A Sensor-Utility-Network Method for Estimation of Occupancy Distribution in Buildings

Sean Meyn

Electrical and Comp. Eng. and the Coordinated Sciences Laboratory
University of Illinois at Urbana-Champaign

Amit Surana, Yiqing Lin, Stella M. Oggianu,
Satish Narayanan and Thomas A. Frewen

United Technologies Research Center
411 Silver Lane, East Hartford, CT.

Abstract— We introduce the sensor-utility-network (SUN) method for occupancy estimation in buildings. Based on inputs from a variety of sensor measurements, along with historical data regarding building utilization, the SUN estimator produces occupancy estimates through the solution of a receding-horizon convex optimization problem. State-of-the-art on-line occupancy algorithms rely on indirect measurements, such as CO_2 levels, or people counting sensors which are subject to significant errors and cost. The newly proposed method was evaluated via experiments in an office building environment. Estimation accuracy is shown to improve significantly when all available data is incorporated in the estimator. In particular, it is found that the average estimation error at the building level is reduced from 70% to 11% using the SUN estimator, when compared to the conventional approach that relies solely on occupancy level or flow measurements.

I. Introduction

A. Objectives & Contributions

The goal of this paper is to introduce a framework and algorithms to analyze and estimate occupancy in buildings using sensor measurements from diverse sources, models of facility use, and historical data. Information regarding occupancy levels and distribution patterns in buildings can be utilized for a number of applications, such as: demand-driven ventilation and lighting controls leading to energy savings [5], to improve security monitoring [9], [13], and to accelerate safe evacuation [3].

The estimation approach proposed in this paper generates occupancy estimates based on a combination of *sensors*, information regarding prior knowledge of building utilization or *utility*, and the *network structure* of the building [7]. An algorithm in the resulting class of algorithms is called the Sensor-Utility-Network (SUN) estimator. Sensors may include CO_2, passive infrared (PIR), video, sound, badge counters, and even measurements of the number of cars in the parking lot. The utility of the building, or zones within, is estimated based on historical data for the same building or a building of similar characteristics, forecast preferences based on room schedules, soft constraints on the speed of occupants, as well as hard constraints on occupancy in each room. Finally, the network structure is specified by decomposing the total building into zones, typically based on individual rooms or groups of rooms. The choice of zones will depend on the primary goal (e.g. ventilation control vs. evacuation), and on the number, location, type and accuracy of the sensors. Estimation is then based on a convex program based on hard and soft constraints, and taking into account the inherent bias and variance of the various sources of information.

B. Background

In typical building applications, CO_2 sensor data are used to estimate occupancy levels for controlling ventilation. Among the drawbacks of this approach is that mixing rates of CO_2 are unpredictable, have slow response (order 10-20 minutes), and depend on external conditions such as open or closed windows and doors.

The SUN algorithm is motivated in part by the *pseudo MAP* estimator in statistics [6], and was created as a refinement of the algorithm introduced in [13], which is based on using an extended Kalman filter. The SUN estimator is much more adaptable than the algorithm considered in [13], in the sense that it can make use of much greater prior knowledge, constraints, and data sources. On the other hand, the algorithm of [13] is much simpler than the SUN algorithm, and it can be constructed in a decentralized fashion so that estimates are based only on local information. The SUN estimator is posed as a global optimization problem, hence decentralizing it is not straightforward, and the subject of ongoing research.

In its simplest form, the convex program that defines the SUN estimator is a type of receding horizon estimation, as described in [4]. Refinements of the algorithm incorporate smoothing as well as prediction to make better use of constraints.

In related research [1], network topology and people behavior is learned based on observed one-bit data streams. This is then used to create a hidden Markov model (HMM) for behavior. Estimation can then be performed using the standard Bayes estimator [6]. In general the complexity of the HMM will be *extremely high*.

Fluid and Markovian networks have been used in queueing theory [8] to estimate people movement during evacuation. Related work includes the 1981 paper by Smith and Towsley [12], the collection of papers [11], and more recent work reported in [13], [14], [2], [10].

C. Organization

This paper is organized in four sections. In Section II we describe the SUN architecture which can be used for

the purpose of occupancy estimation, smoothing and prediction in buildings. Section III deals with experiments in an office building environment which is instrumented with digital video cameras, passive infrared motion detectors and CO_2 sensors. Also contained in this section is a detailed discussion on construction of penalty function for these sensors, and utility for capturing building usage patterns. Estimation accuracy using SUN is shown to improve significantly when compared to the conventional estimation that relies solely on occupancy flow measurements. Finally, we conclude in Section IV with recommendations for future work.

