LIFETIME-DISTORTION TRADE-OFF IN IMAGE SENSOR NETWORKS

Chao Yu, Stanislava Soro, Gaurav Sharma, and Wendi Heinzelman

ECE Dept, University of Rochester, Rochester NY 14627

ABSTRACT

We examine the trade-off between lifetime and distortion in image sensor networks deployed for gathering visual information over a monitored region. Users navigate over the monitored region by specifying a viewpoint that varies with time, and the network attempts to meet the user requirement by synthesizing the desired view using a selection of cameras. We compare two camera selection methods, the first maximizes PSNR for the user's view without considering the cameras' available energy whereas the second uses knowledge of available energy at the cameras to maximize network lifetime. Our simulation results demonstrate a clear trade-off between the two strategies, with selection based on energy alone performing up to 3 dB worse initially than the PSNR based selection yet providing significantly higher coverage lifetime. In addition, we observe that under a fixed total energy constraint, more cameras with lower energy per node are preferable over fewer cameras with higher energy per node. Our results suggest that a hybrid or adaptive camera selection algorithm may provide the optimal lifetime-distortion trade-off.

Index Terms— image sensor network, network lifetime, coverage estimation, energy allocation

1. INTRODUCTION

Image sensor networks have recently emerged as an area of research interest with multiple applications [1, 2, 3]. We examine application scenarios where these networks are deployed in order to allow a user to navigate around the monitored region by specifying a desired viewpoint (position and direction) that varies over time. The user's viewpoint determines the part of the scene that should be captured and displayed. Desired view at this viewpoint is synthesized by combining parts of images from selected cameras. Fig. 1 illustrates such a telepresence system in the context of an art gallery.

While the problem of node (camera) selection in these networks has received some attention based on consideration of coverage and network lifetime [4, 5], the trade-off that these approaches entail in terms of image quality has not been hitherto addressed. In this paper, we examine the trade-off between image quality and network lifetime under different camera selection methods. The problem of energy allocation in such image sensor networks is also explored. Given the total amount of available energy, we compare different energy allocation schemes in terms of network lifetime.

This work is supported in part by the National Science Foundation under grant number ECS-0428157.

2. SYSTEM SCENARIO

We assume that the image sensor network consists of N battery powered cameras and that the monitored region can be decomposed into planes (such as within an art gallery). We focus on a single plane in order to simplify the analysis.

In the initialization step, the geometry of the cameras can be estimated with minimum manual setup by flexible camera calibration techniques [6, 7]. For the individual cameras, coverage on the target is then estimated by projecting world points onto their respective image planes. To provide the image data at user's viewpoint, a subset of cameras is selected and image blocks from the cameras are suitably warped and mosaiced, using the calibration information, to synthesize the desired view. These steps are individually elaborated in the following subsections.

2.1. Initialization

Camera Calibration: Using a homogeneous representation, the image coordinate $\mathbf{m}$ of a $3D$ point $\mathbf{M}$ is given by [8]:

$$\mathbf{m}_{3\times 1} = \mathbf{K}_{3\times 3}[\mathbf{R}|\mathbf{t}]M_{4\times 1} = \mathbf{P}_{3\times 4}\mathbf{M}$$

where $\mathbf{K}$ consists of 5 intrinsic parameters of the camera (such as focal length), $\mathbf{R}$ and $\mathbf{t}$ stand for the rotation and translation of the camera, respectively, and $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$ is called the camera projection matrix. Camera calibration is used to estimate the parameters $\mathbf{K}$, $\mathbf{R}$ and $\mathbf{t}$. The Plane-Based algorithm [6] can be used for camera calibration. Although this algorithm is designed to calibrate a single camera, it can be extended to calibrate multiple cameras with varying intrinsic parameters [7].

Coverage Estimation: The coverage of each camera's view on a target plane can be obtained from the camera geometry. A world point is covered by a camera if the point lies within the FoV (field of view) of the camera. Fig. 2(a) shows an example of coverage estimation

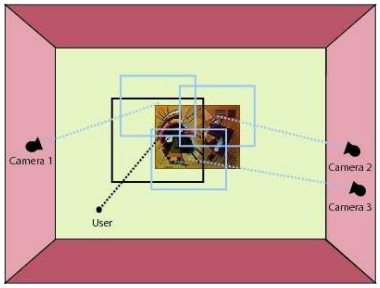
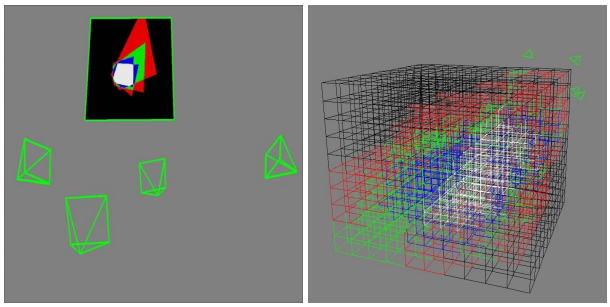


Fig. 1. Gallery monitored by a visual network: the desired view specified by user is synthesized from selected cameras.



