A Novel Energy Efficient Object Detection and Image Transmission Approach for Wireless Multimedia Sensor Networks

Yasar Abbas Ur Rehman[†], *Member, IEEE*, Muhammad Tariq[*‡§], *Member, IEEE*, Takuro Sato[‡], *Fellow, IEEE*

[*] Department of Electrical Engineering, Princeton University, Princeton NJ, USA
[†] City University of Science & Information Technology, Peshawar, Pakistan
[‡] Sato Lab, Waseda University, Tokyo, Japan
[§] tariq@fuji.waseda.jp

***Abstract*—Energy efficient object detection and image transmission are one of the key issues in Wireless Multimedia Sensor Networks (WMSN). Recent approaches in WMSN propose in-node object detection and tracking algorithms. However, a little effort has been made to effectively detect object presence and absence in images in WMSN. Since object detection has a direct relation with image transmission, therefore effective object detection algorithm would provide a reduction in false transmission of image information. In this paper, a novel object presence model and image transmission scheme is proposed for WMSN. This scheme uses the transmission of an image segments rather than a complete image. It guarantees in-node energy conservation and minimal image content transmission to the sink node. The proposed scheme is evaluated based on in-node energy consumption and reconstructed image Peak Signal to Noise Ratio (PSNR). Simulation results show that the proposed approach saves 95% of the node energy with the received image PSNR of 46 db as compared to other state of the art approaches.**

***Index Terms*—Wireless Multimedia Sensor Networks, Object detection, Object tracking, Energy conservation**

I. Symbols and their definition

Symbol	Definition
$r_i(x,y)$	i_{th} Relaying node coordinates
$N_i(x,y)$	i_{th} Monitoring node coordinates
$d_{i,j}$	RF distance between i_{th} and j_{th} node
V_j	Battery state of charge of j_{th} node
B_i	Background for i_{th} frame
I_i	i_{th} current frame
$\Re$	Information set
$P(k_L)$	Probability of L_{th} intensity pixel having intensity K_L
$h(k_L)$	Histogram of L_{th} intensity pixel having intensity K_L
P_{co}	Pixel Coordinates
O_i	Object detection function
δ	Standard deviation
μ	Mean
$\bar{I}$	Reconstructed image
E	Energy consumption

II. Introduction

THE development in miniaturization technology enables us to integrate heterogeneous sensing devices on one sensing platform. One prolific advantage of this miniaturization is the availability of Complementary Metal Oxide Semiconductor (CMOS) cameras and its integration with traditional Wireless Sensor Networks (WSN), which transform the WSN to Wireless Multimedia Sensor Networks (WMSN). This transformation allows to implement multi-dimensional signal processing algorithms on these sensing platforms. Consequently, it offers countless services as compared to the traditional WSN. They can be used in surveillance, habitat monitoring, traffic control, intrusion detection, and health care monitoring. Moreover, they can be used in difficult, unattended and inadequate resourced areas. However, there are certain limitations that need to be addressed when developing algorithms for data processing in WMSN. These are limited memory, limited processing, narrow bandwidth and limited battery power [1]. Therefore, the cognition of amendments in traditional multi-dimensional signal processing techniques is a primitive step before practical implementation in WMSN. To be more specific, the inherent energy hungry visual processing techniques require careful adjustments to coherently adopt with the stringent constraints offered by WMSN. To provide this coherency, the work in this paper focuses on modifying traditional paradigm of data processing in WMSN. Specifically, the objective is to prolong the node lifetime by minimizing the in-node processing cost. Accordingly, visual object identification and tracking processes are carried out at the sink node where the realization of object geometry can enable the remote node to decide about further operation in its sensing area. To provide a formal justification of this modification from image processing perspective, we reveal that background subtraction provides erroneous results, preferably with environmental variations and when camera orientation is deviated from its reference position. This mainly occurs because of a strong wind blowing, waving trees, shadows, seismic movement, water flowing and illumination changes. Eventually, target tracking from image processing perspective demands continuous isolation of objects from the background. Therefore, background subtraction and template matching are typically the healthiest candidates. Most of the algorithms in WMSN grounded on object tracking necessitate the in-node processes to be in the active state for the period in which the target object remains in the vicinity of the sensing node. This implies a swift drain out in the available node energy and therefore in the life time of a sensor node [2].

The main contribution of this work is of two folds. Firstly, to build robust object appearance detector to detect minor and major variations in an image and transmit the image data only when it is required.

Secondly, to remove redundant features from an image to optimize the reduction in the energy consumption during data transmission. Additionally, to fit the data in the available bandwidth keeping accumulative processing cost as low as possible. For this purpose, an object appearance model is developed based on spatiotemporal variations in the image. This model provides energy conservation at the remote node during acquisition and transmission of visual data. The rest of the paper is organized as follows: Section III reviews state of the art work done in WMSN. Section IV defines the Network Model. SectionV describes the in-node processing model. Section VI provides simulation results and finally Section VII concludes the paper.