II. SUN ESTIMATOR

The SUN estimator of occupancy density and flows in buildings is based on information from sensors, historical data for the same building or a building of similar characteristics, and people's preferences and patterns of behavior for diverse situations.

A. Technical challenges

Advances in sensing, embedded systems and wireless technology enable commercial buildings today to have unprecedented access to data (from temperature sensors, CO_2 sensors, smoke detectors, CCTV video cameras, water flow sensors, RFID sensors, and other access control devices). Other sources of information such as phone usage and calendar or scheduling information also offer important insights into occupancy and presence. The resulting large and heterogeneous data streams contain valuable *information* that is largely untapped at present.

The goal of the estimation procedure presented here is to use all these potential information sources, and to address the following issues: (i) Cost-benefit trade-offs involved in the selection of sensors and their placement; (ii) Complement sensor measurements with models of building usage; (iii) Adapt models and algorithms to a changing environment, and an evolving building population.

Occupancy estimation would appear to be a matter of simple book-keeping based on all of this data. However, there are several barriers to reliable estimation, such as:

(i) *Sensor noise*. The selection of the type and quality of sensors will impact the reliability, bias and type of noise of the variables that are in the measurements.

(ii) *Lack of observability*. Consider that sensors only count the number of individuals that pass from one building region to another, then the system is not observable (in a linear system theory sense). This means that a linear filter cannot be used to obtain reliable estimates.

(iii) *Lack of reliable models*. Databases are available that provide insight into occupant behavior in special cases, such as during egress. However, this behavior is highly unpredictable, and is sensitive to culture and environment.

Although any one data source may be unreliable when taken alone, we argue that the data taken together is significantly more informative. This is in fact the motivation for

the SUN estimation procedure that combines data from all available sources for estimation.

B. Estimator architecture

In a discrete-time stochastic model the state process $\phi(t)$ is denoted as:

$$\phi(t) = \begin{pmatrix} x(t) \\ r(t) \end{pmatrix} \tag{1}$$

where $x(t)$ is the vector of occupancy in each zone, and $r(t)$ is a vector of the number of persons moving from one zone to another during time-slot t. The state is subject to the simple mass-balance constraints

$$x_i(t+1) - x_i(t) = \sum_j r_{ji}(t) - \sum_l r_{il}(t). \tag{2}$$

Throughout most of the paper it is assumed that for each time t the vector representing the sequence of observations consists of noisy measurements of a subset of the variables $\{x_i(t), r_j(t)\}$. With (2) taken as a linear state space model with state $\phi = \{\phi(t)\}$, it can be shown that the state process is not observable based on observations of the flow r alone. In fact, observability requires that measurements of each $x_i(t)$ be made available. Fortunately, observability is a purely linear concept. There is a great deal of information relative to building usage that can be used to reduce estimation error. For example, it may be known that a building is empty at 4:00 a.m.

To understand the role of prior information, and to motivate the SUN estimator we turn to a Bayesian setting. Consider the general nonlinear model in discrete time,

$$\begin{aligned} \phi(t+1) &= f_t(\phi(t)) + W(t+1) \\ Y(t) &= h_t(\phi(t)) + V(t+1) \end{aligned} \tag{3}$$

where Y represents the sequence of observations. If the noise in (3) and the initial condition $\phi(0)$ are jointly Gaussian and mutually independent, and the noise is i.i.d., then it is possible to construct a closed form expression for the a posteriori probability $p(\phi \mid y)$. The maximum a posteriori probability (MAP) estimate is then $\hat{\phi}(t) = \arg\max_\phi p(\phi \mid Y_0^t)$.

Under these assumptions the negative of the log of the probability of observations, $-\log(p(\phi_0, \ldots, \phi_T \mid y_0, \ldots, y_T))$, is proportional to

$$\begin{aligned} \|\phi_0 - \bar{\phi}_0\|_{\Sigma_0^{-1}}^2 &+ \sum_{t=0}^{T-1} (\|y(t) - h_t(\phi(t))\|_{\Sigma_{yt}^{-1}}^2 \\ &+ \|\phi(t+1) - f_t(\phi(t))\|_{\Sigma_{dt}^{-1}}^2) \end{aligned} \tag{4}$$

where $(\Sigma_{yt}, \Sigma_{dt})$ are the covariance matrices for (W_t, V_t), Σ_0 is the covariance of the initial condition, and $\bar{\phi}_0$ is its mean. It is useful that the application of variance information on sensors and dynamics arises so naturally in the estimator.

