In the energy-constrained medium of video sensor networks, the objective of much research has been to statistically minimize the number of nodes that will achieve a sufficient degree of coverage. We consider increasing the number of nodes beyond the threshold of full coverage, and cooperatively filtering out the high level of redundant data in the video streams to minimize per-node capacity requirements. The scenario we study is that of a swarm of robots, all with wireless communication capabilities. Some of the robots are equipped with video cameras and are thus considered sensors. A few select robots have sufficient battery and computational power to perform machine vision processing of the video stream. The goal of this scenario is to get the video from the sensors to the video-processing robots, which can then extract high-level surveillance information about the observed environment. We present an optimization framework for minimizing redundant visual data transmissions, while maximizing the throughput from sensors to processing nodes. We also characterize through simulation the performance gain on the sensor network as the video coverage increases.

1. INTRODUCTION

The key result of our work is that cooperative filtering of redundant visual information can increase the amount of surveillance data which can be carried through a communication network. Minimizing excess transmissions is important for the energy-constrained domain of ad-hoc wireless networks. Our work, therefore, touches upon two well studied, but not often combined, fields: the coverage problem [1] and optimal routing [2]. Specifically, we deal with the case when there is excess coverage. We minimize redundant transmissions associated with this excess, and use cooperative routing to utilize the high saturation level of nodes in the arena for increased throughput to the video processing centers. This effect is due to the balancing of the traffic load on the network as the number of data sources increases.

The problem of multi-agent visual coverage is a well-studied problem in computational geometry. Its basic form is known as the art gallery problem [3], which concerns with the minimum number of “guards” needed to visually cover a simple polygon. The problem is NP-complete and APX-hard, however, a tight upper bound on the number of required guards has been proved [4]. Our paper uses a different model of visibility than most of the work related to the art gallery problem, because we seek to remove redundant coverage and maximize unique multi-hop throughput, as opposed to maximizing coverage and minimizing network size. Furthermore, we consider a non-omnidirectional field of vision, which is more realistic for multimedia sensor networks. The coverage problem is also well studied in its application to sensor networks [5, 6]. The focus in majority of such work is on statistical deployment of a minimal set of sensors such as some metric of coverage is sufficiently high. Our work instead looks at the utilization of excessive coverage for increasing throughput of unique visual data on the network.

The goal of our work is to show that increasing the number of sensors while removing redundancy can increase video processed by a sensor network. To demonstrate this we propose a coarse but valid abstraction of both the 3D environment and the transport layer of a video-based sensor network (VBSN). More realistic models of both domains have been extensively studied [7, 8], and are open to the application of methodology proposed in our work. From these studies arise additional requirements and constraints on the visual data needed to achieve acceptable quality of video at the sinks. A good example is demonstrated in the problem of fusing multiple video sequences [9].

The rest of this paper is organized as follows. First,
we introduce the abstraction of the physical environment and the sensor network. We then define the optimization framework for maximizing unique throughput, as well as describing a nonoptimized scheme used to measure the improvement in performance from our algorithm. Lastly, we present results attained from extensive simulation on the effect of increasing the number of sensors on the capacity of the network.

2. SURVEILLANCE NETWORK MODEL

The problem of cooperative surveillance applies to an arena of visual interest and a group of robots with video-sensing and communication capabilities. Elements of visual interest are represented as single-point “landmarks”. Two major assumptions regarding this representation of the environment are made. First, it is assumed that if the arena is sufficiently saturated with landmarks, these visual points represent most of the visual information in the scenario. Second assumption is that the problem of maximizing number of observed landmarks is the same as the problem of video coverage [10]. In other words, full observation of landmarks is equivalent to full coverage of the arena.

