2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

# Saliency Map for Visual Attention Region Prediction Based on Fuzzy Neural Network

Mao Wang
Graduate School of Engineering,
University of Fukui,
Bunkyo, Fukui 910-8507, Japan
Email: mwang@ir.his.u-fukui.ac.jp

Yoichiro Maeda
Faculty of Engineering,
Osaka Institute of Technology,
Omiya, Asahi-ku, Osaka, 535-8585, Japan
Email: maeda@bme.oit.ac.jp

Yasutake Takahashi
Graduate School of Engineering,
University of Fukui
Bunkyo, Fukui 910-8507, Japan
Email: yasutake@ir.his.u-fukui.ac.jp

*Abstract* **Visual attention region prediction has been paid much attention by researchers in intelligent systems recent years because it can make the interaction between human and intelligent agents to be more convenient. In this paper, the prediction method of the visual attention region inferred by using fuzzy neural network (FNN) after extracting and computing of images feature maps and saliency maps was proposed. A method for training FNN is also proposed. A user experiment was conducted to evaluate the prediction effect of proposed method by making surveys for the prediction results. The results indicated that prediction method proposed by us has a better performance in the level of attention regions$'$ position prediction according to different images.**

## I. Introduction

Recently, the intention recognition, recognizing the intention of a user or an agent by analyzing their actions or changes of state, is becoming an important issue in various research fields of intelligent systems. Especially, the intention recognition can make the Human-Computer Interaction (HCI) more convenient. So far, many intention recognition approaches have been proposed.

Much of early work is in the context of speech understanding and response automatically [1]. For example, Pynadath et al. achieved the plan recognition on a problem in traffic monitoring through exploited the context by using a general Bayesian framework [2]. More recently, Pereira et al. [3] described an approach to tackle intention recognition by combining dynamically configurable and situation-sensitive Causal Bayes Networks plus plan generation techniques [4] [5]. Mao et al. have presented a utility-based approach to solve the recognition of intention, which is realized by incrementally using plan knowledge and observations to change state probabilities [6].

In their researches, the probability is the main factor which used to infer the human intention. Another method is use an automatic capture of bottom-up salient stimuli and volitional shifts guided by and top-down context factors [7] [8], where bottom-up salient stimuli are the external factors to user and top-down contexts are the internal factors. The characteristic of all the methods mentioned above is that one or several characteristic values of the image showing before user have been extracted, calculated and combined as their basis for inferring user's intention.

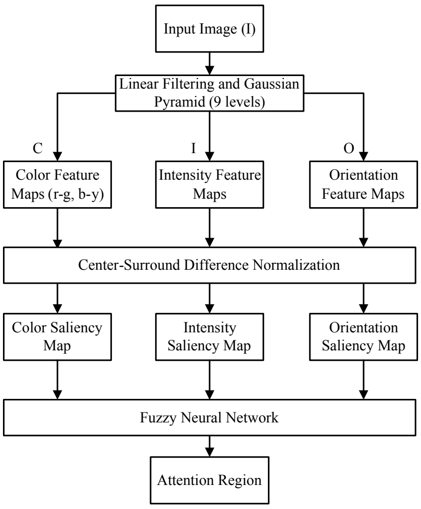


Fig. 1. Overall Procedure for Proposed System

In this paper, a visual attention region prediction system inspired on saliency map is described. The aim of this work is to present a new approach that improves the performance of attention prediction based on saliency map by using fuzzy neural network (FNN). The FNN employing features of image as input allows us to combine features and infer with great flexibility some intuitive decision rules based on the visual perception principles.

## II. Saliency Map by Fuzzy Neural Network

We propose a new approach to predict visual attention region based on image's saliency map which got by using FNN. The overall procedural flow of proposed approach is summarized in Fig.1.

Firstly, intensity, color (red, green, blue and yellow) and orientation (degrees of $0, \pi/4, \pi/2, 3\pi/4$) are extracted in multi-resolution from the Gaussian pyramid by linear filtering. Then a saliency map is generated for each of them by computing the difference between the layers of the pyramids, which imitates the center-surround type receptive field. Finally, the saliency map is generated by a trained FNN using the three saliency maps as inputs. The training method of FNN will be explaining in the following section.

