27

Biologically Motivated Vergence Control System Based on Stereo Saliency Map Model

Sang-Woo Ban[a] & Minho Lee[b,c]
[a]*Dept. of Information and Communication Engineering, Dongguk University South Korea*
[b]*School of Electrical Engineering and Computer Science, Kyungpook National University South Korea*
[c]*Dept. of Brain and Cognitive Science, Massachusetts Institute of Technology USA*

1. Introduction

When the human eye searches a natural scene, the left and right eyes converge on an interesting area by action of the brain and the eyeballs. This mechanism is based on two attention processes. In a top-down (or volitional) processing, the human visual system determines salient locations through perceptive processing such as understanding and recognition. On the other hand, with bottom-up (or image-based) processing, the human visual system determines salient locations obtained from features that are based on the basic information of an input image such as intensity, color, and orientation. Bottom-up processing is a function of primitive selective attention in the human vision system since humans selectively attend to a salient area according to various stimuli in the input scene (Itti et al., 1998). If we can apply the human-like vergence function considered human attention process to an active stereo vision system, an efficient and intelligent vision system can be developed. Researchers have been developing the vergence stereo system. It was known that the two major sensory drives for vergence and accommodation are disparity and blur (Krishnan & Stark, 1977; Hung & Semmlow, 1980). Krotkov organized the stereo system through waking up the camera, gross focusing, orienting the cameras, and obtaining depth information (Krotkov, 1987). Abbott and Ahuja proposed surface reconstruction by dynamic integration of the focus, camera vergence and stereo disparity (Abbott & Ahuja, 1988). These approaches give good results for a specific condition, but it is difficult to use these systems in real environment because the region extraction based on intensity information is very sensitive to luminance change. For mimicking a human vision system, Yamato implemented a layered control system for stereo vision head with vergence control function. This system utilized a search of the most similar region based on the sum of absolute difference (SAD) for tracking. The vergence module utilized a minimum SAD search for each pixel to obtain figure-ground separation in 3D space (Yamato, 1999). But these systems may not give good results when the camera is moving with background because the SAD contains much noise by moving the camera. Jian Peng et al. made that an active vision system enables the selective capture of information for a specific colored object

Source: Scene Reconstruction, Pose Estimation and Tracking, Book edited by: Rustam Stolkin,
ISBN 978-3-902613-06-6, pp.530, I-Tech, Vienna, Austria, June 2007

(Peng et al., 2000). But this system only considered the color information for the selective attention. Thus, the developed active vision only operates for a specific color object and the luminance change deteriorate performance of the system. Bernardino and Victor implemented vergence control stereo system using log-polar images (Bernardino & Santos-Victor, 1996). This work considers only the intensity information. Batista et al. made a vergence control stereo system using retinal optical flow disparity and target depth velocity (Batista et al., 2000). But this system mainly converges on the moving object because of optical flow. Thus, this system only considered the motion information of retina and do not consider intensity, edge and symmetry as retina operation. Moreover, these all approaches take a lot of computation load to get the vergence control. Therefore, we need a new method not only to sufficiently reflect information of images such as color, intensity and edge but also to reduce the computation load during vergence control. Conradt et al. proposed a stereo vision system using a biologically inspired saliency map (SM) (Conradt et al., 2002). They detected landmarks in both images with interaction between the feature detectors and the SM, and obtained their direction and distance. They considered intensity, color, and circles of different radius, and horizontal, vertical and diagonal edges as features. However, they do not consider the occlusion problem. Also their proposed model does not fully consider the operation of the brain visual signal processing mechanism because they only considered the roles of neurons in the hippocampus responding to mainly depth information. On the other hand, the selective attention mechanism allows the human vision system to process visual scenes more effectively with a higher level of complexity. The human visual system sequentially interprets not only a static monocular scene but also a stereo scene based on the selective attention mechanism. In previous research, Itti and Koch (Itti et al., 1998) introduced a brain-like model in order to generate the saliency map (SM). Koike and Saiki (Koike & Saiki, 2002) proposed that a stochastic WTA enables the saliency-based search model to vary the relative saliency in order to change search efficiency, due to stochastic shifts of attention. Timor and Brady (Kadir & Brady, 2001) proposed an attention model integrating saliency, scale selection and a content description, thus contrasting many other approaches. Ramström and Christensen (Ramstrom & Christensen, 2002) calculated saliency with respect to a given task by using a multi-scale pyramid and multiple cues. Their saliency computations were based on game theory concepts. In recent work, Itti's group proposed a new attention model that considers seven dynamic features for MTV-style video clips (Carmi & Itti, 2006) and also proposed an integrated attention scheme to detect an object, which combined bottom-up SM with top-down attention based on signal-to-noise ratio (Navalpakkam & Itti, 2006). Also, Walter and Koch proposed an object preferable attention scheme which considers the bottom-up SM results as biased weights for top-down object perception (Walther et al., 2005). Also, Lee et al. have been proposed a bottom-up SM model using symmetry information with an ICA filter (Park et al., 2002) and implemented a human-like vergence control system based on a selective attention model, in which the proposed model reflects a human's interest in an area by reinforcement and inhibition training mechanisms (Choi et al., 2006). Ouerhani and Hugli proposed a saliency map model considering depth information as a feature (Ouerhani and Hugli, 2000). They insisted that little attention has been devoted so far to scene depth as source for visual attention and also pointed that this is considered as a weakness of the previously proposed attention models because depth or 3D vision is an intrinsic component of biological vision (Ouerhani and Hugli, 2000). Ouerhani and Hugli just used range finder for getting depth information

but did not consider any mechanism about how to deal with binocular vision process. None of the proposed attention models, however, consider the integration of a stereo type bottom-up SM model and a top-down selective attention scheme reflecting human's preference and refusal. In this paper, we propose a new human-like vergence control method for an active stereo vision system based on stereo visual selective attention model. The proposed system reflects the single eye alignment mechanism during an infant's development for binocular fixation, and also uses a selective attention model to localize an interesting area in each camera. The proposed method reflects the biological stereo visual signal processing mechanism from the retinal operation to the visual cortex. Thus, we use a new selective attention model for implementing a human-like vergence control system based on a selective attention mechanism not only with truly bottom-up process but also with low-level top-down attention to skip an unwanted area and/or to pay attention to a desired area for reflecting human's preference and refusal mechanism in subsequent visual search process such as the pulvinar. Moreover, the proposed selective attention model considers depth information to construct a final attention area so that the closer attention area can be easily pop-up as our binocular eyes. Using the left and right saliency maps generated by the proposed selective attention models for two input images from left and right cameras, the selected object area in the master camera is compared with that in the slave camera to identify whether the two cameras find a same landmark. If the left and right cameras successfully find a same landmark, the implemented active vision system with two cameras focuses on the landmark. To prevent it from being a repetitively attended region in the vision system, the converged region is masked by an inhibition of return (IOR) function. Then the vision system continuously searches a new converged region by the above procedure. The practical purpose of the proposed system is to get depth information for efficient robot vision by considering focusing on an interesting object only by training process. Moreover, the proposed method can give a way to solve the occlusion problem. The depth information of the developed system will operate for avoiding an obstacle in a robotic system. Based on the proposed algorithm together with an effort to reduce the computation load, we implemented a human-like active stereo vision system. Computer simulation and experimental results show that the proposed vergence control method is very effective in implementing the human-like active stereo vision system. In Section 2, we briefly discuss the biological background of the proposed model and the proposed stereo visual selective attention model. In Section 3, we explain the landmark selection algorithm in each camera, the verification of the landmarks and depth estimation using eye gaze matching. In Section 4, we explain the hardware setup and describe computer simulation and the experimental results. The discussion and conclusion will be followed in Section 5.

