A New Approach of Feature Combination for Object Detection in Saliency-based Visual Attention

Zahra Kouchaki
Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

Ali MotieNasrabadi
Department of Biomedical Engineering, Shahed University, Tehran, Iran

Keivan Maghooli
Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

ABSTRACT

This paper presents a fuzzy approach of feature maps combination in saliency-based visual attention model proposed by Itti. This strategy applies fuzzy rules to combine three conspicuity maps instead of linear combination in the basic model of visual attention that does not seem reasonable biologically. In this method, in addition to bottom-up features, top-down cues are also considered in the model. As fuzzy rules are designed using target mask information, top-down characteristics of the target are considered helping the model to make the target more conspicuous in the final saliency map. This can be applied in further processing such as object detection and recognition application. The experimental results show the effectiveness of our new fuzzy approach in finding the target in the first hit. A database of emergency triangle in natural environment background is used in this paper to show the results. Moreover, the comparison of this fuzzy combination approach with some other combination methods also proved the priority of the approach over other combination strategies.

General Terms

Your general terms must be any term which can be used for general classification of the submitted material such as Pattern Recognition, Security, Algorithms et. al.

Keywords

Visual Attention, Salient Point, nonlinear Combination, Fuzzy Fusion, Top-Down, Object Detection

1. INTRODUCTION

During the last few decades, machine vision technique which is based on human visual reality has remarkably improved. One of the most important abilities of humans is visual attention system. This ability can direct human vision to the most interesting parts of a scene called salient points and their saliencies are related to how much attention can focus on them. This mechanism has been applied in many applications such as target detection, navigational aids and robotic control [1], [2], [3], [4], [5].

There are some computational models of visual attention which account for bottom-up visual attention [6], [7], [8]. In all these models, low-level visual features such as color, intensity and orientation help to form multi-scale feature maps and then saliency map. The basic computational model of visual attention was proposed by Itti in 1998 which are the basic model for the most of the new models [7]. However, while many researchers have focused on bottom-up features, it has been proved that top-down cues also play an important role in directing attention towards salient regions [9]. As a result, many researchers have also studied in this regards [10], [11]. In top-down approaches, finding a special object or some particular objects are usually taken into consideration which is suitable in object detection application. In terms of combining feature maps to form saliency map, some researchers have presented new ideas [12], [13]. They proposed new combination strategies instead of combining feature maps linearly which does not seem reasonable biologically.

In this paper, we have proposed a new fuzzy combination method for combining conspicuity maps instead of linear summation. Fuzzy logical systems have been employed in many issues such as signal processing, image processing and pattern recognition and etc [14]. They can be replaced for traditional mathematical modeling to model the complex human behavioral systems. Input-output fuzzy sets could be taken into account as a fuzzy model of human behavioral systems. Therefore, they could be a suitable choice for combining feature maps. The fuzzy systems sound suitable as nonlinear function estimator because it can approximate any real continuous function with a high accuracy [15], [16]. So, the training images of each database were applied with their corresponding target mask as input-output fuzzy sets and then designed fuzzy rules as nonlinear combination operator. By doing this, the combination rules were created purposefully in direction of highlighting the target, which, in turn, leads to finding the target as a conspicuous object with a high accuracy.

The rest of this paper is as follows. In section 2, the basic saliency-based visual attention model is briefly explained that is the basis for our model. In section 3, the details of our model will be discussed and the way of designing fuzzy sets, membership functions, rules and designing the *FIS* system. Experimental Results are discussed in section 4. Finally, section 5 concludes the paper.

2. SALIENCY-BASED VISUAL ATTENTION MODEL

In this part, we briefly discuss about the basic saliency-based visual attention model proposed by Itti et al [7]. In this model, the value of every pixel in saliency map indicates the conspicuity value in original input image corresponding to the saliency map. First of all, different spatial scales are generated using Dyadic Gaussian Pyramids which subsample and low-

pass filter the color input image [7]. Then, feature maps in three different channels of color, intensity and orientation are evaluated by linear "center-surround" operator which acts similar to visual receptive fields [7]. Center-surround operator performs as a difference between fine and coarse scales of an image. Finally, 42 feature maps corresponding to color (12maps), intensity (6maps) and orientation (24 maps) are generated. These maps, which take out with various extraction methods, have different dynamic ranges and not-comparable modalities [7]. So, they should be normalized to the same dynamic range to have the capability of combination. After normalization of feature maps, they combine linearly in each channel to generate three conspicuity maps. Ultimately, after linear summation of three conspicuity maps, saliency map are created. The conspicuity at every location of saliency map could guide the attention based on spatial distribution of saliency.