III. Related Work

Image processing in WSN has gained remarkable interest among research communities recently. The availability of low cost CMOS camera has made it possible to incorporate visual sensors with WSN. However, traditional data and image processing techniques for object detection and tracking need modifications, in order to enable them to cope with the aforementioned constraints [1]. This may include fusion of camera and scalar sensors to achieve in-node energy conservation [3]. However, effective operation necessitated the development of middleware for communication between vision module and sensor node [4]. Since these approaches mainly set the initial stage for WMSN operation, therefore, low level computer vision techniques for an object detection and tracking were the choices to express the energy constraints. In WSN, target tracking has been achieved using Unscented Kalman Filter (UKF) [5]. The idea was to maximize the utility information from a set of sensor nodes in a resource constrained environment. Target tracking using particle filter was proposed in [6], [7] and a multi target tracking algorithm using Bayesian filter was proposed in [8]. These algorithms were suitable for scalar WSN and cannot be implemented without modifications for WMSN, where low processing and simplified object tracking algorithms are the suitable choice. In [9], an energy aware dynamic object tracking was proposed using cooperative communication between scalar and camera nodes. However, the visual information processing cost was masked and it was assumed to give accurate results. The amount of data provided by WMSN nodes facilitated the division of data into multiple levels of information [10]. This division enabled a remote node to process information to the extent of an end user requirement [11]. However, since most of the WMSN schemes for object tracking proposed background subtraction scheme as their initial step assuming a static camera [12]. In such cases, there are chances of miss-detection and miss-classification. This mainly occurs because of environmental variations such as illumination changes [13]. Although, background maintenance scheme is one solution, however energy conservation is the main issue in continuous processing of images [14]. Much of the research work in this area was carried out to get an optimal point so that minimum processing and transmission energy are required [15]. To get a good tradeoff between an image quality and transmission energy consumption, a distributed processing using 2 Dimensional Discrete Wavelet Transform (2D-DWT) was studied. Images were decomposed into multiple resolutions using 2D-DWT. A self-adaptive transmission scheme was proposed to get a tradeoff between energy consumption and image quality [16], [17], [18]. Data compression using Discrete Cosine Transform (DCT) and DWT were performed so that less energy was consumed during image transmission [19], [20]. However, conventional compression techniques need to be modified to cope with the low memory and processing requirements of WMSN.

IV. Network Model

In this section, we enlighten upon the topology of the network to be used for image transmission purpose. It encourages to analyze the in-network energy consumption during the transmission of image frames. As the energy consumption during an image transmission is always on the higher side as compared to in-node processing therefore, the main aim of introducing topology is to find the overall energy consumption during the transmission of set of image frames. Consider a random deployment of WMSN nodes in the field. Each WMSN node is equipped with limited resources.We assume that there is no channel impairments in the network. This assumption will be relaxed in Section V. To reliably transmit an image through the network and to avoid any collision due to simultaneous transmission of image information by two or more nodes, each WMSN nodes in this paper are divided into two classes:

1) Relaying Node (RN)
2) Monitoring Node (MN)

The network topology is shown in Fig. 1.

Each RN node maintains a Relaying Node Parameter (RNP) set:

$$RNP = \{r_i(x,y) | r_j(x,y), N_j(x,y), d_{i,j}, V_j \subset r_i\}. \quad (1)$$

where $r_i(x,y)$ is the i^{th} RN node coordinates, $r_j(x,y)$ is the j^{th} RN node coordinate connected to the i^{th} RN node. N_j represents the number of MN nodes connected to i^{th} RN node, $d_{i,j}$ represents the RF distance between the

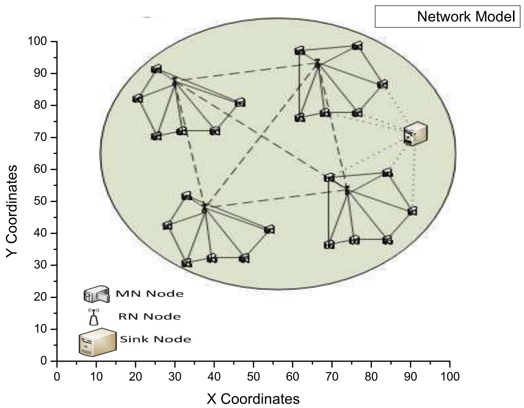

Fig. 1. Network Model

Fig. 2. Test Image

i^{th} and the j^{th} RN node and V_j represents the battery state of the j^{th} MN node. Each MN node also maintains a Monitoring Node Parameter (MNP) set:

$$MNP = \{N_j(x, y)|r_i(x, y), d_{i,j}, V_i \subset r_i\} i \neq j. \quad (2)$$

where $N_j(x, y)$ is the j^{th} MN coordinates, $r_i(x, y)$ is the i^{th} RN node coordinates connected to the j^{th} MN node. $d_{i,j}$ represents the RF distance between the i^{th} MN node and the j^{th} RN node and V_i represents the battery state of the i^{th} RN node. An MN node upon the detection of any activity transmit its data packet to the nearest RN node determined by the distance $d_{i,j}$. Each RN node upon reception of the data packet provides best possible path for the data packet to the sink node through intermediate RN nodes. Since coordinates alone cannot determine the proximity of the two sensor nodes therefore, the distance is incorporated in (1)and(2). Also, by the inclusion of distance parameters in (1)and(2), we are able to find the nearest the relying node to which the monitoring node can send its information with minimal energy consumption.