2. Stereo visual selective attention

2.1 Biological understanding

Fig. 1 shows the biological visual pathway from the retina to the visual cortex through the LGN for the bottom-up processing, which is extended to the extrastriate cortex and the prefrontal cortex for the top-down processing. In order to implement a human-like visual attention function, we consider the bottom-up saliency map (SM) model and top-down trainable attention model. In our approach, we reflect the functions of the retina cells, LGN and visual cortex for the bottom-up processing, and dorsolateral prefrontal, posterior

parietal cortex, the anterior cingulated gyrus, and the pulvinar nucleus of the thalamus for the top-down processing (Goldstein, 1995).

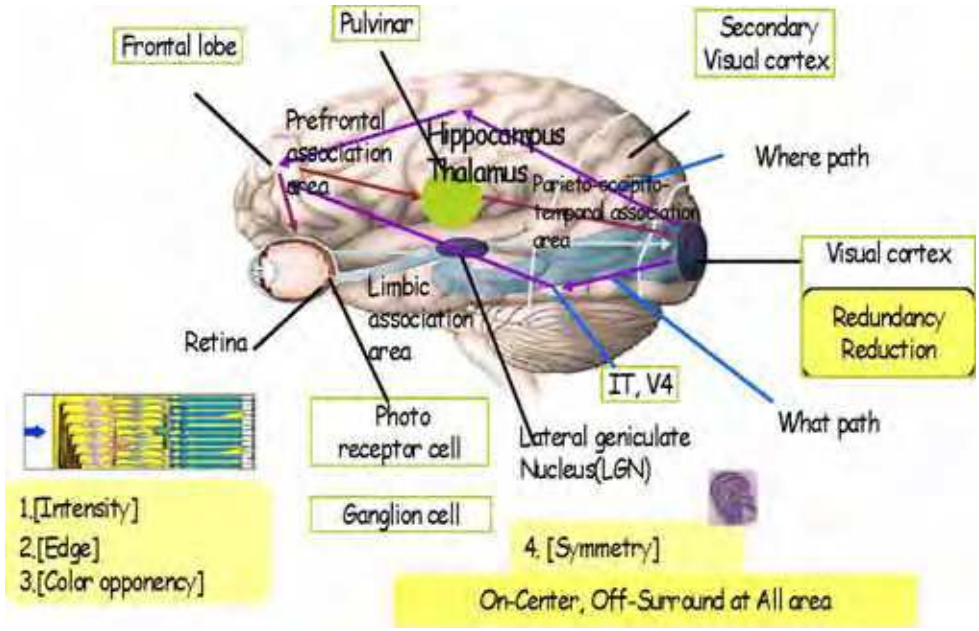


Figure 1. Biological visual pathway of bottom-up and top-down processing

Fig. 2 shows the proposed stereo saliency map model in conjunction with vergence control process based on the simulated biological visual pathway from the retina to the visual cortex through the LGN for the bottom-up processing, which is extended to the limbic system including the pulvinar for the top-down processing. In order to implement a human-like visual attention function, three processes are integrated to generate a stereo SM. One generates static saliency in terms of monocular vision. Another considers low-level top-down process for reflecting human preference and refusal, which mimics the function of the pulvinar in the limbic system. Finally, we can build stereo SM based on two monocular SMs and depth in terms of binocular vision.

2.2 Static bottom-up saliency map

Based on the Treisman's feature integration theory (Treisman & Gelde, 1980), Itti and Koch used three basis feature maps: intensity, orientation and color information (Itti et al., 1998). Extending Itti and Koch's SM model, we previously proposed SM models which include a symmetry feature map based on the generalized symmetry transformation (GST) algorithm

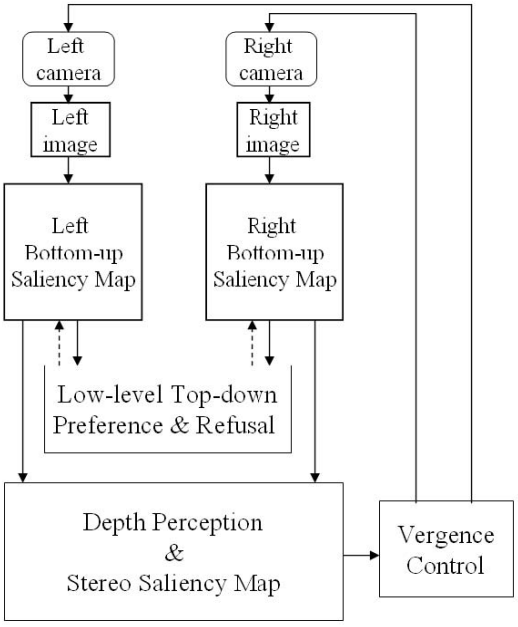


Figure 2. Stereo saliency map model including the static bottom-up SM process, the low-level top-down preference and refusal process and depth perception

and an independent component analysis (ICA) filter to integrate the feature information (Park et al., 2002; Park et al., 2000). In this paper, we investigate through intensive computer experiment how much the proposed symmetry feature map and the ICA filter are important in constructing an object preferable attention model. Also, we newly incorporate the neural network approach of Fukushima (Fukushima, 2005) to construct the symmetry feature map, which is more biologically plausible and takes less computation than the GST algorithm (Park et al., 2000). Symmetrical information is also important feature to determine the salient object, which is related with the function of LGN and primary visual cortex (Li, 2001). Symmetry information is very important in the context free search problem (Reisfeld et al., 1995).In order to implement an object preferable attention model, we emphasize using a symmetry feature map because an object with arbitrary shape contains symmetry information, and our visual pathway also includes a specific function to detect a shape in an object (Fukushima, 2005; Werblin & Roska, 2004). In order to consider symmetry information in our SM model, we modified Fukushima's neural network to describe a symmetry axis (Fukushima, 2005). Fig. 3 shows the static bottom-up saliency map model. In the course of computing the orientation feature map, we use 6 different scale images (a Gaussian pyramid) and implement the on-center and off-surround functions using the center surround and difference with normalization (CSD & N) (Itti et al., 1998; Park et al., 2002). As shown in Fig. 4, the orientation information in three successive scale images is used for obtaining the symmetry axis from Fukushima's neural network (Fukushima, 2005). By applying the CSD&N to the symmetry axes extracted in four different scales, we can

obtain a symmetry feature map. This procedure mimics the higher-order analysis mechanism of complex cells and hyper-complex cells in the posterior visual cortex area,