Most attention models are based on a saliency map and a dynamical process for visiting saliency maxima. Itti et al.

978-1-4799-2072-3/14/$31.00 ©2014 IEEE

[7] introduced a model for the bottom-up selective attention based on serially scanning a saliency map, which is computed from local feature contrasts, for salient locations in the order of decreasing saliency. The saliency map is entirely based on features of an image and was originally designed to explain coverted attention on simple stimuli. A lot of researches on saliency map are getting some features of an image and combining them by simply sum in mathematics[8].

But this method also has its weakness. For example, the saliency map of an image is based on the three features which are color, intensity and orientation. Therefore, the method of simply sum of them gives the same importance at the same time. But based on experiments[9], most people do not pay equal attention to all of them. Actually, for example, the color feature in an image takes more attention from observers than other two features, which shows that the method may not be very reasonable in some situation.

In order to solve this problem, Wang et al.[10] proposed a method to compute saliency map by fuzzy inference based on the features of image. But also mentioned by the author, the method is only suitable for specific images, which means is not universal. In this paper, we proposed a method by using FNN to solve this problem. In this way, the importance of all features can be reflected in fuzzy rule with the human decision making model by the conceptual framework of fuzzy logic.

### A. Feature Maps

In Fig.1, the linear filter is used in order to compute center-surround different of various features at 9 scales. In this paper, the input image is sub-sampled into a dyadic Gaussian pyramid by convolution with a linearly separable Gaussian filter and decimation by a factor of two[11], which means that for example, the third level has a resolution of $\frac{1}{8}$ of the input image′s. After filtering, the three features of images have their values at each position according to the input image, which are divided into 9 levels of pyramid ready to be calculated. Then, in this paper, the color feature is reflected by two values defined by us, which are red-green and blue-yellow opponencies. If *r*, *g*, and *b* are the red, green, and blue values of the input color image respectively, then the color map of one level can be calculated according to the following equations.

$$\begin{aligned} M_{r-g} &= \frac{r-g}{max(r,g,b)}, &(1)\\ M_{b-y} &= \frac{b-min(r,g)}{max(r,g,b)}, &(2)\end{aligned}$$

where $M_{r-g}$, $M_{b-y}$ stand for red-green and blue-yellow opponencies. Red-green and blue-yellow opponencies are central to modeling the contribution of color to saliency because of these two opponency axes can cover the entire visible light [12]. And note that the definitions deviate from the original model by[13].

The intensity map $M_i$ of one level is calculated as:

$$M_i = \frac{r+g+b}{3}. \tag{3}$$

These operations are repeated for each level of the input to obtain an intensity pyramid with also 9 levels.

Local orientation map $M_o$ is obtained by applying steerable filters to the intensity pyramid levels $M_i$[14].

After getting $M_{r-g}$, $M_{b-y}$, $M_i$ and $M_o$, in order to yield the feature maps, we simulate the center-surround receptive fields by subtraction between two maps at the center (*c*) and the surround (*s*) levels in these pyramids. They can be calculated as

$$\begin{aligned} F_{l,c,s} &= N(|M_l(c) - M_l(s)|), &(4)\\ l &\in L = L_C \cup L_I \cup L_O, \end{aligned}$$

where

$$\begin{aligned} L_C &= \{I\},\\ L_I &= \{r-g, b-y\},\\ L_O &= \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}. \end{aligned}$$

Note that *N* is an iterative operator for local nonlinear iterative competition between salient locations within each feature map. And $F_l$ is the feature map summed over the center-surround combinations using across-scale addition while $F_c$ and $F_s$ stand for feather maps at the center and the surround levels in these pyramids, respectively.