V. In-Node Processing Model

To preserve the node energy during image processing and transmission, the in-node process is modeled using a saptio-temporal approach. Before proceeding to further calculations, it is assumed that the camera is initially static and the environmental variations are monotonically modeled. An MN node first captures the image of the scene and stores it in the memory. It also transmits this image to the sink node so that the image variation at the corresponding node can be monitored. The stored image in the MN node is updated with the time by using (3):

$$B_i = \alpha \times B_{i-1} + (1 - \alpha) \times I_i, 0 \leq \alpha \leq 1. \quad (3)$$

where B_i is the background for i^{th} frame I_i, B_{i-1} is the initial background and α control the rate of update. The background model presented in (3) is affected by several factors. The major problem arises in setting the value of α. If the value of α is close to 1 the model will give more preference to the previous background. In this case, the foreground object if remain static, will take more time in becoming part of the background. Whereas, if the value of α is close to 0 the model will give more preference to update the background using current frame. In this case, the foreground object if remain static, will take less time in becoming part of the background. However, in the proposed approach since the images are transmitted to the sink node, therefore, if no image is received for a particular node the sink node assumes that either there is no object or the object in the area is static considering that the node is not depleted of energy.

A. Object Detection

The image captured by each MN node is first divided into four blocks of the same size. Fig. 2 shows an example image which is divided into four equal blocks. These blocks have been shown in Fig. 3. It then forms a set $\Re$ by using (4).

$$\Re = \{i_n | i_n \in I_j, n \leq 4\}. \quad (4)$$

To detect the appearance of the object in the scene, a probabilistic approach is proposed. Considering the distribution of pixels in the image by using (5).

$$P_{(I_j, i_n \in \Re)}(k_L) = \frac{m_L}{R_T}. \quad (5)$$

where $P_{(I_j, i_n \in \Re)}(k_L)$ is the probability that L^{th} intensity pixel having intensity k_L occurs in the n^{th} block of j^{th} frame, m_L is the number of occurrence of k_L intensity pixel, R_T is the total pixels in the n^{th} block of j^{th} frame, i.e., $R_T = R_N \times C_M$. If there is any object appear in the scene, the number of pixels with L^{th} intensity will change from m_L to $m_L^{/}$ and hence the probability from $P_i(k_L)$ to $P_{i+1}^{/}(k_L)$:

$$P_{(I_j, i_n \in \Re)}^{/}(k_L) = \frac{m_L^{/}}{R_T}, \quad (6)$$

$$\triangle P_{(I_j, i_n \in \Re)}(k_L) = \mid P_{(I_j, i_n \in \Re)}(k_L) - P_{(I_{j-1}, i_n \in \Re)}(k_L) \mid . \quad (7)$$

Fig. 3. Image Division into Four Sub-blocks

To simplify calculations, we make use of histogram[21] and relate it with the above equations:

$$h_{(I_j, i_n \in \Re)}(k_L) = n_L, \tag{8}$$

$$h_{(I_j, i_n \in \Re)}(k_L) = R_T \times P_{(I_j, i_n \in \Re)}(k_L). \tag{9}$$

In order to calculate the probability we project each histogram value to a Gaussian distribution by using (10), (11) and (12) as:

$$P_{h_{(I_j, i_n \in \Re)}(k_L)} = \frac{1}{\sqrt{2\pi\delta^2}} e^{\frac{-1}{2\delta^2}(h_{(I_j, i_n \in \Re)}(k_L) - \mu(h_{(I_j, i_n \in \Re)}(k_L)))}, \tag{10}$$

$$\mu(h_{(I_j, i_n \in \Re)}(k_L)) = \frac{1}{L} \sum_{L=0}^{L-1} h_{(I_j, i_n \in \Re)}(k_L), \tag{11}$$

$$\delta(h_{(I_j, i_n \in \Re)}(k_L)) = \sqrt{\frac{\sum_{L=0}^{255} (h_{(I_j, i_n \in \Re)}(k_L) - \mu(h_{(I_j, i_n \in \Re)}(k_L)))^2}{L-1}}, \tag{12}$$

$$\triangle P_{(I_j, i_n \in \Re)}(k_L) = \sum_{L=0}^{L-1} \mid P_{(I_j, i_n \in \Re)}(k_L) - P_{(I_{j-1}, i_n \in \Re)}(k_L) \mid . \tag{13}$$

