ISSN No. 0976-5697

Volume 3, No. 3, May-June 2012

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Region of Interest Detection Based on Dynamic Interaction and Its Evaluation in Perspective of Cognitive Psychology

Md. Rokunuzzaman
Dept. of Mechanical Engineering,
Rajshahi University of Engineering & Technology (RUET)
Rajshahi-6204 , Bangladesh.
rzaman_mctr@yahoo.com

Mahmudul Hasan*
Dept. of Mechanical Engineering,
Rajshahi University of Engineering & Technology (RUET)
Rajshahi-6204 , Bangladesh.
shakil_ruetbd@yahoo.com

Barun Kumar Das
Dept. of Mechanical Engineering,
Rajshahi University of Engineering & Technology (RUET)
Rajshahi-6204 , Bangladesh.
barun_ruet@yahoo.com

Md. Abdul Kader
Dept. of Mechanical Engineering,
Rajshahi University of Engineering & Technology (RUET)
Rajshahi-6204 , Bangladesh.
E-mail: kader_042017@yahoo.com

Abstract: This paper describes an approach to detect the Region of Interest (ROI) in Video Frames where different objects are related with dynamic interaction. A mathematical frame work for determination of the ROI by formulating dynamic interaction in various contexts has also been proposed. This paper introduces a new methodology of evaluating ROI which reflects human psychology of cognitive vision. Various experiments are presented to illustrate the proposed method which gives exciting and promising results.

Keywords: Dynamic interaction, human psychology, evaluating ROI, cognitive value, academic contribution

I. INTRODUCTION

The region of interest (ROI) is a particular region in a scene in which a robotic agent is interested with when considering robotic vision. Automatic ROI selection is very useful and very challenging task for robotic vision due to its psychological behavior. Automatic exploration of unstructured environment, robot cooperation or human robot interaction can be greatly assisted by robotic vision which has the ability of recognizing the world by its psychological behavior. ROI selection is a technique to extract useful data from enormous information. To choose such an ROI, the dynamic interaction can be a good choice. This is because; when the objects are in interaction several important things happen which attracts our attention. However, to determine the best interaction, there need to select a cognitive boundary which covers the important objects with interaction. The selection of this cognitive boundary by human itself is difficult. This is because humans have different psychology of interest and decision making criteria. Then how will such boundary be defined autonomously? What things are to be included and what things are to be excluded from this boundary? The size of ROI is also important. Bigger size of ROI can express more details of the scene but it also increases irrelevance and ambiguity in the scene. In the ROI evaluation part, how will its cognitive value be determined? This cognitive value must represent significant and interesting information. This research answers these questions by relating ROI with cognitive psychology of visual perception.

In order to select the cognitive boundary or ROI the autonomous agents should have the perception of interaction between objects. Interaction can be defined by three factors, namely 'closeness', 'synchronity' and 'causality'. Closeness can be measured by the relative distance between objects. Synchronity is a temporal property. Motion is one of its candidates. Causality is logical or inference based. Before knowing what is happening, the robot must detect the perceptual regions or objects of interest. Visual scenes contain these cues. In order to perceive 'interaction', these cues should be integrated first. Therefore, we need to detect the ROI based on these cues which makes the object interesting. In our research we try to relate the selection of interesting objects with Human interest.

In the ROI evaluation, we have applied the principles of cognitive psychology of visual perception. The term perception, in its broad usage, refers to the overall process of apprehending objects and events in the external environment [1]. We have considered human behavior in visualizing the world by object based perception strategy [2] and relevance [3] for formulation of ROI evaluation based on attention and interaction.

Different methods have been proposed to detect regions of interest in an image or video. Some of the related researches are: Dynamic Interaction of Object- and Space-Based Attention in Retinotopic Visual Areas [4], An Interactive Region-based Image Clustering and Retrieval Platform [5], Automatic target detection and recognition in the process of interaction between visual and object buffers of scene understanding system based on network-symbolic models [6], Analysis of Object Interactions in Dynamic Scenes [7], Dynamic ROI transcoding for multipoint video conferencing [8], automatic detection of salient objects and spatial relations in video [9], A model of saliency-based visual attention for rapid scene analysis [10], dynamic ROI acquisition and face tracking for intelligent surveillance system [11]. In the ROI evaluation part, very few literatures are relevant with our method. Clauss et al. [12] present an evaluation of ROI based on attention algorithm using probabilistic measure which handles situations of unordered ROIs. Huesman et al. [13] has proposed a region of evaluation in computed tomography with the calculation of statistical uncertainty. Paulo estimated video object's relevance using segmentation and evaluated with human observer in [14].