3. MODEL

As it could be observed in Fig.1, after extraction of three conspicuity maps by the basic model [7], we aim to fuse them by *Fuzzy Interface System* (FIS) to form the final saliency map as output of the fuzzy system. After consisting saliency map, the winner-take-all (WTA) network [7] recognizes the most salient point and directs attention towards it. The combination rules are generated through training images that will be discussed in proceeding sections. A look up table method is employed to generate fusion rules, which then will be explained in more detail.

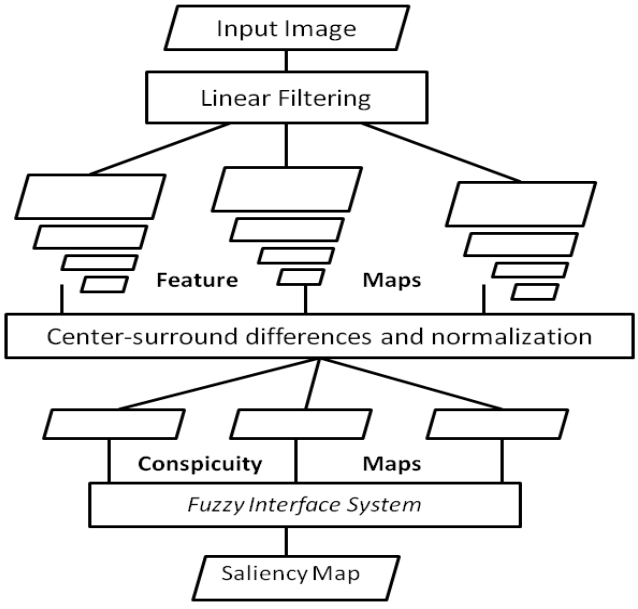


Fig 1: The modified visual attention model using *Fuzzy Interface System* (FIS).

3.1 Database

Image database (emergency triangle database) applied in this study is taken from Itti's Lab at USC. This database contains three groups of images. The first group consists of several color images of the target in natural environment background. The second group contains images of target masks corresponding to the first group Images. As it is shown for some samples in Fig.2, the images of target masks contain white color for the target with totally black background. The first and second groups of images are applied for generating input-output fuzzy sets to design fuzzy rules. The third group consists of other color images of the target in natural environment background. This group is used for test process. Furthermore, compared our method is compared with the basic saliency model and also some relevant model using the same test database.

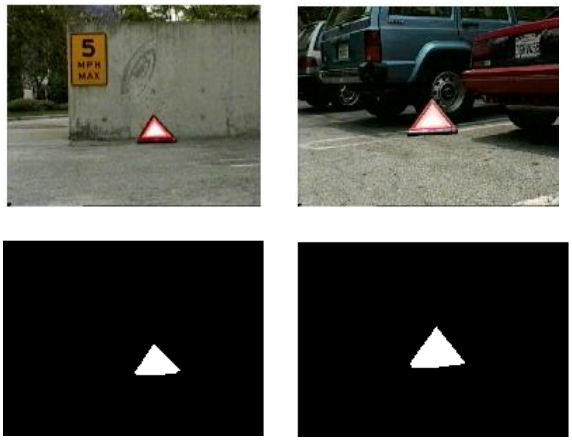


Fig 2: Top: two samples of training images; Bottom: target masks corresponding to the training images in the Top.

3.2 Generating fuzzy input-output pairs using our database

Because we aim to modify the way of combining conspicuity maps, conspicuity maps should be drived in Itti's model for all the images of the first group (training images). As the images are in the size of (640×480) pixels, the conspicuity maps are extracted with the size of (40*30) pixels. As a result, three 1200-dimensional conspicuity maps for all the 32 images of the first group and their corresponding target masks in the second group are applied as the training dataset to generate fuzzy input-output sets. In other words, there are 32 groups of images, each of them containing four maps, three conspicuity maps and one target mask. The pixels of these four images have been employed as numerical input-output set to create fuzzy rules. As shown in Fig .3 for a sample image, 1200 pixels of three conspicuity maps are employed as fuzzy input sets and the pixels of one target mask are applied as fuzzy output sets. These data pairs in all 32 groups are used to create a set of fuzzy if-then rules.

