TELKOMNIKA, Vol.9, No.3, December 2011, pp. 483~488
ISSN: 1693-6930
accredited by DGHE (DIKTI), Decree No: 51/Dikti/Kep/2010

Virtual View Image over Wireless Visual Sensor Network

Nyoman Putra Sastra*[1], Wirawan[2], Gamantyo Hendrantoro[3]
[1,2,3]Electrical Engineering Department, Institut Teknologi Sepuluh Nopember
Kampus ITS, Sukolilo, Surabaya, telp. 031-5947385
[1]Electrical and Informatics Engineering Department, Universitas Udayana
Kampus Bukit Jimbaran – Bali, telp/fax. 0361-703115
e-mail: [1]putra.sastra@unud.ac.id, [2]wirawan@its.ac.id, [3]gamantyo@ee.its.ac.id

Abstrak

Sensor visual umum digunakan untuk membentuk citra arah pandang virtual. Semakin banyak sensor visual yang digunakan maka semakin lengkap informasi yang didapat. Namun pembangkitan citra ini merupakan tantangan tersendiri pada lingkungan Jaringan Sensor Visual Nirkabel karena adanya keterbatasan energi, kompleksitas komputasi, dan lebar pita yang tersedia. Untuk mengatasi keterbatasan bandwidth dan energi tersebut, maka pada paper ini ditampilkan metode pembangkitan citra virtual melalui pemilihan kamera tertentu dari sekumpulan kamera pada Jaringan Sensor Visual Nirkabel. Penggunaan metode ini akan meminimisasi transmisi jumlah citra tanpa mengurangi kualitas informasi yang diperlukan.

Kata kunci: jaringan sensor visual nirkabel, metode pemilihan camera, virtual view

Abstract

In general, visual sensors are applied to build virtual view images. When number of visual sensors increases then quantity and quality of the information improves. However, the view images generation is a challenging task in Wireless Visual Sensor Network environment due to energy restriction, computation complexity, and bandwidth limitation. Hence this paper presents a new method of virtual view images generation from selected cameras on Wireless Visual Sensor Network. The aim of the paper is to meet bandwidth and energy limitations without reducing information quality. The experiment results showed that this method could minimize number of transmitted imageries with sufficient information..

Keywords: *wireless visual sensor network, camera selection method, virtual view*

1. Introduction

Wireless Visual Sensor Network (WVSN) is a system with capability to communicate, to receive and to process signals [1]. In general WVSN architecture, as shown in Figure 1, consists of nodes that contain visual sensor that are spread in area of visual observation. Multi-cameras as visual sensors are used in WVSN to provide multi-view services, multi-resolutions, environmental monitoring, and surveillance system. In WVSN application, visual sensor nodes send captured image and video to sink for further processing suited to application purposes which are designed to meet limited resources such as energy resources and processing capability. However users are still able to get optimal information.

The more sensors to reconstruct scenery, the better results will be obtained. However limitation of energy, processing capability, and transmission bandwidth on WVSNs become obstacles to received maximum information from multi-camera network. To solve those problems, we need a method of virtual sensor camera selection that can give maximum information with minimum amount of data. There are two things to minimize data transmission, i.e. reducing numbers of sensors and maximizing image compression. In other words, we need to select few cameras from all available one, with maximum information.

Algorithms for automatic visual sensor selection on network were designed for any purposes. In study [2] and [3], cameras communicate one another to obtain the amount of active cameras to gain expected spaces on expected scene. Two algorithms were developed in this research, such as (1) distributed processing algorithm, a method used to conduct background and foreground segmentation process that is continued with a method for human face detection, and this method results in lower transmitted data; and (2) centered algorithm, base station uses information obtained from distributed processing to determine the shape on the scene. In this

Received February 1st, 2011; Revised June 15th, 2011; Accepted June 23th, 2011

process it is worth calculating and selecting a camera which may provide the most human face counts on 3D scene. Other cameras which are considered as helper cameras were selected with minimum visual hall method or minimum region of interest (ROI). Meanwhile, in [4] a method for visual sensor selection was designed to determine camera that gives best field of view from one's field of view on the scene. The three methods conduct three counting process on node sensor and this process takes longer time and needs more energy.