The existing algorithms of ROI detections are mostly based on attention algorithm. These algorithms select the salient object individually for each of the frame. Some of the algorithms are based on semantic or graph network.
Interactions are detected by the use of prior knowledge or training algorithms like SVM (Support Vector Machines) or GA (Genetic Algorithm). However, only selection of salient regions has little meaning than relating other objects of context with the salient object. Moreover, existing algorithms lacks of considering psychological behavior in selecting ROI. Furthermore, ROI is evaluated through some statistical measure or entropy based information measure. These algorithms also neglect the psychological behavior in evaluating the ROIs which is very important for intelligent robot systems. Therefore, we have formulated a new technique for ROI detection which is based on dynamic interaction considering all possible human psychological aspects for visual perception. We have proposed an evaluation function for ROI evaluation based on human psychology of relevance and interaction.

The rest of the paper is organized as follows. Section 2 gives an overview system architecture and method on which we implement our algorithm. In Section 3, we give an overview of the proposed framework and explain our approach for dynamic interaction formulation, ROI determination and Evaluation. We present experimental results by using video movie and evaluate the performance of our system in Section 4. We conclude with possible future improvements to the system in Section 5.

II. SYSTEM ARCHITECTURE

A. System Configuration

Table I shows the system configuration that is used to implement the proposed method.

TABLE I
SYSTEM CONFIGURATION

No	*Items*	*Specifications*
1	Vision Processor	Intel Core2 Duo, 2,20 GHz, 2.0 GB of RAM
2	Vision Sensor	Canon PTZ Camera Model: VC50i
3	Development Platform	Microsoft Visual Studio 2005
4	Programming	C++, Visual C++
5	Code development	Intel's Open CV Library

Several videos of different context have been captured in the current work and saved into memory of a Personal Computer (PC). The video frames are further processed through various image processing programs developed with Intel's Open Computer Vision (Open CV) Library. The codes are compiled by Visual C ++ of Microsoft Visual Studio 2005.

B. Overview of the Method

Fig.1 shows the process adopted in the present work to detect the ROI and evaluate it based on dynamic interaction in perspective of cognitive psychology. Video is acquired from camera and processed by image processing techniques to extract object information. The motion of each object is estimated and motion saliency is computed. The dynamic interaction is formulated after motion saliency computation. The ROI is formed by combining salient object and interacting object. The ROI detection is then compared with eye search databases. ROI is evaluated by the evaluation function of the method and compared with human evaluation psychology.

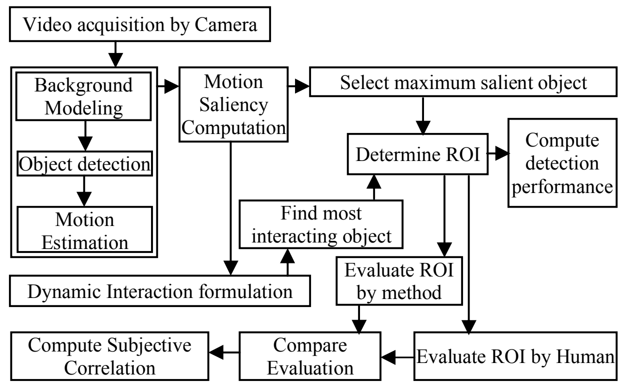

Fig.1. Overview of the method

III. MATHEMATICAL FRAMEWORK

This section gives a mathematical formulation of the ROI selection based on dynamic interaction and evaluation of ROI problem motivated from human psychology of visual perception. This mathematical framework consists of Image Processing formulation, Motion Saliency computation, Dynamic Interaction formulation, ROI Size determination and ROI Evaluation

A. Image Processing

The image processing involves background modeling, object detection by filtering and motion estimation

1. Background Modeling

In dynamic interaction, it is difficult to have a stable background. Therefore, it is essential to update the background with time. The visual scene contains foreground and background. Background is that portion of the scene which appears frequently and has low variance in contrast with foreground. To model such background we use adaptive background model proposed by Stauffer et al [15]. According to his theory, each pixel of a scene is modeled as a mixture of K Gaussian distributions with probability