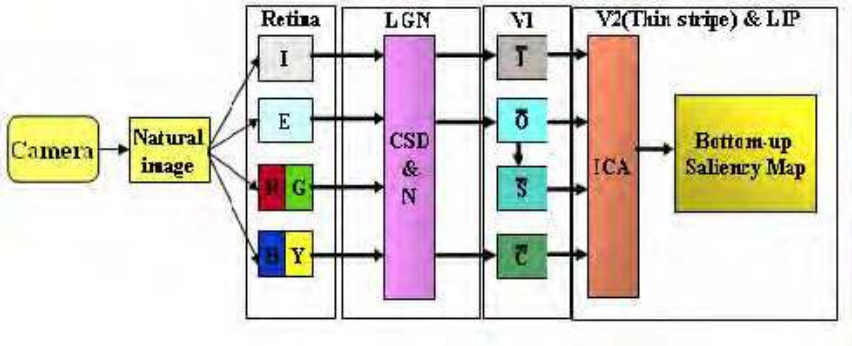


Figure 3. Static bottom-up saliency map model (I: intensity feature, E: edge feature, RG: red-green opponent coding feature, BY: blue-yellow opponent coding feature, LGN: lateral geniculate nucleus, CSD&N: center-surround difference and normalization, I : intensity feature map, O : orientation feature map, S : symmetry feature map, C : color feature map, ICA: independent component analysis)

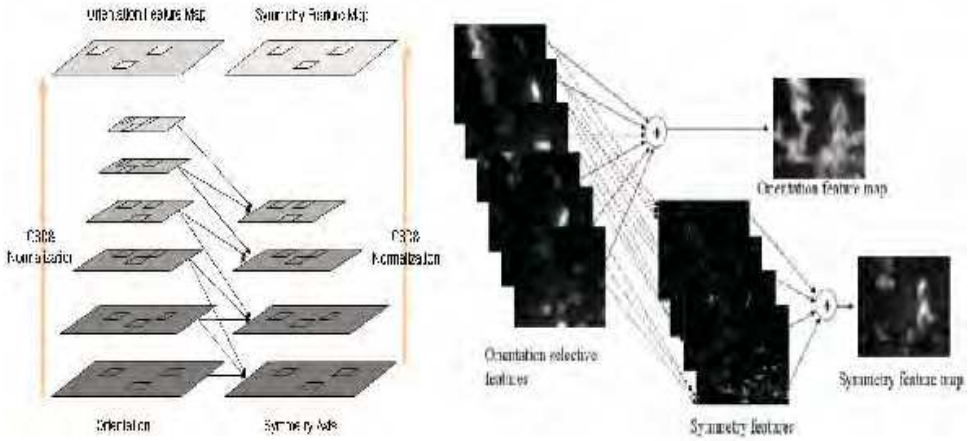


Figure 4. Symmetry feature map generation process

beyond the orientation-selective simple cells in the V1. Using CSD&N in Gaussian pyramidimages (Itti et al., 1998), we can construct the intensity (I), color (C), and orientation (O)feature maps as well as the symmetry feature map (S). Based on both Barlow's hypothesis that human visual cortical feature detectors might be the end result of a redundancy reduction process (Barlow & Tolhust, 1992)and Sejnowski's results that ICA is the best way to reduce redundancy (Bell & Sejnowski, 1997), the four constructed feature maps (*I, C, O, and S*) are then integrated by an independent componentanalysis (ICA) algorithm based on maximization of entropy (Bell & Sejnowski, 1997). Fig. 5 shows the procedure for computing the SM. In Fig. 5, S(x,y) is obtained by

summation of the convolution between the r-th channel of input image(I_r) and the i-th filters(ICs_{ri}) obtained by the ICA learning [9]. A static SM is obtained by Eq. (1).

$$S(x, y) = \sum I_r * ICs_{ri} \quad for\ all\ I \tag{1}$$

Since we obtained the independent filters by ICA learning, the convolution result shown in Eq. (1) can be regarded as a measure for the relative amount of visual information. The lateral-intra parietal cortex (LIP) plays a role in providing a retinotopic spatio-feature map that is used to control the spatial focus of attention and fixation, which is able to integrate feature information in its spatial map (Lanyon & Denham, 2004). As an integrator of spatial and feature information, the LIP provides the inhibition of return (IOR) mechanism required here to prevent the scan path returning to previously inspected sites (Lanyon & Denham, 2004).

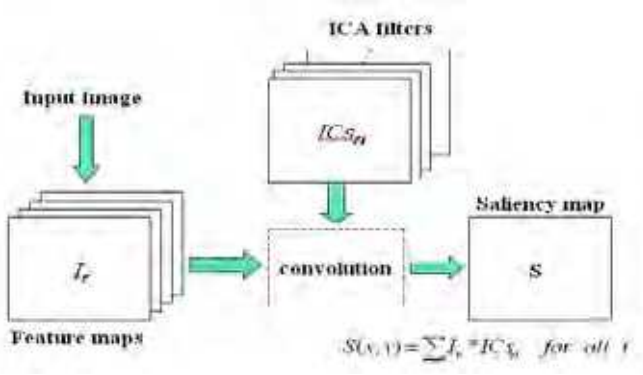


Figure 5. Saliency map generation process using ICA filter

2.3 Low-level top-down selective attention

Although the proposed bottom-up static SM model generates plausible salient areas and a scan path, the selected areas may not be an interesting area for human because the SM only uses primitive features such as intensity, color, orientation and symmetry information. In order to implement a more plausible selective attention model, we need to consider low-level top-down mechanism that can reflect human preference and refusal for visual features. Human beings ignore uninteresting areas, even if they have primitive salient features, and they can memorize the characteristics of the unwanted area. Humans do not pay attention to new areas that have characteristics similar to learned unwanted areas. In addition, human perception can focus on an interesting area, even if it does not have primitive salient features, or if it is less salient than the other areas. We propose a new selective attention model that mimics the human-like selective attention mechanism and that can consider not only primitive input features, but also interactive properties with humans in the environment. Moreover, the human brain can learn and memorize many new things without catastrophic forgetting. It is well known that a fuzzy adaptive resonance theory (ART) network can easily be trained for additional input pattern. Also, it can solve the stability-plasticity dilemma in a conventional multi-layer neural network (Carpenter et al., 1992). Therefore, as shown in Fig. 6, we use a fuzzy ART network together with a bottom-up

