Camera Selection in Visual Sensor Networks *

Stanislava Soro
ECE Department
University of Rochester
Rochester, NY, 14623

Wendi Heinzelman
ECE Department
University of Rochester
Rochester, NY, 14623

Abstract

Wireless networks of visual sensors have recently emerged as a new type of sensor-based intelligent system, with performance and complexity challenges that go beyond that of existing wireless sensor networks. The goal of the visual sensor network we examine is to provide a user with visual information from any arbitrary viewpoint within the monitored field. This can be accomplished by synthesizing image data from a selection of cameras whose fields of view overlap with the desired field of view. In this work, we compare two methods for the selection of the camera-nodes. The first method selects cameras that minimize the difference between the images provided by the selected cameras and the image that would be captured by a real camera from the desired viewpoint. The second method considers the energy limitations of the battery powered camera-nodes, as well as their importance in the 3D coverage preservation task. Simulations using both metrics for camera-node selection show a clear trade-off between the quality of the reconstructed image and the network's ability to provide full coverage of the monitored 3D space for a longer period of time.

1. Introduction

Rapid advances in CMOS technology have enabled the development of cheap (on the order of $10), low-power camera modules, as evidenced, for example, by the ubiquitous cellular phone cameras. We believe that in the near future, these cameras will be combined with low power radios to create visual sensor networks that will provide more suitable solutions, compared with existing networks of high-power and high-resolution cameras, for many image-based applications that assume no infrastructure on site or no time for planning of the cameras placement.

In visual sensor networks, the camera-nodes can be simply stuck on walls or objects prior to use without the need for preplanning of the cameras placement, thereby obtaining arbitrary positions/directions. Furthermore, camera-nodes are powered by batteries, and therefore, they do not require a constant power supply. This makes visual sensor networks suitable for use in applications where temporary monitoring is needed and in applications that require fast deployment and removal of the camera network. For example, a visual sensor network can be quickly deployed in a room that is temporarily used for an exhibition or lecture. These networks can be used for monitoring remote areas, which may be outside and thus not contain any fixed infrastructure. In emergency situations, these visual sensor networks can provide valuable information from unaccessible areas affected by some disaster (e.g., flood, fire, earthquake) and hazardous areas.

In this work, we consider a visual sensor network based telepresence system, which enables the user to take a virtual tour over some remote location through interaction with the system's interface [1]. In such a system, the user is able to virtually *move* through the monitored space, and to see images of the monitored space from any desired viewpoint. In order to provide images from arbitrary viewpoints, it is important that the cameras provide *coverage* of the entire 3D space over time. Each camera "covers" a three dimensional view volume (i.e., field of view). In this paper, we consider 3D coverage as follows: a point in 3D space is considered covered by the network if and only if this point is contained in the view volume of at least one camera.

When there is sufficient coverage, image data from several camera-nodes with overlapped views can be combined together in order to generate an image from any arbitrary viewpoint in the network. However, it is necessary to select the set of camera-nodes that will provide enough images of the scene so that the desired view can be reconstructed. This problem of selection of the best cameras has not been fully addressed in the literature. Therefore, in this paper we explore two different methods to select which parts of the image from each camera should be sent to reconstruct the desired view while avoiding redundant data and conserving the camera-nodes' energy. The first camera selection algorithm is based on the minimum angle between the users's desired view direction and the camera's direction. This heuristic approximates the maximum peak signal-to-noise ratio (PSNR) solution. The second camera selection

algorithm is based on cost metrics that measure the camera's importance to the preservation of full coverage of the 3D monitored space, considering the overlap between the cameras' visible volumes as well as the remaining energy of the nodes.

2. Related work

The problems related with the selection of cameras in a camera-based wireless network have been previously investigated. For example, the authors in [3] investigate the problem of the optimal allocation to each camera-node of a part of a scene that has to be transmitted back to the base station, so that the lifetime of the sensors is prolonged. While they look at 2D coverage, we extend this problem to coverage over a 3D space. Also, we consider not only the battery lifetime of the camera-nodes but also their importance in the 3D monitoring task.

In order to extract images from a specific location, the authors in [4] use a look-up table that covers all locations within the cameras' view volume and determines which camera is most suitable to provide the desired image from a certain location. This is similar to our approach for selecting the favorable cameras based on minimum angle between the user's and the camera's directions. However, we also consider the limited energy of the camera-nodes, and we discuss the implications of such an approach on the network's coverage-time.