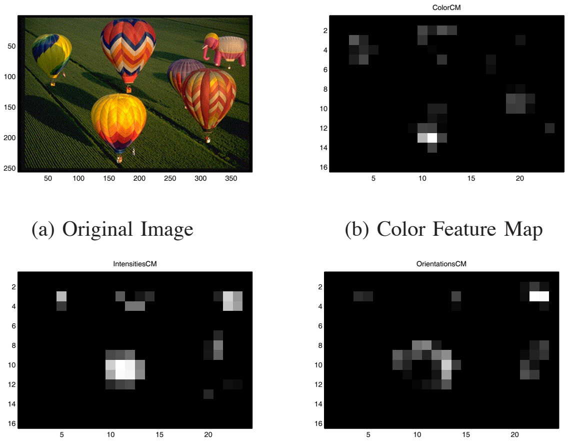
Finally, by summing over the center-surround combinations and normalizing again according the results obtained in Eq.(4), the feature maps of color, intensity and orientation can be obtained according to Eq.(5) as $C_c$, $C_i$, $C_o$, respectively. Here, all the feather maps we need for building region saliency map are already obtained. Fig.2 shows an example of three feature maps mentioned above.

$$\begin{aligned} C_i &= F_i,\\ C_c &= N(\sum_{l\in L_c} F_c), &(5)\\ C_o &= N(\sum_{l\in L_o} F_o). \end{aligned}$$

As shown in Fig.2, the three feature maps make no obvious difference. But when looking at each feature map and considering Fig.5, we know that only local maximas of activity in each feature map are considered. And as shown in Fig.2, the attention region results are based on saliency map, who is combined by the three feature maps. In other words, the results are the comprehensive reflection of the three feature maps.

### B. Fuzzy Neural Network

It has been shown that a fuzzy system can approximate any continuous real function defined on a compact domain by covering its function graph in input-output space using a set of if-then fuzzy rules. Theoretically, these fuzzy rules can always be discovered, but in practice we may have no idea on how to initialize these rules. Thus, it is crucial to have an adaptive fuzzy system which can produce the required rules automatically[15]. A FNN system is a learning machine that finds the parameters of fuzzy rules by exploiting approximation techniques from neural networks.



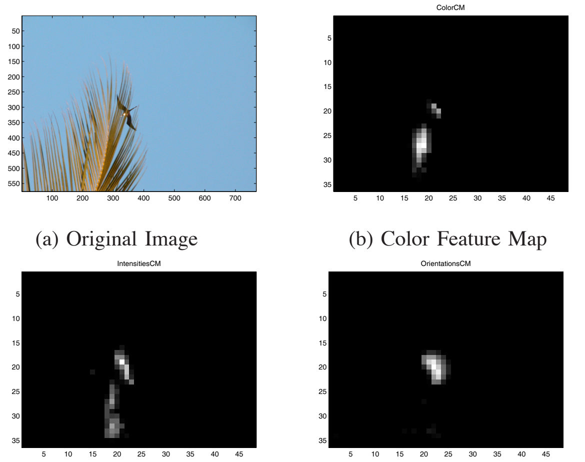
(a) Original Image

(b) Color Feature Map

(c) Intensity Feature Map

(d) Orientation Feature Map

Fig. 2. Example of Feature Maps Got from Image



(a) Original Image

(b) Color Feature Map

(c) Intensity Feature Map

(d) Orientation Feature Map

Fig. 3. Specific Example of Feature Maps

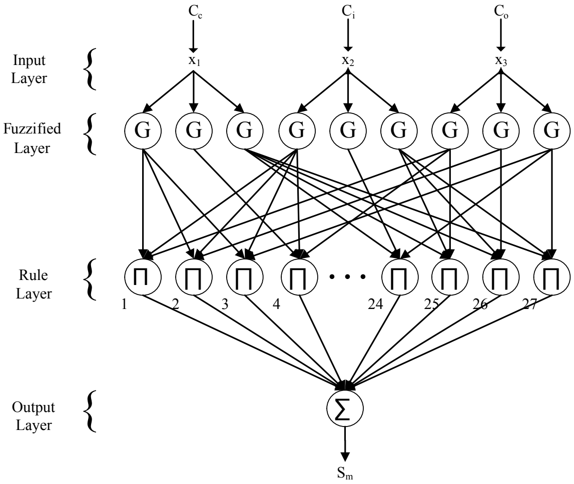


Fig. 4. The Structure of the FNN

A FNN may have 4 layers as follows: 1) the input layer as in a BP-NN or RBF-NN that simply fans out the inputs to the next layer; 2) a hidden layer that fuzzifies the inputs, e.g., into LOW, MEDIUM, and HIGH linguistic variables as antecedent part of fuzzy rule obtained by passing each input value through a fuzzy set at membership function; 3) a rule layer where arrows from certain fuzzifying nodes imply a consequent part of fuzzy rule in this layer; and 4) the defuzzification layer.