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} * \eta\left(X_t, \mu_{i,t}, \Sigma_{i,t}\right) \tag{1}$$

Where K (=1 ~5) is the number of distributions, $\omega_{i,\,t}$ is an estimate of the weight (what portion of the data is accounted for by this Gaussian) of the ith Gaussian in the mixture at time t, $\mu_{i,\,t}$ is the mean value of the ith Gaussian in the mixture at time t, $\Sigma_{i,\,t}$ is the covariance matrix of the ith Gaussian in the mixture at time t, and where η is a Gaussian probability density function

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu_t)^T \Sigma^{-1}(X_t - \mu_t)} \qquad (2)$$

The prior weights of the K distributions at time t, $\omega_{k,t}$, are adjusted as follows

$$\omega_{k,t} = (1-\alpha)\omega_{k,t-1} + \alpha(M_{k,t}) \qquad (3)$$

Where α is the learning rate and $M_{k,t}$ is 1 for the model which is matched and 0 for the remaining models. In this algorithm every new pixel value is checked against existing K-Gaussian distributions until a match is found. A match is defined as a pixel value within 2.5 standard deviations of a distribution.

Background model in this theory is considered as the distribution which has higher evidence (ω) and lower variance (σ).The first distribution is chosen as background model, where

$$B = \arg\min_b \left(\sum_{k=1}^{b} \omega_k > T\right) \qquad (4)$$

Where, T *is* a measure of the minimum portion of the data that should be accounted for by the background.

Due to slow adaptation problems of the Stauffer-Grimson Model, an improved version is developed in [16]. In this improved version, the weight is updated by

$$\omega_{k,t} = \omega_{k,t-1} + \frac{1}{t}(M_{k,t} - \omega_{k,t-1}) \text{ (Sufficient Statistics)} \qquad (5)$$

$$\omega_{k,t} = \omega_{k,t-1} + \frac{1}{L}(M_{k,t} - \omega_{k,t-1}) \text{ (L-recent window)} \qquad (6)$$

Similar update equations can be formed for mean and variance. The L-recent window update equations gives priority over recent data therefore the tracker can adapt to changes in the environment.

2. Object Detection

Human visual system recognizes the world based on perceptual groups of items. These perceptual grouping suggests about object based recognition. For the detection of objects, we have separated background from foreground first and then taken the foreground objects as blobs by sequences of 8 connectivity components. Desired objects were also filtered out by putting a threshold based on its area pixels

3. Motion Estimation

Motion is an important cue to dynamic interaction. In this method a very simple approach has been used to estimate motion. A blob in first frame and its coordinates were detected. When the frame advances, the blobs in the next frame were detected. Then the distance between the blob's coordinates in the next frame and previous frame were estimated. The nearest blob which has minimum distance with the blob of previous frame was then computed. We associate the blobs between frames by the nearest blob. It was assumed that the speed of the object moving from one frame to another frame is not so fast that camera cannot detect. Therefore, we can associate the blob between frames and this method is used to estimate the motion of the blobs. The motion has been estimated by calculating each blob's center to center Euclidean distance

$$M = \sqrt{(dx_n^{n+1}(i))^2 + (dy_n^{n+1}(i))^2} \qquad (7)$$

Where, $dx_n^{n+1}(i)$ and $dy_n^{n+1}(i)$ are the center to center distance of the ith object from n to $n+1$ frame for x-coordinates and y-coordinates respectively.

B. Motion Saliency Computation

Motion saliency denotes the conspicuous state of an object in a video. It is based on the detection of motion, which defines instantaneous speed vectors for each object in a scene. Motion saliency helps to detect moving objects whose motion is discontinuous to its background.