The absolute difference gives a set of threshold values T for each image block, which are arranged sequentially by using (14) as:

$$T = \{\triangle P_{(I_j, i_n \in \Re)}(k_L) | \triangle P_{(I_j, i_n \in \Re)}(k_L) \in (I_j, i_n \in \Re)\}. \tag{14}$$

The resulting value of T is compared with the threshold t empirically calculated. If the value of T is greater than t it will indicate the presence of the object in the scene and vice versa. However, this approach is affected by several environmental variations, therefore, we logically fused the value of threshold values T with background subtracted result. The logical fusion means that the threshold value T and background subtracted results are combined using AND operation. The reason is to check if the result of both the threshold T and background subtracted result is greater than a predefined threshold t, that is empirically calculated, will trigger the transmission of image. The selection of possible value of t depends upon the intended application. For a static camera since the outdoor changes in the scene are not abrupt, therefore a fix value of threshold is sufficient, however for abruptly changing environment or when the scene changes continuously like in the case when the camera moves in an environment to track the object, then the value of t needs to be dynamic [22], [23], [24], [25]. However, in the present case the camera is static therefore we fix the value of threshold t to a certain value depending upon our experiments setting. The background subtracted result will give a conventional foreground. However, in this approach we count the total number of pixels participating in the foreground object. The total number of foreground pixels for each block is calculated using (15).

$$I_{foreground} = \sum_{P_{co}=0}^{N-1} \mid I_{(I_j, i_n \in \Re)}(P_{co}) - I_{(I_{j-1}, i_n \in \Re)}(P_{co}) \mid > I_{th}. \tag{15}$$

where P_{co} is the pixel coordinates. Since the presence and absence of an object is similar to tossing an unbiased coin therefore we define the following function:

$$O_i = \begin{cases} 1, & \text{if } (T, I_{foreground}) > t \\ 0, & \text{otherwise} \end{cases} . \tag{16}$$

Once the presence and absence is predicted by the appearance model, the particular area from the whole image is sent to the 2D-DWT process to further decompose it into multiple sub-bands. The 2D-DWT is applied using 1D-DWT in a row and column approach. Afterwards, images are sent to the sink node in a multi-hop fashion. The in-node process is shown in the Algorithm 1.

Algorithm 1: In-Node Process

Input : $RN = \{r_1 \ldots r_n\}$: Set of RN nodes connected to i^{th} MN node
D_{x_d, y_d}: Sink node location
F_i: Frame captured by N_i node

Output: $i_n = \{i_n| \in F_i, n \leq 4\}$: Image segment from i^{th} MN node

Frame Capture & Transmit `captureTransmit()`

```
capture F_i ;
compute ℜ ;
compute T and I_foreground;
if (T_i, I_{i_n,foreground}) >= t then
    extract(i_n ∈ F_i) ;
    {i_LL, i_LH, i_HL, i_HH} = 2D − DWT(i_n);
    packetize {i_LL, i_LH, i_HL, i_HH} ;
    transmit {i_LL, i_LH, i_HL, i_HH} ;
else
    discard F_i
```

B. Scene Reconstruction

Once image is received at the sink node, it is then superimposed on the reference frame received earlier. Since, in the proposed approach only a portion of the total image is transmitted therefore the pixel coordinates are unaltered at the MN node. This helps in efficiently replacing the pixels in the reference image with a portion of the transmitted image at the sink node. However, the pixel values are susceptible to channel distortion due to the addition of additive noise η.

$$i_r(x,y) = i_n(x,y) + \eta, \quad i_r \in I_o, \quad i_n \in I_j, \quad r \neq j. \tag{17}$$

The quality of reconstructed image is evaluated using Peak Signal to Noise Ratio (PSNR).

VI. Simulation Results and Discussion

Simulations are performed on the Intermittent Object Motion dataset and Baseline dataset provided by [26]. Among the various categories, we choose a scenario for Winter Drive Way (WDW) from Intermittent object motion data set and pedestrians from baseline dataset. The data sets contain 2500 frames of 240×320 resolutions. The algorithm is applied on 200 frames. Each frame is divided into four blocks 120×160 of the same size. For each block the appearance model predicts the presence and absence of an object or activity in the scene. If any activity is detected the image is transmitted in a multi-hop fashion. For image transmission purpose the RN node first form a network topology with its neighboring RN nodes. One example of this topology is shown in Fig. 4. After the formation of relaying node topology, the network topology is initiated. An example of such network topology is shown in Fig. 5. The possible data paths taken by the image blocks are shown in Fig. 6, 7 and 8.These data paths corresponds to the strategy used for image transmission purpose. In Fig. 6, the RN node 2 does not have access to the sink node therefore, it transmits its information to the RN node 1, which transmits the image to the sink node, which is labeled as 30. The advantage of using this type of coordinated transmission increases the reliability of overall network and reduces the energy consumption as selected nodes are used for transmission of image information to the sink node. In Fig. 7, a similar strategy was followed, however, the number of nodes in this case was slightly increased. In Fig. 8, the number of nodes used to transmit the information has further reduced. An important aspect of this strategy is that not all the nodes are utilized for image transmission purpose, which reduces the overall in-network energy consumption. To compare the proposed method with other state of the art approaches, the proposed algorithm is simulated using Imot2 integrated with OV7670 Camera. The Imot2 and the camera parameters are shown in the Table 1.

