Energy Efficiency of Image Compression for Virtual View Image over Wireless Visual Sensor Network

Nyoman Putra Sastra\textsuperscript{a,b}, Wirawan\textsuperscript{a}, and Gamantyo Hendrantoro\textsuperscript{a}

\textsuperscript{a}Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia
Email: \{pirawan, gamantyo\}@ee.its.ac.id
\textsuperscript{b}Department of Electrical Engineering, Universitas Udayana, Bali, 80236 Indonesia
Email: putra.sastra@unud.ac.id

Abstract—Many applications use virtual view to see objects or sceneries in certain directions where there is no angle of visual sensor viewing the directions. Generating indirect virtual view images also save visual sensor node in a network. In addition visual sensors have abilities to give more information related to virtual view images. In fact, generating virtual view images using Wireless Visual Sensor Network is a challenging task, due to the limitation of energy, computation, and bandwidth. Therefore this paper proposes using combination of camera selection method with parameters optimization of JPEG 2000 image compression and image transmission strategy applying mutual information criteria. The camera selection method based on the smallest value of disparity was adopted to reduce transmission of number of images. In each node of visual sensor, the optimization of JPEG 2000 image compression using parameters selection such as DWT level and bit rate was applied to reduce the size of the images. However image quality in PSNR has to be preserved. By implementing the sensor selection, the image compression, and the mutual information strategic, the study results that the total energy required to send one compressed image can be reduced to 10.27\% of energy consumption of uncompressed image. Thus the image will be faster delivered.

Index Terms—Wireless Sensor Network, virtual view images, image transmission, image compression, Linux embedded

I. INTRODUCTION

WIRELESS Sensor Network (WSN) is a wireless communication system with limited resources, especially on energy, computation, and bandwidth [1], [2]. The use of WSN platform in visual data transmission, such as image and video, is known as WVSN [3]. Visual sensor or camera generates information in the form of images, or video, to be further processed for various visual applications at a distributed processing unit or a central processing. WVSN is widely used in many applications such as environmental observation and surveillance system, which is designed with limited resources to provide optimal information to users. This limitation of resources, mainly on energy source of WVSN, creates a challenge in system design since the size of data is larger than scalar data.

Visual applications require various angle of view in order to observe an object. Even though visual sensors are distributed in many places, certain FoV may not be available. Moreover, observation of one particular direction is changing in time because of the movement or displacement of objects and weather condition. To address user demand of viewing an object from certain angle, more than one camera is required.

More visual sensor means better view or scenery reconstruction, but limitation of energy, computation, and bandwidth become issues to be addressed. Thus, an image compression technique is required to save those resources. A visual sensor selection method is also needed to generate virtual image with the lowest size of data transmission.

In wireless network, energy required to transmit information is larger than energy to process data [3]. To save energy, there are some issues to be considered in this paper: 1) selecting images to be transmitted which has differences of the information from predefine image of each visual sensor, 2) Applying distributed compression to the captured images, 3) Reducing number of images or visual sensors as basic image to generate image on certain FoV.

The first issue will be addressed in Section 2.A.4 that discuss implementation of mutual information strategy. Section 2.A.1 up to Section 2.A.3 will meet the second issue by optimizing JPEG 2000 parameters, which is adapted in Imote2 with its OS Linux Embedded. Last issue will be responded in Section 2.C and Section 2.D, which will discuss selection technique of visual sensor, or camera who provide maximal information of images, in other words, from all available visual sensors, only few are selected to provide maximal information.

The method of visual sensor selection is required to reduce number of visual sensors. There are many visual selection methods that had been developed by some researchers. In [4]–[6], cameras are communicating to find number of active cameras to create space that is desired in scene. They developed distributed processing and centered algorithms. These three researches require three processes on sensor nodes and these processes require extended time that lead to significant energy consumption. None of them is aimed to directly reconstruct virtual image in WVSN.
platform.