Rui and Akyildiz in their research [5] designed a method for visual sensor selection based on spatial and entropy correlation among overlap images to determine differences of information content from the overlap images. This camera selection method aims to combine images from several visual sensors to widen the view of the scene. Researches conducted on [6] and [7] are to select a camera leading to the most active man on a human group in a room based on semantic feature criterion involving body gestures and lips movement (talk). Thus, in a teleconference, only the camera that leads to people who are talking that will be active.

Studies on visual sensor or camera selection as mentioned above aim to reduce the transmission amount of data towards a centralized processing unit in different goal and method according to the above explanation. On the other hand, not all of those studies have demonstrated minimum resource utilization. Another study discussing about camera selection is explained in [8]. The main purpose of the camera selection here for object tracking, such as for tracking a walking man. In this study object movement can be predicted using fuzzy method so that it can give order to activate other cameras according to the object movement. Study on camera selection for object tracking application was also conducted in [9] and [10]. Study in [9] uses a geometry calculation method for object movement to visual sensor on network, while in [10] is based on spatial probability values of object movement to visual sensor nodes. Camera selection on application category of object tracking is to activate other cameras and to disable cameras which are currently active according to direction of the object movement. Both studies do not discuss visual sensor selection if it is used to generate images to particular field of view according to user's wishes and being not applied in WVSN environment

Images of virtual view in [11] are generated from images that result from eight visual sensors. This study does not relate to process of image selection criterion that is used to generate virtual view image. Ray Space method has been applied to build the virtual image. The advantage of the method applied in this study is its ability to overcome problem warping 3D image considered as the base to generate virtual view image. On the other hand, this study does not take into account the use of energy consumption as a result of activating all nodes contributing on the expected scene. The problem is energy consumptions of transmission will be higher than those from image processing on node. Therefore this method will be difficult to be applied in WVSN environment. Hence it is important to find out a new image generating method that uses minimum amount of sensor to reduce the use of resources.

A study that is conducted in [12] applied a method to replay a foot ball scene. The method describes that virtual view image is generated from interpolation of two and three cameras close to virtual images that are selected by the user, but it is not significantly determined by the way of choosing the used cameras. The football scene can be divided into three or four areas depend on the scene characteristics. Particular projection geometry is used on view interpolation for each area. By separating offline process with online one, the foot ball scene can be completely and effectively rendered. The use of resources is not the main purpose of this study.

From all the studies discussing virtual image generating, none is clearly discussing the effect of resource efficiency to the quality of resulted images. The resources such as energy and bandwidth in WVSN are main limitation in enhancing quality of images. On the other hand, preliminary research has been conducted on low energy consumption by modelling WVSN environment [13]. As a result, the paper focuses on minimizing energy and bandwidth utilization by selecting the right visual sensor in WVSN to generate virtual images while keeping quality threshold.

2. Research Method

2.1 Problem statement

Virtual view image is needed when a user wants to see a particular object which is no visual sensors available in that field of view. To have the virtual image, then a procedure of virtual view image generation is required. The procedure is shown in Figure 2. Let C be collection of visual sensor in certain area, $C=\{C_1, C_2, C_3, \ldots, C_N\}$ that are already calibrated. Images set created by C is $I = \{I_1, I2, I3, \ldots, I_N\}$. The number of visual sensors are used to generate a virtual view image is limited to a number of M with M < N. N is number of all visual sensor available.