Fig. 9 shows the received image PSNR for the two datasets. Fig.10 shows the received image PSNR for the two datasets in case of channel distortion. Although, in the proposed approach the PSNR ratio is not stable,

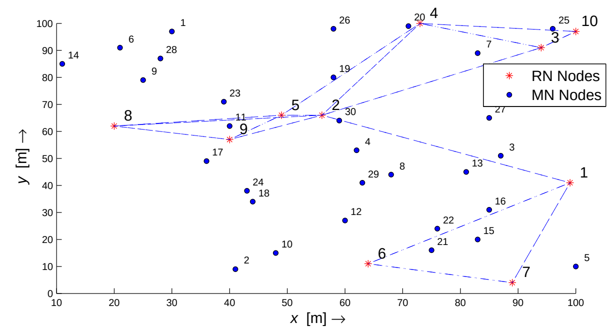

Fig. 4. Relaying Node Topology

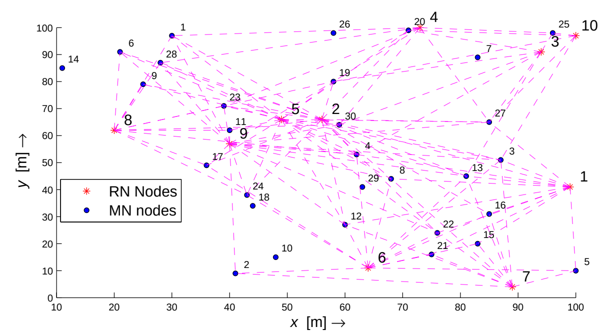

Fig. 5. Network Topology

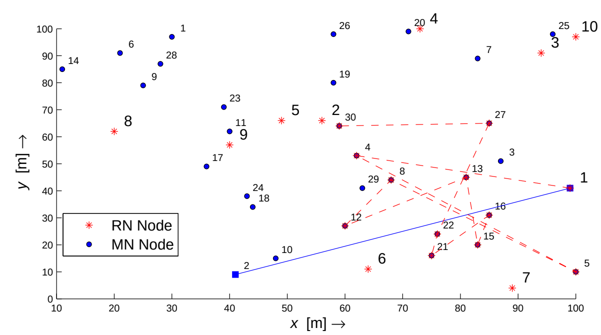

Fig. 6. Data Path 1

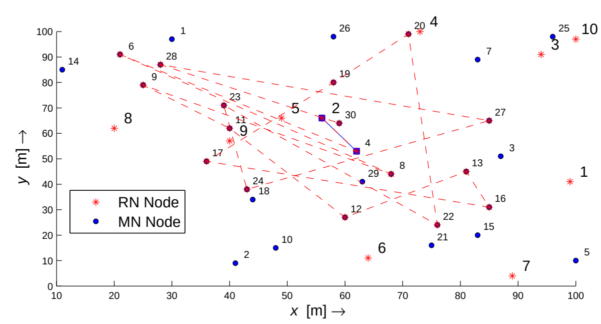

Fig. 7. Data Path 2

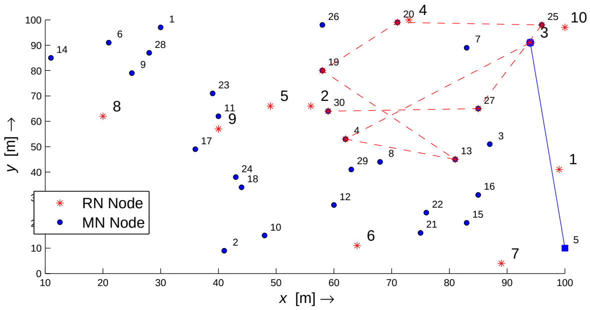

Fig. 8. Data Path 3

TABLE I
IMOTE2 INTEGRATED WITH OV7670

Symbol	Description	Value
$I_{t,on}$	Current drawn in active mode T_x/R_x in (mA)	44 mA
$I_{c,on}$	Current drawn when camera is active	18 mA
T_{tx}	Time to transmit a single bit	0.4 usec
V_b	Supply voltage	4.5
Clk	Clock period	24 nsec

however, the maximum value is well above 40 db. The energy consumption is calculated by using (18).

$$E_{consumed} = \{(V_b \times ClK \times I_{c,on}) + (V_b \times I_{t,on} \times T_{tx})\} \times i_n, \ i_n \in I_j, n \leq 4. \tag{18}$$