In [7] and [8], researches of multiview image from overlapping images captured by a group of visual sensors in WVSN environment have been conducted to broaden the view. In [7], visual sensor selection is based on spatial correlation and entropy between images. The implementation of image joining in WVSN or WMSN in [8] is mainly about the effort of resources utilization reduction, especially on memory usage, by reducing less important information and transmitting only image’s portion of a visual sensor that overlapped with images from other sensor visual. There is no explanation of energy consumption in that research. In [8] also nothing specifically addresses the selection of visual sensors to evoke the image of a particular FoV. Ref. [9] is research of generating duplicates of foot ball play scene. In this method, image is constructed from two and three visual sensors located near the FoV without explicitly describing the criteria of visual sensor selection.

In our previous research [10], we have developed a method of image construction in WVSN. This method has successfully generated virtual image with low PSNR compared to capture image by visual sensor in the FoV. The research had not implemented image compression after capturing using visual sensor. This causes the image file size becomes larger, such that sending one image to central processing requires up to 10 minutes of time, with the cost of high energy consumption. Additionally, the research did not make effort of reducing data transmission. Moreover, In this study did not implement the mutual information criterion for the selection of images to be sent.

Based on the discussion of the above studies, the main contribution of this research is to give solution for energy saving in distributed processing. The distributed processing by means of image compression and image selection which have different information to the information already stored in each visual sensor node. The use of energy on each process is measured. The quality of virtual view image is improved and compared to the result in [10]. WVSN platform XScale PXA 271 processor, visual sensor SOC OV 7670, and sensor board with Linux Embedded OS [11]–[13] are used in this research. This paper is organized as follows. Section II describes proposed scheme and system design. Section III presents testing, measurement and analysis of the system. Finally, section IV concludes the research.

II. THE PROPOSED SCHEME

Multivisual sensors are used in WVSN to provide multiview and multi resolution services, and also for environment surveillance. With the availability of resources, image can be constructed by sending all image captured by visual sensor. In WVSN application, visual sensor node processes the image or video before transmission.

Virtual view image is image on peculiar FoV where no camera is directly available to serve users. Many applications have taken advantages of the virtual view image, for examples, (1) object recognition as a result of captured object from different point of view, (2) view widen of specific scenery, and (3) virtual reality such as to see player and ball position in foot ball game.

WVSN inherently has limited resources; as a result a method is required to save resources by sending the image that was selected based on some defined criteria. Fig. 1 shows the platform of generating virtual view image in WVSN. Initially, system is setup whereas sensor node position is calibrated. Each sensor node will send its captured image to central processor as a referenced image. When an image is requested based on FoV, the system will calculate and calibrate the requested FoV according to all images in the central processor. A sensor node, then, is selected to send the image which is employed as a virtual view image foundation. Next step, the sensor node will capture and compressed new images. The new images are compared to the referenced image. Mutual information is used as decision basis in sending the image. Finally the image will be foundation in building virtual view image.

A. Image Processing in Visual Sensor Node

1) Wireless Sensor Network Platform: In this research we use Imote2 as WSN platform. Imote2 has modules of radio board and visual sensor board. The radio board uses Intel Xscale PXA271 processor with 32 MB SDRAM, 32 MB flash, IO port as GPIO, 2x SPI, 3x UART, PC, SDIO, USM host, USB client, and integrated wireless communication IEEE 802.15.4 (CC2420). The Visual sensor board provides miniature speaker and line output, color image and video camera chip, omni Vision OV7670, audio capture and playback CODEC WM890, onboard microphone and line input, and PIR (Passive Infrared) motion sensor. Table I displays the Imote2 consumption energy parameters.

2) Embedded Linux Operating System: Linux embedded has been successfully developed in Imote2 board [12], [13]. With this Linux embedded system, development of
for tiles on image boundaries. Each tile is compressed into uniform size of non-overlapping squared tiles, except tiling process is the process of dividing source image set, and Irreversible Colour Transform (ICT) [14]. The formed after preprocessing, including tiling, level off-processes of JPEG 2000 compression method are performed below.

The DWT

3) Image Compression on Visual Sensor: The DWT processes of JPEG 2000 compression method are performed after preprocessing, including tiling, level offset, and Irreversible Colour Transform (ICT) [14]. The tiling process is the process of dividing source image into uniform size of non-overlapping squared tiles, except for tiles on image boundaries. Each tile is compressed independently using its own parameters. The aim of tiling is to reduce memory utilization.