We have pointed out the defect of ordinary combination method of feature maps in the beginning of this section. It has not an importance distinction between various features, especially when a feature is more important comparing with others. For example, when the color feature values are lower compared with the other two, the method mentioned above will decide the saliency values according to the higher two features. In general cases they will have correct results, but when special cases such as the relative difference value is larger than other ones' although has lower feather values itself. As we can see in Fig.3, the color and orientation feature values are higher than intensity one's in this case, but according to the image and our experience, we know that the attention object of human should be the bird in the image, which also means that the saliency region should be decided by the intensity feature.

In the feature maps building stage, it has little sense to use fuzzy theory as a classifier, but in the combination stage, using FNN to infer visual attention region can lead to the results which can reflect the importance of each feather map in the saliency feature that images have. The greatest difference between mathematical sum method and FNN method is that the importance reflected in every feature map is different. These are tuned by fuzzy rules and connection weights of the network according to feature values. In FNN method, the importance of each feature map is different according to different images.

In this study, we use the feature variables from color feature map ($C_c$), intensity feature map ($C_i$) and orientation feature map ($C_o$) as input while the output is a value of region saliency map ($S_m$). Every value of region saliency map is decided by FNN as shown in Fig.4 where G stands for Gaussian Function [16].

As shown in Fig.4, the FNN structure with three input variables, three input nodes of fuzzified layer for each input variable, 27 rule nodes of hidden layer, and one output node of output layer. A typical format for a fuzzy rule base consists of a collection of fuzzy IF-THEN rules in the following form:

$$IF\, x_1\, is\, A^j_{111}, \cdots, and x_n\, is\, A^j_{nml}, THEN\, S^j_m\, is\, \beta^j \tag{6}$$

where $A^j_{nml}$ and $\beta^j$ are fuzzy sets and $x_i$ and $S^j_m$ are the input and output of the fuzzy inference rule, respectively. And $n$, $m$, $l$ stand for the node number in fuzzified layer of ($C_c$), ($C_i$) and ($C_o$) while $i$ is the input node number for rule layer and $j$ is the output node number of rule layer.

*1) Fuzzified Layer:* This layer uses a Gaussian function as a membership function, so the output of the $i$th term node associated with $x_i$ is:

$$\mu_{A_{ijk}} = exp(-(\frac{x_i - m_{ijk}}{\sigma_{ijk}})^2) \tag{7}$$

where $m_{ijk}$ and $\sigma_{ijk}$ denote the mean (center) and variance (width) of $A_{ijk}$, respectively. And $i$, $j$, $k$ have the similar meaning as $n$, $m$, $l$ in Eq.(6).

*2) Rule Layer:* This layer implements the links relating preconditions (fuzzified layer) to consequences (output layer). The connection criterion is that each rule node has only one antecedent link from a fuzzified node of a linguistic variable. Hence there are 27 rule nodes in the initial form of FNN structure. We mention that there is still no weight adjustment in this layer. The output of the $j$th rule node is:

$$out_j^3 = \prod_{i=1}^{n} \mu_{A_{ikl}}(x_i) \tag{8}$$

where the superscript 3 of out stands for the input number is 3, and $i$, $k$, $l$ have the similar meaning as $n$, $m$, $l$ in Eq.(6). Only noted that $l$ is determined by the connection criterion.

*3) Output Layer:* All consequence links are fully connected to the output nodes and interpreted directly as the strength of the output action. This layer performs defuzzification to obtain the numerical output:

$$S_m = \sum_{j=1}^{m} \beta^j \prod_{i=1}^{n} \mu_{A_{ijk}}(x_i) \tag{9}$$

where $m$ is the number of fuzzy IF-THEN rules and $n$ is inputs number.

### C. Supervised Learning of Fuzzy Neural Network

The adjustment of the parameters in the proposed FNN can be divided into two tasks, corresponding to the IF (antecedent) part and THEN (consequent) part of the fuzzy inference rules. A simple and intuitive method of initializing the center and width for Gaussian functions is to use normal fuzzy sets to fully cover the input space. In this paper, we initialized these singletons based on the method mentioned in [10].