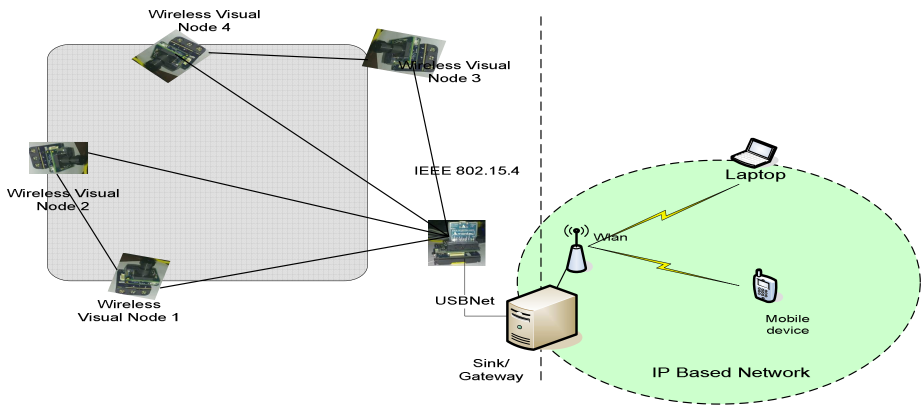


Figure 1. Wireless Visual Sensor Network Architecture

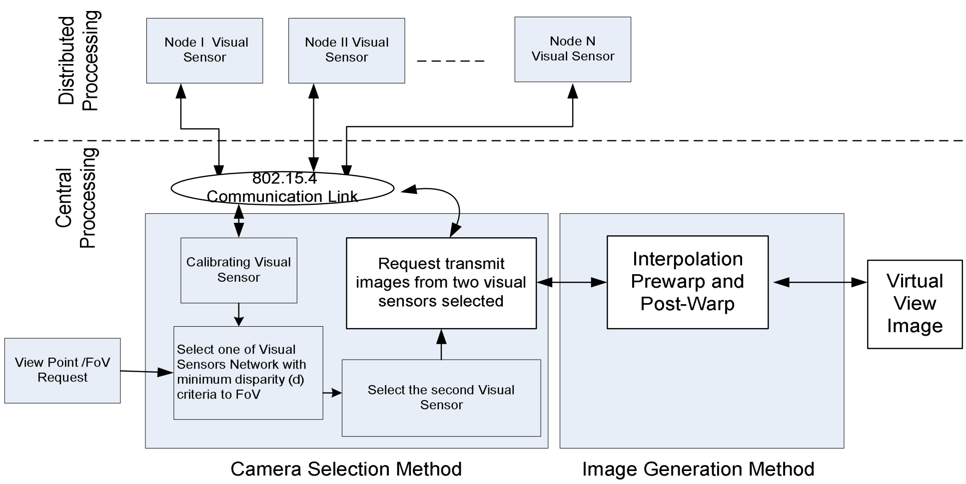


Figure 2 Virtual View Image Generation in WVSN Platform Mechanism.

2.2 System Model

System model as shown in Figure 3 describes that five sensor nodes are installed with embedded Linux operation system with visual ability [1]. General specification of visual sensor is IMB 400 Imote2 with chipset OV7670 that results images with resolution 640x480 pixels. File is data raw format, portable pixel map (ppm), with image size 901 kB. As an object is a statute with 5cm thick which is installed in the wall. Distance between camera and image field center is constantly 2 meters.

To conduct visual sensor selection, position of each visual sensor needs to be recognized. The position has known by conducting a calibration. Calibration aims to obtaine intrinsic and extrinsic parameters of visual sensor. Pointer laser links image field center point and the cameras . Then, value of extrinsic parameter is calculated by using method in [14] and is specifically explained in [15] .Visual sensor only rotates along X axes (horizontal direction). In our experiment, we pick one camera to be a basis of virtual view generation.

If there is demand to show an image in particular FoV (field of view) where there have no visual sensors, the system will calculate to look for the nearest disparity value from FoV to the existence visual sensor. First visual sensor will be chosen based on minimum disparity value between existence visual sensor and expected FoV. If the first visual sensor is located in the left of expected FoV, the second visual sensor located in the right side will be chosen. Then each selected node is requested to send the captured images. Calculation in the centralized system may minimize communication with node sensor. This deals to resource efficiency especially in the node sensor with limited resources. Two images from selected node sensors are then employed as the base to generate virtual view image.

2.3 Effect of Camera Location on the Virtual Field of View

Disparity is difference of depth of a point P in the scenery, but different field of view. Disparity changes as rotation and translation are difference of one field of view to other view. By this disparity, the point P is projected on image plane that produced by camera *C.* The image from camera *C'* will be placed on different pixel location. Lower disparity value from two fields of view indicates closer pixel location for the same point on different images. With this assumption, disparity value can be used as a reference in selecting virtual sensors.