Preprocessed image source becomes input for blocks of wavelet transform. DWT is a two dimensions filter separable along its columnand row. The filter works by applying low pass filter (L) and high pass filter (H) on each sample, row by row, and columnby column. The output is then applied again to the same filter. As the result, image is divided into sub-bands: low-low (LL1) low-high (LH1), high-low (HL1) and high-high (HH1) [15]. High pass sub-band refers to the residues of original image, which is required in the perfect reconstruction of the image from its low resolution version. DWT level can be raised to create transformation level two with two sub bands LL2, LH2, HL2, and HH2, and so on.

The implementation of JPEG 2000 in this research is using library OpenJPEG [16] on WSN platform. The JPEG 2000 parameters are altered for system optimization that is reducing energy consumption while keeping the image quality. From our preliminary experiments, as shown in installation step of Linux embedded implementation, the parameters of JPEG 2000 being used in this implementation is as in II. These parameters have low energy consumption with adequate PSNR.

4) Mutual Information to Determine Image Transmission: Joint histogram of two images can be used to estimate joint probability distribution from grey values of image by dividing each value of histogram with its total value. Shannon entropy for joint distribution is defined as:

\[ H = - \sum_{i,j} p(i, j) \log_2 p(i, j) \]  

(1)

The value of joint entropy of two images defines their degree of correlation. In registry image, low joint entropy means that both images are similar. Indifferent to registry image, for camera selection, large value of joint entropy is preferable to get more information.

Another technique to see information contain within two or more images is by computing mutual information. Mutual information is one of many quantities to state how much one random variable provides information about another random variable. The mutual information of two images, I and J, is defined as:

\[ M(i, j) = H(i) + H(j) - H(i, j); \]  

(2)

Where \( H(i, j) \) is joint entropy of two images I and J. The equation shows that low joint entropy means high mutual information. In other words, whenever mutual information...
of two images taken from visual sensors tends to be similar, there is no information is required to be sent from sensor node to central processing. The sensor only sends image, after being compressed, when there is a significant change in mutual information value.

B. Visual Sensor Selection

Disparity is difference of depth of a point $P$, on the same scenery subjected to different field of view, which is shown in 2. Disparity changes as rotation and translation difference of one field of view to other field of view [17]. To simplify the implementation, virtual view movement in this study only the horizontal rotation, which are in the left and right point $P$.

Two visual sensors used to generate images with the same scenery, as in 2, and according to [17], one point in the first camera $C$ has projection point below:

$$x = f \frac{y}{z} \quad y = f \frac{z}{z}$$

Projection that point on the second camera $C'$, is stated by below equation:

$$x' = f' \frac{y'}{z'} \quad y' = f' \frac{z'}{z'}$$

Coordinate system $C'$ can be expressed as coordinate system $C$ rotated by $R$ and followed by translation $T$, $[tx \ ty \ tz]$:

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \bar{R} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} tx \\ ty \\ tz \end{bmatrix}$$

From this system model, there is no rotation around axes $X$ and $Z$ (rotation on horizontal direction only). Rotation equation becomes [15]:

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

Position of point $P$ on image plane that produced by visual sensor $C$ can be located on image plane produced by visual sensor $C'$ by substituting coordinate equation between visual sensor $C$ and visual sensor $C'$ into projection equation:

$$x' = f' \frac{X'}{Z'} = \frac{\cos \beta L' x + \sin \beta L' y}{-f' \frac{d}{d} + \cos \beta}$$

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

We can get disparity value from reference point on geometry of two visual sensors. For visual sensor $i$ and visual sensor $j$ in 2, the disparity, $\delta$, between two images at camera $i$ and camera $j$ with position parameters $(\delta_i, r_i, \theta_i)$ and $(\delta_j, r_j, \theta_j)$ are calculated as [6], [18]:

$$\delta = \frac{1}{4} \left| \frac{d \sin \beta}{d + \cos \beta} + \frac{d \sin \beta}{d - \cos \beta} \right| + \frac{1}{4} \left| \frac{d \cos \beta}{d + \sin \beta} - 1 + \frac{-d \cos \beta}{d - \sin \beta} + 1 \right|$$