Now, There are 1200 three-input one-output pairs of data for one sample image as the following:

$$(i_1^{(1)}, i_2^{(1)}, i_3^{(1)}; t^{(1)}), (i_1^{(2)}, i_2^{(2)}, i_3^{(2)}; t^{(2)}), \ldots, (i_1^{(1200)}, i_2^{(1200)}, i_3^{(1200)}; t^{(1200)}) \quad (1)$$

Where, i_1, i_2, i_3 are inputs and t is output. Thus, 32 groups of 1200-dimentional data as demonstrated in (1), with three inputs and one output are employed as numerical fuzzy sets to design fuzzy rules which can effectively model the input-output pairs of data extracted from 32 images dataset.

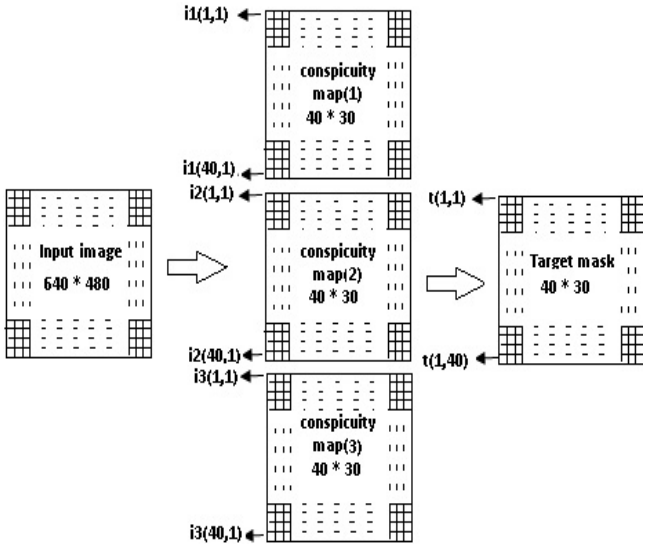


Fig 3: Pixels of conspicuity maps (40×30=1200 pixels) extracted from input image and pixels of its corresponding target mask (1200 pixels)

3.3 Designing fuzzy membership functions

This $\{(A_p^j, B^j)$: p = (1: Ni), j = (1:Nj)$\}$ could be considered as membership functions for the inputs and output. Where, $N_i$is the number of inputs and $N_j$is the number of membership functions which is allocated to each input. Here three conspicuity maps are existed that their pixels are regarded as three inputs. So, $N_i$is equal to 3. Moreover, each input image, which is conspicuity map, is normalized between zero and one. Therefore, the inputs domain are considered between [0, 1], which are divided into three regions, signified by *L* (*Low intensity*) for A_p^1, *M* (*Medium intensity*) for A_p^2, *H* (*high intensity*) for A_p^3. So, N_j is also considered 3. In other words,three fuzzy trapezoid membership functions are allocated to each input. Besides, in order to achieve the fuzzy output, three trapezoid membership functions (*Lo*, *Mo*, *Ho*) for B^j(j=1:3) were considered. It should be mentioned that other types of membership functions with different divisions could be also applied. Since there are many pairs of input-output and three membership functions using for each input, 3D input space was divided into (3×3×3=27) regions to cover all the input space. Furthermore, one rule is assigned to each region with look-up table method. As a result, 27 rules were designed. Each input $\{(i_1^{(x)}, i_2^{(x)}, i_3^{(x)})$: ($x$=1:38400)$\}$ is allocated to one of these 27 regions with its corresponding rule.

3.4 Generating fuzzy rules

Creating fuzzy rules is a very important part of designing a fuzzy logical system. First of all, a rule should be designed for each input-output pairs of data demonstrated in equation (1). For each input-output pair of $(i_1^{(x)}, i_2^{(x)}, i_3^{(x)}; t^{(x)})$ (x=1:38400), membership values of $i_p^{(x)}$ (p=1:3) in fuzzy membership functions of A_p^j (j=1:3) should be determined. Then, the membership values of $t^{(x)}$ in the fuzzy membership functions of B^j (j=1:3) should be calculated. In other words, $\mu_{A^j p}(i_p^{(x)})$ for (j=1:3) and (p=1:3) and $\mu_B{}^j(t^{(x)})$ for (j=1:3) should be calculated. After that, for each input and output variable, the fuzzy membership functions should be determined in which $i_p^{(x)}$and $t^{(x)}$ have the most values. Finally, an if-then rule for each input-output fuzzy pair could be designed.