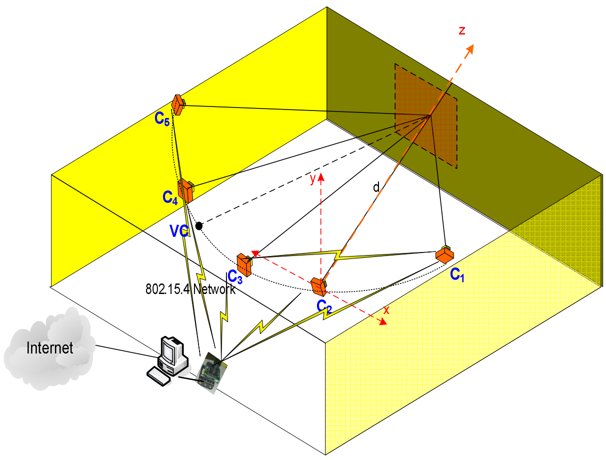


Figure 3. Modeling visual sensor placement and field of view locations

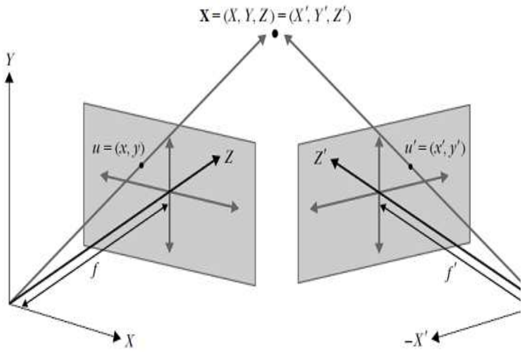


Figure 4 Geometry of two cameras [14]

2.4 Geometry of Two Cameras

Two cameras were used to generate images of the same scenery, as shown in Figure 4. According to Aghajan and Cavallaro [14], one point in first camera *C* has projection point as below:

$$x = f\frac{X}{Z} \qquad y = f\frac{Y}{Z} \tag{1}$$

a projection point on the second camera *C'*, is stated by below equation:

$$x' = f'\frac{X'}{Y'} \qquad y' = f'\frac{Y'}{Z'} \tag{2}$$

Coordinate of system *C'* can be expressed as coordinate of system *C* which is rotated by R and followed by translation T, $\lfloor t_x \quad t_y \quad t_z \rfloor$ as:

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \hat{R}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{3}$$

Then there is no rotation around axes X and Z. The rotation is only for horizontal direction. The rotation equation becomes [14]:

$$R = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \tag{4}$$

Point *P* on image plane that is produced by visual sensor *C* can be placed on image plane that is produced by visual sensor *C'* by substituting coordinate equation between visual sensor *C* and visual sensor *C'* into projection equation below

$$x' = f'\frac{X'}{Z'} = \frac{\cos\beta \frac{f'}{f} x + \sin\beta \frac{f'}{f}}{\frac{-\sin\beta}{f} + \cos\beta} \tag{5}$$

$$y' = f'\frac{Y'}{Z'} = \frac{1}{\frac{-\sin\beta}{f} + \cos\beta} \tag{6}$$

Whenever field of view of visual sensor *C'* is field of view of desired virtual view, this geometry of two cameras model can be used to get scene for virtual view image.

2.5 Disparity

We can get disparity value from a reference point on geometry of two cameras. Equation (7) determines a reference point as below [9]:

$$d = \frac{1}{4}\left(\left(\left|\frac{d\sin\beta}{d+\cos\beta}\right| + \frac{d\sin\beta}{d-\sin\beta} + \frac{d\cos\beta}{d+\sin\beta} - 1 + \frac{-d\cos\beta}{d-\sin\beta} + 1\right)\right) \tag{7}$$

where *d* is mean of distance of four vectors that change value with rotation.