(a) (b)

Fig. 2. (a) $2D$ coverage, and (b) $3D$ coverage where the $3D$ space is divided into cubes. Black: 0-coverage, Red: 1-coverage, Green: 2-coverage, Blue: 3-coverage, White: 4-coverage. The reader is referred to the electronic version of this paper for the best visual effect.

for a planar object. This method is readily extended to $3D$ space, as shown in Fig. 2(b).

2.2. Camera Selection

PSNR Cost (MaxPSNR): The first method of camera selection aims to maximize the resulting mosaiced image's quality, which is measured by PSNR. The desired view is divided into blocks for ease of analysis. For each block, the corresponding regions in the other cameras are transformed and compared with the ground-truth (available in the simulation). For each block, the camera that provides the highest PSNR is selected for transmitting the corresponding region. Although this method is feasible only in simulation, it provides an upper-bound on the achievable PSNR of the reconstructed images and helps quantify the trade-off between image quality and lifetime of the network.

View Angle Cost (MinANG): The MaxPSNR approach requires use of the ground-truth image of the desired view, yet this only exists in simulation. We approximate MaxPSNR by selecting the cameras that have the most similar viewing directions with user's viewpoint at each block. This approximation is justified since our simulated art gallery scenario sets all cameras at the same distance from the target plane so that the resolution of the cameras on the target plane are similar, and thus image quality is affected primarily by viewing direction.

Coverage Cost (MaxCOV): The coverage cost metric considers the camera's importance to the monitoring task. Assuming each camera s_i has remaining energy $E_r(s_i), i \in 1 \ldots N$, the set of cameras that cover the point (x, y) is described as $\{s_j \mid (x, y) \in FoV(s_j), j \in 1..N\}$. We define the total energy available for the monitoring task at (x, y) as [4]:

$$E_{total}(x, y) = \sum_{\{s_j | (x,y) \in FoV(s_j)\}} E_r(s_j) \qquad (1)$$

The coverage cost of a camera s_i is then defined as the sum of inverses of the energies at all points covered by s_i:

$$COST(s_i) = \sum_{\{(x,y) | (x,y) \in FoV(s_i)\}} \frac{1}{E_{total}(x, y)} \qquad (2)$$

Given this definition of coverage cost, those cameras that have large overlapping FoVs with other cameras have small cost and therefore are selected more frequently than those cameras that solely cover a part of the monitored scene. Thus selection of cameras with minimal cost results in a longer network lifetime.

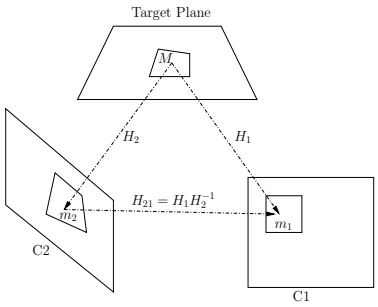


Fig. 3. Homography between multiple views.

2.3. Image Mosaicing

Once the image parts are selected from different cameras, they must be transformed to the desired view and mosaiced together. Our image mosaicing application differs from conventional applications (for example [9]) because we select regions from different images to generate the desired view instead of stitching together images.

Without loss of generality, we assume the monitored plane is at $\mathbf{Z} = 0$, and thus the image coordinate $\mathbf{m}$ of a $3D$ point $\mathbf{M}$ is: $\mathbf{m} \sim \mathbf{H}_{3\times3}\mathbf{M}$, where $\sim$ indicates equality up to a scale factor and $\mathbf{H}$ is the homography [8] between the camera plane and the target plane $\mathbf{Z} = 0$. $\mathbf{H}$ can be calculated from the result of camera calibration, as discussed in Sec. 2.1. Multiple projections of the same point are connected as shown in Fig. 3:

$$\mathbf{M} = \mathbf{H}_1^{-1}\mathbf{m}_1 = \mathbf{H}_2^{-1}\mathbf{m}_2,$$

$$\mathbf{m}_1 = \mathbf{H}_1\mathbf{H}_2^{-1}\mathbf{m}_2 = \mathbf{H}_{21}\mathbf{m}_2$$

This relation allows the image plane at user's viewpoint to be rendered from corresponding regions in the cameras selected according to the metrics described in Sec. 2.2.

