A Weighted FMM Neural Network and Its Application to Face Detection*

Ho-Joon Kim[1], Juho Lee[2], and Hyun-Seung Yang[2]

[1] School of Computer Science and Electronic Engineering
Handong University, Pohang, 791-708, Korea
hjkim@handong.edu
[2] Department of Computer Science, KAIST
Daejeon, 305-701, Korea
{jhlee, hsyang}@paradise.kaist.ac.kr

Abstract. In this paper, we introduce a modified fuzzy min-max(FMM) neural network model for pattern classification, and present a real-time face detection method using the proposed model. The learning process of the FMM model consists of three sub-processes: hyperbox creation, expansion and contraction processes. During the learning process, the feature distribution and frequency data are utilized to compensate the hyperbox distortion which may be caused by eliminating the overlapping area of hyperboxes in the contraction process. We present a multi-stage face detection method which is composed of two stages: feature extraction stage and classification stage. The feature extraction module employs a convolutional neural network (CNN) with a Gabor transform layer to extract successively larger features in a hierarchical set of layers. The proposed FMM model is used for the pattern classification stage. Moreover, the model is utilized to select effective feature sets for the skin-color filter of the system.

1 Introduction

Fuzzy min-max (FMM) neural networks were introduced by Simpson [1] using the concept of hyperbox fuzzy sets. A hyperbox defines a region of the n-dimensional pattern space that has patterns with full class membership using its minimum point and its maximum point. The fuzzy min-max neural networks are built by making one pass through the input patterns and forming hyperboxes into fuzzy sets to represent pattern classes. Gabrys and Bargiela have proposed a General Fuzzy Min-Max (GFMM) neural network which is a generalization and extension of the FMM clustering and classification algorithm [2]. In GFMM method, input patterns can be fuzzy hyperboxes or crisp points in the pattern space. We present a modified FMM, called the weighted FMM (WFMM) neural network that takes the weights into account. The

* This research was supported as a 21st Century Frontier R&D Program and Brain Neuroinformatics Research Program sponsored by Ministry of Information and Communication and Minister of Commerce, Industry and Energy in KOREA.

I. King et al. (Eds.): ICONIP 2006, Part II, LNCS 4233, pp. 177 – 186, 2006.

rationale for this idea is that a feature of a particular hyperbox can cover many more training patterns than other features of the same hyperbox and features of other hyperboxes. A weight value is assigned to each of the dimensions of each hyperbox so that membership can be assigned considering not only the occurrence of patterns but also the frequency of the occurrences within that dimension.

Growing interest in computer vision has motivated a recent surge in research on problems such as face recognition, pose estimation, face tracking and gesture recognition. However, most methods assume human faces in their input images have been detected and localized [3-5]. Color usually presents a strong intuitive cue in complex scene images. Recently, skin detection has emerged as an active research topic in several practical applications including face detection and tracking. Various generic skin models in a number of color spaces have been presented [6-7]. However, we can expect variations when images are taken in various settings, with different kinds of camera hardware, and under a wide range of lighting conditions [8]. Therefore the generic skin model may be inadequate to accurately capture the wide distribution of skin colors in an individual image. In this paper we present a FMM-based feature analysis technique for the face detection system. Two kinds of relevance factors are defined to analyze the relationships between features and pattern classes. Through the feature analysis, we can select the most relevant features for the skin-color filter as well as the pattern classifier. Moreover, the training process can make it possible to adaptively adjust the feature ranges of the skin-color filter.

2 A Weighted FMM Neural Network

In our previous work, a weighted fuzzy min-max (WFMM) neural network has been introduced [9]. The model employs a new activation function which has the weight value for each feature in a hyperbox. In this paper, we introduce an improved structure of the WFMM neural network and its application to a face detection problem.