SM model to implement a selective attention model with preference and refusal process that can interact with a human supervisor. During the training process, the fuzzy ART network "learns" and "memorizes"the characteristics of uninteresting and/or interesting areas decided by a human supervisor. After the successful training of the fuzzy ART network, an unwanted salient area isinhibited and a desired area is reinforced by the vigilance value ρof the fuzzy ART network, as shown in Fig. 6. As shown in Fig. 6, corresponding four feature maps from the attended area obtained fromthe SM are normalized and then represented as one dimensional array X that are composedof every pixel value a_i of the four feature maps and each complement a_i^c computed by $1-a_i$, which are used as an input pattern of the fuzzy ART model. Then, the fuzzy ART model consecutively follows three processes such as a choice process, a match process and an adaptation process.

During a choice process, for every node Y_j in the F2 layer, a net activity y_j is calculated using a fuzzy conjunction operator ($\wedge$) as shown in Eq. (2), which can be seen as the degree of prototype bottom-up weight vector W_j, being a fuzzy subset of input pattern X

$$y_j = \frac{|X \wedge W_j|}{\alpha + |W_j|} \tag{2}$$

where the fuzzy conjunction $\wedge$ is computed by component wise min operator and the magnitude operator $|\cdot|$ of a vector is calculated by its L_1-norm defined by the sum of its components. And the parameter α works for avoiding a floating point overflow. Node Y_j in the F_2 layer with the highest value y_j is chosen as the winner node. After selecting the winner node for input pattern X, the fuzzy ART checks the similarity of input pattern X and the top-down weight vector W_J of the winner node Y_J as shown in Eq. (3)

$$\rho \le \frac{|X \wedge W_J|}{|X|} \tag{3}$$

where a vigilance parameter ρ is defining the minimum similarity between input pattern and the prototype of the winner node. The synaptic top-down weight vector W_J are identical to the bottom-up weight vector W_j. If the similarity is lager than the vigilance value, then the vector W_J is adapted by moving its values toward the common MIN vector of X and W_J as shown in equation (4)

$$W_J^{(new)} = \eta(X \wedge W_J^{(old)}) + (1-\eta)W_J^{(old)} \tag{4}$$

where η is a learning rate. When Eq. (3) is satisfied, we call resonance is occurred. However, if the similarity is less than the vigilance, the current winning F_2-node is removed from the competition by a reset signal. The fuzzy ART searches again a node with the next most similar weight vector with the input pattern X before an uncommitted prototype is chosen. If none of the committed nodes matches the input pattern well enough, search will end with the recruitment of an uncommitted prototype (Frank et al., 1998).

As the number of training patterns increases, however, the fuzzy ART network requires more time to reinforce or inhibit some selected areas. For faster analysis in finding an inhibition and/or reinforcement area, we employed the hierarchical structure of this

network. Fig. 7 shows the modified hierarchical structure model of the fuzzy ART. The hierarchy of this network consists of a five-layer concatenated structure, in which each layer represents a different hierarchical abstract level of information. The highest level of the model stores the most abstract information that represents a highly abstract cluster. The lowest level of the model stores more detailed information. The input of the higher level in the model is generated by dimension reduction of the input of the lower level by averaging operator. For example, if the dimension of input for the lowest level is 32 by 32, the dimension of the next higher level becomes 16 by 16. The input pattern comparison with the memorized patterns of the model starts from the highest level, then the proposed model progress to the lower level according to the resonance result at the fuzzy ART module for the level. In the highest level, the input dimension is so small that it takes short time to finish the process of the fuzzy ART module. If the current input pattern has no resonance in the highest level, the hierarchical model can finish the process without considering the other lower levels, through which it can reduce the computation time. After the training process of the model is successfully finished, it memorizes the characteristics of the unwanted or desired areas in order to reflect human's preference and refusal. If a salient area selected by the bottom-up SM model of a test image has similar characteristics to the fuzzy ART memory, it is ignored by inhibiting that area in the SM or it is magnified by reinforcing that area in the SM according to human interest.

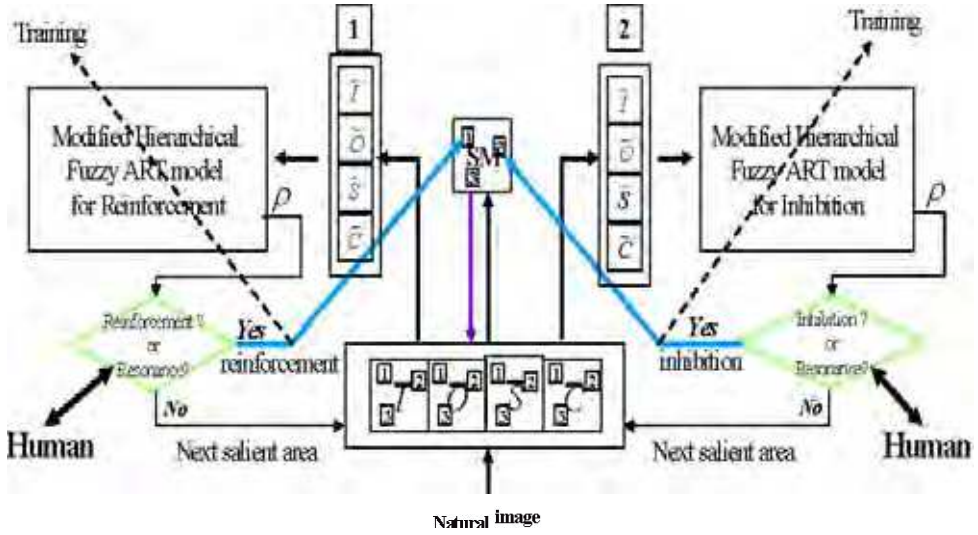


Figure 6. The architecture of the proposed low level top-down attention model with reinforcement(preference) and inhibition(refusal) property: (I : intensity feature map, O : orientation feature map, S : symmetry feature map, C : color feature map, SM: saliency map). Square block 1 in the SM is an interesting area, but block 2 is an uninteresting area

