

Construction of a non-overlapping panoramic mosaic in wireless multimedia sensor networks

B.W. Kim^{1,2} J.-H. Jung¹ B.G. Kim² D.-J. Park¹

¹Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), 335 Gwahangno, Yuseong-gu, Daejeon 305-701, Republic of Korea

²Department of Computer Science and Engineering, College of Engineering, Sunmoon University, Kalsan-ri, Tangeong-myeon, Asan-si, Chungnam 336-708, Republic of Korea
 E-mail: byungwook.kim@kaist.ac.kr

Abstract: As a result of the rapid development of hardware whose operation has low power requirements, a great deal of interest has been shown in wireless multimedia sensor networks for various applications. In sensor networks in which the sensor nodes are distributed randomly and have low-power cameras, the fields of view of some adjacent sensor nodes may overlap. The authors introduce a scheme for constructing a non-overlapping panoramic mosaic by transmitting partial images. In particular, a solution of the joint cost function for network lifetime and video quality is used to select the boundary lines between adjacent images. The experimental results show that the proposed scheme increases the network lifetime while providing better video quality.

1 Introduction

Traditional wireless sensor networks (WSNs) [1] consist of thousands of interconnected sensors that retrieve only scalar data, such as temperature, pressure or humidity. However, this situation has recently changed for the better. The considerable advances have been made in developing low-cost miniature hardware solutions, such as smart cameras and motes. These sensors which can ubiquitously capture multimedia contents, such as video and audio streams and still images, have led to the development of wireless multimedia sensor networks (WMSNs) [2–4]. It is widely thought that the WMSNs will not only improve existing sensor network applications, but will also enable the development of new technology and applications.

In WMSNs, sensor nodes are equipped with miniature battery-powered cameras and wireless low-power transceivers that are capable of transmitting, receiving and processing video streams [5, 6]. These transceivers are placed randomly over a large area and, via the retrieval of video streams, complement the functions of existing surveillance systems. Given that a large number of sensors are distributed around an event area, the fields of view of several cameras may overlap. The redundant information in the overlapping image region contains the spatial relations between the adjacent images and hence enables the construction of a panoramic mosaic. It increases the field of view of a camera by allowing several views of the area of interest to be combined into a single image. Given the information about the spatial relations between the adjacent images, we can construct a panoramic mosaic that retains only non-redundant information by connecting partial images, rather than connecting whole images. Partial images can be obtained by eliminating the

overlap between images. This will allow the construction of a seamless non-overlapping panoramic mosaic.

To construct a mosaic image, each sensor node should send video streams to a centralised sink, via a different wireless path, so that the video data can be aggregated. The quality of the reconstructed video will be degraded by packet loss caused by the transmission error or the late arrival of packets. The probability that packets will be lost is determined strongly by the distance between the sensor node and the sink [7]. In order to improve the video quality at sinks, the sensor nodes that are close to the sinks should provide more data than others. This can be achieved by allocating a bit size for the partial image for each sensor node.

Sensor nodes usually operate in an unsupervised area; hence, the battery cannot be recharged or replaced. This limits the lifetime of the network and affects the overall operation of a network. As a result, network lifetime has become a key performance metric for WSNs. In order to prolong network lifetime, sensors should use energy as little as possible for data transmission, because it is the major cause of energy dissipation.

In this paper, we present a method for optimising video transmission over WMSNs, in which we construct a non-overlapping panoramic mosaic that takes video quality into account and increases the lifetime of the network. The remainder of this paper is organised as follows. Section 2 presents the network model. Section 3 describes the proposed scheme. In Section 4, the simulation results are provided. Section 5 draws a conclusion.

2 Network model

We consider a WMSN in which, for the monitoring and detection of objects, sensor nodes are distributed randomly

over a large area. As shown in Fig. 1, several adjacent sensors may have overlapping sensing areas, due to the field of view of the cameras in the sensors. Since the overlapping image region contains information about the alignment of the image, we can construct a panoramic mosaic of an event area that has a larger field of view. Conventional object detection and tracking methods are only applicable in a single field of view. However, if a mosaic image is available for detecting and tracking objects, we can obtain more information about the object in the larger and continuous scene.