2.1 Structure and Behavior

The weighted fuzzy min-max(WFMM) neural network is a modified version of Simpson's FMM model[1]. The model consists of three layers: input layer, hyperbox layer and class layer. In the model, the membership function of a hyperbox is defined as Equation (1).

$$B_j = \{X, U_j, V_j, C_j, F_j, f(X, U_j, V_j, C_j, F_j)\} \quad \forall X \in I^n \tag{1}$$

In the equation, U_j and V_j mean the vectors of the minimum and maximum values of hyperbox *j*, respectively. C_j is a set of the mean points for the feature values and F_j means the a set of frequency of feature occurrences within a hyperbox. As shown in Equation (2) and (3), the model employs a new activation function which has the factors of feature value distribution and the weight value for each feature in a hyperbox.

$$b_j(A_h) = \frac{1}{\sum_{i=1}^{n} w_{ji}} \bullet \sum_{i=1}^{n} w_{ji}[\max(0, 1-\max(0, \gamma_{jiv} \min(1, a_{hi} - v_{ji}))) + \max(0, 1-\max(0, \gamma_{jiu} \min(1, u_{ji} - a_{hi}))) - 1.0] \quad (2)$$

$$\begin{cases} \gamma_{jiU} = \dfrac{\gamma}{R_U} & R_U = \max(s, u_{ji}^{new} - u_{ji}^{old}) \\ \gamma_{jiV} = \dfrac{\gamma}{R_V} & R_V = \max(s, v_{ji}^{old} - v_{ji}^{new}) \end{cases} \quad (3)$$

The hyperbox membership function has weight factor to consider the relevance of each feature as different values. In the equation, the w_{ij} is the connection weight between i-th feature and j-th hyperbox. The weighted FMM neural network is capable of utilizing the feature distribution and the weight factor in learning process as well as in classification process. Since the weight factor effectively reflects the relationship between feature range and its distribution, the system can prevent undesirable performance degradation which may be caused by noisy patterns. Consequently the proposed model can provide more robust performance of pattern classification when the training data set in a given problem includes some noise patterns or unusual patterns.

2.2 Learning Algorithm

The learning process of the model consists of three subprocesses: hyperbox creation, expansion, and contraction processes.

If the expansion criterion shown in Equation (4) has been met for hyperbox B_j, f_{ji}, u_{ji}, v_{ji} and c_{ij} are adjusted using Equation (5) and (6).

$$n\theta \geq \sum_{i=1}^{n} (\max(v_{ji}, x_{hi}) - \min(u_{ji}, x_{hi})) \quad (4)$$

$$\begin{cases} f_{ji}^{new} &= f_{ji}^{old} + 1 \\ u_{ji}^{new} &= \min(u_{ji}^{old}, x_{ki}) \\ v_{ji}^{new} &= \min(v_{ji}^{old}, x_{ki}) \end{cases} \quad (5)$$

$$c_{ji}^{new} = (c_{ji} * f_{ji}^{old} + x_{hi}) / f_{ji}^{new} \quad (6)$$

As shown in the equations, the frequency value is increased by 1 at every expansion and the feature range expansion operation is similar to the fuzzy intersection and fuzzy union operations [6]. The mean point value, c_{ji}, is updated by Equation (6). During the learning process the weight values are determined by Equation (7) and (8).

$$w_{ji} = \frac{\alpha f_{ji}}{R} \tag{7}$$

$$R = \max\left(s, v_{ji} - u_{ji}\right) \tag{8}$$

As shown in the equations, the weight value is increased in proportion to the frequency of the feature. In the equations, s is a positive constant to prevent the weight from having too high value when the feature range is too small. The value of f_{ji} is adjusted through the learning process. The contraction process is considered as an optional part of our model. The contraction process is to eliminate the possible overlappings between hyperboxes that represent different classes. We can expect that the weights concept of the model replace the role of overlapping handling because the weights reflect the relevance of feature values and hyperbox as different values. We define a new contraction method including the weight updating scheme. To determine whether or not the expansion has created any overlapping, a dimension by dimension comparison between hyperboxes is performed. If one of the following four cases is satisfied, then overlapping exists between the two hyperboxes.