In the current method, motion saliency is expressed as a value which is a difference in motion of each object with minimum among all the objects at that instant. If M_i is the motion of the ith object and $M_{\min}$ = min { M_i ….. M_n }, where, i = 1, 2,....n are the number of objects in the frame at that instant, then motion saliency value can be expressed as

$$M_{sv} = (M - M_{\min}) \qquad (8)$$

To obtain the value as a factor ranging from 0 to 1, we normalize it with its maximum value as

$$M'_{sv} = M_{sv} / \max(M_{sv}) \qquad (9)$$

Insignificant motion saliency value can cause the system to irresponsive to interaction. Therefore, we need to set a weight for this. This weight can be pixel information of the object in motion. This is because; one of the aspects of human vision system is that it attains objects with larger area as it covers most of the portion of the retina. Based on this concept, we introduce a term of this weight as information density defined as

$$I_D = A_O / A_{OBB} \quad , 0 \le I_D \le 1 \qquad (10)$$

Where, A_O is the area measured by number of pixels inside the object and A_{OBB} is the rectangular area of the box which fits the periphery of the object.

Hence, the weighted motion saliency value is

$$M''_{sv} = M'_{sv} \times I_D \qquad (11)$$

C. Selection of Maximum Motion Salient Object

Motion salient objects are scored according to the weighted motion saliency value computed with Eq.11. Then the maximum value was selected from all the values of the object. Finally, the object which corresponds to maximum motion saliency value was determined and selected as a maximum motion salient object.

D. Dynamic Interaction Formulation

Interaction between objects depends on their motion saliency and their closeness. In visual information, only physical interaction is viewed. Therefore, causal effect cannot be experienced. In order to select the ROI, we need to detect what is interesting. Usually, human vision system is very sensitive to motion objects when dynamic situation is considered and get interested when there are some interactions. Based on this psychological behavior we devise an interaction detector what we name it as "interaction factor". The value of this factor will determine how much the object is interacting. The higher the value, the more is the interaction. The object with maximum interaction factor will be in focus of the attention and will be confined by the ROI. In order to formulate Interaction factor we need to define closeness. Closeness can be mathematically defined by the relative distance between objects. Usually closer objects are relevant also. In perspective of psychology, human tries to observe relevant things. Relevant things become interesting when it is nearer to its maximum salient things. Therefore, relative distance is a measure of relevancy as well as closeness for interaction. If we denote, D_R as relative distance then it is simply an Euclidean distance of the surround objects from the maximum salient object and formulated as

$$D_R = \sqrt{(Cx_{MaxSalObj} - Cx_i)^2 + (Cy_{MaxSalObj} - Cy_i)^2} \quad (12)$$

Where, $Cx_{MaxSalObj}$ is the center x coordinate of the maximum salient object and Cx_i is the ith object's center x coordinate except maximum salient object. The second term applies for y coordinate.

Now, we can define an interaction factor as a ratio of weighted motion saliency value to relative distance. If we denote Interaction Factor as I_F then it is expressed as

$$I_F = M''_{sv} / D_R \quad (13)$$

Where, M''_{sv} and D_R are defined in Eq.11 and Eq. 12 respectively.

E. ROI Size Determination

Our algorithm selects the ROI when there is a dynamic interaction in the scene, therefore includes most salient object and most interacting object by the maximum value of interaction factor. We assume our ROI as rectangle which covers two objects, namely most salient and most interactive. The reason behind using two objects is based on human psychology of attention. For example, if two people want to talk to another person at the same time, the person who is listening cannot pay attention to both at the same time. He or she prefers to listen from the person who is more salient. Similarly our ROI is taking the objects which are more salient than others.

Let us have *n* number of salient objects. The most salient object is determined by the maximum saliency value and will be in the ROI first. Then from *n-1* objects, I_F is calculated and the object which corresponds to maximum I_F is determined. Let R_1 is the rectangle which selects maximum salient object and R_2 is the most interactive object based on I_F. Fig.2 illustrates the ROI size, S_{ROI} by $R_1$and R_2, where, the sides of the ROI rectangle are computed as

$$W_{ROI} = \{(Max\, x \mid x \in R_1, R_2) - (Min\, x \mid x \in R_1, R_2)\} \quad (14)$$

$$H_{ROI} = \{(Max\, y \mid y \in R_1, R_2) - (Min\, y \mid y \in R_1, R_2)\} \quad (15)$$

Hence, the ROI size is determined by

$$S_{ROI} = W_{ROI} \times H_{ROI} \quad (16)$$

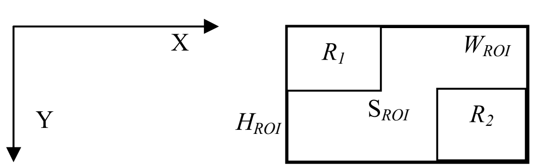

Fig.2 ROI size determination

F. ROI Evaluation

ROI evaluation is necessary to determine the most interesting information in each video frames. Interesting information should be important as well as interactive. Since saliency gives important information and saliency of the most interacting object gives interactive information, therefore, we can combine this information to form an interest value for ROI. Moreover, bigger size of ROI contains more information relative to its image size. Therefore, this interest value should be weighted by this size ratio, S_r to formulate an Evaluation function for ROI.