As there are lots of input-output pairs and a lot of rules, many created rules with the same "IF" part and different "THEN" part conflict with each other. As a result, a degree should be assigned to each rule to ignore the rules with lower degree to reduce the number of rules. For instance, in the supposed rule "IF i_1 is A_1 and i_2 is A_2 and i_3 is A_3, THEN t is B", the degree of the rule is considered as"μ_{A1} (i_1) μ_{A2} (i_2) μ_{A3} (i_3) μ_B (t)". After calculating the degree of all rules, the ones with maximum degree can be kept.

After allocating a degree to each rule and omitting the conflicting rules, all of the twenty seven spaces were not covered. Therefore, the interpolation method was used to design rules for the empty spaces. Finally, twenty seven fuzzy rules were created to cover all the input space. Fuzzy rules are considered as the following general from:

If input1 (Conspicuity map1) is (L1, M1 or H1) and input2 (conspicuity map2) is (L2, M2or H2) and input3 (conspicuity map3) is (L3, M3orH3) then output (saliency map) is (Lo, Mo or Ho).

3.5 Designing fuzzy system based on fuzzy rules

The proposed system was designed based on Mamdani Fuzzy Interface System after generating if-then rules with look-up table method. Moreover, for "and" and "or" functions, "min" and "max" operators were applied respectively. It should be noted that the Center of Gravity defuzzifier was utilized to obtain the exact output. Three inputs are given pixel by pixel to the fuzzy system in order to be combined through fuzzy rules. The saliency map is formed as the output of the system. Three conspicuity maps are combined pixel by pixel by fuzzy rules to get the corresponding pixel of the output saliency map. After generating all pixels of the saliency map in the output of the FIS, the salient point can be found with winner-take-all network.

4. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the results of implementing the modified fuzzy model on emergency triangle database have illustrated to show the effectiveness of our modified model. This database comprises *32 training images* with their corresponding target masks for designing rules and another *32 test images* of emergency triangle for testing. The MATLAB Programming language was applied for implementing the algorithms. The size of all images is (640×480) pixels and the conspicuity maps are extracted in the size of (40×30) pixels. In addition, the results of comparing the fuzzy combination method with other combination methods are presented in this section. Our proposed strategy not only applies nonlinear data fusion instead of simple superposition, which is more convincing biologically, but also combines top-down information with bottom-up features. On the other hand, due to creating combination rules using target masks available for targets, our model can modulates a competition for task-relevant objects in object recognition purpose.

4.1 Fuzzy combination results

Two images out of total 32 test images are illustrated in the top row of the Fig.4. First of all, three conspicuity maps are

extracted with the basic saliency model [7] that could be seen on the next rows below each figure on the top row. The pixels of these three conspicuity maps are fused through *Fuzzy Interface System* with 27 designed rules instead of linear summation in basic model [7]. The final saliency map is generated in the output of the FIS. As mentioned before, the designed rules are in direction of highlighting the target more than surrounding that makes the target more conspicuous in the final saliency map. On the other hand, the top down information or previous knowledge of the viewer is considered in the model. This leads to a goal-directed search among the target and other clutter.