A gradient-descent-based BP algorithm is employed to adjust FNN's parameters [15] [17]. The goal is to minimize the error function:

$$E = \frac{1}{2}(d - S_m)^2 \tag{10}$$

where $S_m$ is the output of the FNN and $d$ is the desired output for the input pattern. If $w_{ijk}$ is the adjusted parameter, then the learning rule is:

$$w_{ijk}(t+1) = w_{ijk}(t) - \eta \frac{\partial E}{\partial w_{ijk}} + \alpha \Delta w_{ijk}(t) \tag{11}$$

and

$$\Delta w_{ijk}(t) = w_{ijk}(t) - w_{ijk}(t-1) \tag{12}$$

where $\eta$ is the learning rate and $\alpha$ ($0 < \alpha < 1$) is the momentum parameter.

TABLE I. INITIAL PARAMETERS OF THE MEMBERSHIP FUNCTIONS

| Weights | Value | Weights | Value | Weights | Value |
|---|---|---|---|---|---|
| A111 | 0.2396 | A211 | 0.6741 | A311 | 0.6741 |
| A112 | 0.3389 | A212 | 0.3536 | A312 | 0.3389 |
| A113 | 0.2396 | A213 | 0.2396 | A313 | 0.2396 |
| A121 | 0.3536 | A221 | 0.5000 | A321 | 0.6741 |
| A122 | 0.5000 | A222 | 0.3536 | A322 | 0.9533 |
| A123 | 0.3536 | A223 | 0.2396 | A323 | 0.6741 |
| A131 | 0.6741 | A231 | 0.3389 | A331 | 0.3536 |
| A132 | 0.9533 | A232 | 0.2396 | A332 | 0.5000 |
| A133 | 0.6741 | A233 | 0.9533 | A333 | 0.3536 |

In this paper, the sample data for training is a McGill calibrated color Image Database [18]. The database provides a large number of color images of natural scenes, calibrated, for use in biological and computer vision research. In order to obtain the teaching signals, we proceeded in the following manner. First, we got input and output data by doing lots of experiments by using the image database mentioned above. Then got the color, intensity and orientation feature values of each image. Second, we showed the images to subjects and asked the region they intended in the experiment process. Actually, they were asked to give three regions in order. Then saliency value of pixel in users intention region was raised according to regions order. At last, by analyzing the data obtained in previous steps, we generated the teaching signals. The feature maps of image are calculated as explained above as input data. And the output data for training is the saliency value of the image calculated based on Itti's model [8] but adjusted according to the actual attention region given by user who look over the sample images.

## III. EXPERIMENTAL RESULTS

The initial structure of the FNN uses three input nodes for $x_1$, $x_2$ and $x_3$, which stand for $C_c$, $C_i$ and $C_o$, respectively. So in this case we have 3×3×3 initial rules. Suppose one epoch of learning takes 16×24 points. The supervised learning is continued for 500 epochs of training. The fuzzy sets for these linguistic term nodes are normally and uniformly initialized. We choose $\eta$ = 0.02 and $\alpha$ = 0.85 for supervised learning. The desired error $d$ is got from the adjusted saliency map value calculated by Itti's method. The parameters of the initial and final membership functions are illustrated in Table I. And Table II listed the weight value of the FNN after training. In the two tables $A_{ijk}$ standstill for the weight for nodes $i$, $j$, $k$ in rule layer. Finally, the mean squared error (MSE) is 0.000497. The learning curve is illustrated in Fig.5. From the figure we can see that the learning speed is very fast. This is because there are only 20 groups sample data used as inputs in this time. And the number of values of each group is 16×24 points. As we all know, good production parameters can accelerate the learning speed of FNN. We also did the experiment with different settings of initial parameters by setting all to 0.5. And in this case, with the same setting of $\eta$, $\alpha$ and $d$, after 500 epochs of training, the MSE is 0.002318, which is worse than the last one.