The group of robots deployed for surveillance form an ad-hoc video sensor network. For this reason, they are often referred to as nodes in this paper. Each robot optionally has two capabilities (possibly both): (a) video sensing and (b) video processing. The ability to sense implies that a robot has a video camera on it. The ability to process video is a general term that means the robot is able to convert a video stream to semantic information that relates to the goal of surveillance scenario. This could be either automated processing (i.e. object tracking, face recognition) or human-assisted monitoring of the video stream. The processing nodes are so called for their need of processing power and thus costly energy requirements, implying characteristics of limited mobility and high per-unit price. All robots have wireless communication capability. Therefore, robots which can neither sense nor process video serve as relay nodes in the communication network.

2.1. Visibility Graph

The arena representation used is a two-dimensional abstraction of the three-dimensional world which is under surveillance. Inside the world are obstacles which represent 3D structures obstructing LOS communication. Examples of such obstacles can be seen as large shaded polygon shapes in Figure 1.

Figure 1 also shows a set of 100 visual landmarks for the given example. They are represented in the figure as small circles. In this example, a landmark model that focuses on physical structure is used. Therefore, the landmarks are spread evenly along the perimeter of the arena and the perimeter of each obstacle. We have implemented other variations of the landmark placement method, such as random placement in the arena free space, and achieved similar performance results.

The triangles in Figure 1 represent robots with video sensing capabilities, while squares represent robots with video processing capabilities. Finally, big circles represent communication relay nodes with neither capability. Shown in the figure is a visibility graph which is formed by connecting each video sensing robot (or “observer”) with the set of landmarks which satisfy the following visibility requirements:

1. The landmark is in the observer’s field of vision
2. The observer has line-of-sight with the landmark

3. Distance between landmark and observer is less than $R(m)$ for a predefined $m$

In Figure 1, the field of vision is set to 360 degrees, therefore the first constraint cannot be violated. In Figure 1(a), the third constraint cannot be violated either, as $m = \infty$. An example of a 55 degree field of vision is shown on the left part of Figure 2, where an observer can only see a subset of the landmarks because of the first constraint.

2.2. Communication and Visibility Radius

There are two constraints on distance in the proposed model. The first is the radius of communication, which is the radius within which communication between robots is feasible. In other words, if a channel distance is less than this radius, the assumption is that a high enough signal-to-interference-noise ratio (SINR) can be achieved to support the required capacity of transmission on this channel. The second constraint on distance is in determining whether a robot can adequately observe a landmark. If it has LOS with the landmark, a radius-limited constraint requires that it also is close enough to have sufficiently high resolution video of it.

In order to define this constraint in a way that is independent to network and arena size, we introduce a unit of distance which scales with these two scenario-defining parameters:

$$R(m) = \frac{m}{\pi} \sqrt{\frac{A \log n}{n}},$$

where $R(m)$ is used as a single scalable unit distance, $A$ is the area of the arena, and $n$ is the number of robots. This equation comes from the analytical result from [11] which shows asymptotic connectivity of a network of nodes placed with a 2D Poisson process for $m = 1$.

In Figure 1(b), a radius constraint of $R(1)$ is used to determine whether a landmark is seen by the robot even if it satisfies the LOS requirement.

2.3. Video Data Partitioning

The set of landmarks visible to the robot is a 2D representation of the video feed received by it. The line between the robot and each landmark serves as a spatial partitioning guide for the snapshot in the video. An example of this process is shown in Figure 2. On the left side of the figure is a visibility graph with one video-sensing robot having 55 degree field of vision (FOV). On the right side of the figure is a rotated snapshot from a sample video that is split into columns based on the positions of the landmarks relative to the observer. Because in this case the landmarks are approximately evenly spaced, the expected partition of the image is that of approximately equal-width columns.

Each of the columns resulting from the partitioning of the video snapshot is treated as single package of data. In some cases transmission of only a fraction of this package is required, which we assume is done by converting the column into a stream of pixels and transmitting the properly-sized in-order sequence of those pixels.