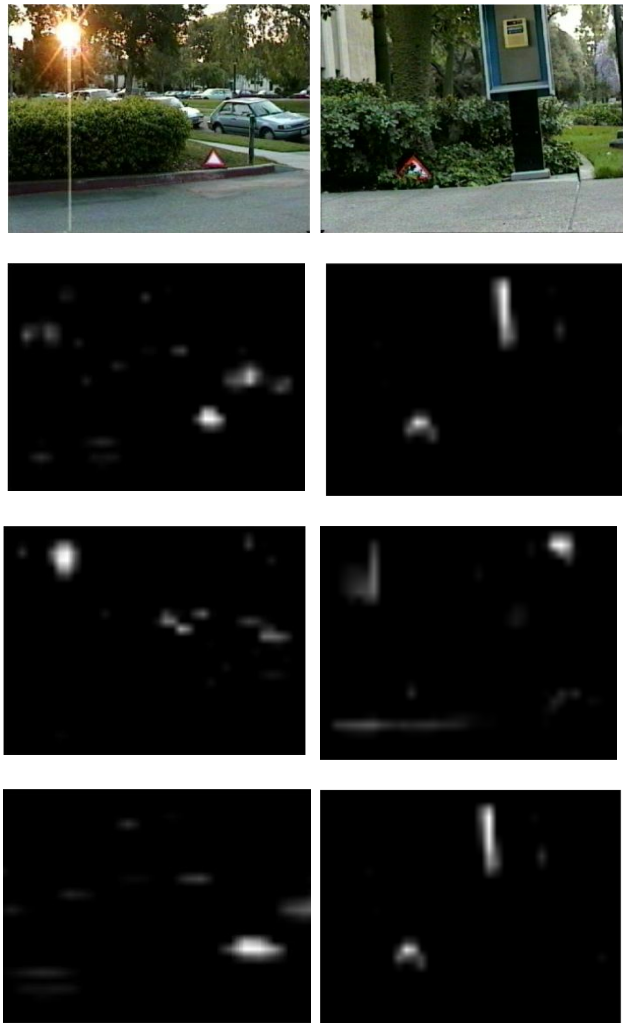


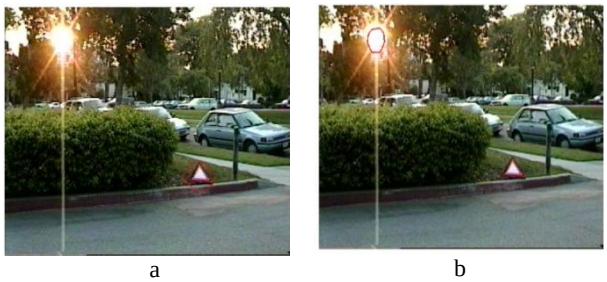
Fig 4: Two of the total 32 test images and their three conspicuity maps extracted by basic visual attention model [7]

As could be seen in the Fig.5, after combining three conspicuity maps in the Fig.4 for two test images with *basic model* and *fuzzy rules*, the saliency maps were generated. Fig.5.c, g and Fig.5.d, h, show the saliency maps that were generated by *Fuzzy Interface System* and by the *basic saliency model* [7] respectively. As it is shown in Fig.5.a, b, the fuzzy model found the target (emergency triangle) in the first hit even with the existence of sun brightness, while the basic model found the sun as the first salient point that is shown with the red contour. As the fusion rules were designed in fuzzy model using training images with their corresponding target masks, these rules are in direction of strengthening the target in the final saliency map and weakening the distracters. As it is illustrated in the Fig.5.c, in the saliency map generated by fuzzy combination rules, the brightness of the sun has weakened so that the model has found the emergency triangle as the salient point in the first hit. As could be seen in the Fig.5.e, f, the target was detected in the first hit with fuzzy fusion combination rules while in the basic model the emergency triangle was not detected in the first hit.

Table.1 shows the results of fuzzy combination rules and naive superposition with two parameters. The number of trials in all test images in which the target was detected in the first hit is called the *No. of .Zeros* parameter, and the number of trials in which the target was not detected before five hit (five times of running the algorithm) is called *No. of. UST* (number of unsuccessful trials). As could be seen in Table.1, the naive summation of conspicuity maps in the basic model of visual attention [7] has detected the emergency triangle in the first hit in *9 images* of 32 total test images. Moreover, in 5 images the target could not find in the first hit with naive superposition. However, with fuzzy combination rules, in *17 images* of the 32 total images, the target was detected in the first hit, and also there were 4 unsuccessful trials. The *No. of Zeros* parameter is so valuable in object detection and recognition application because it indicates the number of trials in which the target has detected in the first hit. As there was a very remarkable improvement in this parameter, this modified model is suitable as object detection application.