Computation of the estimates $\{\hat{\phi}(0), \ldots, \hat{\phi}(T)\}$, given the observations $\{Y(0), \ldots, Y(T-1)\}$, reduces to a quadratic program (QP) when h_t and f_t are linear, and there are no hard constraints on ϕ. In this special case, the MAP estimator coincides with the Kalman filter.

Since the model is constrained, it makes little sense to assume that any of these random variables are Gaussian. Regardless, the resulting estimation algorithm defined as the optimization (4) subject to state space constraints, is very appealing and flexible.

The estimator introduced in this paper generalizes these ideas through the introduction of utility and penalty functions that model behavior. These functions are denoted as follows:

- $\mathcal{U}_x$: Occupancy utility function defined for each room or connections between rooms, such as corridors.
- $\mathcal{P}_0$: Penalty function that models confidence in the initial estimate $\bar{\phi}_0$.
- $\mathcal{P}_y$: Penalty function based on sensor information
- $\mathcal{P}_d$: Penalty function based on temporal dynamics.

Each of these is a function of $\phi = (x, r)$, and the specific form of the penalty and utility functions will be determined based on various factors. Some suggestions are provided below.

This list is not meant to be exhaustive. For example, a utility function for rates could be used to model path formation. Moreover, a utility or penalty function may depend on the time of day. For example, the utility of a room will change throughout the day and night.

The convex program that defines the SUN estimator is given as,

$$\operatorname*{arg\,min}_{\phi_0,\ldots,\phi_T} \left\{ \mathcal{P}_0(\phi_0) + \sum_{t=0}^{T-1} \left(\mathcal{P}_y(\phi(t), y(t)) \right. \right.$$
$$\left. \left. + \ \mathcal{P}_d(\phi(t+1), \phi(t)) - \mathcal{U}_{x\,t}(\phi(t)) \right) \right\} \quad (5)$$

Given $\{y_0, \ldots, y_{T-1}\} = \{Y(0), \ldots, Y(T-1)\}$, the maximizer in (5), subject to given hard constraints, is taken as the estimates $\{\hat{\phi}(0), \ldots, \hat{\phi}(T)\}$.

Interpretation of the results of the estimator requires special considerations. At time T the SUN estimator gives the $T+1$ estimates $\{\hat{\phi}(0;T), \ldots, \hat{\phi}(T;T)\}$ after solving the optimization problem (5). $\hat{\phi}(T;T)$ is the current estimate of the state. For $t < T$, $\hat{\phi}(t;T)$ is interpreted as the *smoothed* estimate of $\phi(t)$. We can also obtain *predictions* of future state values. Define $\hat{\phi}(t;T)$ for $t > T$ through a second optimization procedure *after* $\hat{\phi}(T) = \hat{\phi}(T;T)$ is obtained. Fix $T_1 > T$ and consider,

$$\operatorname*{arg\,min}_{\phi_T,\ldots,\phi_{T_1}} \left\{ \sum_{t=T}^{T_1-1} \left(\mathcal{P}_d(\phi(t+1), \phi(t)) - \mathcal{U}_x(\phi(t)) \right) \right\} \quad (6)$$

and subject to $\phi_T = \hat{\phi}(T)$. In some situations smoothing and prediction should be performed simultaneously. For example, if it is known that a very large meeting is scheduled at 3:00 p.m., then this information should be made available at time 2:30 p.m. In this case, at time T we solve the optimization problem,

$$\operatorname*{arg\,min}_{\phi_0,\ldots,\phi_T,\ldots,\phi_{T_1}} \left[\mathcal{P}_{\text{smooth}} + \mathcal{P}_{\text{predict}} \right] \quad (7)$$

where, $\mathcal{P}_{\text{smooth}}$ is the summand in (5), and $\mathcal{P}_{\text{predict}}$ is the summand in (6). Thus, the SUN estimator can be employed in different modes: smoothing, estimation and prediction.

C. Construction of Utility and Penalty Functions

This section describes the mathematical construction of a utility function for occupancy and flow rate; and a penalty function for initial estimates, temporal dynamics and sensors.

1) Penalty functions: The construction of a penalty function $\mathcal{P}_y$ is assumes that the observations are linear, $Y(t) = C\phi(t) + N(t)$ for some matrix C, with $\mathbf{N}$ as the sensor noise, so it takes the general form,

$$\mathcal{P}_y(\phi(t), y(t)) = \mathcal{P}_y(y(t) - C\phi(t)) \quad (8)$$

where $\mathcal{P}_y : \mathbb{R}^m \to \mathbb{R}_+$ vanishes only at the origin.