The energy consumption analysis has been done by comparing the proposed approach with the direct approach. In the direct approach a whole image is transferred to the sink node in a multi-hop fashion. Fig.11 shows the In-node energy consumption using both approaches applied on the aforementioned datasets. As can be seen, there is a significant amount of reduction in the node energy consumption. Table II shows a comparison of performance with other state of the art approaches. The work presented in [18],

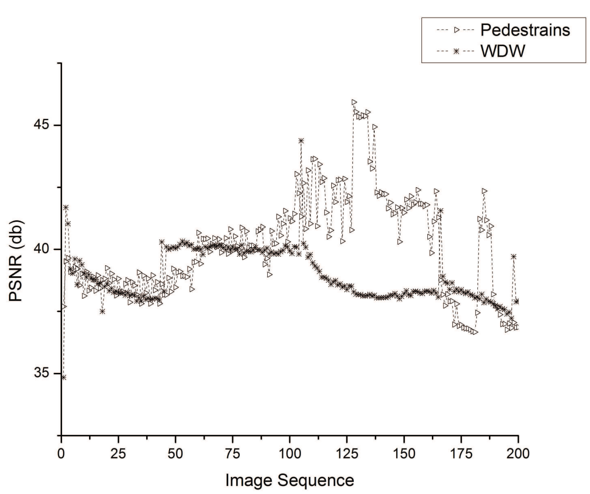

Fig. 9. Received PSNR

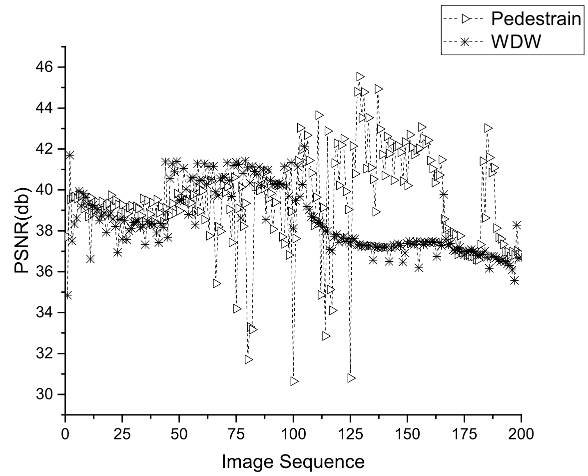

Fig. 10. Received PSNR in case of Channel Distortion

TABLE II
COMPARISON OF TECHNIQUES

Technique	$PSNR_{db}$	Energy Consumption (mJ)	% Energy Saving
Lucire et. al[18]	31	1252.6	87.47
Nasri et. al [16]	20	1252.6	87.47
Proposed Approach	46	411.3	95.88

[16] used direct approaches in transmitting images through the sink node. In their work, the image is transmitted as a whole to sink node traversing the path through all available network nodes, which is considered as a direct approach. The network energy consumption in this case is on the higher side. The reason for this increase in energy consumption is because of the two main assumption: First, all nodes in the sensor network has been utilized for the transmission of image information, that makes the in-network energy consumption on the higher side. Secondly, the image information has been transmitted as a whole utilizing 2D DWT approach, however, no intelligence has been incorporated in the system which makes the nodes energy on the higher side. Fig. 12 shows the in-network energy consumption for the transmission of 200 images. As can be seen in Fig. 12, in a coordinated transmission the energy consumption of the whole network can be greatly reduced. Further, the work presented in the proposed approach, provides energy conservation both in-node and in-network along with the additional features of object detection in the remote area. Fig.13 and Fig. 14 show the simulation results from the two datasets. It can be concluded from the Fig.13,14, that since a portion of the image information is being received at the sink node with an acceptable range of PSNR values, therefore, the proposed approach can be used in real time object tracking and localization systems.

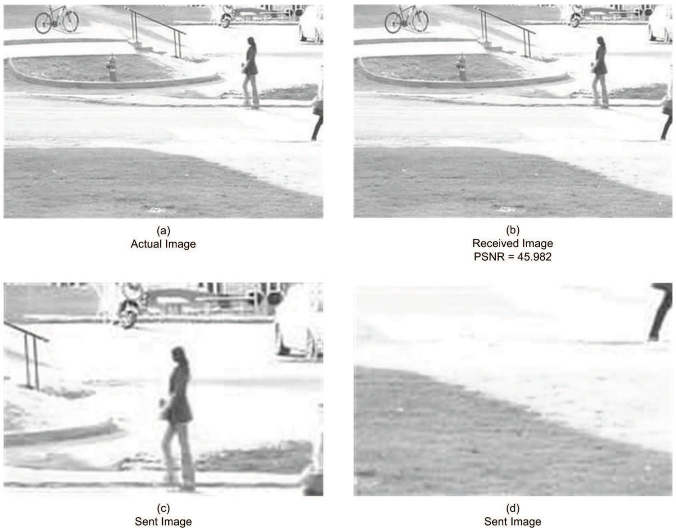

Fig. 14. (a) Actual image from Pedestrians dataset (b) Reconstructed image (c) The upper right portion of the image in (a) is sent to the sink node (d)The bottom right portion of the image in (a) is sent to the sink node