2.4. Network Flow Formulation

The robots form an ad-hoc communication network with the goal of transmitting video data from video-sensing nodes to nodes which are able to process the video. We use a flow network abstraction [12] to formulate this problem. A flow network $G(V, E)$ is a directed graph, where $V$ is the set of nodes and $(i, j) \in E$ is a set of all edges with non-zero capacity $c(i, j)$. We define two sets on $V$, the source nodes $S \subset V$, and the sink nodes $T \subset V$, such that $S \cap T = \emptyset$. A valid flow $f : E \to \mathbb{R}$ is one that satisfies the following two constraints:

1. Capacity constraint: $f(i, j) \leq c(i, j)$ \quad $\forall(i, j) \in E$
Visibility graph

Communication graph

Surveillance graph

Fig. 3. Example of constructing a surveillance flow graph from a visibility graph, with the intermediate step of building a communication graph based on channel requirements. The single large shaded polygon is an obstacle. The polygon shapes are source nodes, the squares are the sink nodes, the large circles are relay nodes and the small circles in the visibility graph are landmarks.

2. Flow conservation:
\[
\sum_{(i,j) \in E} f(i,j) = \sum_{(j,k) \in E} f(j,k) \quad \forall j \in V \setminus (T \cup S)
\]

The first constraint states that the flow on an edge may not exceed the capacity of that edge. The second constraint states that for all nodes that are not sources or sinks, the incoming flow has to equal the outgoing flow.

We use the network flow formulation to form a communication network amongst the robots, and further to form a surveillance network that treats landmarks as "sources" of information to be delivered to video-processing nodes. We use the visibility graph in Figure 3(a) to show a construction example of a communication and a surveillance network.

2.4.1. Communication Graph

A communication graph is constructed with robots as vertices. A bidirectional edge is added between two robots if the distance between them is less than \(R(m)\) from (1). Figure 3(b) shows a communication graph constructed over the robots in Figure 3(a). The robots with video-sensing capabilities are designated as sources (pentagon shapes), while the robots with video-processing capability are designated as sinks (squares). Robots with neither capability are the relay nodes (circles). The communication graph is a subset of the surveillance graph described in the next section, and in the optimization framework serves only as an intermediate step in the construction of a surveillance graph.

2.4.2. Surveillance Graph

In the communication graph, the sources are the robots with video-sensing capability. However, from the perspective of data, the true sources of information are the landmarks. The surveillance graph is formed by adding the landmarks to the communication graph, and making them the sources, while keeping the video-processing robots as sinks. An edge from a landmark to a video-sensing robot is added if this landmark is visible to the robot. Figure 3(c) shows a surveillance graph formed from the visibility graph in Figure 3(a). The pentagon shapes again designate the sources, and the square shapes designate the sinks. Note that the video-sensing robots have now become relay nodes.

The capacity on the all edges between a landmark and an observer is by default 1. A flow value on these edges is in the [0,1] range, and designates the fraction of the landmark which is observed by this robot. More specifically, this value is the fraction of observed pixels in the video snapshot partition column associated with the landmark. It is assumed that a video-sensing robot has sufficient capacity to transmit all of the landmarks which it can observe. It distributes this capacity evenly among its outgoing edges. The outgoing capacity of re-
flows on \((i,j)\) from a landmark node that doesn’t exceed one: communication network. Unique flow is defined as flow video-processing nodes under the limited capacity of the minimize the number of unique visual data that reaches the node set. mark. If it cannot, then it is defined as part of the relay it is not an observer unless it can see at least one land- other words, even if a node has video-sensing capability it is not an observer unless it can see at least one land- mark. If it cannot, then it is defined as part of the relay node set.

The objective of surveillance problem is to maxi-
mize the number of unique visual data that reaches the video-processing nodes under the limited capacity of the communication network. Unique flow is defined as flow from a landmark node that doesn’t exceed one:

\[
U(f) = \sum_{i \in L} \sum_{(i,j) \in E} \min(1, f(i,j)) \tag{2}
\]

where \(f\) is the set of all flows.