$$\begin{aligned}
&case1: u_{ji} < u_{ki} < v_{ji} < v_{ki} \\
&\qquad \delta^{new} = \min(v_{ji} - u_{ki}, \delta^{old}) \\
&case2: u_{ki} < u_{ji} < v_{ki} < v_{ji} \\
&\qquad \delta^{new} = \min(v_{ki} - u_{ji}, \delta^{old}) \\
&case3: u_{ji} < u_{ki} < v_{ki} < v_{ji} \\
&\qquad \delta^{new} = \min(\min(v_{ki} - u_{ji}, v_{ji} - u_{ki}), \delta^{old}) \\
&case4: u_{ki} < u_{ji} < v_{ji} < v_{ki} \\
&\qquad \delta^{new} = \min(\min(v_{ji} - u_{ki}, v_{ki} - u_{ji}), \delta^{old})
\end{aligned}$$

For each of these cases, contraction process is performed. If $\delta^{old} - \delta^{new} > 0$, then $\Delta = i$, $\delta^{old} = \delta^{new}$, signifying that there was an overlap for Δth dimension. Otherwise, the testing is terminated and the minimum overlap index variable is set to indicate that the next contraction step is not necessary, i.e. $\Delta = -1$. If $\Delta > 0$, then the Δth dimension of the two hyperboxes are adjusted. Only one of the n dimensions is adjusted in each of the hyperboxes to keep the hyperbox size as large as possible and minimally impact the shape of the hyperboxes being formed.

As illustrated in Equation (9), we have defined new adjustment schemes from the new definition of hyperbox for the four cases. The frequency values, the mean points and the feature ranges are updated for the four cases. Consequently the frequency factor is increased in proportion to the relative size of the feature range, and the mean point value is adjusted by considering the expanded feature range.

$$case1: u_{j\Delta} < u_{k\Delta} < v_{j\Delta} < v_{k\Delta}$$

$$\begin{cases} v_{j\Delta}{}^{new} & = & v_{j\Delta}{}^{old} - \dfrac{f_{k\Delta}}{f_{j\Delta} + f_{k\Delta}}(v_{j\Delta}{}^{old} - u_{k\Delta}{}^{old}) \\ u_{k\Delta}{}^{new} & = & u_{k\Delta}{}^{old} + \dfrac{f_{j\Delta}}{f_{j\Delta} + f_{k\Delta}}(v_{j\Delta}{}^{old} - u_{k\Delta}{}^{old}) \\ f_{j\Delta}{}^{new} & = & f_{j\Delta}{}^{old} * (\dfrac{v_{j\Delta}{}^{new} - u_{j\Delta}{}^{new}}{v_{j\Delta}{}^{old} - u_{j\Delta}{}^{old}}) \\ f_{k\Delta}{}^{new} & = & f_{k\Delta}{}^{old} * (\dfrac{v_{k\Delta}{}^{new} - u_{k\Delta}{}^{new}}{v_{k\Delta}{}^{old} - u_{k\Delta}{}^{old}}) \\ c_{j\Delta}{}^{new} & = & u_{j\Delta}{}^{new} + (c_{j\Delta}{}^{old} - u_{j\Delta}{}^{old}) * (\dfrac{v_{j\Delta}{}^{new} - u_{j\Delta}{}^{new}}{v_{j\Delta}{}^{old} - u_{j\Delta}{}^{old}}) \\ c_{k\Delta}{}^{new} & = & u_{k\Delta}{}^{new} + (c_{k\Delta}{}^{old} - u_{k\Delta}{}^{old}) * (\dfrac{v_{k\Delta}{}^{new} - u_{k\Delta}{}^{new}}{v_{k\Delta}{}^{old} - u_{k\Delta}{}^{old}}) \end{cases} \tag{9}$$

3 A Face Detection Method Using the WFMM Model

As shown in Fig.1, our face detection system consists of three modules: preprocessor, feature extractor and pattern classifier. Through the skin color analysis and training process, the system can generate an adaptive skin model and a relevant feature set for the given illumination condition. The feature extractor generates numerous features from the input image. The number of features and the relevance factors of the features affect the computation time and the performance of the system. Therefore we propose a feature analysis technique to reduce the amount of features for the pattern classifier.