Sensor devices are interconnected by two-tier sensor network architecture, that is, sensor nodes and sinks. The available sensors have different capabilities and power requirements. Sensor nodes consist of cameras and motes with resource-constrained and low-power operation. Cameras that consume little power are capable of taking low-resolution images and motes are responsible for sending images. Sinks are equipped with more capable and higher-power operating motes to collect images from sensor nodes and to transmit images to a centralised server for more complex processing. Owing to the energy limitation of WMSNs, it is difficult to implement the sophisticated video coding techniques used in the moving picture experts group (MPEG) or H.26x series. However, WMSNs can support image coding and compression standards, such as joint photographic experts group (JPEG), thanks to its simple encoding structure.

The scenarios of use for a WMSN might include frequent and scheduled image capture and transmission, and asynchronous requests for imagery. In both scenarios, sensor nodes initially send their whole captured images to a sink to obtain information about the alignment of images. The sink should also provide data that contain information on how to eliminate the overlapping image region, to each sensor node at the initialisation phase. Then, sensor nodes can produce partial images and transmit them to a sink at the communication phase to construct a panoramic mosaic without redundant information. When sensor nodes are not required to operate, they can stay in a sleep mode, using minimum power to extend their lifetime. Here, network lifetime is defined as the time duration taken until one of the sensor nodes in the network runs out of battery power.

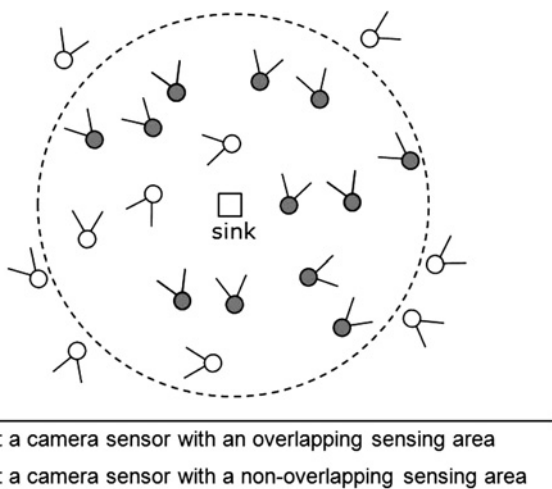


Fig. 1 Distributed sensors around a sink in WMSN

3 Proposed scheme

3.1 Generation of boundary line candidates

A sink initially receives whole, often overlapped, images from the nearby sensor nodes. In order to obtain non-redundant information about the sensed area, we need to remove unnecessary data from the panoramic mosaics. To connect sets of sequential images for constructing a mosaic image, we use Lowe's scale invariant feature transform (SIFT) algorithm [8], which can detect and describe local features in an image. Given the correspondence features, random sample consensus (RANSAC) [9], an algorithm to build robust estimates for parameters of a mathematical model from a set of matched features, estimates the best image transformation parameters, that is, a projective matrix T_j . Here, we select sets of four best feature correspondences and compute an eight-parameter projective matrix T_j .

We present image I_j as the image taken from the sensor node j . The sink can determine the overlapped image region by transforming I_{j-1} to I_j with a projective matrix T_j , as depicted in Fig. 2. For simplicity, we assume that one of the sequential images has left- or right-side overlapped images captured from neighbouring sensor nodes. By properly removing the overlapping image region, we can obtain partial images that can be used to construct a panoramic mosaic without redundant information.

As shown in Fig. 2, the boundary lines in the overlapping image region determine the shape of the partial images for each sensor node. In order to express boundary lines, we use the horizontal pixel indices of the top and the bottom of an image. The generation of boundary line candidates is explained in Fig. 3, where $a[i_j]$ and $b[i_j]$ are the horizontal pixel indices where the boundary line meets the top and bottom of the image I_j and can express the boundary line. tr_j and br_j are the top-right and bottom-right horizontal location of the warped I_{j-1} on the I_j domain. Here, N is used to express the larger horizontal pixel indices of the overlapped image region and incremented by the ratio between tr_j and br_j . Among the boundary line candidates, we should select the best one that minimises the given cost function which is derived in Section 3.4. Note that the boundary line candidates are generated only once in the system initialisation since it requires high computational power.