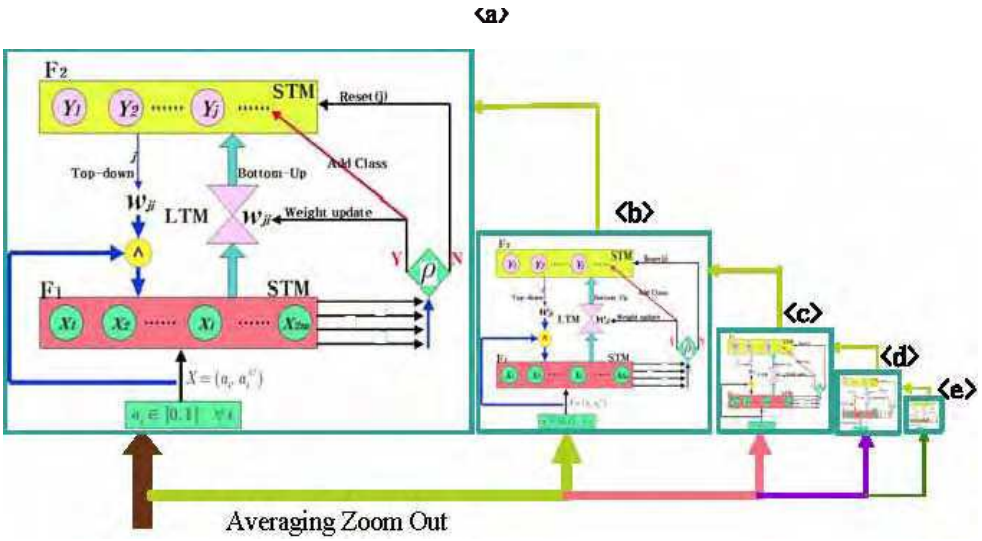


Figure 7. The modified hierarchical fuzzy ART network, where <a> represents the lowest level in fuzzy ART network and <e> does the highest level one

2.4 Stereo saliency

In this paper, we now utilize the depth information obtained by the vergence control vision system to construct the stereo saliency map model, which can then support pop-up for closer objects. In our model, the selective attention regions in each camera are obtained from static bottom-up saliency in conjunction with the low-level top-down preference and refusal, which are then used for selecting a dominant landmark. After successfully localizing corresponding landmarks on both the left image and the right image, we are able to get depth information by a simple triangular equation described in Section 3. Then, the proposed stereo SM model uses depth information as a characteristic feature in deciding saliency using a decaying exponential function. The final stereo SM is obtained by $S(x,y)\cdot\exp^{-z/\tau}$, where z is the distance between the camera and an attend region, and τ is a time constant.

3. Vergence control using the stereo selective attention model

3.1 Selection and verification of landmarks

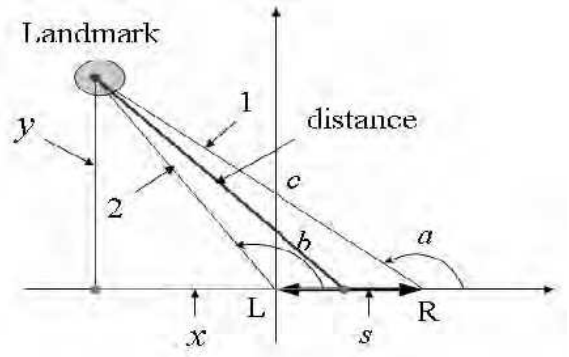
During an infant's development, binocular disparity by binocular fixation is decomposed into three different mechanisms; alignment of eyes, convergence and sensory binocularity (Thorn et al., 1994). According to this fact, the single eye alignment should be the first factor considered regarding convergence that needs binocular fixation. In order to accomplish the single eye alignment, we use successive attention regions selected by the selective attention model in each camera image. Most of the stereo vision systems fix one of two cameras in the

master eye. Humans, however, will probably not perform single eye alignment in this manner. The eye that has the dominant landmark may be considered the master eye, and the other eye is the slave eye that aligns itself to the landmark of the master eye. In our model, the trainable selective attention model generates the maximum salient value for the dominant landmark in each camera image. Comparing the maximum salient values in two camera images, we can adaptively decide the master eye that has a camera with a larger salient value. As shown in Fig. 2, the selective attention regions in each camera are obtained by the low-level top-down SM model for reflecting human's preference and refusal in conjunction with the bottom-up static SM model, which are used for selecting a dominant landmark. The low-level preference and resual SM model can reinforce an interesting object area in the bottom-up saliency map, which makes the interesting object area to be the most salient region even if it is less salient than another area in the bottom-up saliency map. Moreover it can inhibit an unwanted object area in the bottom-up saliency map, which makes the unwanted object area to be the least salient region even if it is more salient than another area in the bottom-up SM model. Therefore, the proposed attention model can have an ability to pay an attention to an interesting object area by the low-level top-down attention process together with the bottom-up SM model. Although the position of a salient region in the left and right cameras is almost the same, there exists a misalignment case due to occlusion and the luminance effect. In order to avoid this situation, we compare the difference of y coordinates between a dominant salient region in the master eye and successive salient regions in the slave eye because one of the successive salient regions in the slave eye may be in accordance with the salient region of the master eye. When the difference of the y coordinates is smaller than the threshold, we regard the salient region as a candidate for a landmark. In order to verify the candidate as a landmark, we need to compare the salient region of the master eye with that of the slave eye. The regions obtained by the IOR function, which is to avoid duplicating the selection of the most salient region, are compared in order to decide on a landmark. If the IOR region of the master eye is similar to that of the slave eye, we regard the IOR regions as a landmark to make convergence. The comparison of values of the IOR regions between the left and right cameras is used for the verification of a landmark.

3.2 Depth estimation and vergence control

After the landmark is successfully selected, we are able to get depth information. Fig. 8 shows the top view of verged cameras. First, we have to obtain the degrees of two camera angles to be moved to make a focus on a land mark. Considering the limitation of the field of view (F) in the horizontal axis and motor encoder resolution (U), we can get the total encoder value (E) to represent the limited field of view of the horizontal axis. The total encoder value (E) can be obtained by Eq. (5). As shown in Eq. (6), the total encoder value (E) is used to calculate the encoder value (x_t) of the horizontal axis motor for aligning of each camera to a landmark. In Eq. (6), R denotes the x-axis pixel resolution of the image and T denotes the relative pixel coordinate of the x-axis of a landmark from the focus position. In other words, T represents the disparity of x- axis. The x- axis encoder value (x_t) that uses to move each camera to the landmark point is translated into the angel (x_d) by Eq. (7). As a result, the angles a and b are obtained by Eq. (7) by substituting T for the x coordinates of the left and right cameras. Obtained depth information is used to generate a stereo SM. Finally, the proposed model decides the most salient area based on the obtained stereo SM

and makes two cameras to focus on the same area by controlling motors of them, which is called vergence control.