After the training of FNN, We conduct several experiments to demonstrate the inference result of the proposed method and

TABLE II. FINAL PARAMETERS AFTER LEARNING

| Weights | Value | Weights | Value | Weights | Value |
|---|---|---|---|---|---|
| A111 | -0.1406 | A211 | 0.0024 | A311 | 0.3015 |
| A112 | 0.3324 | A212 | 0.2227 | A312 | 0.4174 |
| A113 | 0.3159 | A213 | 0.0605 | A313 | 0.4369 |
| A121 | 0.0161 | A221 | 0.0386 | A321 | 0.2487 |
| A122 | 0.0541 | A222 | 0.2268 | A322 | 0.3591 |
| A123 | 0.0928 | A223 | 0.0182 | A323 | -0.1952 |
| A131 | 0.4472 | A231 | 0.4159 | A331 | 0.8196 |
| A132 | 0.7299 | A232 | 0.5501 | A332 | 1.2614 |
| A133 | 0.4659 | A233 | 0.2168 | A333 | 1.1184 |

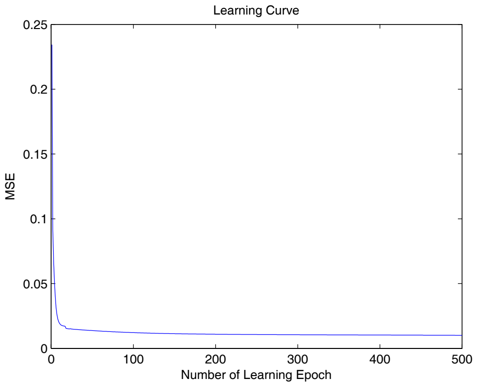
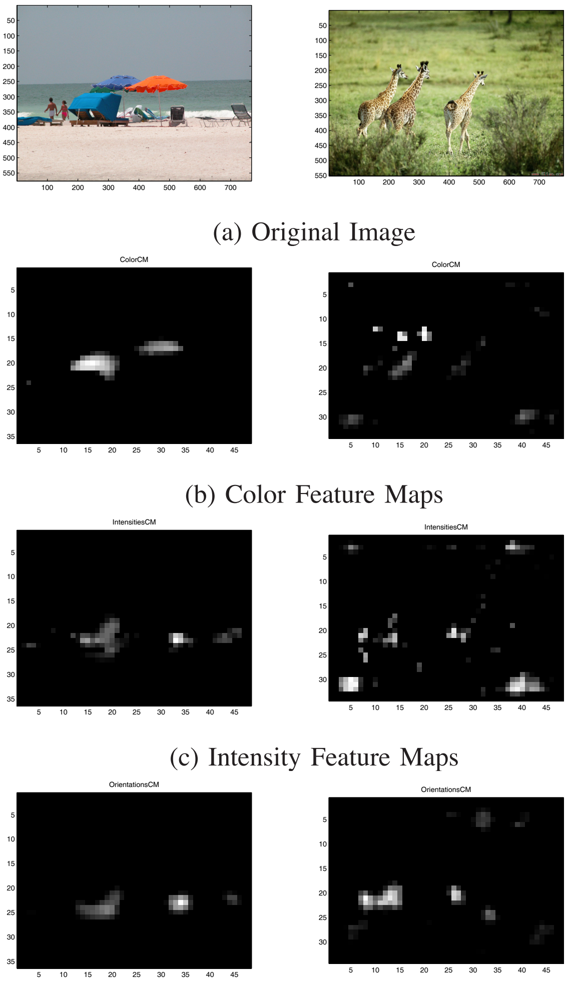