Table 1.The results of naive superposition and fuzzy combination methods with *No. of Zeros* (the number of trials in which the target was detected in the first hit) and *No. of UST* (the number of trials in which the target was not detected before five hit)

Combination method	No. of. Zeros	No. of.UST
NaiveSuperposition	9	5
Fuzzy Combination Rules	17	4



a b

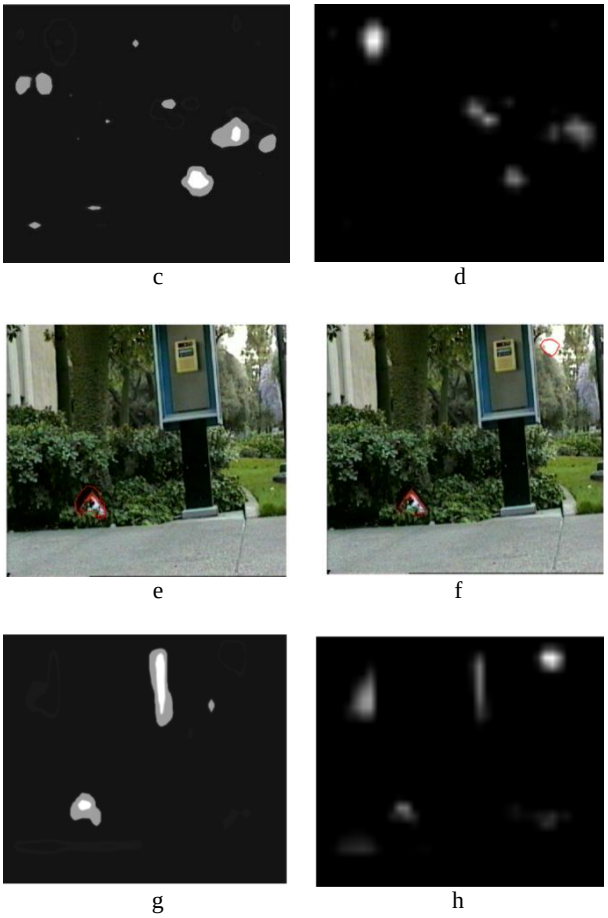


Fig 4: (a),(e) Two samples of test images in which the target has detected in the first hit using fuzzy combination rules. (c), (g) the saliency maps of two test images created using fuzzy fusion rules. (b), (f) two samples of test images in which the target has not detected in the first hit using basic saliency model (d), (h) the saliency maps of two test images created using basic saliency model.

4.2 Comparison of fuzzy combination method with other proposed combination methods

Our proposed modified model is compared with five other computational models of visual attention which have been studied on the feature map combination methods [7], [12], [17]. These approaches are *"Naive superposition*" [7], "*Genetic Algorithm combination*" [17], *"Harmonic Mean combination*" [12]. The results of comparison are illustrated in Table .2. Two parameters are used as comparison parameter: mean and standard deviation of false detection before finding the target that is called *Mean* and *STD* respectively [14].

As illustrated in Table.2, there was considerable improvement in finding the desired object (emergency triangle) in the mean and standard deviation of false detection before finding the target. As could be seen in the Table.1, the genetic algorithm combination method [17] had the better results compared to naive superposition. Genetic algorithm method weighs the feature maps in direction of finding the target using target mask information [17]. Because of weighing the feature maps purposefully using target mask information, it had better results compared to naive superposition. Moreover, harmonic mean combination approach [12] had the better result than naive superposition and genetic algorithm with 1.15 and 1.51 for *Mean* and *STD* respectively. In harmonic mean approach the feature maps are combined with harmonic mean formula [12] that is a kind of parallel data combination. As illustrated in Table.2, fuzzy combination approach had the best results among other mentioned data combination methods with 0.85 and 1.20 for *Mean* and *STD* respectively. As a result,there was fewer numbers of false detections in fuzzy fusion approach compared to other combination strategies.

Table2:Three combination strategies compared to fuzzy combination method with *Mean* (Mean of false detection before finding the target) and*STD* (standard deviation of false detection before finding the target) parameters

Combination Method	Mean	STD
Naive Superposition	2.44	2.20
Genetic Algorithm	1.92	2.01
Harmonic Mean	1.15	1.51
Fuzzy Combination Rules	0.85	1.20