3.2 Network lifetime criterion

If we are to take into account the network lifetime without any information about the residual energy of each sensor node, the amount of transmission energy consumed by each sensor node

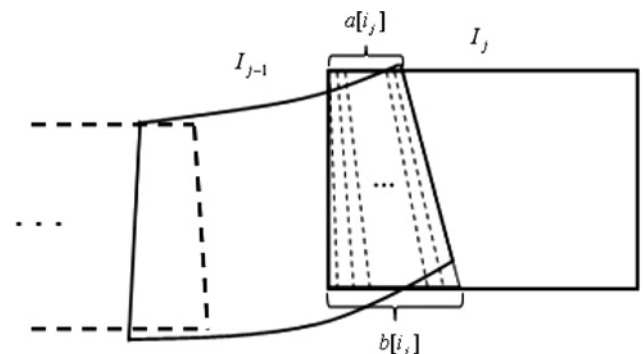


Fig. 2 Overlapping image region with boundary line candidates

Algorithm 1

```

Receive  $I_1, I_2, \dots, I_m$  from  $m$  sensor nodes
for  $j = 2$  to  $m$  do
  Find a projective matrix  $T_j$  using  $I_{j-1}$  and  $I_j$ 
  Find  $br_j$  and  $tr_j$  on  $I_j$ 
   $x_j \leftarrow \min(tr_j, br_j)$ ,  $N = 1.0$ 
  for  $i_j = 1$  to  $x_j$  do
    if  $tr_j \geq br_j$  then
       $a[i_j] \leftarrow \text{floor}(N + 0.5)$ 
       $b[i_j] \leftarrow i_j$ 
       $N \leftarrow N + tr_j/br_j$ 
    else
       $a[i_j] \leftarrow i_j$ 
       $b[i_j] \leftarrow \text{floor}(N + 0.5)$ 
       $N \leftarrow N + br_j/tr_j$ 
    end if
  end for
end for
Obtain  $x_j$  number of boundary line candidates

```

Fig. 3 Generation of boundary line candidates

must be set to be as similar as possible. Since the bit size of the images captured from all the sensor nodes is different, choosing the best boundary line between images results in the same transmission energy being consumed by all sensor nodes that are dedicated to sending data. Note that the selection of the boundary line can determine the bit size of the corresponding partial images, due to the block-based Huffman decoding in the JPEG algorithm. Once we have obtained the bit sizes of the partial images according to the boundary line, we can identify the one that is best for minimising the cost function. Here, we employ a simple model for the j th sensor node's transmission energy consumption C_j shown as

$$C_j = R_j E_j = R_j (e_a + e_r d_j^2) \quad (1)$$

where R_j is the bit size of the image, E_j is the j th sensor node's transmission energy consumption per bit, e_a is a distance-independent constant term, e_r is a coefficient term associated with the distance-dependent term and d_j is the distance between the j th sensor node and the sink. In this paper, the bit size denotes the total packet size in bits corresponding to the image. Here, we obtain the optimal bit size for each sensor node by calculating the bit size of the non-overlapping mosaic image divided by the number of sequential images. Let C^{no} and C^{o} denote the total transmission energy consumed for the non-overlapping and overlapping regions in the panoramic mosaic, respectively. Considering n sequential images, C^{no} and C^{o} can be expressed as

$$C^{\text{no}} = \sum_{j=1}^n R_j^{\text{no}} E_j C^{\text{o}} = \left(R_1^{\text{o}} E_1 + \sum_{j=2}^{n-1} (\tilde{R}_j^{\text{o}} + R_j^{\text{o}}) E_j + \tilde{R}_n^{\text{o}} E_n \right) / 2 \quad (2)$$