3.1 WFMM-Based Feature Analysis Technique

This section describes a feature analysis technique for the skin-color filter and the classifier. We define two kinds of relevance factors using the proposed FMM model as follows:

$RF1(x_j, C_k)$: *the relevance factor between a feature value* x_j *and a class* C_k

$RF2(X_i, C_k)$: *the relevance factor between a feature type* X_i *and a class* C_k

The first measure *RF1* is defined as Equation (9). In the equation, constant N_B and N_k are the total number of hyperboxes and the number of hyperboxes that belong to class k, respectively. Therefore if the $RF1(x_i, k)$ has a positive value, it means an excitatory relationship between the feature x_i and the class k. But a negative value of $RF1(x_i, k)$ means an inhibitory relationship between them. A list of interesting features for a given class can be extracted using the *RF1* for each feature.

$$RF1(x_i, C_k) = \left(\frac{1}{N_k} \sum_{B_j \in C_k} S(x_i, (u_{ji}, v_{ji})) \cdot w_{ij} - \frac{1}{(N_B - N_k)} \sum_{B_j \notin C_k} S(x_i, (u_{ji}, v_{ji})) \cdot w_{ij}\right) / \sum_{B_j \in C_k} w_{ij} \tag{9}$$

In Equation (9), the feature value x_i can be defined as a fuzzy interval which consists of min and max values on the *i*-th dimension out of the n-dimension feature space. The function ***S*** a similarity measure between two fuzzy intervals.

The second measure *RF2* can be defined in terms of *RF1* as shown in Equation (10). In the equation, L_i is the number of feature values which belong to *i*-th feature.

$$RF2(X_i, C_k) = \frac{1}{L_i} \sum_{x_l \in X_i} RF1(x_l, C_k) \tag{10}$$

The *RF2* shown in Equation (10) represents the degree of importance of a feature in classifying a given class. Therefore it can be utilized to select a more relevance feature set for skin color filter.

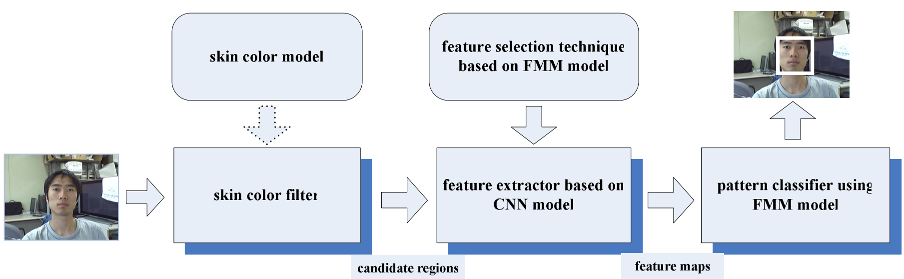

Fig. 1. The face detection system using hybrid neural networks

3.2 Feature Extraction and Face Classification

The most advantageous feature of convolutional neural network is invariant detection capability for distorted patterns in images [2-3]. The underlying system employs a convolutional neural network in which a Gabor transform layer is added at the first layer. As shown in Fig. 2, the first layer of the network extracts local feature maps from the input image by using Gabor transform filters.