5. CONCLUSION

We have proposed a novel fuzzy feature combination strategy in this paper that could be a suitable substitution for previous traditional linear combination which does not seem reasonable biologically. This combinational approach combines three conspicuity maps in basic model of visual attention not only with using bottom-up features but also with applying top-down information. These top-down cues are considered through designing fuzzy combination rules using target mask information for training images. Three conspicuity maps were extracted from each training images. The pixels of these conspicuity maps were applied as input-output fuzzy sets to generate fuzzy rules. 27 fuzzy if-then rules were created with look up table method and the *MamdaniFuzzy Interface System* was designed. This system acts as a nonlinear combination operator and it takes three conspicuity maps as its inputs and generates the final saliency map in the output. Because of designing fuzzy rules using target mask information available for each training image, these rules are designed in direction of strengthening the target and weakening the distracters. The experimental results showed the effectiveness of this new fusion rules and its comparison with other combination rules. It also demonstrated the remarkable improvement of fuzzy fusion approach in finding the target in the first hit with high accuracy in comparison with other methods, which is valuable in object detection and recognition application.

6. ACKNOWLEDGEMENT
Datasets are downloaded from Itti's Lab at USC.

7. REFERENCES
[1] N. Ouerhani and H. Hugli, "Multi scale attention-based pre segmentation of color images", *4th International Conference on Scale-Space theories in Computer Vision*, Springer Verlag, LNCS, Vol. 2695, 2003, pp. 537-549.

[2] S. Baluja and D.A. Pomerleau, "ExpectationBased Selective Attention for Visual Monitoring and Control of a Robot Vehicle," *Robotics and Autonomous Systems*, vol. 22, no. 3-4, 1997, pp. 329-344.

[3] D. Walther, U. Rutishauser, Ch. Koch, and P. Perona. "Selective visual attention enables learning and recognition of multiple objects in cluttered scenes," *Computer Vision and Image Understanding*, Vol. 100 (1-2), 2005, pp. 41-63Tavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd.

[4] D. Walther, "Interactions of Visual Attention and Object Recognition", PhD Thesis, *California Institute of Technology*, 2006.

[5] D. Walter, L. Itti, M. Riesenhuber, T. Poggio, and C. Koch, "Attentional selection for object recognition – a gentle way", *LNCS*, Vol. 2525, 2002, pp. 472-479.

[6] C. Koch and S. Ulman ."Shifts in selective visual attention: towards the underlying neural circuitry", *Human Neurobiology*, 4, 1985, pp. 219–227.

[7] L. Itti and C. Koch,ENiebur."A model of saliency-based visual attention for rapid scene analysis".*IEEETransactions on Pattern Analysis and Machine Intelligence*. 1998.

[8] L. Itti and C. Koch, "Computational modeling of visual attention.",*Nature Reviews Neuroscience*, 2(3), 2001, pp. 194–203.

[9] J. Wolfe. "A revised model of visual search", *Psyconomic Bulletin Review*, 1(2), 1994, pp. 202–238.

[10] S. Frintrop. "VOCUS: A Visual Attention System for Object Detection and Goal-directed Search", *LNAI*, 3899, Springer Berlin/Heidelberg. ISBN: 3-540-32759-2, 2006.

[11] V. Navalpakam, J. Rebesco, and L. Itti. "Modeling the influence of task on attention", *Vision Research*, 45(2), 2005, pp.205–231.

[12] H. Bahmani, A.M. Nasrabadi, and M.R. HashemiGholpayeghani. "Nonlinear Data Fusion in Saliency Based Visual Attention." *4th International IEEE Conference in Intelligent System*, 2008.

[13] L. Itti and C. Koch. "Feature combination strategies for saliency-based visual attention systems". *Journal of Electronic Imaging*, 2001,10(1):161–169.

[14] S. Mitra, and S. K. Pal."Fuzzy sets in pattern recognition and machine intelligence" *Fuzzy Sets and Systems* .156, 2005, pp.381–386.

[15] M. Sugeno, "An Introduction Survey of Fuzzy Control". *Information Sciences*. 36, 1985, pp 59-83. 1992, pp 1414-1427.

[16] E. H. Mamdani, "Advances in the linguistic Synthesis of Fuzzy Controllers", *International J. Man-Machine Stud*., 8(6), 1976, pp.669-678.

[17] Z. Armanfard, H. Bahmani, A. M. Nasrabadi. "A Novel Feature Fusion Technique in Saliency-Based Visual Attention".*Advances in Computational Tools for Engineering Applications*, 230-233, 2009.