3. SIMULATION RESULTS

We simulate a monitored plane of size $4m \times 3m$ (typical size of a wall). To simulate ad hoc deployment, N cameras are placed randomly within a $4m \times 3m$ field located $3m$ from the target plane, and the cameras are pointed toward the target plane with a random rotation within ± 0.1 radian along each of the X, Y, Z axes to simulate practical variability in camera placement. Pixel values at non-integer locations are generated by bilinear interpolation. All cameras (including the user's viewpoints) are assumed to have images of 200×200 (in pixel units), with a focal length $f_0 = 218.75$ (in pixel units). For an image sensor with size $20mm \times 20mm$, this would correspond to focal length of $f_0 = 21.875mm$.

We first conduct a Monte Carlo simulation in order to determine the number of cameras required in order to provide adequate coverage of the target plane. Fig. 4 shows the average (over 200 simulations) coverage percentage achieved over the target plane upon initial deployment of the cameras as a function of the number of cameras for different focal lengths f. We see that using a focal length $f = f_0 = 218.75$ (in pixel units), a minimum of 18 cameras are necessary in order to ensure that the target plane is covered

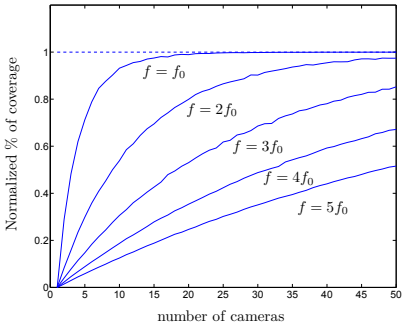


Fig. 4. Average percentage of coverage on the target plane vs. the number of cameras. An increase in focal length results in a decrease in the percentage of coverage.

(i.e., each point in the target plane is within the FoV of at least one camera) with a confidence of 99.5%. As expected, an increase in focal length results in a decrease in the percentage of coverage for a given number of cameras. For the remainder of the simulations, we use a focal length $f = f_0$.

The user's viewpoints are represented in our simulations by a random walk on a 16×16 grid in the plane of the cameras starting at the center $(x = 2m, y = 1.5m, z = 3m)$. Subsequent viewpoints are chosen from the neighboring 8 grid points and the current position (9 choices in total) with equal probability, and the user's views are assumed to be directed toward the target plane with random rotation within ± 0.1 radian along each of the X, Y, Z axes. We generate 100 user's viewpoints in each simulation. All images in the simulation are rendered by ray tracing. Fig. 9 shows a snapshot of the images generated in the simulation.

3.1. Coverage

In this simulation, all cameras begin with $3J$ of energy, which corresponds to each camera being able to transmit 3 full frames of images. In order to increase coverage redundancy and prolong network lifetime, we use twice the minimum number of cameras necessary for full coverage, i.e., $N = 36$. The results presented represent averages over 10 simulations.

Fig. 5 shows the percentage coverage on the target plane over time for the different camera selection methods. From the figure we see that MaxCOV camera selection prolongs the 95% coverage lifetime of the network by a factor of 1.5 compared to MinANG and 1.3 compared to MaxPSNR. MaxCOV distributes energy consumption more evenly across the cameras, which keeps the percentage of coverage higher. However, MaxCOV may generate a sharp drop in coverage when cameras die out since multiple cameras may run out of energy at the same time.

3.2. PSNR

The PSNR of the mosaiced image using the three camera selection methods is shown in Fig. 6 as a function of time. Initially, MaxCOV results in the lowest SNR, with MinANG performing about $1dB$ better and MaxPSNR performing the best (roughly $3dB$ better than MaxCOV). However, as time progresses the MaxPSNR and MinANG strategies lose coverage earlier than MaxCOV, and MaxCOV in fact provides the better PSNR as compared to the other methods. This demonstrates the advantage of using the MaxCOV metric in an energy-limited scenario.