To realize this equation according to the idea mentioned above let's formalize it as follows:

Let the Image size S_I, most interesting information I_{MI}, andsize ratio, S_r are defined as follows

$$S_I = \mathrm{Im}_w \times \mathrm{Im}_H \quad (17)$$

Where, Im_w and Im_H are image width and image height in pixels respectively. The maximum interactive value of object is determined by its weighted motion saliency value corresponding to maximum I_F denoted by $M''_{sv}|_{\max(I_F)}$. Making the summation we obtain,

$$I_{MI} = \max(M''_{sv}) + M''_{sv}|_{\max(I_F)} \quad (18)$$

$$I_{MI} = \max(M''_{sv}) + M''_{sv}|_{\max(I_F)} \quad (19)$$

Where S_{ROI} is the size of ROI defined in eq. (16), Evaluation of ROI is quantified by EF_{ROI} which can be expressed

$$EF_{ROI} = I_{MI} / S_r$$
$$= (\max(M''_{sv}) + M''_{sv}|_{\max(I_F)}) \times S_{ROI} / S_I \quad (20)$$

IV. EXPERIMENTAL RESULTS

A. ROI Detection Based On Interaction

Fig.3 shows some video results of ROI detection based on interaction. This figure illustrates interactions between hand and cup, hand and mouse, hand and pen etc. The ROIs are selected dynamically, when there is maximum interaction with these objects. When there is no interaction in the video, the ROI is absent showing that the video is no longer interesting.