As for $\mathcal{P}_d$, the mass-balance constraint (2) is imposed as a hard constraint in the optimization (5), so that

$$\widehat{x}_i(t+1) = \widehat{x}_i(t) + \sum_j \hat{r}_{ji}(t) - \sum_k \hat{r}_{ik}(t). \quad (9)$$

Soft constraints on $\hat{r}(t)$ can be imposed via,

$$\mathcal{P}_d(\phi(t+1), \phi(t)) = \mathcal{P}_r(r(t+1) - Ar(t)) \quad (10)$$

with A a square matrix, and $\mathcal{P}_r$ similar to $\mathcal{P}_y$. For example, $\mathcal{P}_r$ is a continuity constraint if $A = I$.

2) Utility functions: The utility function for the SUN estimator can be constructed based on historical data or knowledge relative to rooms occupancy and people flow.

Note that the optimization procedure to obtain the estimates $\{\hat{\phi}(t)\}$ will impose hard constraints on occupancy levels: non-negativity and upper limits, as well as mass-balance constraints in equation (9). Hence the utility function can be designed without regard to state space constraints.
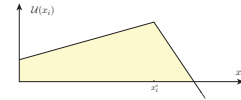

Fig. 1. Candidate utility function for a room with reservation for x_i^0 people.
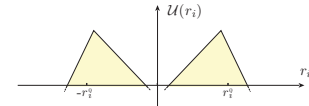

Fig. 2. Candidate utility function for occupancy flow

The following is a non-exclusive list of possible factors included in the utility function:

Occupancy: For example, there may be prior knowledge regarding a meeting in an office building; or in a park, a music event may be scheduled in advance. Figure 1 illustrates a typical piecewise linear utility function that could represent these events. The particular shape depends upon the uncertainty on the a-priori expected occupancy.

Rates: In a building, the utility of a hallway is low from the point of view of occupation, while the utility function for the same hallway from the point of view of transportation is high. Since traffic can go in one of two directions, the utility function for a given rate would appear as shown in Figure 2. To obtain a convex
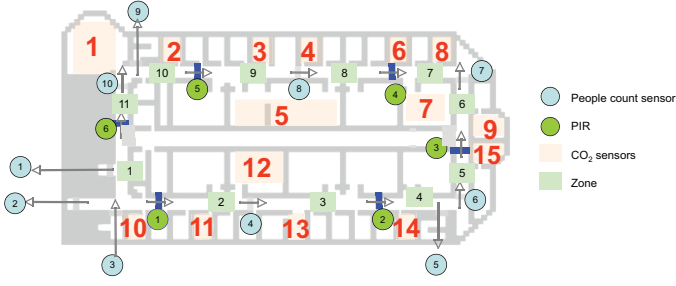
Fig. 3. Sensor layout for people traffic estimation.

program, the state space must be enlarged to include the positive and negative parts of r_i as separate variables.

Patterns: In applications over a large region with limited information a utility function that favors certain patterns of behavior might be considered, including clustering and lane formation.

Seasons or times of the day: The utility functions may vary seasonally, with time of the day or even days of the weeks and holidays. For example, a park utility function is usually higher during the day than over night. This changes on a 4th of July, when big crowds gather to watch the fireworks.

III. EXPERIMENTAL RESULTS

In this section we describe occupancy estimation results obtained in experiments in an office building.

A. Sensor and sensor layout

Figure 3 shows the building floor layout as well as locations of the sensors. Three classes of sensors were used in the experiment: (i) Digital Video Cameras, (ii) Passive Infrared (PIR) Detectors and (iii) CO_2 sensors. In order to monitor occupancy flow, 10 video cameras were installed across pre-defined lines located at the 5 entrances of the floor, and in the middle of the hallways. In order to cover the gaps, twelve non-directional PIR sensors were located in pairs between the video cameras. Additionally, 15 CO_2 sensors were distributed in different rooms.

1) Video Cameras: Video cameras provide information regarding people count and the direction of the flow. These cameras were found to exhibit significant error in the form of variance and bias, arising from three main factors. Firstly, during early and late hours, when the lighting condition is poor, a single person crossing may be counted multiple times. Secondly, turning a light switch on or off may trigger a sensor count. Lastly, the video system may count several crossings during a time interval in which there are occupants loitering close to the camera. Such events lead to a significant positive bias.

2) PIR Detectors: PIR detectors provide indication of motion within sensor range. The PIR detectors employed in this study were non-directional. By using them in pairs, they were rendered *directional*. Consequently, the twelve PIR sensors were grouped into six pairs. Preprocessing of

PIR sensor data was performed to eliminate drift in data logging time stamps over the course of days. In addition the directional information was inferred based on the time difference between event time for each pair of detectors.