where R_j^{no} is the bit size of the non-overlapping region, R_j^{o} and \tilde{R}_j^{o} are the bit sizes of the overlapping image region on the right and left side in I_j , respectively. Note that R_j^{o} and $\tilde{R}_{j-1}^{\text{o}}$, which contain an identical overlapping sensing area, are not always the same due to the different locations of sensor nodes capturing images. Then, the optimal transmission energy consumption for each node can be simply expressed as $C^{\text{avg}} = (C^{\text{no}} + C^{\text{o}})n$. Using C^{avg} , we can determine the bit sizes of n sensor nodes to minimise the following problem as

$$\begin{aligned} \arg \min_{0 \leq R_j^{\text{o}} \leq R_j^{\text{t}}} G(R_j^{\text{o}}) &= \arg \min_{0 \leq R_j^{\text{o}} \leq R_j^{\text{t}}} (R_j E_j - C^{\text{avg}})^2 \\ &= \arg \min_{0 \leq R_j^{\text{o}} \leq R_j^{\text{t}}} (R_j^{\text{o}} E_j + R_j^{\text{no}} E_j - C^{\text{avg}})^2 \end{aligned} \quad (3)$$

where $1 \leq j \leq n$, R_j^{t} is the maximum bit size of the overlapping image region.

3.3 Video quality criterion

Considering deterioration of the received video quality at sinks that is caused by transmission error, we should select the boundary line that minimises the channel distortion effect of the perceived image. Here, we use the channel distortion model in [10] to formulate the problem. In addition, we use the packet loss model in [7] to present the probability of packet loss, which can be shown as a function of the distance between a sink and a sensor node. Assuming a linear relation between the number of pixels and the size of images in bits, the problem of selecting the boundary line, given channel distortion, can be

expressed as

$$\arg \min_{R_1, R_2, \dots, R_n} \sum_{j=1}^n \left(\frac{\alpha P(d_j)}{1 - P(d_j)} N_p(R_j) F_d \right) \quad (4)$$

where $P(d_j)$ is the packet loss probability depending on the distance, $N_p(R_j)$ is the number of pixels depending on the size of the image in bits, α is the energy loss ratio of the encoder filter and F_d is the expectation of the mean square error (MSE) between frames. It should be noted that in (4), boundary lines can be selected by considering a number n of adjacent images together. However, this results in calculations of high complexity. Since R_j^{no} is a constant, we can obtain a simplified equation as follows

$$\arg \min_{0 \leq R_j^o \leq R_j^t} U(R_j^o) = \arg \min_{0 \leq R_j^o \leq R_j^t} \left(\frac{\alpha P(d_j)}{1 - P(d_j)} N_p(R_j^o) F_d + \frac{\alpha P(d_{j+1})}{1 - P(d_{j+1})} N_p(R_{j+1}^t - \eta R_j^o) F_d \right) \quad (5)$$

where η is R_{j+1}/R_j and R_{j+1}^o can be approximated as $R_{j+1}^t - \eta R_j^o$. Note that only two adjacent images are considered to obtain the solution of (5).

3.4 Joint optimisation problem

Our goal is to increase network lifetime and to enhance video quality by choosing the best boundary line between adjacent images in order to construct a non-overlapping panoramic mosaic. We can write this optimisation problem by combining the criteria of the previously presented problems

as follows

$$\hat{R}_j^o = \arg \min_{0 \leq R_j^o \leq R_j^t} G(R_j^o) + \lambda U(R_j^o) \quad (6)$$

where \hat{R}_j^o is the optimal size in bits of the overlapping image region in I_j . The constant λ denotes the relative importance weight that we attach to network lifetime and video quality. Since (6) is a function of R_j^o , we can calculate the derivative by R_j^o , expressed as

$$\begin{aligned} \frac{\partial G(R_j^o)}{\partial R_j^o} + \lambda \frac{\partial U(R_j^o)}{\partial R_j^o} &= 2(R_j^o E_j + R_j^{no} E_j - C^{avg}(R_j^o))(E_j - C^{avg}(R_j^o)) \\ &\quad + \lambda \beta (D_j - \eta D_{j+1}) \\ &= 2(R_j^o E_j + R_j^{no} E_j - C^{avg}(R_j^o)) \left(E_j - \frac{(n-1)\bar{E}_j}{2n} \right) \\ &\quad + \lambda \beta (D_j - \eta D_{j+1}) = 0 \end{aligned} \quad (7)$$