The authors in [5] describe the architecture of a RealityFlythrough telepresence system that uses video streams from mobile cameras in order to acquire visual information from a monitored site. The main limitation of such a system is the incomplete and perturbing coverage provided by the head-mounted cameras attached to personnel, since they cannot monitor every part of the space over time. Our work differs from the work presented in [5] in several aspects. First, we assume that all cameras in the system are static after deployment, thereby providing fixed (unchangeable) coverage of the monitored area. Also, we are more concerned with the energy constraints of such a system, which influences the selection of the "best" cameras to provide the desired view and increase the system lifetime.

In [6] we analyzed the problem of selecting camera-nodes to cover a planar scene and selecting routers, utilizing an application-specific cost metric that considered coverage of the planar scene. Here, we extend this application-specific cost metric for the 3D coverage case.

3. System Scenario

The camera-based network in our scenario consists of the camera-nodes s_m, $m \in 1..N$, mounted at random locations on the four vertical walls of a room (an art gallery, for example). The direction of each camera c, which is represented
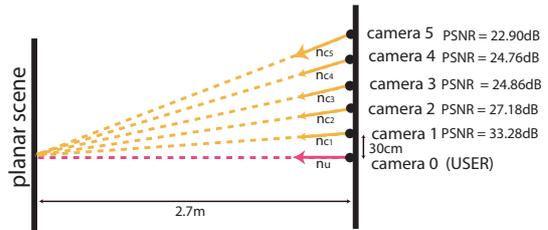

Figure 1: Experiment with aligned cameras.

by a vector in 3D space $\vec{n_c}$, is arbitrarily chosen toward the room's interior.

We assume that a user is able to "move" through the room, meaning that the user can change position and viewing angle over time. As the user virtually moves through the monitored space, the system periodically receives queries that contain the user's 3D location in the room and viewing direction (represented by $\vec{n_q}$). From the system perspective, a user can be replaced by a *virtual camera* that has the same intrinsic parameters as the cameras used in the system.

Our initial scenario assumes that the room monitored by the camera-node system does not contain objects that could partially or fully occlude the view of some cameras. Such a scenario is a simplified version of the more realistic case, when objects appear in the monitored scene. We will come back to this more complex problem in Section 7 and provide insights on how this problem can be approached by using the work presented here. In the absence of objects that occlude the scene, the user's view of an arbitrary scene is just the view of the planar scene from the desired viewpoint. The planar scene is projected according to the perspective projection model [2] onto the user's image plane, forming the user's requested image.

4. Camera Selection Metrics

As the angle between the directions of a selected camera and the user's desired view (which corresponds to the spatial angle between $\vec{n_c}$ and $\vec{n_q}$) becomes larger, it is expected that the difference in the image obtained by this camera and the desired user's image (ground truth image) is larger. In order to evaluate this intuition, we conducted an experiment with several cameras aligned as illustrated in Figure 1. Each camera captures an image of the planar scene in front. The angle between each camera's direction and the user's direction (camera 0) increases with the distance of the camera to the user. We aligned the images taken from each camera to the image taken by the user camera, by finding the homography mapping [2] between the user's image and each camera's image, and we measured the peak signal-to-noise ratio (PSNR) of the rendered images. We use the same sets of feature points, the projective model and bilinear interpolation of any missing pixels in the reconstruction of the

warped images from all cameras. We found that the PSNR of the aligned images does in fact decrease with an increase in the angle between the user's and the camera's viewing directions. Therefore, the angle between the user's and the camera's directions can be used as an approximate measure of the quality (PSNR) of the reconstructed image.

Thus, if the camera-nodes are not constrained by limited energy, the preferable way to select cameras that jointly provide the user's desired image is by choosing those cameras that contain parts of the scene a user is interested in, and that have the smallest angle between their directions and the user's direction. However, since the camera-nodes are battery-operated, this camera selection method should be modified so that it considers the remaining energy of the camera-nodes as well. Also, another constraint for camera selection comes from the fact that the monitored space is non-uniformly covered (monitored) by the cameras, due to the random placement of the camera-nodes on the walls.