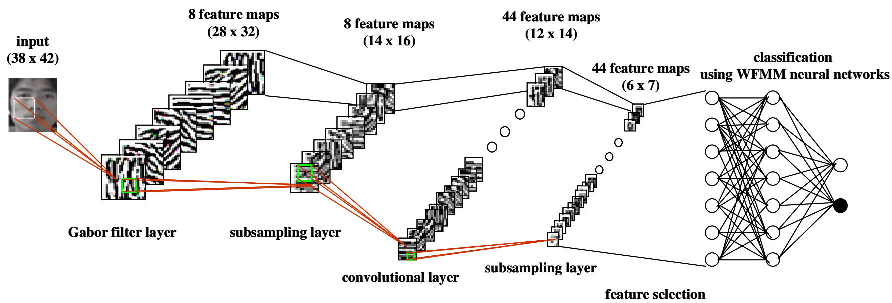

Fig. 2. Face detector using hybrid neural networks

The other layers of the feature extractor include two types of sub-layers called *convolution layer* and *sub-sampling layer*. Each layer of the network extracts successively larger features in a hierarchical set of layers. Finally a feature set is generated for the input of the pattern classifier. The number of the features can be reduced by the feature analysis technique using the FMM model described in the previous section. For the feature extractor, a set of (38 x 42) candidate areas are selected as input images. The first layer of the feature extractor, Gabor filter layer, extracts eight feature maps in which the size of feature map is (28 x 32). Each unit in each feature map is connected to a 11 x 11 neighborhood into the input retina. In the subsampling layer, the feature map has half the number of rows and columns of the input data. Therefore the layer has eight feature maps of size 14 x 16. The convolutional layer generates 44 feature maps. Each unit is connected to 3 x 3 neighborhood at identical locations in a subset of the feature maps of the Gabor transform layer. 1,848 feature values are generated and inputted into the input layer of the WFMM-based classifier. The aforementioned feature analysis technique can be used to reduce the number of these features.

4 Experimental Results

Two types of experiments have been conducted for a set of real images. For the training of skin-color filter, the system considers eleven color features, *Red, Green, Blue, Intensity, Cb, Cr, Magenta, Cyan, Yelleow, Hue and Saturation*. Fig. 3 shows two input images captured under different illumination conditions. Table 1 shows the skin-color analysis result and the feature range data derived from the training process. As shown in the table, different kinds of features can be adaptively selected for a given condition, and the feature ranges of skin-color filter can be also adjusted by the training process.

Table 1 shows four features which have the highest value of the relevance factor *RF1*. As shown in the table, a number of hyperboxes for face and non-face patterns have been generated and the relevance factors are also adjusted through the training process. Therefore the system can select more effective feature set adaptively for the given environment.

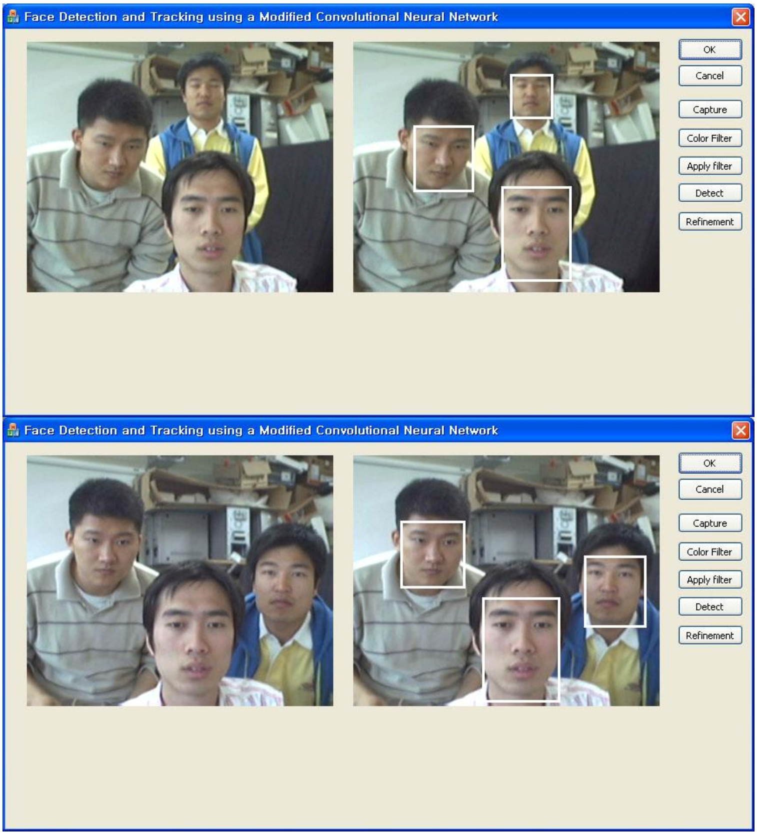

Fig. 3. Two training data captured under different illumination conditions

Table 1. Feature analysis results for the two images

image - 1			*image - 2*		
feature	feature range	*RF1*	feature	feature range	*RF1*
Hue	0.833 ~ 0.992	9.1103	*Cb*	0.589 ~ 0.772	0.8888
Saturation	0.019 ~ 0.135	9.0104	*Yellow*	0.433 ~ 0.632	0.7832
Cb	0.761 ~ 0.964	8.8212	*Saturation*	0.056 ~ 0.243	0.7204
Cr	0.053 ~ 0.234	6.7320	*Blue*	0.437 ~ 0.627	0.6929

