also, the final output can be evaluated by substituting equation (4) into 3-layer feed-forward neural network or multilayer perception (MLP) as follows,

$$O(u,v)=g(W_j^O\sum_{j=1} f(W_j^H\sum_{i=1} f(z\otimes w_i^I+b_i^I)+b_j^H)+b^O))) \quad (8)$$

Where *O* is the final output and W^O, W^H, W^I are refers to the weights of output, hidden and input layer.

The final output is a scalar but a matrix of dimension (S-m+1) by (T-n+1), where S and T are respectively the rows and columns of the input image and m and n are respectively the rows and columns of sliding sub image. In Fig. 6, given an input image the neural network performs the function of cross correlation operation to detect and localize human face and non-human object. The output of the detector is a neural network matrix of values +1, -1 and 0, representing human, non-human object and non-object location, respectively.

V. Multi-Scale Detection

The FNN described above is trained to detect and locate human faces and non-human object images from still images. As the neural network is trained on a 25x25 pixels, it would detect human faces and non-human object of only this size. However, the size of the human face and non-human object in real situation are usually larger than 25x25 pixels. Thus we have scaled-down the input image by a factor of 1.2 for each step in the pyramid. Scanning an input image at different resolutions allows human face and non-human object detection by sub-sampling the whole test image at several scales before feeding to the neural network. The sub-sampling section forms an integral part of the multi-scale detector. Each input image is being processed in neural networks at various resolutions. The output is then fed into an arbitrator structure which decides at which resolution an object has been detected. During the computation of the cross correlation, the sub-sampling can be entirely performed using the following scaling property of the Fourier Transform as,

$$f(ax,by) \Leftrightarrow \frac{1}{|ab|}F\left(\frac{u}{a},\frac{v}{b}\right) \quad (9)$$

Where, f(ax,by) is the original image and $F\left(\frac{u}{a},\frac{v}{b}\right)$ is its Fourier transform with *a* and *b* are the scaling factor. For an example, we chose a = b = 2 in order to get an image reduced to half of its original size. The main cross correlation in (6) can be modified to accommodate multiple scales resolution in the detector as follows,

$$z|_{a,b} \otimes w_i = F^{-1}\left[\frac{1}{|ab|}F\left(z\left(\frac{u}{a},\frac{v}{b}\right)\bullet \overline{F}(w_i)\right)\right] \quad (10)$$

VI. Verification of Object Region

In the previous section, we have discussed a fast multi-scale face detection scheme by calculating the Fourier transform of the image and of the neural network filter, and then process the image in the Fourier space. By using this single network, we observe that this approach always suffer from uncontrolled lighting conditions that will generate more false alarms. Thus, we have used a classical neural network to select a real face region and reject the false detection. Our CNN is trained using the pre-processed dataset with an image of 25x25 pixels as its input. Detail architecture for CNN is described in Table 1. The verification procedure is carried out step by step as follows:

1. The extracted possible face regions in all level of scaling are sub-sampled and interpolated to a resolution of 25x25.

2. To reduce variability due to lighting and camera characteristics, we perform a simple lighting normalization approach. The first technique tries to correct the intensity values of the extracted candidate regions by subtracting with best-fit linear function. Then, histogram equalization is performed to enhance the contrast in the image.

3. To verify the face regions, the pre-processed data is then mapped into CNN and an output will be produced. Any output below a threshold will be rejected; otherwise the face regions will be mapped into the original image.

VII. Experimental Results

The system uses 2-data sets of images in the experiment to test the detection performance of human face and non-human object, which are distinct from the training sets. The first set has 253 test images of wide variety of complex background in various environment and varying scale with some occlusions and variations in lighting. 25 human face image of interest were taken from the 253 test image. The second data set contains 112 test images that have been collected from 7 non-human object of interest. The systems undergoes the bootstrapping cycle with ending up between 4500 to 9500 zero samples, to evaluate the performance of true detection of the test images and the rate of false detection. The zero samples do not contain any human face or non-human object. The algorithm is tested optimal implementation of the proposed approach requires minimum

specifications of Intel Pentium 4, 3.0 GHz with 512 MB RAM.