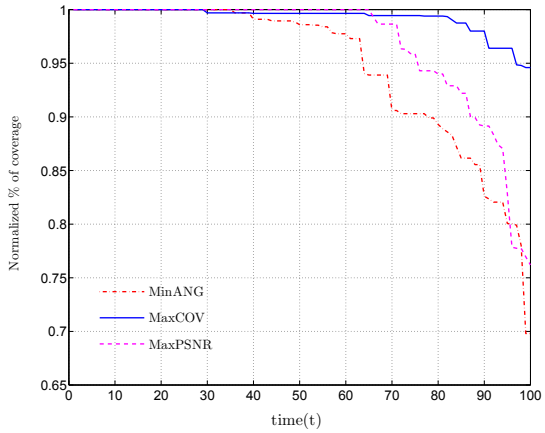


Fig. 5. Percentage of coverage on the target plane over time. The values on the X-axis are the numbers of observations, which is equivalent to time.

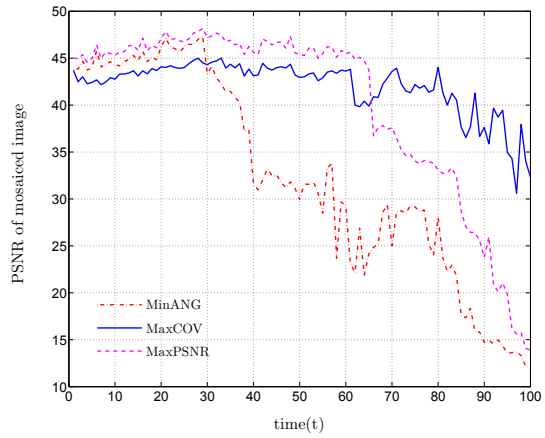


Fig. 6. PSNR of the mosaiced image at the user's viewpoint.

Fig. 7 shows PSNR values as a function of time for the scenario where cameras have infinite energy. Since all three methods retain full coverage here, MaxPSNR gives better image quality, by about $3dB$ compared to MaxCOV and $1dB$ compared to MinANG. Figs. 7 and 6, clearly demonstrate the trade-off between the quality of the mosaiced image and the lifetime of the network. Fig. 9 is a snapshot of the mosaiced output using each of the camera selection methods.

3.3. Energy Distribution

We also explore the impact of allocating a fixed amount of energy ($108J$) evenly among 18, 36, 54 and 108 cameras in our simulations and the resulting impact on network lifetime, which is defined as the time duration for which coverage is maintained for at least 95% of the target plane. Fig. 8 shows the network lifetimes obtained for these different allocations.

We observe that by increasing the number of cameras, the energy is more evenly distributed in the monitored space and thus the network lifetime is prolonged assuming the user's viewpoints are also evenly distributed. In all cases, MaxCOV performs better in terms of network lifetime then MaxPSNR and MinANG. The variances in MaxCOV are smaller than those in MinANG and MaxPSNR, which

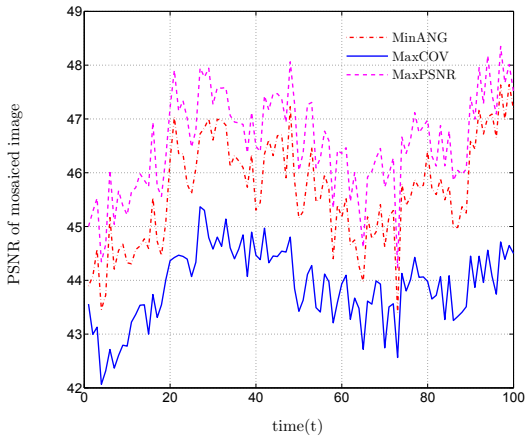


Fig. 7. Image quality when cameras have infinite energy.

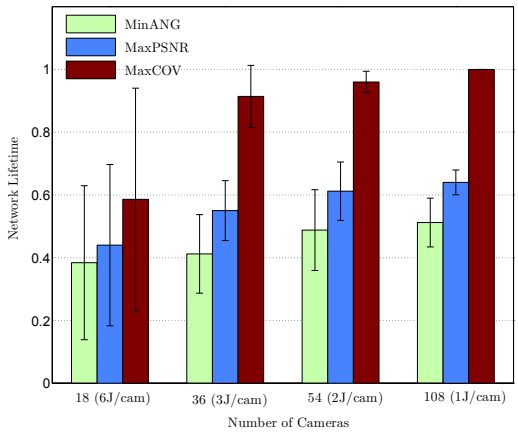


Fig. 8. Different energy allocation with a total energy of $108J$. Simulation results are averaged on different camera geometries, while the user's viewpoints remain the same.

means the performance of the visual network is less sensitive to the user's viewpoints using MaxCOV. The variances of network lifetime of all three methods decrease when more cameras are deployed, and thus increasing the number of cameras not only prolongs network lifetime, it also reduces the uncertainty in lifetime obtained under different use instances.