3. OPTIMIZATION FRAMEWORK

In order to formulate the flow optimization problem, we first formalize the surveillance graph \(G(V,E)\). \(L \subseteq V\) are the landmark nodes. They are the sources of the flow data. \(O \subseteq V\) are the video-sensing nodes, \(P \subseteq V\) are the video-processing nodes, and \(R \subseteq V\) are the relay nodes. Of these four sets only \(O\) and \(P\) can have a non-empty intersection. By construction, for all \(j \in O\) there exists at least one edge \((i,j) \in E\) such that \(i \in L\). In other words, even if a node has video-sensing capability it is not an observer unless it can see at least one landmark. If it cannot, then it is defined as part of the relay node set.

The objective of surveillance problem is to maxi-
mize the number of unique visual data that reaches the video-processing nodes under the limited capacity of the communication network. Unique flow is defined as flow from a landmark node which exceeds

\[
f(i,j) = f^*(j) \frac{X(i,j)}{\sum_{(k,j) \in E} f^*(k,j)} \quad \forall j \in O \tag{3}
\]

where \(X(i,j) \sim \text{Uniform}(0,1)\) is a uniformly distributed random variable defined for all \((i,j) \in L \times O\). Optimal outgoing flow \(f^*(i)\) is defined as \(\sum_{(i,j) \in E} f^*(i,j)\), where \(f^*(i,j)\) is the flow achieved on edge \((i,j)\) at maximum network throughput. In other words, \(f^*(i)\) is a reasonable estimate for the “actual” outgoing capacity of node \(i\).

A single realization of \(U(f)\) is generated by sampling a set of flows on \(L \times O\) using (3), and from that use (2) to compute the unique throughput. To construct an empirical distribution of \(U(f)\), we generate a large number of realizations and use them to form a histogram.

3.2. Unique Throughput Maximization

The maximization of throughput on the surveillance graph is achieved by solving the following classical definition of the max flow linear program [13]:

\[
\begin{align*}
\max & \sum_{i \in L} \sum_{(i,j) \in E} f(i,j) \\
\text{s.t.} & \quad f(i,j) \leq c(i,j) \quad \forall (i,j) \in E \\
& \quad \sum_{(i,j) \in E} f(i,j) = \sum_{(j,k) \in E} f(j,k) \quad \forall j \in V \setminus (L \cup P) \\
\end{align*} \tag{4}
\]

where the objective is the total flow outgoing from the landmark nodes. The constraints are those defined in the section on the network flow formulation. They are the capacity constraint, and the flow conservation con-
straint, respectively. We denote the solution to (4) as \(f^*\). This value is used in (3) to compute the unique through-
put of the nonoptimized transmission scheme.

The objective of (4) does not measure redundant flow, which is a flow out of a landmark node which exceeds the value of 1. In order to maximize unique throughput, instead of net throughput, the objective needs to be:

\[
\max U(f) \tag{5}
\]

Since \(U(f)\) is a piecewise linear function (PLF) that is both continuous and convex, it can be converted to a linear program [14]. We ensure that the framework re-
ains a linear program by maximizing flow as in (4) but
with an additional constraint that the total flow outgoing from a landmark node cannot exceed the value of 1:

\[ \sum_{(i,j) \in E} f(i,j) \leq 1 \quad \forall i \in L \]

This constraint enforces that zero excess visual data is transmitted on the surveillance network. The solution to this linear program is computed by automatically generating the problem in AMPL [15] and using LP Solve [16] to solve it.

Once each video-sensing robot knows the fraction of each landmark it needs to transmit, it partitions its video stream spatially according to this flow allocation and discards the excess data.

4. SIMULATION AND RESULTS

First, we define the key model and simulation parameter values used to generate the results. Then, we show the empirical distribution of the network performance under a nonoptimized scheme. Lastly, we compare optimized and nonoptimized performance as it relates to the relative number of robots with video-sensing capabilities.