2.6 Camera Selection Method Based on the Geometry of Two Cameras

We select the first visual sensor by searching smallest rotation angle *β* that has minimum disparity between camera *C* and desired field of view. Second camera is selected by smallest disparity with opposite direction to *β*, and is seen from virtual view. For example, if position of the first camera is on the right side of virtual view, so the position of the second camera is on the left side of virtual view. From these two points, we formulate disparity field of view when is seen from both cameras's position by equation below:

$$d_1' = \frac{d_1}{d_1 + |d_2|} \text{ and } d_2' = \frac{d_2}{d_1 + |d_2|} \tag{8}$$

where d_1 and d_2 are disparities of the two cameras. They are selected against the desired FoV.

2.7 Virtual View Generation

Pixel correspondence between two images from two selected sensors required to interpolate intermediate image. To make a fast computation, we can use method of correspondence pixel pairs search along epipolar line from both images. Epipolar line correspondence is:

$$\begin{cases} l = F^T p \\ l' = Fp' \end{cases} \tag{9}$$

where *l* and *l'* are a pair of epipolarline, and *F* is fundamental matrix [14] as a result from this equation:

$$p^T = Fp' = 0 \tag{10}$$

where *p* and *p'* are a pair of point from two-dimensional coordinate that corresponds from two image. *F*, is fundamental matrix that has redundancy value. F can be solved by least square method from equation below:

$$\sum_{i=1}^{n}\left(p_i^T F p_i'\right)^2 \tag{11}$$

Furthermore, interpolation process aims to generate intermediate image from two available images. This process suited to disparity viewpoint of two selected cameras. To find new image points, we do interpolation process below:

$$p_v = d_1' p + d_2' p' \tag{12}$$

Post-warp process aims to estimate epipolar lines coordinate on virtual view and warps interpolated image by mean of placing those lines on correct position on virtual view. By then, virtual view is produced.

2.8 Peak Signal-to-Noise Ratio (PSNR)

Peak Signal-to-Noise Ratio (PSNR) is the maximum signal power ratio with noise signal power. PSNR in image is used to seek image quality generated from a communication process with the original image/expected image. Every pixel has changeable value of the communication process, can be compressed image or transmission image. If the original image is stated as I_1 and the result of processing system is I_2, PSNR is defined as below:

$$PSNR = 10\log_1\left(\frac{Max_{I_1}^2}{MSE}\right) \tag{13}$$

The MSE value can be calculated as follow

$$MSE = \frac{1}{m.n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I_1(i,j) - I_2(i,j)\right]^2 \tag{14}$$

with m x n as the size of image.