L: Left camera focus center, R : Right camera focus center, a : Right camera angle, b : Left camera angle, c : an intercept of the line 1, s : The distance between the two cameras focus, 1 and 2 : straight line from right and left cameras to a landmark

Figure 8. Top view of verged cameras

$$E = (F \times 360°)/U \tag{5}$$

$$x_t = -E + (E \times T)/R \tag{6}$$

$$x_d = 90° - (R \times x_t)/U \tag{7}$$

The vertical distance (y) is obtained by the following Eqs. (8-10).

$$\tan(a) \cdot x - s \cdot \tan(a) = y \tag{8}$$

$$\tan(b) \cdot x = y \tag{9}$$

$$y = \frac{\tan(a) \cdot \tan(b)}{\tan(a) - \tan(b)} \cdot s \tag{10}$$

Eqs. (8) and (9) show the equation of straight lines between the cameras and the landmark, respectively. In Eq. (8), x and y denote the disparities for x-axis and y-axis respectively between a land mark and a current focus position and s represents the distance between each focal axis of two cameras. Eq. (10) is the equation to calculate the vertical distance (y) in Fig. 8.

Fig. 9 shows three different cases for differently calculating depth information. If the angle of a and b shown in Fig. 8 are above 90° (case 1), the distance is $\sqrt{\dot{x}^2+y^2}$ because of $x = y/\tan(b)$ and $\dot{x} = |y/\tan(b)| + s/2$. If angle a is above 90° and angle b is less than 90° (case 2), the distance is almost y because the error is very small if the vergence point is not very close. If the angle of a and b are under 90° (case 3), the distance is $\sqrt{\ddot{x}^2+y^2}$ because of $x = y/\tan(b)$ and $\ddot{x} = |y/\tan(b)| - s/2$.

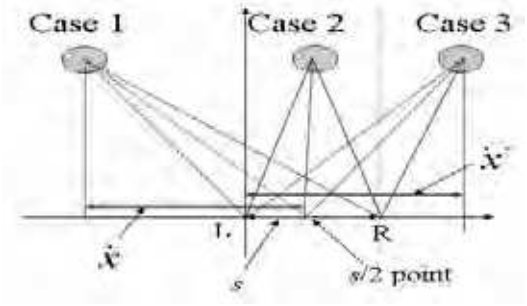


Figure 9. Three cases of obtaining the depth information

4. Implementation & Experimental results

4.1 Hardware implementation

We implemented a stereo vision robot unit for vergence control of two cameras. Fig. 10 shows the implemented system called by SMART-v1.0 (Self Motivated Artificial Robot with a Trainable selective attention model version 1.0). The SMART-v1.0 has four DOF and two 1394 CCD camera and Text to Speech module (TTS) to communicate with humans to inquire about an interesting object, and tilt sensor to set offset position before starting moving. We use the Atmega128 as the motor controller and zigbee to transmit motor command from a PC to SMART-v1.0. The SMART-v1.0 can search an interesting region by the selective attention model and vergence control which are explained in section 2 and 3.

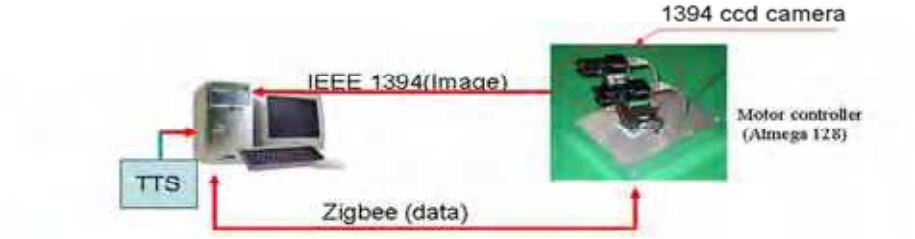


Fig. 10. SMART-v1.0 platform

4.2 Experimental results

4.2.1 Static saliency

Fig. 11 shows an example in which the proposed bottom-up SM model generates more object preferable attention by using symmetry information as an additional input feature and ICA for feature integration. The numbers in Fig. 11 represent the order of the scan path according to the degree of saliency. As shown in Fig. 11, the symmetry feature map is effective in choosing an attention area containing an object. The ICA filter successfully reduces redundant information in feature maps so that the final scan path does not pay attention to sky in the input image. Table 1 compares the object preferable performance of three different bottom-up SM models using hundreds of test images. The bottom-up SM model considering both the symmetry feature and ICA method for integrating features shows the best object preferable attention without any consideration of top-down attention.

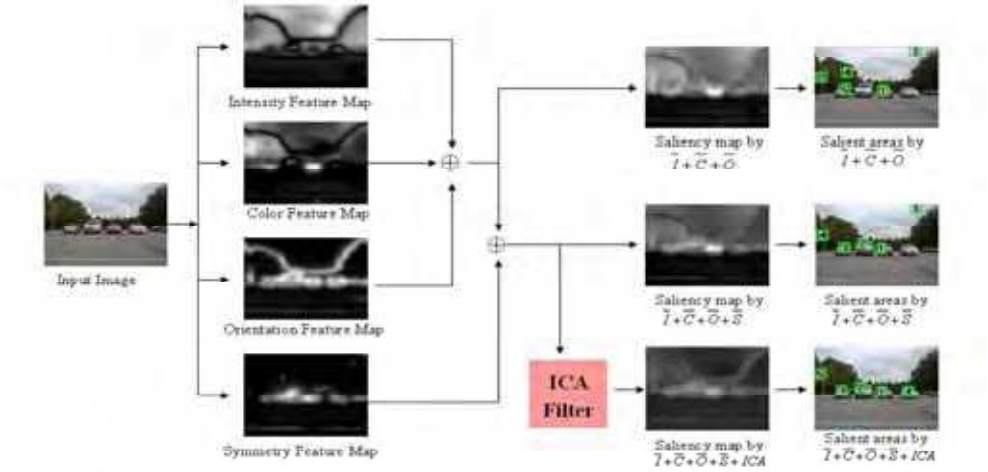


Fig. 11. Comparison of attention results by three different models for a same input scene. One is a result by $\bar{I}+\bar{C}+\bar{O}$ model. Another is a result by $\bar{I}+\bar{C}+\bar{O}+\bar{S}$ model. The other is a result by $\bar{I}+\bar{C}+\bar{O}+\bar{S}+ICA$ model

Salient area(SA)	$\bar{I}+\bar{C}+\bar{O}+\bar{S}+ICA$	$\bar{I}+\bar{C}+\bar{O}+\bar{S}$	$\bar{I}+\bar{C}+\bar{O}$
1st SA	165	150	143
2nd SA	106	104	103
3rd SA	68	77	64
4th SA	66	57	47
5th SA	46	32	28
Total	451	420	385
(%)	90.2 %	84.0 %	77.0 %

Table 1. Comparison of three different bottom-up SM models for object preferable attention