As the cameras' visible volumes are overlapped, the volume of one camera can be partially or fully contained in the visible volume of other cameras. In the absence of objects, the scene viewed by a camera may be recovered from the images taken by the cameras with overlapping views (albeit with differing quality levels depending on the angles of view). Therefore, the loss of a redundantly covered camera will not prevent a user from seeing the part of the scene that is covered by this camera. On the other hand, the case when the system looses an "important" camera—one that solely monitors some part of the space—can be prevented (delayed) when the selection of the active camera-nodes is done based on a metric that combines information about the remaining energy of the camera-nodes with information of how redundantly each camera's visible volume is covered by the rest of the cameras. Since this metric does not consider the angle between the directions of the selected camera and the user, it is expected that the images from the cameras selected based on this metric differ more from the image expected by the user, compared to images obtained from the cameras selected based on the "minimum angle" method.

For a given position and direction of the user's desired view, there is a group of camera-nodes that can provide images of the scene in response to the user's query. We label this group of cameras as a set of candidate cameras (CC). However, to prevent the selection of the cameras that provide very different images from the image expected by the user, we exclude from the set CC every camera for which the angle between its optical axis $\vec{n_c}$ and the user's directional view $\vec{n_q}$ is larger than some threshold angle α_{th}.

Based on these ideas, we introduce three methods for the selection of cameras as described next.

4.1. Minimum Angle Camera Selection

Based on the experimental results shown in the previous section, in this minimum angle selection approach, the cameras are chosen by minimizing the angle between the camera's axis and the user's view direction. Although this method is straightforward and will minimize the distortion between the reconstructed image and the desired image, there is a drawback — it does not consider the importance of the camera-node to the task of coverage preservation over the monitored space. Thus it may cause a premature loss of the nodes important to monitoring areas that are not redundantly covered by other camera-nodes' viewing volumes.

4.2. Volumetric Camera Cost (VCC)

In order to define this cost metric, we use the volumetric description of the scene, which is a concept commonly used in 3D computer graphics for the reconstruction of a scene or an object based on joint consideration of all cameras' available views. In the simplest case, the monitored space is divided into small equidistant cubical elements called voxels [1].

Knowing the field of view of each camera, for each voxel we can find the group of cameras that contain this voxel in their view volumes. If each camera-node has remaining energy $E_r(s_m)$, $m \in 1..N$, we can find the total energy of each voxel as the sum of the remaining energies of all the cameras that contain this voxel:

$$E_{total}(c(i,j,k)) = \sum_{\{s_m | c(i,j,k) \in VV(s_m)\}} E_r(s_m) \quad (1)$$

where $c(i,j,k)$ is the center of the voxel, and $VV(s_m)$ is the visible volume of camera-node s_m.

The volumetric camera cost (VCC) measures the camera's importance to the monitoring task, and it is defined as the sum of the energies of all voxels (defined in equation 1) that belong to this camera's viewing volume:

$$C_{VCC}(s_m) = \sum_{c(i,j,k) \in VV(s_m)} \frac{1}{E_{total}(c(i,j,k))} \quad (2)$$

4.3. Direction Based VCC (DVCC)

Although the cameras can share the same 3D space, the information content of their images may be completely different. For example, two cameras on opposite walls can have overlapped visible volumes, but they image completely different scenes. Based on this observation, we define a direction dependent volumetric camera cost metric (DVCC), which considers not only the fact that the cameras share the same visible volume, but also whether or not they view the scene from similar viewing directions. DVCC considers only those cameras that share the same 3D space with this camera and for which the angle between their direction and this camera's direction is smaller than $90°$.

For every camera s_m, $m \in 1..N$, we can find a subset of the cameras that satisfy these requirements, labeled as $Ss(m)$. As seen from camera s_m, the total energy of the voxel $c(i,j,k)$ is equal to the energy of all cameras from the subset $Ss(m)$ that contain this voxel:

$$E_{total}(c(i,j,k))\{m\} = \sum_{\{s_t | c(i,j,k) \in VV(s_t), s_t \in Ss(m)\}} E_r(s_t)$$

(3)

The direction based cost of the camera is thus:

$$C_{DVCC}(s_m) = \sum_{c(i,j,k) \in VV(s_m)} \frac{1}{E_{total}(c(i,j,k))\{m\}}$$

(4)

5. Block-Based Camera Selection

The low-power camera-nodes used in this work are envisioned to have the ability to send only a part of the captured image instead of the entire image. Using inputs from the user about the desired view and any of the proposed cost metrics as a criteria for camera selection, the main processing center (MPC) runs a camera selection algorithm to determine the set of active cameras that take part in the reconstruction of the user's desired view along with the specific image parts needed from each active camera.