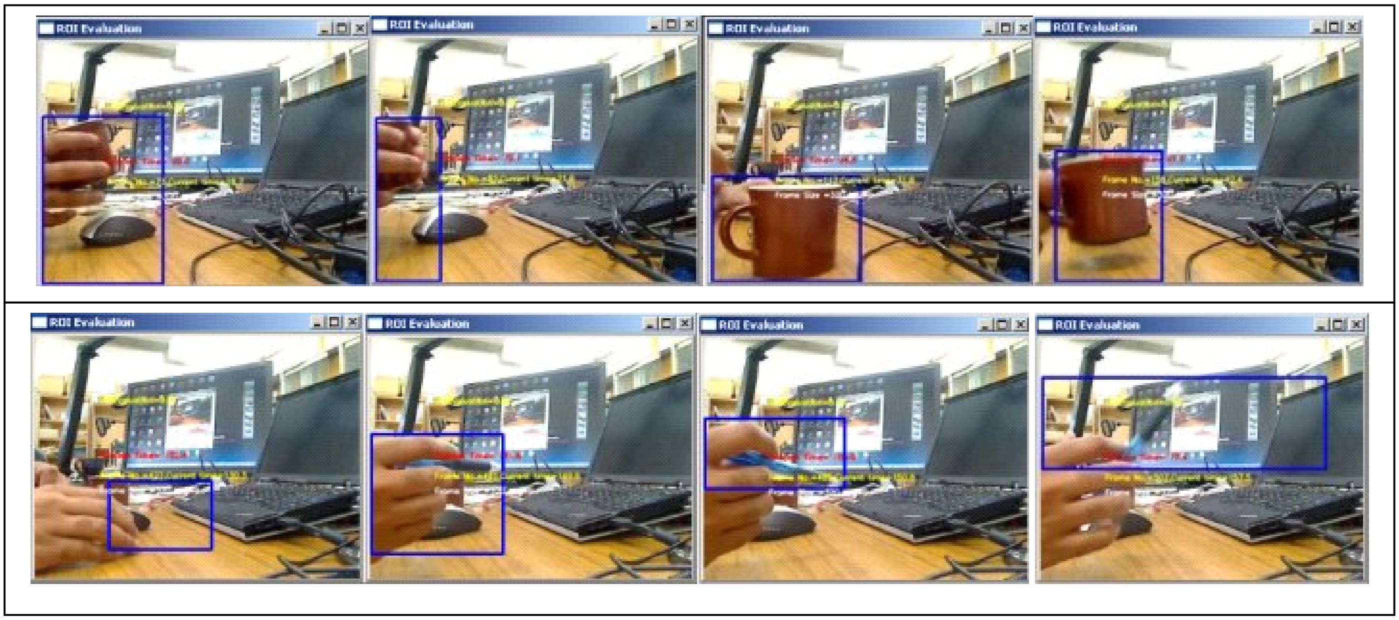

Fig.3 ROI detection based on interaction (Video output)

B. ROI Evalution Results

The selected ROIs are evaluated in real time by the Evaluation Function of the algorithm. Another interacting video results are shown by Fig. 4. The evaluation function quantifies the interesting information at each frame sequence for each selected ROI instantly. From the evaluation value at each frame sequences has significant meaning. When the ball is thrown to wall and the ball interacts with wall the evaluation value becomes higher than when it is in hand due to less interaction. If we observe this simple scenario in the perspective of cognitive psychology then we can find that our interest grows when the interaction becomes most salient.

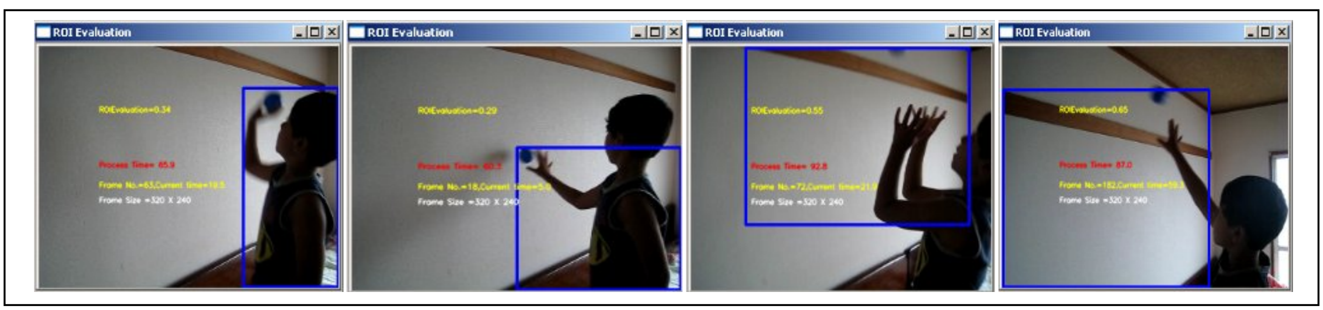

Fig.4 ROI evaluation results of a video showing a kid is throwing a ball on the wall and catching sequences as an example of interaction.

C. Detection Accuracy Calculation Using ROC Analysis

We have used ROC (Receiver Operating Characteristic) curve to determine the performance of our ROI detection method. We set the ground truth like this:

True Positive (*TP*): If the ROI is detected at the locations where we are interested, then it is TRUE positive.

False Positive (*FP*): If the ROI is detected at wrong location then it is FALSE positive.

Right and wrong locations of ROI are determined by the detection result of eye tracker and our decision. We count the total number of *TP*s and *FP*s among several frames of a test video. Then we calculate Hit Rate, H and False Alarm Rate, F by the following equations:

$$H = \frac{TP}{N} \tag{21}$$

$$F = \frac{FP}{N} \tag{22}$$

In Fig.5 data points for ROC graph are obtained by plotting H and F value for each test videos.

Due to the nature of discrete classifier, we obtain only one data point for each video. However, with only one point we cannot determine the area under the curve in usual way. To complete the ROC curve, we draw a line passing through origin and data point and another line from upper right (1, 1) and data point for each result as proposed by a method [17].