4.2.2 Low-level top-down selective attention

Fig. 12 shows the simulation results using the low-level top-down saliency component of our proposed model. Fig. 12 (a) shows the scan path generated by the static SM model, where the 3rd salient area is deemed a refusal area according to the human's preference and refusal, and it is trained by the low-level top-down SM model for refusal. Also, the 2nd salient area is changed after training the low-level SM model for preference. The bottom image shows the modified saliency map by reflecting human's preference and refusal during training process. Fig. 12 (b) shows the modified scan path by the preference and refusal low-level top-down attention process after training human's preference and refusal as shown in Fig. 12 (a). Fig. 13 shows an example for lip preferable attention, and Table 2 shows the performance comparison between the bottom-up SM model and the low-level

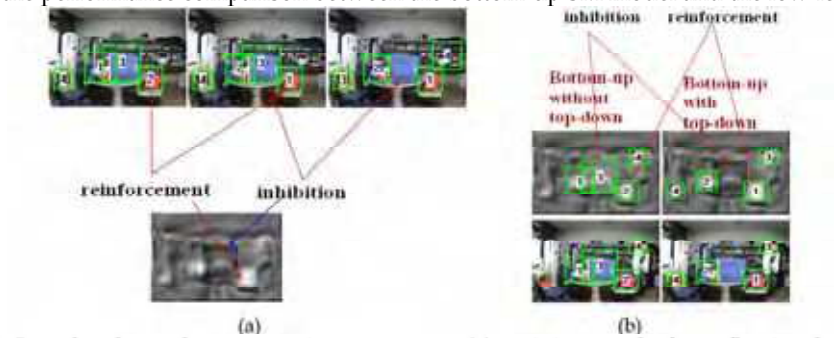


Fig. 12. Low-level top-down attention processes; (a) training mode for reflecting human's preference or refusal by reinforcing or inhibiting saliency map (b) experimental results reflecting human's preference and resual after training like (a)

Fig. 13. Simulation results for lip preferable low-level top-down attention; (a) without considering the low-level top-down process, (b) considering the low-level top-down process, (c) saliency map after considering the low-level top-down process

	96 Training images	90 Test images	
		Bottom-up saliency map	Low-level top-down saliency map
# of images containing lip in 1st salient area	96	35	88
# of images containing lip in 1st salient area	0	55	2
	100%	38.9%	97.8%

Table 2. Low-level top-down attention performance

4.2.3 Stereo saliency and vergence control

Fig. 14 shows a simulation result using stereo saliency. As shown in Fig. 14, by considering the depth feature, the proposed model can make closer attened objects mostly pop out.

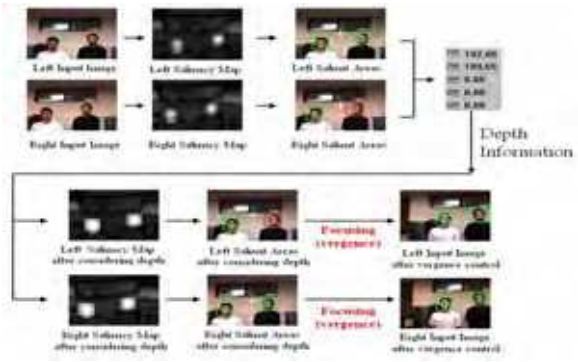


Fig. 14. Stereo saliency and vergence control experimental results

5. Conclusion

We proposed a new biologically motivated vergence control method of an active stereo vision system that mimics human-like stereo visual selective attention. We used a trainable selective attention model that can decide an interesting area by the low-level top-down mechanism implemented by Fuzzy ART training model in conjunction with the bottom-up static SM model. In the system, we proposed a landmark selection method using the low-level top-down trainable selective attention model and the IOR regions. Also, a depth estimation method was applied for reflecting stereo saliency. Based on the proposed algorithm, we implemented a human-like active stereo vision system. From the computer simulation and experimental results, we showed the effectiveness of the proposed vergence control method based on the stereo SM model. The practical purpose of the proposed system is to get depth information for robot vision with a small computation load by only considering an interesting object but by considering all the area of input image. Depth information of the developed system will operate for avoiding an obstacle in a robotic system. Also, we are considering a look-up table method to reduce the computation load of the saliency map for real-time application. In addition, as a further work, we are now developing an artificial agent system by tracking a moving person as main practical application of the proposed system.

6. Acknowledgment

This research was funded by the Brain Neuroinformatics Research Program of the Ministry of Commerce, Industry and Energy, and the sabbatical year supporting program of Kyungpook National University.

7. References

Abbott, A. L. & Ahuja, N. (1988). Surface reconstruction by dynamic integration of focus, camera vergence, and stereo, Proceedings of IEEE International Conference on Computer Vision, pp.532 -543, ISBN: 0-8186-0883-8

Barlow, H. B. & Tolhust, D. J. (1992). Why do you have edge detectors?, Optical society of America Technical Digest, Vol. 23. 172, ISBN-10: 3540244212, ISBN-13: 978-3540244219

Batista, J.; Peixoto, P. & Araujo, H. (2000). A focusing-by-vergence system controlled by retinal motion disparity, Proceedings of IEEE International Conference on Robotics and Automation, Vol. 4, pp.3209 -3214, ISBN: 0-7803-5889-9, April 2000, SanFrancisco, USA

Bell, A. J. & Sejnowski, T. J. (1997). The independent components of natural scenes are edge filters, Vision Research, Vol. 37., 3327-3338, ISSN: 0042-6989

Bernardino, A. & Santos-Victor, J. (1996). Vergence control for robotic heads using log-polar images, Proceedings of IEEE/RSJ International Conference. Intelligent Robots and Systems, Vol. 3, pp.1264 -1271, ISBN: 0-7803-3213-X, Nov. 1996, Osaka, Japan

Carpenter, G. A.; Grossberg, S.; Markuzon, N. J.; Reynolds, H. & Rosen, D. B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps, IEEE Trans. on Neural Networks, Vol. 3, No.5, 698-713, ISSN: 1045-9227

Carmi, R. & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes, Vision Research, Vol. 46, No.26, 4333-4345, ISSN: 0042-6989

Choi, S. B.; Jung, B. S.; Ban, S. W.; Niitsuma, H. & Lee, M. (2006). Biologically motivated vergence control system using human-like selective attention model, Neurocomputing, Vol. 69, 537-558, ISSN: 0925-2312

Conradt, J.; Pescatore, M.; Pascal, S. & Verschure, P. (2002). Saliency maps operating on stereo images detect landmarks and their distance, Proceedings of International Conference on Neural Networks, LNCS 2415, pp.795-800, ISBN-10: 3540440747, ISBN 13: 978-3540440741, Aug. 2002 , Madrid, Spain

Frank, T.; Kraiss, K. F. & Kuklen, T. (1998). Comparative analysis of Fuzzy ART and ART-2A network clustering performance. IEEE Trans. Neural Networks, Vol. 9, No. 3, May 1998, 544-559, ISSN: 1045-9227