At system start-up, the image plane of each camera is projected onto the plane (wall) in front of the camera. These visible regions of the scene are labeled as B_m, $m \in \{1..N\}$). The user's image plane is divided into equal size blocks of pixels. Based on the current position and direction of the user, the system projects the user's image plane onto the plane (wall) that the user currently sees. The cells of the projected user's grid onto the wall are labeled as GPu.

When the VCC or DVCC metrics are used, from all cameras in CC that see a part of the scene the user is interested in, the MPC first chooses the camera c with the smallest cost. Then, the MPC determines all the grid cells from GPu that are contained in the viewing region B_c of this camera. This subset of grid cells from GPu is then mapped back to the camera image plane, determining the region of the image captured by camera c that will be transmitted back to the MPC. All cells from GPu that belongs to the viewing region B_c of this camera are mapped as covered. For the rest of the still uncovered cells from GPu, the selection algorithm repeats the same procedure. The algorithm stops once either all the cells of the user's projected grid GPu are covered or there are no more cameras from CC that can be considered by this algorithm.

When the cameras are chosen based on the "minimum angle" criteria, the selection algorithm has to consider a perspective projection of the scene onto the cameras' image planes, where the angle between a ray from the camera to some point on the user's projected grid and a ray from the
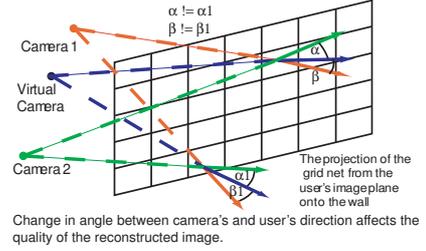

Figure 2: A change in the viewing direction of the camera and the user across the planar scene.

user to the same point changes over the planar scene (wall), as illustrated in Figure 2. The MPC finds the best camera from the set of candidate cameras CC for each cell from GPu individually. Among all cameras that contain this cell from GPu in their field of view, the selection algorithm chooses the camera-node with the smallest angle between the ray that passes from the camera through the center of this cell and the ray from the user to this cell's center.

6. Simulation Results

We performed simulations for 10 different scenarios with the proposed camera selection metrics. Each scenario uses a visual network of 40 camera-nodes, mounted on the four vertical walls of a room of size $10 \times 10 \times 4$ meters. Each wall contains 10 camera-nodes, and the height and directions of the cameras are chosen randomly. The selection of the camera-nodes, which together reconstruct the user's desired view, is repeated in every iteration, where in each iteration the user moves to a different position in the room. The cameras provide images with a resolution of 320×240 pixels, and the horizontal viewing angle (field of view) for all cameras is equal to $40°$. The image plane of the user is divided into blocks of 8×8 pixels. We assume that the energy needed for transmission of an image part from the camera node to the MPC is proportional to the size of the transmitted image part.

6.1. Coverage vs. Quality Trade-off

Figure 3(a) shows how the coverage (expressed as the percentage of all voxels that are in the view volume of at least one alive camera-node) of the monitored 3D space changes over time for different cost metrics. The simulations show that over time, a larger part of the 3D monitored space is considered covered when the VCC or the DVCC costs are used to find the set of cameras, compared with using the "minimum angle" metric. Since both the VCC and the DVCC metrics consider whether the view volume of a camera is covered by the view volumes of other cameras, these metrics direct the camera selection algorithm to avoid the selection of cameras that are not redundantly covered, thus

(a) Coverage of 3D space over time, $\alpha_{th} = 90°$.

(b) Average angle between the user's and the selected cameras' directions, $\alpha_{th} = 90°$.

(c) Coverage of 3D space over time, $\alpha_{th} = 60°$.