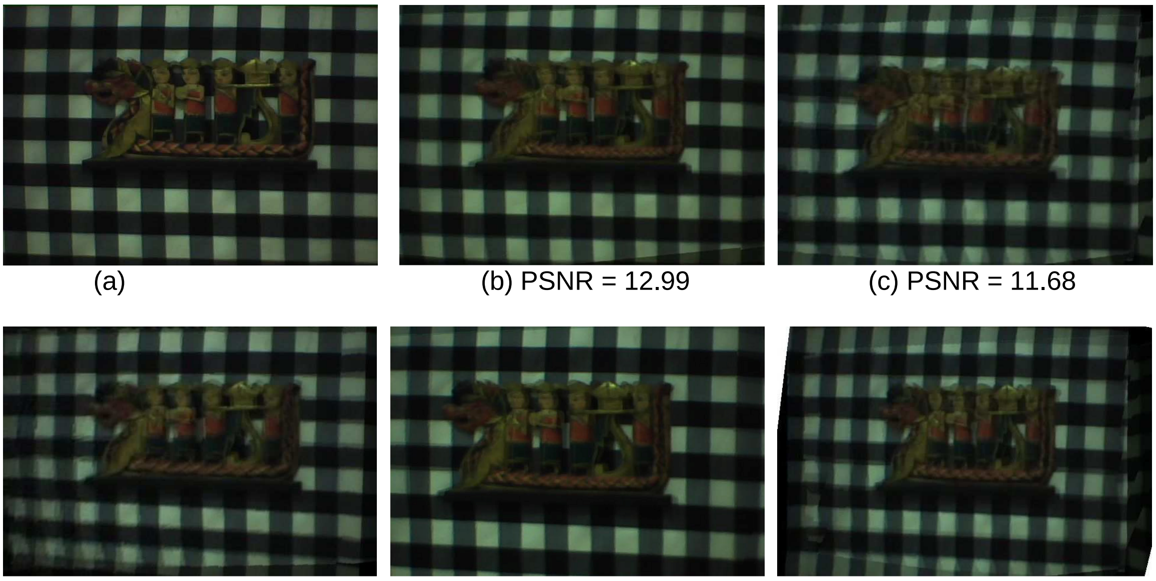
3. Experimental Result and Discussion

Two factors are compared in this experiment, i.e. the number of reference points and the angle between desired field of view with visual sensors are selected. In the first experiment, we use 17 and 35 reference points. The rotation angle for the first selected visual sensor 30^o and 45^o, while the rotation angle of second visual sensor is fixed, i.e., 30^o. Furthermore, the image is taken using visual sensors that have the same intrinsic characteristics in the field of view virtual view.

Figure 5 (a) represents the desired image, while (b) is an image formed from 17 reference points and the first and second camera rotates 30^o and -30^o (disparity = 0.5) to the desired field of view. Figure 5. (c) is a virtual image generated from 17 reference points while the first and second visual sensor rotates at 45^o and -30^o. The first experiment showed that, the virtual image disparity affects its outcome. The smaller the disparity, the virtual image will be more desirable. However, the greater the disparity, the more artifacts will occur. To handle the large-value disparity, it would require more reference points.

We used 17 reference points in the second experiment and the first visual sensor rotates starting from 20^o, 50^o, and 65^o, while the second visual sensor angle is fixed, of. 10^o. We also find more clearly, that the disparity is very influential in forming a virtual image. As can be seen in fig. 5 (d), (e), and (f) that more artifacts appear in the virtual image. We still investigated methods of interpolation between corresponding points from two selected images to improve the quality of virtual view images (reduce or even eliminate artifacts).

The resulted image based on PSNR value shows that the more disparity between expected FoV of the first and the second selected visual sensors, the lower PSNR value we obtain. This can be seen in PSNR value in Figure 5 b, c, d, e, and f and summarized in table 1. This is clearly shown on image generated in Figure 5 d, e, and f. The three images are attained by changing disparity value of the second sensor with value 20^o, 50 ^{o} and 65^o. However, the disparity value of the first sensor is constantly kept.



(a) (b) PSNR = 12.99 (c) PSNR = 11.68

(d) PSNR = 13.33 (e) PSNR = 12.10 (f) PSNR = 10.79

Figure 5. Virtual view result, image resolution: 640x480 pixels. (a) desired image (b) ref. points=17, β_1=30 and β_2=-30 (c) ref. points=17, β_1=45^0 and β_2=-30 (d) ref. points=17, β_1=10^o and β_2=-20^o (e) ref. points=17, β_1=10^0 and β_2=-50^o (f) ref. points=17, β_1=10^0 and β_2=-65^o

Table 1 PSNR value againts disparity of expected FoV.

Image	d1	d2	d1'	d2'	PSNR
b	30	-30	5	0.5	12.99
c	45	-30	6	0.4	11.68
d	10	-20	33	0.67	13.33
e	10	-50	17	0.83	12.10
f	10	-65	0.13	0.87	10.79

Table 2. Average delivery time vs No. Of transmitted images

Number of images transmitted	Average time to deliver	
	seconds	Minutes
1	138	2.30
2	336	5.60
3	489	8.15
4	605	10.08

Table 2 demonstrates the experimental results of image transmission from the node to the sink. The results show that the delivery time is significantly influenced by the number of images that are sent. These occur, as the size of image of the visual sensor is 901 kB, communication speed limits to 250 kbps and the size of header in each packet.