To review a complete characterization of the detection scheme, we generate receiver operating characteristic (ROC) curve that illustrate the accurate detection rate versus false positive rate tradeoffs, rather than providing a single performance result. This is accomplished by varying the detection threshold in the neural network. Fig. 7 shows ROC curve of the human face and non-human object detection system using FNN and (FNN + CNN) methods, which are measured on a logarithmic scale. It can be seen that the performance of the (FNN+CNN) method corresponds to a 90% detection rate at a false detects of 0.35% and 0.4% of human face and non-human object, respectively compared to the FNN, at a false detects rate of 1.1% and 1.25%. The ROC curve also shows that the higher penalties for miss positive examples may result in better performance.

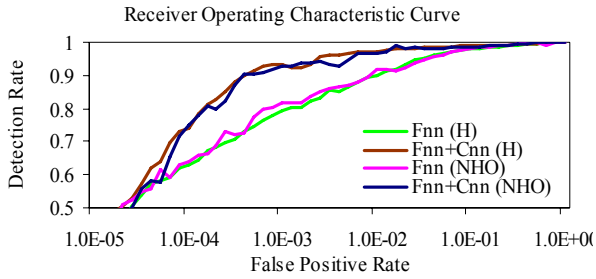

Fig. 7 Detection rate against false positive rate of Human and non-human object using FNN and (FNN+CNN)

Fig. 8 shows the results of human face and non-human object detection system over the test images. Fig. 8 (a) shows the output of FNN at different scales human face detection. However, overlapping windows corresponding to human face region and false positive alarm have been detected. These incorrect detections are due to the complex background, higher degree of rotations and variation of illuminations than were present in the training database. With further training after analyzing using CNN, the incorrect detections are eliminated as shown in Fig. 8 (b). Similarly, Fig. 8 (c) and 8 (d) exhibit some typical images of non-human object detection system using FNN and CNN.

Table 2 shows the performance of human face detection results of various methods on test set 1 and compare with other systems in terms of the number of detected faces, mis-detected faces, false detection and computation time. The success rate of the proposed method is 97.6 %, with 6 false alarms. It should be noted that the number of false alarms is quite small when compared to methods of Yacoub *et al.* 1999 and Fasel *et al.* 1998 which had 347 and 278 false alarms, respectively [19, 20]. This may show the capability of the combination of two networks to highly separate human face from non-object examples. The higher performance of Rowley *et al.* 1998 is likely due to the size of training data. In this work, we have used a total of 7344 human face images and 8000 non-object examples, while Rowley *et al.* 1998 system was trained with 16000 face images and 9000 non-faces images [11].

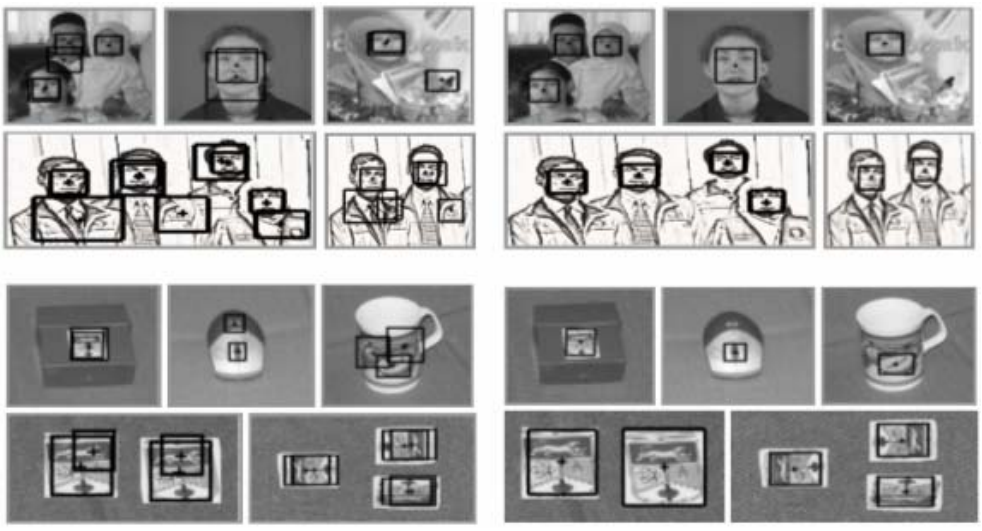

Fig. 8 Human Face and Non-Human Object Detection using a) FNN b) (FNN+CNN) c) FNN d) (FNN+CNN)