where

$$D_j = \frac{\alpha P(d_j)}{1 - P(d_j)} F_d, \quad \bar{E}_j = (1/n) \sum_{j=1}^n E_j \quad \text{and} \quad N_p(R_j^o)$$

is set to βR_j^o , where β is a constant. $C^{avg}(R_j^o)$ denotes that C^{avg} is a function of R_j^o . Then, we can simply calculate \hat{R}_j^o by

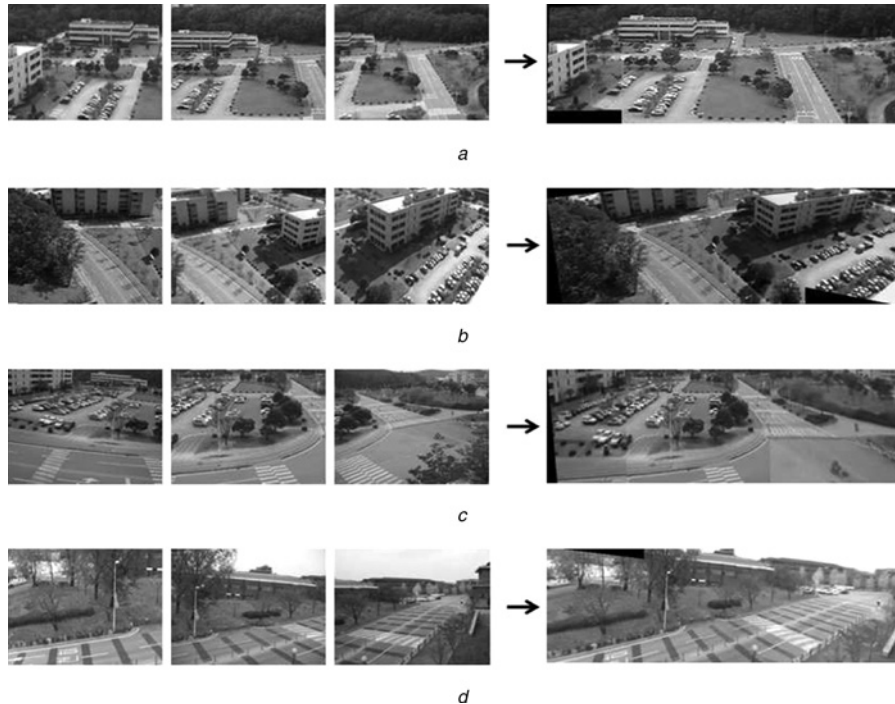


Fig. 4 Whole images producing overlapping panoramic mosaics

- a Image set 1
- b Image set 2
- c Image set 3
- d Image set 4

rewriting (7) as

$$\hat{R}_j^o = \frac{C^{\text{avg}}(R_j^o)}{E_j} - R_j^{\text{no}} - \frac{\lambda\beta(D_j - \eta D_{j+1})}{2E_j^2 - ((n-1)E_j\bar{E}_j/n)} \quad (8)$$

The above equation tells us the optimised size in bits of a partial image, considering the network lifetime and the video quality. Then, the selection of the boundary line among the generated candidates, expressed by $a[i_j]$, $b[i_j]$, can be done by finding the closest partial image's bit size to \hat{R}_j^o . According to the boundary line, the number of macroblocks in the JPEG coded partial image is changed, that is, we can obtain the different bit size of the partial image and select the closest one to \hat{R}_j^o . After selecting the boundary line on image I_j , we can obtain the horizontal pixel indices \tilde{a}_{j-1} , \tilde{b}_{j-1} for presenting the boundary line on image I_{j-1} , simply by using the inverse projective matrix T_j^{-1} . The sink sends the indices to the corresponding sensor nodes in order for them to construct partial images. Note that boundary line information is necessary only when the sensor network is being initialised. Even though the captured image contains moving objects, the size of the image in bits depends mainly on the texture of the scene and does not change very much. Given the information about boundary line for each sensor node, partial images can be transmitted by all sensor nodes to construct a mosaic image until one of the sensor nodes runs out of energy.