First visual sensor is selected by searching smallest rotation angle $\beta$ that has minimum disparity between visual sensors $C$ and desired FoV. Second visual sensors selected by smallest disparity with opposite direction to $\beta$, seen from virtual view. As example, let position of first visual sensor be on the right side of virtual view, so the position of second visual sensor is on the left side of virtual view. From this two points, the FoV disparity when seen from both visual sensors position is defined by equation below:

$$\delta_i' = \frac{d_1}{d_1 + |d_2|}, \quad \delta_j' = \frac{d_2}{d_1 + |d_2|}$$

Where $\delta_i$ and $\delta_j$ are the disparities of the two visual sensors that are selected against the desired direction of view.

C. Virtual View Generation

There are three stages to generate a virtual image view items, namely, interpolation, pre-wrap, and post-wrap. In addition, interpolation is the step to generate intermediate image by using pixel correspondence between two images from two selected sensors. A method of correspondence pixel pairs search along epipolar line from both images is used to make a fast computation. Moreover, epipolar line correspondence is:

$$\begin{cases} \ell = F^T p \ell' = F p' \end{cases}$$

Where $\ell$ and $\ell'$ are a pair of epipolar line, and $F$ is fundamental matrix [10] as a result from this equation:

$$p^T = F p' = 0$$

Where $P^T$ and $P'$ are pair of point from two-dimensional coordinate that corresponds from two images. $F$ is fundamental matrix that has redundancy value. The parameter $F$ can be solved by least square method by minimizing value from equation below:

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

© 2015 ACADEMY PUBLISHER
Interpolation process suited to disparity viewpoint of two selected visual sensors. Furthermore, interpolation process to find new image points is defined below:

\[ p_v = d_i^p + d_i^p' \]  
(14)

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

D. Energy Consumption

Since energy source of WVSN solely from battery with limited power, thus energy consumption becomes a critical criteria on implementation [19]. This means that the main priority of one particular model or any implementation method of image virtual view generator is to produce low energy consumption. In general, there are three tasks that affect energy consumption of WVSN, i.e.: 1) energy for images capturing, 2) energy for images processing, and 3) energy for images transmitting. Those three energy consumptions are formulated as below:

\[ E_{WVSN} = E_{capt} + E_{proc} + E_{transmit} \]  
(15)

\( E_{capt} \) is images capturing energy, \( E_{proc} \) is images processing energy consumption, and \( E_{transmit} \) is images transmitting energy consumption. Furthermore, we can manually compute energy consumption by measuring voltage (\( V \)), current (\( I \)), and time of the whole process i.e. images capturing time (\( t_{capt} \)), images processing time (\( t_{proc} \)), and images transmitting time (\( t_{transmit} \)), with or without any compression methods, as:

\[ E = V \times I \times (t_{capt} + t_{proc} + t_{transmit}) \]  
(16)

Energy consumption of JPEG 2000 image compression can be estimated from DWT processes. In this research, OpenJPEG library used as JPEG 2000 representation using Le Gall filter model 5–3 [15], [20]. Equations regarding the filter are:

\[ L_0(z) = \frac{1}{8}(-z^2 + 2Z + 6 + 2z^{-1} - z^{-2}) \]  
(17)

\[ H_0(z) = \frac{1}{2}(z + 2 + z^{-1}) \]  
(18)

\( H_0(z) \) defines low frequency sub band of DWT filter and \( L_0(z) \) defines high frequency subband of DWT filter. The impulse responses of the filter are:

\[ L[2n] = \frac{-x[2n - 2] + 2x[2n - 1]}{4} + \frac{6x[2n] + 2x[2n + 1] - x[2n + 2]}{4} \]  
(19)

\[ H[2n + 1] = \frac{-x[2n] + 2x[2n + 1] - x[2n + 2]}{2} \]  
(20)

Analyzing energy consumption of DWT image compression requires size and type of some basic operations of the given process. Basic operations such as shifting and adding can derive from (19) and (20). In (19), filter decomposition requires 8 shifts and 8 adds operations to convert 1 pixel of an image into low pass coefficient, meanwhile (20) requires 2 shifts and 4 adds operations to convert into high pass coefficient.