Fig. 5. The Learning Curve of the FNN



(a) Original Image

(b) Color Feature Maps

(c) Intensity Feature Maps

(d) Orientation Feature Maps

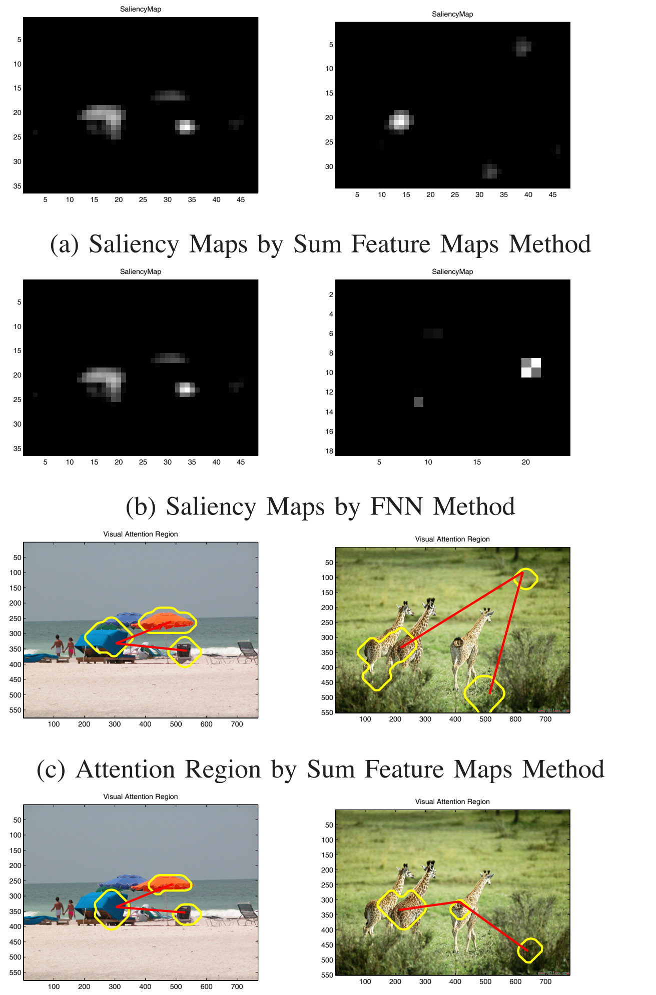
Fig. 6. Two Examples of Feature Saliency Maps

also compare with the performance of the proposed method with Itti's model[8]. As mentioned in the last section, we get the various feature saliency maps at first. Here, two different images are used as the input images. The input image is processed for low-level features at multiple scales, and center-surround differences are computed according to Eq.(4). Then, the resulting feature maps are combined into feature saliency maps according to Eq.(5), which is shown in Fig.6.

After getting the feature saliency maps, the region locations in the saliency map compete for the highest saliency value by FNN method proposed by us. After segmentation around the most salient region location, this saliency map is used for obtaining a smooth object mask at image resolution and for object-based inhibition of return. The parameters of fuzzy rules are shown in Table II. The results of saliency map and attention region by both sum feature maps method and FNN method are shown in Fig.7. The attention regions are marked by yellow lines while red lines express the order easy to be paid attention of them.

As we can see in the figure, there are only little differences between the saliency maps of two methods. And for the first example, the approximate locations of attention regions and the orders are basically the same between two methods while the little difference is the shape and size of region. This is because the color, intensity and orientation feature are all reflect obviously in regions marked of it compared with the rest regions, which also means that the differences of importance for the three features are small. So the method we proposed has not functioned very efficiently. But from the result of the second example we can see that the attention regions of our proposed method have better result. But only from these results we cannot yet say whether our proposed method is better than Itti's[8] or not definitely. So we conduct another experiment to verify it.

The following experiment is asked 5 males who are between 20 to 30 years old to look over 20 images. After all the experiments they will be showed the results of attention region got by the two methods and asked to compare with the ones they actually attending and looking at in the experiment process. Finally, an evaluation of user's attitude to the results is carried on and shown in Fig.8. Every factor of them has five ranks and represented by 1∼10 from worst to best in these figures. We can see from the evaluation results that the performance of our proposed method is higher than the conventional method, who obtained saliency map just by summing every feature maps. This also illustrated that proposed FNN method can improve the performance of attention region prediction at some aspect. It is worth to note that in this research, user's intention regions are obtained only by processing of image features. In other words, the proposed method is just for the unconscious intention. When subsets have conscious intention, for example, with special interests at something or finding something, the intention region will also depends on the subjective factors.



(a) Saliency Maps by Sum Feature Maps Method

(b) Saliency Maps by FNN Method

(c) Attention Region by Sum Feature Maps Method

(d) Attention Region by FNN Method

Fig. 7. Two Examples of Saliency Maps and Attention Regions

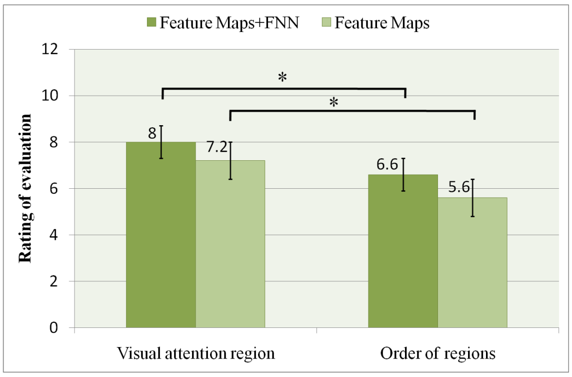


Fig. 8. Evaluation Value and Standard Deviation of Attention Regions Obtained by Two Methods