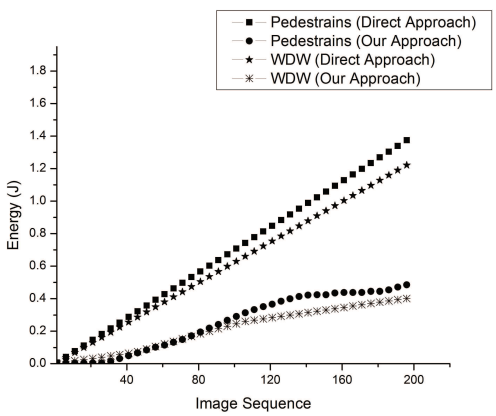

Fig. 11. In-Node Energy Consumption

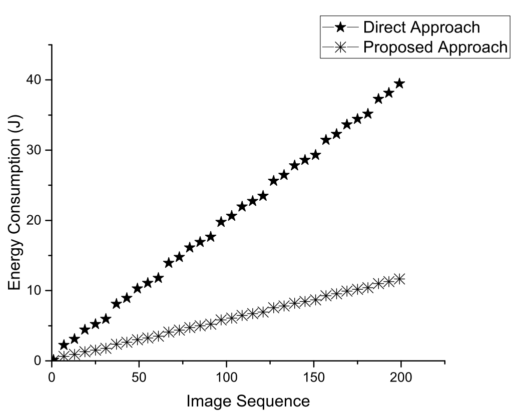

Fig. 12. In-Network Energy Consumption

VII. Conclusion

In this paper, we have presented a scheme for efficient object detection and image transmission in WMSN nodes. Our scheme provides energy conservation during transmission of image frames. These images can be transmitted through the available bandwidth because of the size reduction. Besides, conserving the in-node energy it has been shown that the object presence model effectively locate the object when it appeared in the scene. Additionally, transmission of particular image fragments provides a better image reconstruction at the sink node. The scheme

Fig. 13. (a) Actual image from WDW dataset (b) The upper left portion of the image in (a) is sent to the sink node (c) The reconstructed image after superimposing the received image on the stored image for a particular sensor node

has been evaluated using a custom made hardware WMSN node and other state of the art online image dataset. The received image PSNR is taken into account during the reconstruction process at the sink node. Simulation results report 95%of energy saving and a PSNR of 46 db as compared to other popular image transmission techniques.

References

[1] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Computer networks*, vol. 51, no. 4, pp. 921–960, 2007.

[2] W. Fang, D. H. Wang, and Y. Wang, "Energy-efficient distributed target tracking in wireless video sensor networks," *International Journal of Wireless Information Networks*, pp. 1–11, 2015.

[3] M. Magno, F. Tombari, D. Brunelli, L. Di Stefano, and L. Benini, "Multimodal video analysis on self-powered resource-limited wireless smart camera," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems,*, vol. 3, no. 2, pp. 223–235, 2013.

[4] M. Casares, M. C. Vuran, and S. Velipasalar, "Design of a wireless vision sensor for object tracking in wireless vision sensor networks," *in Proc. Second ACM/IEEE International Conference on Distributed Smart Cameras*, Sep 2008.

[5] P. Pahalawatta, T. Pappas, and A. Katsaggelos, "Optimal sensor selection for video-based target tracking in a wireless sensor network," *in Proc. International Conference on Image Processing, 2004. ICIP '04.*, 2004.

[6] P. M. Djuric, M. Vemula, and M. F. Bugallo, "Target tracking by particle filtering in binary sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, pp. 2229–2238, Jun 2008.

[7] O. Ozdemir, R. Niu, and P. Varshney, "Tracking in wireless sensor networks using particle filtering: Physical layer considerations," *IEEE Transactions on Signal Processing*, vol. 57, pp. 1987–1999, May 2009.

[8] S. Oh, "A scalable multi-target tracking algorithm for wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, pp. 1–16, 2012.

[9] I. Boulanouar, A. Rachedi, S. Lohier, and G. Roussel, "Energy-aware object tracking algorithm using heterogeneous wireless sensor networks," *in Proc. IFIP Wireless Days (WD)*, Oct 2011.

[10] Z.-J. Zhang, C.-F. Lai, and H.-C. Chao, "A green data transmission mechanism for wireless multimedia sensor networks using information fusion," *IEEE Wireless Communications*, vol. 21, pp. 14–19, Aug 2014.

[11] M. S. Alhilal, A. Soudani, and A. Al-Dhelaan, "Low power scheme for image based object identification in wireless multimedia sensor networks," *in Proc. International Conference on Multimedia Computing and Systems (ICMCS)*, pp. 927–932, Apr 2014.

[12] D. M. Pham and S. M. Aziz, "Object extraction scheme and protocol for energy efficient image communication over wireless sensor networks," *Computer Networks*, vol. 57, pp. 2949–2960, Oct 2013.