(d) Average angle between the user's and the selected cameras' directions, $\alpha_{th} = 60°$.
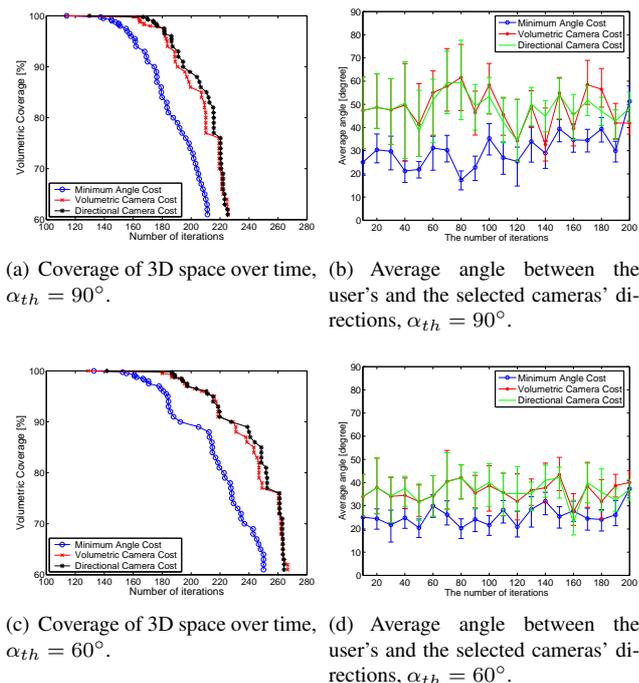
Figure 3: Simulation results for the different cost metrics used for camera selection.

prolonging the lifetime of these high cost camera-nodes. Also, as the camera-node's remaining energy gets smaller, the cost of the camera-node increases significantly, again with the purpose of keeping the camera-node from being selected as an active node whenever the selection algorithm can find another suitable camera.

In order to estimate the quality of the reconstructed image, we measured the average angle between the user's direction and the direction of the selected cameras, as shown in Figure 3(b). The advantage of using "minimum angle" camera selection is that the images are, on average, less warped compared to the images from the cameras selected using VCC or DVCC. The smaller angle between the user's direction and the selected cameras' directions equates to a higher PSNR of the images compared to the ground truth image. Thus, by combining the results provided in Figures 3(a) and 3(b), we can see that there is a clear trade-off in the time during which the monitored space is completely covered by the visual network, and the quality of the reconstructed images requested by the user of this system.

6.2. Influence of α_{th} on 3D Coverage

The simulation results discussed in the previous section are obtained for the case when the set of cameras CC is chosen based on threshold angle $\alpha_{th} = 90°$. For smaller values of α_{th}, the average angle between the cameras' and the user's

direction gets smaller, as can be seen by comparing Figure 3(b) with Figure 3(d) where the angle α_{th} is set to $60°$.
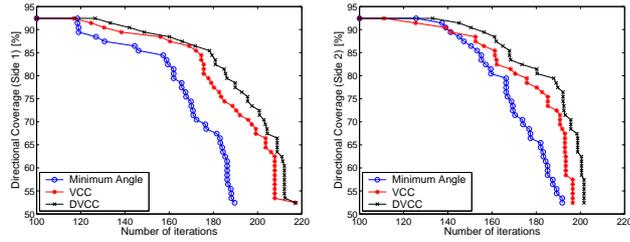
Surprisingly, we notice that partial (less than 100%) 3D coverage is preserved for a longer period of time when α_{th} has a smaller value. In order to explain these coverage results, we compared the amount of data transmitted from all selected cameras to the MPC during the simulations. We found that once the coverage drops below 100%, for smaller α_{th} it happens more often during the simulations that the user's image cannot be fully reconstructed, since there are not enough cameras that can provide all the required image parts. The cameras on average thus send less data to the MPC, which results in less energy spent over time. This is the reason for the prolonged partial coverage over time compared with the case when α_{th} is equal to $90°$.

6.3. Camera Direction Dependent Coverage

The previous results show that the DVCC metric achieves slightly better performance in terms of prolonged coverage over the 3D space compared to the VCC metric. DVCC more specifically determines the cost of the camera-node, since it considers the fact that the information content from the camera's image depends on the camera's direction, so that cameras that share the same information content should have smaller cost, and vice versa.

Following this logic, every camera-node can measure the 3D coverage of the space in its view volume from its direction. Since the measurement of the 3D coverage from each camera's direction is complex, we measure the direction dependent 3D coverage in the following way. The directional coverage represents the percentage of the space (in terms of voxels) that is seen by at least one camera-node from a certain direction. We choose four directions ($n_{w_i}, i \in [1..4]$) in the room, which correspond to the normals of the four walls of the room. The cameras choose their groups according to the angle between their directions and n_w. Each group of cameras observe the monitored space from a different direction, and each of them see different facets of the voxels. We measured the percentage of the voxels' facets contained in the field of view of at least one camera from each group of cameras.