## IV. Conclusions

In this paper, we proposed a FNN method based on color, intensity and orientation feature maps of images to predict the visual attention regions. We also conducted a series of attention region predict experiments. The prediction accuracy of our proposed approach was evaluated in experiments, and the results confirmed the effectiveness of our method in visual attention region prediction. The problem still existing is the input image data for training of FNN is still less and the advantages of FNN have not reflected very well. And another problem is the method of getting the training data is not very improvement yet. In the future work, we will work on proposing a method to get more sample data for training FNN effectively and verify the method.

## References

[1] F. Sadri, *Logic-Based Approaches to Intention Recognition, Handbook of Research on Ambient Intelligence: Trends and Perspectives,* 2010.

[2] D. V. Pynadath and M. P. Wellman. "Accounting for Contextin Plan Recognition, with Application to Traffic Monitoring," *Proceedings of the Eleventh International Conference on Uncertainty in Artificial Intelligence*, pp.472-481, 1995.

[3] L. M. Pereira and H. T. Anh, "Intention Recognition via Causal Bayes Networks Plus Plan Generation," *Progress in Artificial Intelligence*, pp. 138-149, 2009.

[4] K. A. Tahboub, "Intelligent Human-Machine Interaction Based on Dynamic Bayesian Networks Probabilistic Intention Recognition," *Journal of Intelligent and Robotic Systems*, vol.45, pp.31-52, 2006.

[5] J. W. Harris and H. Stocker, *Handbook of Mathematics and Computational Science,* Springer-Verlag New York, 1998.

[6] W. Mao and J. Gratch, "A Utility-Based Approach to Intention Recognition," *Proceedings of the AAMAS 2004 Workshop on Agent Tracking: Modeling Other Agents from Observations*, 2004.

[7] L. Itti, C. Koch and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.20, no.11, pp.1254-1259, 1998.

[8] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision Research*, vol.40, pp.1489-1506, 2000.

[9] D. Walther, U. Rutishauser, C. Koch and P. Perona, "On the usefulness of attention for object recognition," *Workshop on Attention and Perfromance in Computational Vision*, pp.96-103, 2004.

[10] M. Wang, Y. Maeda and Y. Takahashi, "A Fuzzy Inference Method Based on Saliency Map for Visual Attention Region Prediction," *Fuzzy System Symposium 2013*, pp.495-500, 2013.

[11] L. Itti, *Models of bottom-up and top-down visual attention,* PhD thesis, California Institute of Technology, 2000.

[12] L. M. Hurvich and D. Jameson, "An opponent-process theory of color vision," *Psychological Review*, vol.63, pp.384-404, 1957.

[13] R. Manduchi, P. Perona and D. Shy, "Efficient deformable filter banks," *IEEE Transactions on Signal Processing*, vol.46, no.4, pp.1168-1173, 1998.

[14] W. O. Lee, J. W. Lee, K. R. Park, E. C. Lee and M. Whang, "Object recognition and selection method by gaze tracking and SURF algorithm," *2011 International Conference on Multimedia and Signal Processing*, pp.261-265, 2011.

[15] G. B. Huang, Q. Y. Zhu and C. K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference*, vol.2 pp.985-990, 2004.

[16] C. M. Lin, C. F. Hsu, "Supervisory recurrent fuzzy neural network control of wing rock for slender delta wings," *IEEE TRANSACTIONS ON FUZZY SYSTEMS*, vol.12, no.5, pp.733-742, 2004.

[17] C. T. Chao, C. C. Teng, "Implementation of a fuzzy inference system using a normalized fuzzy neural network," *Fuzzy Sets and Systems*, vol.75, pp.17-31, 1995.

[18] A. Olmos, F. A. A. Kingdom, "A biologically inspired algorithm for the recovery of shading and reflectance images," *Perception*, vol.33, no.12, pp.1463-1473, 2004.