4. DISCUSSION AND CONCLUSION

In this paper we highlighted the trade-off between the network lifetime of an image sensor network and the distortion in the reconstructed images using simulations. In order to keep the complexity manageable, our simulations in the present work use a relatively simple application scenario and simple camera and scene geometry models that allow us to analytically perform mosaicing. The simulation results quantify the performances of various camera selection approaches under ideal scenarios and lead to several useful qualitative observations:

i) Camera selection for optimal image quality can offer significant improvements when energy considerations do not predominate (up to $3dB$ PSNR improvement in our simulations). However, in scenarios where energy is limited, the performance of these methods

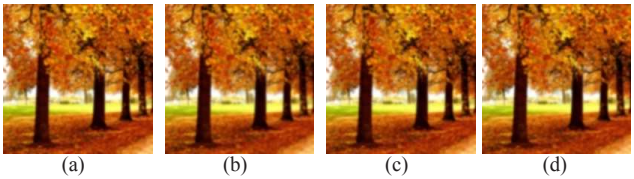


Fig. 9. A snapshot of images rendered in the simulation. (a) Real image at a given viewpoint, (b) Mosaiced image using MinANG, PSNR = 43.956, (c) Using MaxCOV, PSNR = 43.432, and (d) Using MaxPSNR, PSNR = 45.053.

deteriorates rapidly after the initial period, once coverage loss occurs. In these cases, using an energy-aware camera selection scheme significantly prolongs network lifetime at the cost of some sacrifice in the initial image quality.

ii) For a constant total energy budget, the use of a larger number of camera nodes (with lower initial energy per node) is preferable over a smaller set of camera nodes (with higher initial energy per node). This indicates that there is good reason to deploy larger numbers of small battery-based image sensors as compared to a smaller subset of cameras with more battery power.

iii) The relatively large disparity between the MaxPSNR and MaxCOV methods, along with the relatively poor lifetime performance of the MinANG heuristic, emphasize the need for research on camera selection algorithms that offer better choices in the lifetime-distortion trade-off. For example, a hybrid camera selection method that combines the MinANG and MaxCOV metrics or an adaptive camera selection algorithm that switches from MinANG to MaxCOV when camera energy diminishes, may better balance the lifetime-distortion trade-off than the techniques discussed here.

In actual practice, we recognize that applications will often require 3D coverage of complex scenes, and for actual sensors with imaging distortions, greater sophistication and computation will be required in image processing (e.g., for mosaicing). Our future work aims at extending our current analysis along these directions.

5. REFERENCES

[1] A. Keshavarz, A. Maleki-Tabar, H. Aghajan, "Distributed vision-based reasoning for smart home care," *ACM SenSys Workshop on Distributed Smart Cameras (DSC'06)*,

[2] P. Kulkarni, D. Ganesan, P. Shenoy Q. Lu, "SensEye: A multi-tier camera sensor network," *Proc. of ACM Multimedia*, 2005, p.229.

[3] M. Wu, C. Chen, "Collaborative Image Coding and Transmission over Wireless Sensor Networks," *EURASIP Journal on Advances in Signal Processing, special issue on visual sensor networks*, vol. 2007

[4] S. Soro, W. Heinzelman, "Camera selection in visual sensor networks," *Accepted by IEEE Advanced Video and Signal based Surveillance'07*.

[5] J. Dagher, M. Marcellin, and M. Neifeld, "A method for coordinating the distributed transmission of imagery," *IEEE Trans. Image Proc.*, vol. 15, no. 7, 2006, p.1705.

[6] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *ICCV'99*, p.666.

[7] C. Yu and G. Sharma, "Plane-based calibration of cameras with zoom variation," in *Proc. of SPIE Visual Communication and Image Processing'06*. San Jose, CA, p.352

[8] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge University Press, 2000.

[9] H.-Y. Shum and R. Szeliski, "Construction and refinement of panoramic mosaics with global and local alignment," in *ICCV'98*, Washington, DC, USA, 1998, p.953.