4.1. Simulation Setup

4.1.1. Physical Parameters

The scenario is generated on a square 1 km² arena. It is filled with 10 simple polygons, placed and shaped randomly such that the area occupied by the obstacles is approximately 30% of the arena area. We use a 2D Poisson process to place 50 robots in the arena. If a robot falls inside an obstacle, its position is generated again. 100 visual landmarks are placed in the environment with the method described in the section on the visibility graph. Unless otherwise specified, we provide 10% of the robots with video-processing capability, and 60% of robots with video-sensing capability, allowing for any of the robots to have both. The rest are marked as relay nodes.

4.1.2. Network Parameters

The scalable distance \( R(m) \) controls two radius constraints. We use \( R(1.5) \) for the radius of feasible node-to-node communication. The visibility model used in simulation is a radius-constrained one with \( R(2) \) as the radius defining sufficiently high-resolution visibility.

Fig. 4. Empirical distribution of unique throughput attained on a sample scenario with the nonoptimized transmission scheme.

4.2. Nonoptimized Performance

A reasonable measure of unique throughput for the nonoptimized scheme in which sensors essentially flood the network with the observations they collect is an empirical distribution constructed from generating realizations of possible flow allocations on \( L \times O \). Figure 4 shows an empirical distribution formed from a sample network topology generated with the parameters described in the Setup section. The independent variable is the measure of unique throughput \( U(f) \) for the nonoptimized scheme over the optimized. Therefore, in this case the nonoptimized scheme on average delivers about 15% less visual data to the video-processing nodes.

This is a conservative estimate of performance for the nonoptimized scheme because it does not consider the cost associated with flooding the network. It is assumed that data which exceeds the capacity of the network simply does not reach the video-processing nodes and does not negatively affect network performance. This is an unrealistic assumption, as the transmission of the excess data is likely to increase interference and thus reduce the capacity of the network, leading to a convex decline of the unique throughput. We take this assumption to focus on the benefit attained strictly at the transport layer, without considering the significant benefits at the PHY and MAC layers expected when redundant data transmission is minimized [17].
4.3. Optimized Performance

The intuitive increase in unique throughput from cooperative filtering of redundant visual data is seen in the gap between the two lines of Figure 5. The independent variable is the fraction of robots which have videosensing capability. The dependent variable is the unique throughput divided by all the landmarks in the arena. The plot is generated in the following way:

1. Provide 10% of robots with video sensing capability.
2. Generate obstacles, landmarks, and robot topology.
3. Compute nonoptimized and optimized unique throughput.
4. Repeat steps 2 and 3 one thousand times, and average the result to make one data point in the plot.
5. Increase the fraction of robots with video sensing capability and go to step 2.

The less intuitive result of our work, seen in Figure 5, is that as the number of video-sensing robots increases, keeping the total number of robots constant, the amount of visual data which can reach the video-processing robots increases. In other words the more robots can see the same set of landmarks, the more capable they are of decreasing the clustered overload of the network which leads to bottlenecks. This result goes against the conventional wisdom that in a sensor network the ultimate goal is to achieve coverage with minimum number of nodes. We show that increasing the number of sensors increases the capacity of the network, and thus the “multi-hop coverage” or the amount of data which reaches the processing centers.

5. CONCLUSION

We propose an algorithm of cooperative surveillance of an obstacle-ridden arena by a group of robots who form a video sensor network. Some of the robots have video cameras, and others have the computational and battery power to process the video. Our goal is to maximize the delivery of unique visual information from the sensing to the processing agents. Delivering sufficient quality of live video feed across an energy-constrained ad hoc network has received much attention. Spatial and temporal compression, coding, and optimal sensor deployment are some of the well-studied approaches to help improve the throughput on such networks. We propose a method for balancing the flow on the network by cooperatively removing redundant visual data at the sources. We show that as the density of video sensors increases, the throughput of unique data also increases because of the reduction on the amount of distinct visual information requiring transmission at each sensor.

Acknowledgments

This work is funded by the U.S. Army Communications–Electronics Research, Development and Engineering Center (CERDEC), under contract #DAAB-07-01-9-L504.

6. REFERENCES