To have an Index of discriminability of detection performance, we calculate the area under curve (AUC) according to this method as:

$$AUC_{video1} = 0.75+0.25\times (H-F) - (1-F) \times H$$
$$= 0.75+0.25\times (0.7-0.3) - 0.3\times (1-0.7) = 0.76$$
$$AUC_{video2} = 0.75+0.25\times (0.8-0.2) - 0.2\times (1-0.8) = 0.86$$

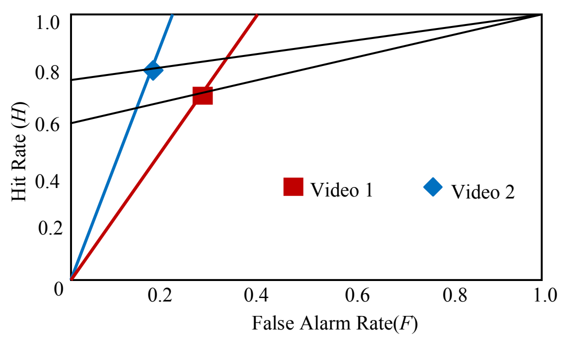

Fig.5 ROC curve for ROI detection accuracy (Video Output)

With Human eye fixation data, these results can be compared to other existing methods. The comparative results are shown by table II as follows:

TABLE II
COMPARATIVE DETECTION ACCURACIES

No	*Method*	*ROC area*
1	**Proposed Method**	**0.86**
2	Neuro-Vision Tool	0.75
3	Itti et.al	0.69
4	Informax	0.72
5	Saliency Tool Box(STB)	0.74

D. Subjective Correlation of ROI Evaluation

ROI evaluation or justification of its selection is entirely subjective. This is because, the human have different psychologies of interest in ROI selection. The same ROI can be evaluated by different scores by different people. Then how can we examine the effectiveness of this method of ROI evaluation? As ROI selection or evaluation is based on human psychology, we have to compare our results with evaluation of ROIs by Human evaluators. To make this correlation, we first design a scoring system for evaluation of each ROI at each frame as shown by Table III.

TABLE III
SCORING SYSTEM FOR ROI EVALUATION

Point range	*Assessment (A)*	CV_{assess}
0.0	Unjustified	0
0.1~0.4	Poor	25
0.50~.55	No discrimination	35
0.56~0.65	Good	60
0.66~0.75	Very good	70
0.76~0.85	Best selection	80
0.86~0.95	Excellent	90
0.96~1.0	Fully Justified	100

To compare different assessments (A) we need a correlation between assessments. This we define as Correlation value on assessment denoted by CV_{assess}. Assuming a 100 scale we assign its value for comparative assessment. To evaluate ROI we also assign some point range from 0~1 so that it can be compared with our EF_{ROI}.

Subjective Correlation is a quantitative measure of comparative assessment between human evaluation and systems evaluation of ROI. Let Subjective Correlation is denoted by CV_{subj}, assessment by Method is A_M and assessment by Human is A_H and difference in correlation value correspond to assessment is ΔCV_{assess}, then

Subjective correlation, $CV_{subj} = S - (\Delta CV_{assess})$ (23)

Where, S is the scale we choose. Here we have chosen the scale as, S=100. Therefore, subjective correlation is found in percentage.

If $A_H = A_M$, then the subjective correlation is 100. Otherwise it is calculated by Eq. (23) using the correlation values on assessment from Table III. To compare ROI evaluation by the method, 20 human evaluators evaluates the same video sequences and give a justification value ranges from 0~1 for each ROI of each sequence based on attention and interactiveness. The Evaluation Factors are not provided to them so that their decision is not influenced by it.

To determine the ROI evaluation performance, we compare ROI evaluation by the method (EF_{ROI}) with this human justification value (JV) and compute a subjective correlation as an assessment performance. The reason behind this comparison is to see how our method is imitating human decision.

We have tabulated the evaluation results for video2 as a sample result, by examining several selected video sequences in Table IV.

TABLE IV
SUBJECTIVE CORRELATION AS ROI EVALUATION PERFORMANCE

ROI Index	*ROI Evaluation*		*Subjective Correlation*	*Average CVsubj(%)*
	Human JV	*Method* EF_{ROI}	*CVsubj (%)*	
1	0.40	0.20	100	
2	0.50	0.60	75	
3	0.53	0.59	75	
4	0.55	0.10	90	
5	0.68	0.71	100	
6	0.61	0.10	65	
7	0.28	0.21	100	85
8	0.42	0.10	100	
9	0.51	0.55	100	
10	0.58	0.30	65	
11	0.66	0.35	55	
12	0.52	0.40	90	

The average subjective correlation of ROI evaluation is found 85%. This means that our method is imitating human decision making most of the time. Therefore, we can conclude that the system is consistent with human evaluation psychology when it is evaluating ROIs in real time.