Overall the computation of an image on DWT decomposition is acquired by counting the whole operations. The computation of one image of size \( MN \) with \( L \) DWT level is estimated step by step as follows. Assume that the first step is horizontal decomposition. Since the whole pixels at even position are decomposed into low pass coefficient and pixels at odd positions are decomposed into high pass, which total weight at horizontal is \( \frac{1}{2} MN(10S + 12A) \). The vertical decomposition has similar weight. Image’s size will reduce with factor of 4 from initial size for every transformation level. Then, the total computation cost becomes:

\[ C_{DWT}(M, N, L) = MN(10S + 12A) \sum_{i=1}^{L} \frac{1}{4^{i-1}} \]  
(21)

Apart from arithmetic operations, on transformation steps, energy is significantly consumed from memory access. Within memory access processes, each pixel goes through read and write processes 2 times, thus memory access for \( M \times N \) image is:

\[ C_{Read}(M, N, L) = C_{Write}(M, N, L) = 2MN \sum_{i=1}^{L} \frac{1}{4^{i-1}} \]  
(22)

Total processing load is computed as sum of computational load (\( C_{DWT} \)) and memory access load, (\( C_{read} \)) and (\( C_{write} \)), from Eq. (21) and Eq. (22) and becomes:

\[ C_{proc} = C_{DWT} + C_{read} + C_{write} \]  

\[ C_{proc}(M, N, L) = MN(10S + 12A) \sum_{i=1}^{L} \frac{1}{4^{i-1}} + 2MN \sum_{i=1}^{L} \frac{1}{4^{i-1}} \]  
(23)

Furthermore, if \( E_{shift} \) denote energy consumption for shift (\( S \)) operation, \( E_{add} \) denote energy consumption for add (\( A \)) operation, \( E_{read} \) denote energy consumption for read memory access, and \( E_{write} \) denote energy consumption for write memory access, then the overall energy consumption for processing becomes:

\[ E_{proc} = E_{DWT} + E_{read}C_{read} + E_{write}C_{write} \]  

\[ E_{proc}(M, N, L) = MN(10E_{shift} + 12E_{add}) \sum_{i=1}^{L} \frac{1}{4^{i-1}} + E_{write}2MN \sum_{i=1}^{L} \frac{1}{4^{i-1}} + E_{read}2MN \sum_{i=1}^{L} \frac{1}{4^{i-1}} \]  

\[ E_{proc}(M, N, L) = MN(10E_{shift} + 12E_{add} + 2E_{write} + 2E_{read}) \sum_{i=1}^{L} \frac{1}{4^{i-1}} \]  
(24)
Image transmission is generally in the form of data packets. If the number of packets is \( m \), average packets size is \( t \), and energy required to transmit 1 bit data is \( E_b \), transmission energy is defined as:

\[
E_{\text{transmit}} = m \cdot t \cdot E_b \tag{25}
\]

The IEEE 802.15.4 communication system on Linux embedded OS uses the tosmac structure data packet [1], with the payload of each packet being 28 bytes. From this model and substituting (24) and (25) into (15), the energy consumption for transmitting a JPEG 2000 compressed image becomes:

\[
E_{\text{wvsn}} = E_{\text{capt}} + E_{\text{proc}} + E_{\text{transmit}} \\
E_{\text{wvsn}}(M, N, L) = E_{\text{capt}} + MN(10E_{\text{shift}} + \\
12E_{\text{add}} + 2E_{\text{read}} + 2E_{\text{write}}) \sum_{i=1}^{L} \frac{1}{4i-1} + m \cdot t \cdot E_b
\]

According to (26), the captured energy remains fixed for the same devices. Therefore total consumption energy depends on both processing energy \( E_{\text{proc}} \) and transmission energy \( E_{\text{transmit}} \). The transmission energy \( E_{\text{transmit}} \) is affected by image size in bits.