[13] S. Yoshinaga, A. Shimada, H. Nagahara, and R.-i. Taniguchi, "Object detection based on spatiotemporal background models," *Computer Vision and Image Understanding*, vol. 122, pp. 84–91, May 2014.

[14] M. Chen, Q. Yang, Q. Li, G. Wang, and M.-H. Yang, "Spatiotemporal background subtraction using minimum spanning tree and optical flow," *Lecture Notes in Computer Science*, pp. 521–534, 2014.

[15] J. Martínez-de Dios, A. Jiménez-González, and A. Ollero, *Localization and Tracking Using Camera-Based Wireless Sensor Networks*. INTECH Open Access Publisher, 2011.

[16] M. Nasri, A. Helali, H. Sghaier, and H. Maaref, "Adaptive image transfer for wireless sensor networks (wsn)," in *Proc. 5th international conference on Design and technology of integrated systems in nanoscale era (DTIS)*, pp. 1–7, IEEE, 2010.

[17] V. Lecuire, C. Duran-Faundez, and N. Krommenacker, "Energy-efficient transmission of wavelet-based images in wireless sensor networks," *EURASIP Journal on Image and Video Processing*, vol. 2007, 2007.

[18] V. Lecuire, C. Duran-Faundez, and N. Krommenacker, "Energy-efficient image transmission in sensor networks," *International Journal of Sensor Networks*, vol. 4, no. 1, pp. 37–47, 2008.

[19] H. Wu and A. A. Abouzeid, "Energy efficient distributed image compression in resource-constrained multihop wireless networks," *Computer Communications*, vol. 28, no. 14, pp. 1658–1668, 2005.

[20] Y. M. Chi, R. Etienne-Cummings, G. Cauwenberghs, P. Carpenter, and K. Colling, "Video sensor node for low-power adhoc wireless networks," in *Proc. 41st Annual Conference on Information Sciences and Systems, 2007. CISS'07.*, pp. 244–247, IEEE, 2007.

[21] R. C. Gonzalez, *Digital image processing.* Pearson Education India, 2009.

[22] Y. Sugaya and K. Kanatani, "Extracting moving objects from a moving camera video sequence," in *Proc. 10th Symposium on Sensing via Imaging Information*, pp. 279–284, 2004.

[23] X. Fang, W. Xiong, B. Hu, and L. Wang, "A moving object detection algorithm based on color information," *Journal of Physics: Conference Series*, vol. 48, no. 1, p. 384, 2006.

[24] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999.*, vol. 2, IEEE, 1999.

[25] L. Unzueta, M. Nieto, A. Cortés, J. Barandiaran, O. Otaegui, and P. Sánchez, "Adaptive multicue background subtraction for robust vehicle counting and classification," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 13, no. 2, pp. 527–540, 2012.

[26] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "Cdnet 2014: an expanded change detection benchmark dataset," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 387–394, 2014.

Yasar Abbas Ur Rehman received his M.Sc Degree in Electrical Engineering from National University of Computer and Emerging Sciences (NUCES), Pakistan in 2015, and his B.Sc Degree in Electrical Engineering from City University of Science and Information Technology, Peshawar, Pakistan in 2012. Currently, he is working as Lab Engineer at the Department of Electrical Engineering, City university of Science and Information Technology, Peshawar, Pakistan. His research interests include object and sensor node localization in Wireless Sensor and Wireless Visual Sensor Networks, Visual Communication, Computer Vision, Signal Processing and Smart Grids.

Muhammad Tariq received his BS degree in Computer Systems Engineering from University of Engineering Technology Peshawar, Pakistan in 2006, his MS degree in Electronics and Computer Engineering from Hanyang University, Seoul, South Korea in 2009, and his Ph.D. degree in Wireless Communication Systems from Waseda University, Tokyo, Japan in 2012. Currently, he is Post Doctoral Fellow in the Department of Electrical Engineering, Princeton University, NJ, USA and a visiting researcher at SatoLab, Waseda University, Japan. His research interests include mathematical modeling, designing, and analysis of wireless ad-hoc and sensor networks, wired and wireless communication, control, and automation in smart grid. He is the editorial board member of various peer reviewed journals and TPC member of distinguished conferences. He is a member of IEEE, IEICE, JSST, and PEC.

Takuro Sato received the B.E. and Ph.D. degrees in Electronics Engineering from Niigata University, Niigata, Japan, in 1973 and 1994, respectively. He is currently a Professor of Graduated School of Global Information Telecommunication and Studies in Waseda University since 2004. Previously, he worked at Research and Development Laboratories, Oki Electric Industry Co., Ltd., Tokyo, Japan from 1983 to 1996 and as Professor in the Department of Information and Electronics Engineering, Niigata Institute of Technology in 1995. His current research is next generation mobile communication system, smart grid/energy and social information infrastructure network. He is a fellow of IEEE.