V. CONCLUSION

This paper introduced a new technique of ROI detection. Interaction is very useful phenomenon in real life. For interaction detection, this method does not require prior knowledge or training or motion history. This method is more psychological based compared to other ROI detection techniques. In robot-cooperation, there need to share a ROI which should be most interesting to all agents. For this purpose there need to extract significant features in real time. This problem becomes more crucial when the interacting objects are viewed from different perspectives by the robotic vision system. The existing methods use conventional image processing techniques which are not efficient in terms of processing speed and detection accuracies. However, current method is independent to all viewing perspectives and very intelligent to select the ROI in real time. Moreover, our method of ROI evaluation is unique from other evaluation algorithm in terms of quantifying perceptual and cognitive value of ROI in real time. Therefore, this research has an academic contribution in the field of cognitive psychology where concrete formalization is rarely available for applied systems. In future, it is expected to extend the current approach to generate common region of interest generation among different robots with different cognitive ability.

VI. REFERENCES

[1] D. J. Levitin, "Foundations of cognitive psychology: core readings," in *MIT Press*, Illustrated ed., 2002, pp. 133-188.

[2] E. A. Styles, "The Psychology of Attention," in *Psychology Press*, Illustrated ed., reprint, 1997, pp. 87-112.

[3] S. D. Cara, "Relevance theory explains the selection task," *Cognition*, vol.57, 1995, pp. 31-95.

[4] N. G. Muller and A. Klienschmidt, "Dynamic Interaction of Object-and Space-Based Attention in Retinotopic Visual Areas," *J. Neuroscience*, vol.23, no.30, pp. 9812-9816, 2003.

[5] Y. Liu, X. Cheng, C. Zhang and A. Sprague, "An Interactive Region-based Image Clustering and Retrieval Platform," in *Proc. Int. Conf. on Multimedia & Expo,* Canada, 2006, pp. 929–932.

[6] K. Gary, "Automatic target detection and recognition in process of interaction between visual and object buffers of scene understanding system based on network-symbolic models," in *Automatic target recog. conf.,* USA, 2006, vol. 6234, pp. 62340A.1–62340A.11.

[7] B. Möller and S. Posch, "Analysis of Object Interactions in Dynamic Scenes," in *Springer Berlin*, Pattern Recognition, vol. 2449, 2002, pp. 361-369.

[8] C. W. Lin, "Dynamic Region of Interest Transcoding for Multipoint Video Conferencing," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 10, pp. 982–992, 2003.

[9] T. Sevilmis and M. Bastan, "Automatic detection of salient objects and spatial relations in videos for video database system," *Int. J. Image and Vision Computing*, vol.26, pp. 1384-1396, 2008.

[10] L. Itti and C. Koch, "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.

[11] Y. Kim and S. Park, "Dynamic region-of-interest acquisition and face tracking for intelligent surveillance system," in *Proc. SPIE,* vol.5299, pp. 929–932, 2006.

[12] M. Clauss, P. Bayerl and H. Neumann, "A Statistical Measure for Evaluating Regions-of-interest Based Attention Algorithms," *Pattern Recognition*, vol.3175, pp. 383-390, 2004.

[13] R. H. Huesman, "A new fast algorithm for the evaluation of regions of interest and statistical uncertainty in computed tomography," *Phys. Med. Biol.*, vol.29, no. 5, pp. 543-552, 1984.

[14] P. Correia and F. Pereira, "Estimation of Video Object's Relevance," in *Proc. of the European Signal processing conf. (EUPSICO),* 2000, pp. 925–928.

[15] C. Stauffer and W. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," in *Proc. IEEE Int. conf. of Computer Vision and Pattern Recognition,* 1999, pp. 246–252.

[16] P. KaewTraKulPong and R. Bowden, "An Improved Adaptive Background Mixture Model for Real time Tracking with Shadow Detection," in *Proc. 2nd European Workshop on Video Based Surveillance System,* 2001, pp. 135–144.

[17] S. Mueller and J. Zhang, "Upper and Lower bounds of area under ROC curves and index of discriminability of classifier performance," in *Proc. Int. Workshop on ROC Analysis in Machine Learning,* 2006, pp. 41–46.