E. Peak Signal to Noise Ratio (PSNR)

Error metrics often used to compare various image compression techniques are the Mean Square Error (MSE) and the Peak Signal to Noise Ratio (PSNR). The MSE is the cumulative squared error between the compressed and the original image, whereas PSNR represents a measure of the peak error. The mathematical formula for PSNR and MSE are:

\[
MSE = \frac{1}{m \cdot n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \tag{27}
\]

\[
PSNR = 20 \log_{10} \left( \frac{\text{MAX}_I}{\sqrt{MSE}} \right) \tag{28}
\]

Where \( I \) represents matrix of original images, \( K \) is compressed image, \( m \) is number of pixel row of image, \( i \) is index of that row, \( n \) represents number of image’s pixel column and \( j \) is index of the column.

F. Procedures

The following flowchart of virtual view image generation procedures in accordance with 3, is described as follows:

1) A number of visual sensors are installed and deployed in the desired area of coverage. These visual sensors send visual information to the central processing. The central processing unit is in charge of collecting information from each of the visual sensor and received a request from a user for a FoV.

2) Calibration to determine the position and direction of view and visual coverage of each sensor or point of reference to sensor node is done before any other task because the direction of visual sensor is assumed not to be changed for energy savings. The disparity between cameras is obtained by determining the direction of view of each sensor. An initial image of each visual sensor is distributively stored in each sensor node and also in a central processing to be used in generating virtual image.

3) JPEG 2000 image compression method is implemented on all visual sensors to compress the captured image. The purpose of image compression is to save energy in the image delivery as the transmission of data requires much higher energy than computing the data.

4) After the system recognizes the position of each sensor node in accordance to reference point, request of each user to any desired FoV can be performed.

5) Desired FoV will be compared with the position / direction of view of any visual sensors. Based on mutual information criteria and the disparity, the system decides which camera to be used.

6) Since environmental conditions might change at any time, due to weather, new additional object, changing the position of objects and others, then on every FoV request, a new image will be captured from selected visual sensors. The new image will be compared with the initial image to find their differences.

7) The system calculates mutual information of images. Furthermore, only two images with the largest mutual information are transmitted to a central processing to save energy for transmitting.

8) Image generation process is done at central processing. The Image generation steps are 1) Interpolation of correspondence points, 2) pre-wrapping process, 3) post-wrapping process.

III. RESULTS AND DISCUSSION

Initial stage is begun when all visual sensor nodes capture and store images. Then the nodes send the compressed images to central processing. Each node has calculated and stored mutual information of the images. When particular FoV is requested, the central processing will select two visual sensor nodes based on disparity method as seen in Fig. 4.

In the initial stage, the mutual information computation for the first image and the second image to themselves applies (2), that are \( H(I_{11}, I_{11}) = 7.0800 \) and \( H(I_{21}, I_{21}) = 5.3526 \) (See TABLE III). Condition 2 is the condition when images are captured with no differences to the initial condition, Fig. 4c and Fig. 4d, so the mutual information of images is the same. The mutual information are \( H(I_{11}, I_{12}) = 7.0800 \) and \( H(I_{21}, I_{22}) = 5.3536 \). On the other hand, the mutual information, as seen in TABLE III is obtained from calculation of image in Fig 4c to the image in Fig. 4a, and image in Fig 4d to the image in Fig 4b.
Whenever any differences occurred, as in condition 3, the mutual information between initialized image and captured image at that time, becomes smaller, that are
\[ H(I_{11}; I_{13}) = 1.3599 \] and \[ H(I_{21}; I_{23}) = 1.1110. \] These are the criteria that are used to determine whether it is necessary to send new image to the central processing. When mutual information is relatively the same, the visual sensor node only sends one bit data, 0, to inform the central processing that there is no new image available. When the object or scene has experienced any modification, visual sensor node sends another bit, i.e. bit 1, followed by new captured image.