The results for this directional based 3D coverage are shown in Figure 4, for two arbitrarily chosen groups of cameras (the cameras from each group see the 3D space from one main direction). These results show that since the information content of the images depends on the cameras' directions, the 3D coverage can be seen as a measure of the amount of 3D space covered by the cameras from certain directions. Therefore, the 3D space not only needs to be covered by the cameras, but it also has to be covered uniformly from all possible directions, and the DVCC metric can be used to achieve this.

(a) Coverage from the first direction.

(b) Coverage from the second direction.

Figure 4: Directional coverage: coverage measured for two arbitrarily chosen directions and with different cost metrics.

(a) "Minimum angle" criteria (2 cameras selected).

(b) VCC criteria (4 cameras selected).

Figure 5: The user's desired image obtained by mosaicing the images parts.

6.4. Reconstruction of the User's Image

Upon reception of the image parts from the selected camera-nodes, the MPC must map these image pieces to the virtual image plane of the user, and it must stitch the image parts together. In this section, we present the results of an experimental set-up with 7 cameras. The goal of this experiment is to mosaic the parts of the images provided based on the different camera selection metrics, and to show what the final images would look like. The cameras are mounted in one plane, and they point toward the observed plane.

Figure 5(a) shows the mosaiced image from the cameras that have the most similar directions with the user. Figure 5(b) shows the same reconstructed user's view, generated from the cameras that have the smallest volumetric cost. By visually comparing these two images we can notice that the image in Figure 5(b) has more distortion than the image in Figure 5(a), since the chosen cameras lie further away from each other. However, this result is provided for illustrative purposes, and it presents only a rough estimation of the quality of the final images obtained by both metrics. The characterization of the quality of the images obtained by the different proposed camera selection criteria is further discussed in [7].

7. Conclusions and Future Work

In this work we analyze the problem of selecting cameras based on different cost metrics in order to reconstruct a view of the monitored space from an arbitrary viewpoint. Our goal was to explore camera selection methods that will lead to prolonged coverage of the 3D monitored space. By considering the redundancy in the coverage of the cameras' view volumes as well as the remaining energy of the camera-nodes, we define several new selection metrics (VCC and DVCC), and we show that these metrics can provide full coverage of the 3D space for longer than using a "minimum angle" selection criterion, which approximates a maximum PSNR selection method. Comparing all of these metrics shows clearly the trade-off in the quality of the re-

constructed images versus the time during which the network is able to maximally cover the monitored 3D space. Also, since the information captured by a camera depends on the camera's direction, we show that 3D coverage depends on and can be measured with respect to the cameras' view directions.

The scenario presented in this work is a special case of a more general scenario, in which there may be one or more objects in the monitored area. The work presented here will serve as a cornerstone for exploring the problem of camera selection in the presence of objects in the 3D scene. In order to reconstruct an image when the 3D space contains an object, it is important to locate the object in 3D space. Instead of finding the full shape of the object, which is computationally expensive, each object can be approximated by a bounding box. Then, each side of the bounding box can be seen as a plane (planar scene), so we can use the same cost metrics and algorithms presented here to find the best cameras that see each plane of the object's bounding box.

References

[1] O. Schreer, P. Kauff, T. Sikora, "3D Video Communication," John Willey & Sons, 2005.

[2] R. Hartley, A. Zisserman, "Multiple View Geometry in Computer Vision," Cambridge University Press, 2000.

[3] J.C. Dagher, M.W. Marcellin, M.A. Neifeld, "A Method for Coordinating the Distributed Transmission of Imagery," IEEE Transactions on Image Processing, vol.15, no.7, 2006.

[4] J. Park, P.C. Bhat, A.C. Kak, "A Look-up Table Based Approach for Solving the Camera Selection Problem in Large Camera Networks," In Proceedings of the International Workshop on Distributed Smart Cameras (DCS '06), 2006.

[5] N. J. McCurdy, W. Griswold, "A System Architecture for Ubiquitous Video," in Proceedings of Mobysis'05.

[6] S. Soro, W. Heinzelman, "On the Coverage Problem in Video-based Wireless Sensor Networks," In Proc. 2nd Workshop on Broadband Advanced Sensor Networks (BaseNets '05).

[7] C. Yu, S. Soro. G. Sharma, W. Heinzelman, "Lifetime-Distortion Trade-Off in Image Sensor Networks," in Proceedings of ICIP'07.