However, the technique is less efficient than our techniques in terms of false detection and time. On the other hands, Yacoub *et al.* 1999 [19] shows a very fast time processing but have a drawback of higher false alarms.

TABLE II
DETECTION RATE OF SET 1 ON DIFFERENT METHODS

Method	Human Object Detection		No. of False Detect	Process Time (s)
	Correct (%)	Incorrect (%)		
FNN+CNN	97.63	2.37	6	2.3
Rowley et al.	97.86	2.14	13	0.78
Yacoub et al.	84.31	15.69	347	0.7
Fasel et al.	96.8	3.2	278	3.1

Similarly, Table 3 shows the summarized results of non-human object of test set 2 compared to other systems. We found that the non-human object detection rate is 96.42%, which mean 108 out of 112 numbers of non-human objects were detected correctly.

TABLE III
DETECTION RATE OF SET 2 ON DIFFERENT METHODS

Method	Non Human Object Detection		No. of False Detects	Process Time (sec)
	Correct (%)	Incorrect (%)		
FNN+CNN	96.4	3.6	4	2.9
Agarwal et al.	94	6	30	3.6
Mahmud & Hebert	82	18	187	4.0
Viola & Jones	95	5	71	0.7

The false detection rate is 3.58%, which is lower that Fasel's and other methods. However, the average processing time is almost same with the others providing additional calculation on CNN. Based on the results shown in Tables 2 and 3, we can conclude that both human face and non-human object detection system make acceptable tradeoffs between the number of false detection and detection rate.

Fig. 9 demonstrates the detection capability of the proposed human face and non-human object detection scheme using the test image database. The top row of Fig. 9 showed the detection result for human face in a car environment under various lighting conditions and background. The second row depicts the single and multi face detection results for various poses. Additionally the third and forth rows displayed the human and nonhuman detection capabilities within various contrast.

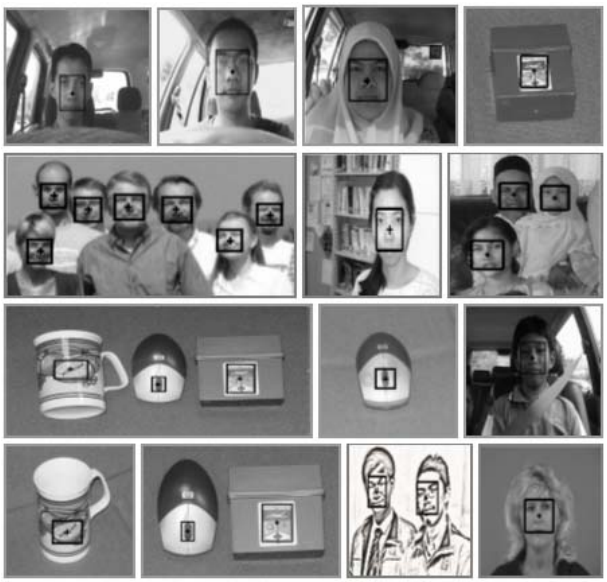

Fig. 9 Human face and non-human object detection

VIII. CONCLUSION

Many research has been geared towards face and object detection individually. However, we have implemented and evaluated both human face and non-human object detection in a frame and the results are very encouraging results. In this paper, we have presented a combination of fast neural network and classical neutral network based algorithm for fast and robust detection system of static images that is able to detect human face and non-human object. The framework described here is applicable to any other domains detection besides the proposed one. The system performs the detection by means of a two level hierarchical process. On the first level, fast neural network independently detects high level of accuracy using cross correlation. In second level, classical neural network verify the object region and performs the final detection step. Experimental result shows that proposed method does boost and improved the performance of both human face and non-human object detection system. Further work of the project is to generalize the current system for detection and recognition in the application of real time vehicle occupancy detection.