Energy consumption as seen in TABLE IV is obtained from (26). For condition 2 and condition 3, the processing energy consumption (\( E_{\text{proc}} \)) and transmission energy consumption (\( E_{\text{transmit}} \)) are change as follows:

\[
E_{\text{WVSN}} = m \cdot t \cdot E_b + MN \sum_{i=1}^{L} \frac{1}{2^{rac{i-1}{4}}\text{t}} (10E_{\text{shift}} + 12E_{\text{add}} + 2E_{\text{read}} + 2E_{\text{write}})
\]

Assumed that active processor is a condition when signal processing takes place, then \( E_{\text{shift}}, E_{\text{add}}, E_{\text{read}}, \) and \( E_{\text{write}} \) can be substituted with the energy consumption when the processor is active, as shown in Table I. According to this table, the energy consumption of active processor is 192.3 mW. Then by encoding time duration, the \( E_{\text{DWT}} \), the processing energy consumption (\( E_{\text{proc}} \)) becomes:

\[
E_{\text{Proc}} = MN \sum_{i=1}^{L} \frac{1}{2^{rac{i-1}{4}}\text{t}} (10E_{\text{shift}} + 12E_{\text{add}} + 2E_{\text{read}} + 2E_{\text{write}}) = P_{\text{proc}} \cdot t
\]

where \( t \) is the time required for encoding processes. Based on JPEG 2000 parameters in TABLE II, the average encoding time for 0.1 bpp and DWT level 3 is 18.5 s and the energy consumption becomes:

\[
E_{\text{WVSN}} = 0.04J + 192.3mW \times 18.5s + 864 \times 3,291.5 = 0.04J + 3.558J + 131,9J = 135,52J
\]

At condition 2, when two selected sensor nodes send two images, then total energy consumption of sending two compressed images by JPEG 2000 as base of virtual image construction is 271.04 J. Meanwhile, at condition 3, when the mutual information value is the same as the value at initial stage, only information bit is sent. Therefore the energy is required for transmission process is:

\[
E_{\text{WVSN}} = E_{\text{capt}} + E_{\text{proc}} + E_{\text{transmit}} = 0.04J + 1.10J + 0.02J = 1.16J
\]

Hence total energy is required for transmitting information from two selected sensors of 2.32J.

On the contrary, transmitting image without compression makes file size of 900 KB. This large file is split into 32,915 packets with average of 28 byte payload per packet. By the use of Imote2 parameter on TABLE II and computed using (26), energy consumption becomes:

\[
E_{\text{WVSN}} = E_{\text{capt}} + m \cdot k \cdot E_b = 0.04J + 32,915 \cdot 864 \cdot 46.39\mu J = 1,319.35J
\]
Finally, the energy consumption to send one compressed image is reduced to only 10.27% compared to energy consumption without compression (see Table 4). When there is no modification of object or scene, i.e., condition 2, the energy consumption reduces to 3% when compared with the whole compressed image is transmitted.

The total energy consumption of sending two compressed images by JPEG 2000 as base of virtual image construction is 271.04 J, and 2639.70 J for uncompressed images. Virtual image of disparity of 0.5 is shown in Fig. 5a and the captured image from visual sensor for FoV as requested by user is shown in Fig. 5b.

### IV. Conclusions

Virtual view image from one particular direction that is demanded by user requires at least two images from visual sensors. This application becomes a challenge in WWSN environment due to its limited resources. One method to save energy consumption is by applying image compression, such as JPEG 2000 on each visual sensor. In this research, we conclude that by applying such compression method, energy consumption on each sensor node increases as much as 3.35 J. This cost is compromised by significant reduction of energy consumption on image transmission. Total energy required to send one compressed image reduces to only 10.27% of energy consumption of uncompressed image.

The mutual information computation as detector for scene modification also increases energy consumption on each node. However when mutual information shows no change on the scene, the total energy required is down to 1.16 J, or 0.9% of energy of uncompressed image.

### ACKNOWLEDGMENT

The authors are grateful to the anonymous referees for their valuable comments and suggestions to improve the presentation of this paper. This work is supported by Indonesian Ministry of Education and Culture through the BPIT scholarship and 2013 Doctoral Research Grant, contract number: 175.45/UN14.2/PNL.01.03.00/2013, May 16, 2013.

### REFERENCES