4 Experimental results

In order to evaluate the performance of the proposed scheme, we ran computer simulations of the network lifetime and the video quality. The number of sensors was set to 12 and captured images from every three sensors were used to construct overlapping panoramic mosaics, shown as Fig. 4.

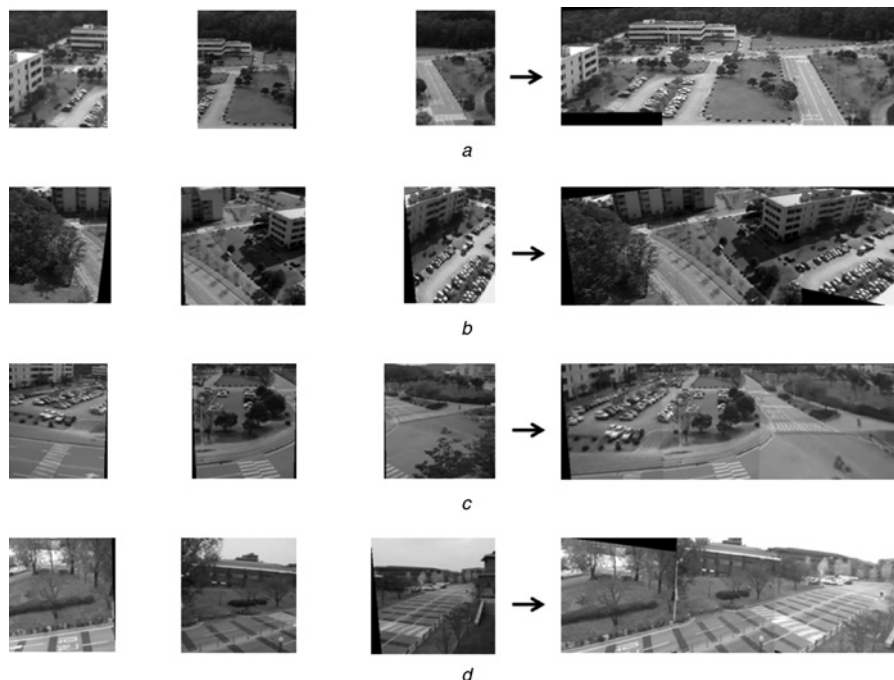


Fig. 5 Partial images producing non-overlapping panoramic mosaics

- a Image set 1
- b Image set 2
- c Image set 3
- d Image set 4

Table 1 Distance between the sink and the sensor nodes

Image set no.	Sensor node 1, m	Sensor node 2, m	Sensor node 3, m
Image set 1	14	23	19
Image set 2	20	22	18
Image set 3	20	18	15
Image set 4	18	13	21

As may be seen from the figure, three sensor nodes capture overlap views at four different locations. The distance between the sink and the sensor node is based on the geometrical locations and presented in Table 1. Since sensor nodes are energy-constrained, we do not encode the original partial image using the JPEG algorithm. Instead, the differential JPEG image between the current image and the previous image is utilised.

We used 320×240 (1/4 VGA) 8-bit resolution original images. The values of the parameters of the energy consumption model are the typical values $e_a = 50$ nJ/bit and $e_c = 100$ pJ/bit/m², as used in [11]. This system uses the IEEE 802.15.4 network standard, which provides the highest achievable data rate of 250 Kbps. The frame rate is 10 fps and the refreshing frame is used every five frames to prevent compression error and accumulation of the error in the difference of temporal images. Since we use the captured images from the real cameras which provide lower peak signal-to-noise ratio (PSNR) values than the standard test images, high refreshing rate is required. The packet size is set at 133 bytes and the payload is 127 bytes [12]. If a packet containing blocks in an image is not decoded, the sink simply copies the blocks at the same location from the previous decoded frame. The mosaic image obtained by connecting partial images is shown in Fig. 5. As may be