4. Conclusion

In this paper, we model and utilise WVSN where positions of virtual sensors are calibrated. Those visual sensors generate virtual view as desired viewpoint. From desired viewpoint, we select two visual sensors with opposite location from virtual view, by smallest disparity criteria seen from virtual view. The captured images by both visual sensors are used to generate virtual view image. The experimental results show that the use of camera selection method can reduce quite significantly the number of images delivery of visual sensors in WVSN, i.e. from 4 images to be 1 image. In addition the results show that the method reduced average delivered time of images by 7.38 minutes. As a result, the method can accelerate the process of generating virtual view. In our future work, we will investigate other factors in generating virtual image over WVSN, such as distributed pre-processing in the sensor nodes and energy consumption of collaborations among the sensors.

Acknowledgement

The authors gratefully acknowledge their team in Multimedia Communication Labs (ITS) for the many insightful discussions and contributions. This work has been supported by Directorate General of Higher Education (DHGE) of Indonesia Government.

Reference

[1] N. P. Sastra, Wirawan, G. Hendrantoro. *Design and Implementation of Wireless Multimedia Sensor Network Nodes Based on Linux OS.* 13th Seminar on Intelligent Technology and Its Applications. Surabaya. 2010;1: 98-101
[2] M. Morbee, L. Tessens, L. Huang, W. Philips, H. Aghajan. *Optimal camera selection in vision networks for shape approximation*. IEEE 10th Workshop on Multimedia Signal Processing. Cairns, Queensland, Australia. 2008; 1: 46-51.
[3] H. M. Toshihiro Matsui, Akira Iwata. *Dynamic Camera Allocation Method Based on Constraint Satisfaction and Cooperative.* 2nd International Conf. on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing. Nagoya, Japan. 2001; 1: 955–964.
[4] L. Tessens, M. Morbee, L. Huang, W. Philips, H. Aghajan. *Principal view determination for camera selection in distributed smart camera networks*. 2nd ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC 2008). Stanford University. California, USA. 2008; 1: 1-10.
[5] D. Rui dan I. F. Akyildiz. *Joint Effect of Multiple Correlated Cameras in Wireless Multimedia Sensor Networks.* IEEE International Conference on Communications (ICC '09). Dresden, Germany. 2009; 1: 1-5.
[6] B. Hornler, D. Arsic, B. Schuller, G. Rigoll. *Boosting multi-modal camera selection with semantic features*. IEEE International Conference on Multimedia and Expo (CME 2009). New York City, USA. 2009; 1: 1298-1301.
[7] M. Al-Hames, B. Hornler, R. Muller, J. Schenk, G. Rigoll. *Automatic Multi-Modal Meeting Camera Selection for Video-Conferences and Meeting Browsers*. IEEE International Conference on Multimedia and Expo. Beijing, China. 2007; 1: 2074-2077.
[8] K. Morioka, S. Kovacs, J. H. Lee, P. Korondi, H. Hashimoto. *Fuzzy-based camera selection for object tracking in a multi-camera system*. Conference on Human System Interactions. Krakow, Poland. 2008; 1: 767-772.
[9] E. Monari, K. Kroschel. *A knowledge-based camera selection approach for object tracking in large sensor networks.* 3rd ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC 2009). Como, Italy. 2009; 1: 1-8.
[10] K. Jiman, K. Daijin. *Probabilistic camera hand-off for visual surveillance.* 2nd ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC 2008). 2008; 1; 1-8.
[11] Y. Mori, N. Fukushima, T. Fujii, M. Tanimoto. *View Generation with 3D Warping Using Depth Information for FTV*. 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video; 2008; 1; 229-232.
[12] N. Inamoto, H. Saito. Virtual Viewpoint Replay for a Soccer Match by View Interpolation From Multiple Cameras. *IEEE Transactions on Multimedia*. 2007; 9(6): 1155-1166.
[13] N. P. Sastra, I. M. O. Oka Widyantara, D. M. Wiharta, Wirawan, G. Hendrantoro. *Modeling Wireless Visual Sensor Network with a Low Energy Consumption for Desired Image Quality and View Point.* 6th Kumamoto University Forum, Surabaya - Indonesia. 2008; 1: 94-95.
[14] H. Aghajan, A. Cavallaro. Multi-Camera Networks: Principles and Applications. ELSEVIER; 2009.
[15] J. Heikkila. Geometric Camera Calibration using Circular Control Points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000; 22(10): 1066-1077.