Fukushima, K. (2005). Use of non-uniform spatial blur for image comparison: symmetry axis extraction, Neural Network, Vol. 18, 23-22, ISSN: 0893-6080

Goldstein, E. B. (1995). Sensation and perception, 4th edn., An international Thomson publishing company, ISBN-10: 0534539645, ISBN-13: 978-0534539641, USA

Hung, G. K. & Semmlow, J. L. (1980). Static behavior of accommodation and vergence: computer simulation of an interactive dual-feedback system, IEEE Trans. Biomed. Eng., Vol. 27, 439-447, ISSN: 0018-9294

Itti, L.; Koch, C. & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Patt. Anal. Mach. Intell., Vol. 20, No. 11., 1254-1259, ISSN: 0162-8828

Kadir, T. & Brady, M. (2001). Scale, saliency and image description, International Journal of Computer Vision, 83 -105, ISSN: 0920-5691 Koike, T. & Saiki, J. (2002) Stochastic guided search model for search asymmetries in visual search tasks, Lecture Notes in Computer Science, Vol. 2525, 408-417, ISSN: 0302-9743

Krishnan,V. V. & Stark,L. A. (1977). A heuristic model of the human vergence eye movement system, IEEE Trans. Biomed. Eng., Vol. 24 , 44-48, ISSN: 0018-9294

Krotkov, E. (1987). Exploratory visual sensing for determining spatial layout with an agile stereo camera system, University of Pennsylvania Ph.D. Dissertation also available as a Tech. Rep, MS-CIS-87-29

Lanyon, L. J. & Denham, S.L. (2004). A model of active visual search with object-based attention guiding scan paths, Neural Networks Special Issue: Vision & Brain, Vol. 17, No. 5-6, 873-897, ISSN: 0893-6080

Li, Z. (2001). Computational design and nonlinear dynamics of a recurrent network model of the primary visual cortex, Neural Computation, Vol.13, No.8, 1749-1780, ISSN: 0899-7667

Navalpakkam, V. & Itti, L. (2006). An integrated model of top-down and bottom-up attention for optimal object detection, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.2049-2056, ISBN: 0-7695-2597-0

Ouerhani, N. & Hugli, H. (2000). Computing visual attention from scene depth, Proceedings of 15th International Conference on Pattern Recognition, Vol. 1, pp. 375-378, ISBN: 07695-0750-6, Oct. 2000, Barcelona, Spain

Park, C. J.; Oh, W. G. S.; Cho, H. & Choi, H. M. (2000). An efficient context-free attention operator for BLU inspection of LCD production line, Proceedings of IASTED International conference on SIP, pp. 251-256

Park, S. J.; An, K. H. & Lee, M. (2002). Saliency map model with adaptive masking based on independent component analysis, Neurocomputing, Vol. 49, 417-422, ISSN: 0925-2312

Peng, J.; Srikaew, A.; Wilkes, M.; Kawamura, K. & Peters, A. (2000). An active vision system for mobile robots, Proceedings of IEEE International Conference. Systems, Man, and Cybernetics, Vol. 2, pp. 1472 – 1477, ISBN: 0-7803-6583-6, Oct. 2000, Nashville, TN, USA

Ramstrom, O. & Christensen, H. I. (2002). Visual attention using game theory, Lecture Notes in Computer Science, Vol. 2525, 462-471, ISSN: 0302-9743

Reisfeld, D.; Wolfson, H. & Yeshurun, Y. (1995). Context-free attentional operators : The generalized symmetry transform, Internatioanl Jouranl of Computer Vision, Vol. 14, 119-130, ISSN: 0920-5691

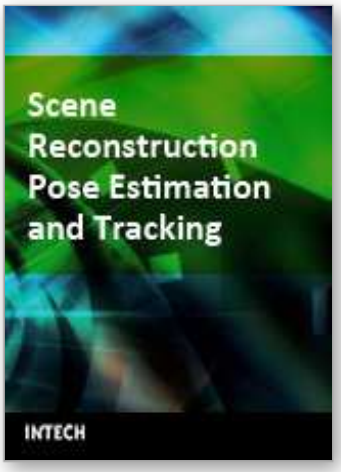
Thorn, F.; Gwiazda, J.; Cruz, A. A. V.; Bauer, J. A. & Held, R. (1994). The development of eye alignment, convergence, and sensory binocularity in young infants, Investigative Ophthalmology and Visual Science, Vol. 35, 544-553, Online ISSN: 1552-5783, Print ISSN: 0146-0404

Treisman, A. M. & Gelde, G. (1980). A feature-integration theory of attention, Cognitive Psychology, Vol. 12, No. 1, 97-136, ISSN: 0010-0285

Walther, D.; Rutishauser, U.; Koch, C. & Perona, P. (2005). Selective visual attention enables learning and recognition of multiple objects in cluttered scenes, Computer Vision and Image Processing, Vol. 100, No.1-2, 41-63, ISSN: 1077-3142

Werblin, F.S. & Roska, B. (2004). Parallel visual processing: A tutorial of retinal function, Int. J. Bifurcation and Chaos, Vol. 14, 83-852, ISSN: 0218-1274

Yamato, J. (1999). A layered control system for stereo vision head with vergence, Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, Vol. 2, pp. 836 -841, ISBN: 0-7803-5731-0, Oct. 1999, Tokyo, Japan



Scene Reconstruction Pose Estimation and Tracking

Edited by Rustam Stolkin

ISBN 978-3-902613-06-6

Hard cover, 530 pages

Publisher I-Tech Education and Publishing

Published online 01, June, 2007

Published in print edition June, 2007

This book reports recent advances in the use of pattern recognition techniques for computer and robot vision. The sciences of pattern recognition and computational vision have been inextricably intertwined since their early days, some four decades ago with the emergence of fast digital computing. All computer vision techniques could be regarded as a form of pattern recognition, in the broadest sense of the term. Conversely, if one looks through the contents of a typical international pattern recognition conference proceedings, it appears that the large majority (perhaps 70-80%) of all pattern recognition papers are concerned with the analysis of images. In particular, these sciences overlap in areas of low level vision such as segmentation, edge detection and other kinds of feature extraction and region identification, which are the focus of this book.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Sang-Woo Bana and Minho Lee (2007). Biologically Motivated Vergence Control System Based on Stereo Saliency Map Model, Scene Reconstruction Pose Estimation and Tracking, Rustam Stolkin (Ed.), ISBN: 978-3-902613-06-6, InTech, Available from: http://www.intechopen.com/books/scene_reconstruction_pose_estimation_and_tracking/biologically_motivated_vergence_control_system_based_on_stereo_saliency_map_model

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821