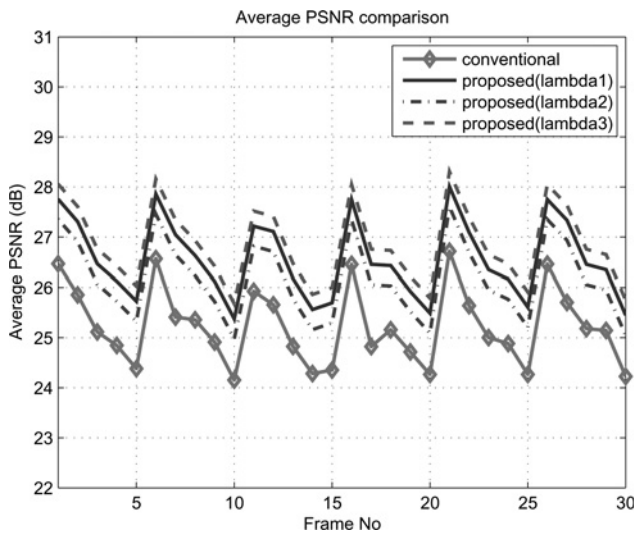


Fig. 6 Average PSNR performance with four image sets

seen, the difference between the non-overlapping panoramic mosaics and the overlapping images is barely visible to the human eye. The energy that is saved by effectively removing the overlapping image region for transmission can be used for other purposes.

We examined the effect of the proposed algorithm on the video quality. Here, we regard the conventional scheme as the method which transmits whole images and results in overlapping mosaic images. Fig. 6 illustrates the average performance of the proposed and conventional schemes with respect to PSNR, using four image sets. The PSNR formula is defined as follows

$$PSNR = 10 \log_{10} \left(\frac{255^2}{MSE} \right) \quad (9)$$

where MSE is the sum of the square of the difference between the constructed mosaic image and the reference mosaic image divided by the number of pixels. Even if the non-overlapping panoramic mosaics and the overlapping images provide the identical information to the human eye, they are apparently different when two images are compared pixel by pixel. Since the overlapping image region from the warped partial images and the whole images is not exactly the same, the reference mosaic image is determined by averaging the overlapping and non-overlapping mosaic images prior to JPEG encoding for the fair PSNR comparison according to the packet loss probability. It can be seen that the average performance of the proposed scheme with respect to PSNR is better than that of the conventional scheme. By allocating more information, that is, by transmitting a larger partial image via a wireless path with a lower packet loss probability, we can obtain a panoramic mosaic providing better image quality. We also ran a simulation with a different constant λ . Here, λ_1 is set to give the same importance to each of the network lifetime and the video quality. The weight parameters λ_2 and λ_3 give more importance to the network lifetime and the video quality, respectively.

To evaluate the network lifetime of the proposed scheme and the conventional scheme, we define approximated residual energy $\tilde{E}_j^O = \tilde{E}_j^i - aC_j$, where \tilde{E}_j^i is the initial energy of the j th sensor node and a is the number of frames