ACKNOWLEDGMENT

The authors would like to thank the Malaysian Ministry of Science, Technology and Innovation (MOSTI) for funding this work through research grant IRPA: 03-02-02-0017-SR0003/07-03.

REFERENCES

[1] S. Mahamud and M. Hebert, "The Optimal Distance Measure for Object Detection" *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR'03), 2003, Vol. 1, pp. 248-255.

[2] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-Based Object Detection in Images by Components", *IEEE Transaction on Pattern Analysis and machine Intelligence*, 2001.Vol. 23, No. 4, pp. 349-361.

[3] S. Ullman, E. Sali, and M. Vidal-Naquet, "A Fragment-Based Approach to Object Representation and Classification", *Proc. Fourth Int'l Workshop Visual*, 2001, pp. 85-100.
[4] S. Ullman, M. Vidal-Naquet, and E. Sali, "Visual Features of Intermediate Complexity and Their Use in Classification", *Nature Neuroscience*, 2002, Vol. 5, No. 7, pp. 682-687.
[5] M. Weber, M. Welling and P. Perona, "Unsupervised Learning of Models for Recognition", *Proc. Sixth European Conf. on Computer Vision*, 2000, pp. 18-32.
[6] D. Roth, M-H. Yang and N. Ahuja, "Learning to Recognize Three-Dimensional Objects," *Neural Computation*, 2002, Vol. 14, No. 5, pp. 1071-1103.
[7] S. Agarwal, A. Awan, and D. Roth, "Learning to Detect Objects in Images via a Sparse, Part-Based Representation", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2004, Vol. 26, No. 11, pp. 1475-1490.
[8] W. Wang, Y. Gao, S. C. Hui and M. K. Leung, "A Fast and Robust Algorithm for Face Detection and Localization", *Proc. of the 9th International Conference on Neural information Processing* (ICONIP'02), 2002, Vol. 4, pp. 2118-2121.
[9] J. Colmenarez and T. S. Huang, "Face Detection With Information-Based Maximum Discrimination", *IEEE Conference on Computer Vision and Pattern Recognition*, 1997, pp. 782-787.
[10] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1997, Vol. 19, No. 7, pp. 696-710.
[11] H. A. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection", *IEEE Transaction on Pattern Analysis and Machine Intellignce*, 1998, Vol. 20, No. 1, pp. 23-38.
[12] K-K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1998, Vol. 20, No. 1, pp. 39-51.
[13] H. M. El-bakry, "Fast Cooperative Modular Neural Nets for Human Face Detection", *Proc. of IEEE International Conference on Image Processing*, 7-10 Oct.,2001, Thessaloniki**,** Greece, pp. 1002-1005.
[14] L-L. Huang, A. Shimizu, Y. Hagihara and H. Kobatake, "Face detection from clustered images using polynomial neural network", *Proceedings of the IEEE International Conference on Image Processing*, 2001, pp. 669-672.
[15] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: An application to face detection", *Proc. of Computer Vision and Pattern Recognition,* 1997, pp. 130-136.
[16] C. Papageorgiou and T. Poggio, "A trainable system for object detection", *International Journal of Computer Vision*, 2000, Vol. 38, No. 1, pp. 15–33.
[17] P. Viola and M. Jones, "Rapid object detection using a boosted cascacd of simple features", *Proc. Computer Vision and Pattern Recogntion,* 2001, Vol. 1, pp. 511-518.
[18] R. Crane, "A Simplified Approach to Image Processing", Prentice Hall, 1997.
[19] S. Ben-Yacoub, B. Fasel and J. Luettin, "Fast Face Detection using MLP and FFT", *Proc. Second International Conf. On Audio and Video-based Biometric Person Authentication* (AVBPA '99), 1999.
[20] B. Fasel, S. Ben-Yacoub and J. Luettin, "Fast Multi-Scale Face Detection", *IDIAP-Com 98-04*, 1998, pp. 1-87.