We have selected face patterns from the real images and non-face patterns from the background images. 100 face patterns and 100 non-face patterns have used for training process. Fig. 4 shows the change of detection rate and false alarm rate by varying the number of training patterns, respectively. The result shows that the detection rate increases as more training patterns are used, and the false alarm rate decreases as more non-face counter examples are used for training.

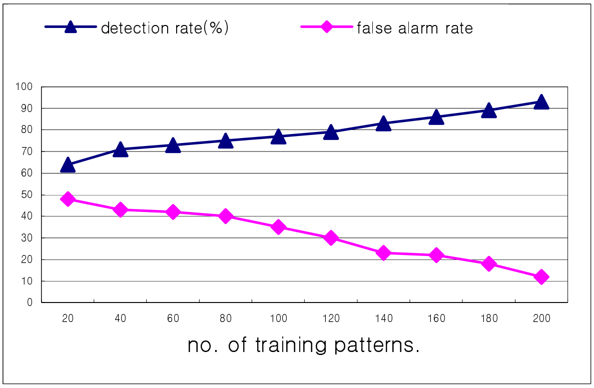

Fig. 4. Detection rate and false alarm rate as varying the number of training patterns

5 Conclusion

The proposed WFMM can provide at least two advantages over the original FMM: (a) it would work better for pattern classification than the original scheme, especially for data sets with highly uneven distribution of features or noisy features since the hyper-boxes in the WFMM model is not too sensitive to a few occurrence of unusual/noisy features in input patterns, (b) the learned weights for each feature during training process can be used to identify the relevance of the feature to the given class, which can be easily used for possible rule generation. The feature relevance measures computed through the feature analysis can be also utilized in designing an optimal structure of the classifier. We have applied the proposed model to a real-time face detection system in which the illumination conditions are frequently changed.

References

1. Simpson, P. K.: Fuzzy Min-Max Neural Networks Part 1: Classification. IEEE Transaction on Neural Networks, Vol.3. No.5. (1997) 776-786
2. Gabrys, B. and Bargiela A.: General Fuzzy Min-Max Neural Network for Clustering and Classification. IEEE Transaction on Neural Networks, Vol.11. No.3. (2000) 769-783
3. Feraud, R., Bernier, O. J., Viallet J. E. and Collobert, M.: A Fast and Accurate Face Detector Based on Neural Networks, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.23, No.1.(2001) 42-53.
4. Garcia, Cristophe and Delakis, Manolis: Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.26, No.11, (2004) 1408-1423
5. Lawrence, Steve, Giles, C. L., Tsoi, A. C. and Back, Andrew D.: Face Detection: A Convolutional Neural-Network Approach, IEEE Transaction n Neural Networks, Vol.8, No.1, (1997) 98-113

6. Hsu, Rein-Lien, Mohamed Abdel-Mottaleb and Jain, Anil K.: Face Detection in Color Images," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.24, No.5, (2002) 696-706
7. Storring, Moritz, Kocka, Tomas, Andersen, Hans J., Granum, Erik: Tracking Regions of Human Skin Through Illumination Changes, Pattern Recognition Letters, Vol.24, (2003) 1715-1723
8. Zhu, Q, Cheng, K. T., Wu, C. T., Wu, Y. L.: Adaptive Learning of an Accurate Skin-Color Model, Proceeding of the Sixth IEEE International Conf. on Automatic Face and Gesture Recognitin, 1 (2004)
9. Kim, Ho-Joon, Yang, Hyun-Seung, "A Modified FMM Neural Network for Pattern Classification and Feature Analysis," Proceeding of the ISATED International Conference on Artificial Intelligence and Applications, 1 (2005)