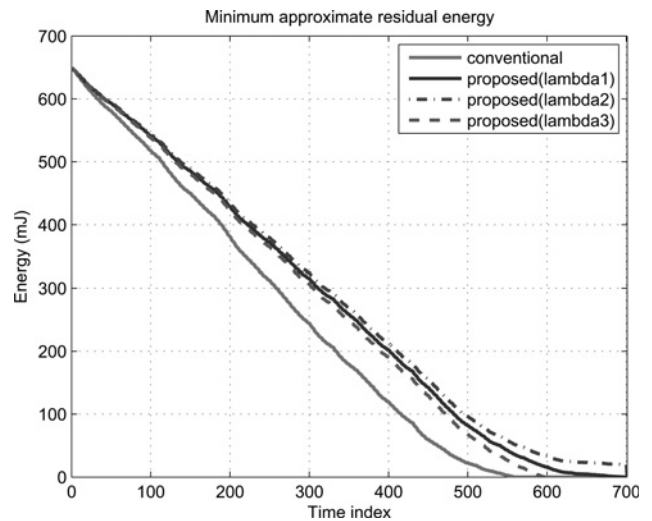


Fig. 7 Network lifetime performance

to transmit. As we mentioned above, the energy consumed for transmission is a critical factor for battery-operated sensor nodes. To compare \tilde{E}_j^O between the two schemes, two or three adjacent sensor nodes are randomly chosen from four image sets for image transmission. We assume that all the sensor nodes have the same residual energy and that the chosen sensor nodes are responsible for transmitting 10 frames at each request. This simulation is repeated 1000 times to calculate the average of the network lifetime performance. The minimum of approximated residual energy of the network for each scheme is shown in Fig. 7. We assume that a certain request for image transmission is made in every time index. We can see that the proposed scheme significantly improves the network lifetime, even for varying λ values. This increase is due to the method used to adjust the size of the partial image based on the remaining node energy. It is observed that the proposed scheme with λ_1 shows a $(679 - 554) / 554 \times 100 = 22\%$ increase in network lifetime compared to the scheme that transmits whole images.

5 Conclusion

We have proposed a scheme for constructing a non-overlapping panoramic mosaic for wireless multimedia sensor networks. The selection of the boundary line which results in partial images for a seamless mosaic image that retains only non-redundant information is based on the joint optimisation problem considering two criteria: increasing network lifetime while improving video quality simultaneously. Simulation results show that sensor networks that use the proposed scheme have a longer lifetime and better video quality than those using the conventional scheme.

6 References

- 1 Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: 'A survey on sensor networks', *IEEE Commun. Mag.*, 2002, **40**, (8), pp. 102–114
- 2 Akyildiz, I.F., Melodia, T., Chowdury, K.R.: 'A survey on wireless multimedia sensor networks', *Comput. Netw. J.*, 2007, **51**, pp. 921–960
- 3 Huang, H.C., Huang, Y.M., Ding, J.W.: 'An implementation of battery-aware wireless sensor network using ZigBee for multimedia service'. Proc. IEEE ICCE'06, 2006, pp. 369–370
- 4 Shu, L., Zhou, Z., Hauswirth, M., Phuoc, D.L., Yu, P., Zhang, L.: 'Transmitting streaming data in wireless multimedia sensor networks with holes'. Proc. MUM'07, 2007, pp. 24–33

- 5 Rahimi, M., Baer, R., Iroezi, O., *et al.*: 'Cyclops: in situ image sensing and interpretation in wireless sensor networks'. Proc. ACM Conf. Embedded Networked Sensor Systems, November 2005, pp. 192–204
- 6 Feng, W., Code, B., Kaiser, E., Shea, M., Feng, W.: 'Panoptes: scalable low-power video sensor networking technologies'. *ACM TOMCCAP*, 2005, **1**, (2), pp. 151–167
- 7 Zhao, J., Govindan, R., Estrin, D.: 'Computing aggregates for monitoring wireless sensor networks', *ACM TOMCCAP*, 2005, **1**, (2), pp. 151–167
- 8 Lowe, D.: 'Distinctive image features from scale-invariant keypoints', *Int. J. Comput. Vis.*, 2004, **60**, (2), pp. 91–110
- 9 Hartley, R., Zisserman, A.: 'Multiple view geometry in computer vision' (Cambridge University Press, 2004, 2nd edn.)
- 10 He, Z., Cai, J., Chen, C.W.: 'Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding', *IEEE Trans. Circuits Syst. Video Technol.*, 2002, **12**, (6), pp. 511–523
- 11 Shebli, F., Dayoub, I., Rouvaen, J.M., Zaouche, A.: 'A new optimization approach for energy consumption within wireless sensor networks'. Proc. IEEE AICT'07, May 2007, p. 14
- 12 Burchfield, T.R., Venkatesan, S., Weiner, D.: 'Maximizing throughput in ZigBee wireless networks through analysis, simulations and implementations'. Proc. LOCALGOS'07, June 2007, pp. 15–29

Copyright of IET Image Processing is the property of Institution of Engineering & Technology and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.