3) CO_2 Sensors: CO_2 sensors provide concentration readings in parts per million (ppm), which is indicative of the occupancy. However, reliably correlating CO_2 levels with the actual occupancy is difficult due to high variability and the slow response time of the CO_2 sensors. This variability arises due to fluctuations in ambient CO_2 levels, HVAC system settings and door open/close status. Additionally, CO_2 sensors suffer from poor response time (about 10-20 minutes for sensors employed in this study) due to delay between occupancy level increase and build up of CO_2 concentration.

In next section we show how the bias and variance in the occupancy and flow measurements, can be systematically handled in the SUN framework.

B. SUN Implementation

The estimation algorithm used in the experiments was a special case of the convex program (5), with the objective function,

$$\|\phi_0 - \bar{\phi}_0\|^2_{\Sigma_0^{-1}} + \sum_{t=0}^{T-1} \big(\mathcal{P}_y(\phi(t), y(t))$$
$$+ \|\phi(t+1) - A_t\phi(t)\|^2_{\Sigma_d^{-1}} - \mathcal{U}_x(\phi(t))\big) \quad (11)$$

where, $\mathcal{P}_y$ and $\mathcal{U}_x$ are quadratic functions of $\phi = (x, r)$. In addition bound constraints were imposed on occupancy and flow. As a result the convex program, simplifies to a quadratic optimization problem.

In our current implementation, the building floor is divided into 11 zones which is dictated by the location of video and PIR sensors (see Figure 3 for zone definition). The matrix A_t in the dynamics penalty term in the objective (11) is taken to be the identity matrix. Incorporating CO_2 measurements in above framework requires additional modifications. This is described in detail below, along with the construction of $\mathcal{P}_y$, $\mathcal{U}_x$ and the choice of bounds.

1) Penalty function for Video Cameras and PIR Motion Detectors: It was pointed out in section III-A that video sensors suffer from large bias and variance. PIR detectors on the other hand, completely miss on the people count information. Statistical models are required in order to adaptively correct for such errors. For instance, consider PIR sensors and assume that we have a prior probability distribution $p_{ij}(r)$, that the flow from zone i to zone j is $r_{ij} = r$, given that the PIR indicates movement. Let $\{p_i : i = 0, 1\}$ be two such prior distributions (see Figure 4 top two panels), where p_0 is the probability distribution when the PIR sensor does not indicate any activity and p_1 is the distribution when it is activated. Assume that these probabilities are log-concave:

$$p_i(r) = e^{-P_i(r)}$$

where, P_i is a linear piecewise utility or soft penalty function, introduced in section II-C.2. By convolving these penalty
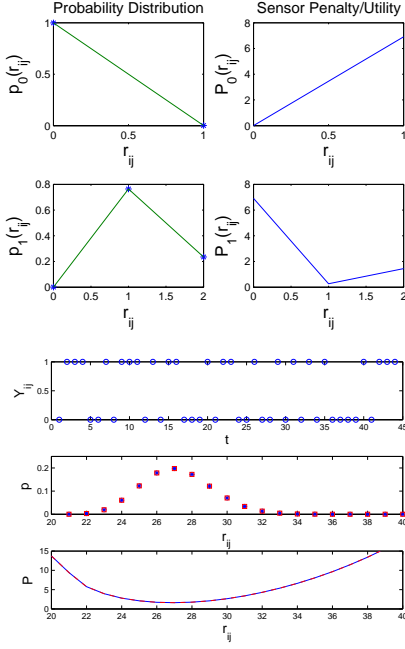
Fig. 4. When $\{P_i, i = 0, 1\}$ has a small support (Top two panels), the convolutions preserve the convexity (Bottom Panel, last figure). The convolved utility P can be approximated by a quadratic.

functions, we can obtain the probabilistic model at any time scale desired for occupancy estimation. The convolved function, which we would denote by P, is no longer guaranteed to be convex. However, from numerical experiments we found that, when P_i has small support, the convolutions preserve this convexity, and in fact P can be very well approximated by a quadratic as shown in the bottommost panel in Figure 4. Indeed, if P_i in the convolution were Gaussian probability distributions, the resulting convolution would be Gaussian with mean equal to the sum of the means, and variance equal to the sum of the variances. Motivated by this, we take the sensor penalty term $\mathcal{P}_y$ in (11) to be of the form

$$\mathcal{P}_y(\phi, y) = \sum_{i=1}^{N_f} \sum_{j>i} P_{y_{ij}}, \quad (12)$$

where, N_f is the total number of flow sensors (video and PIR) and

$$P_{y_{ij}}(r_{ij}, y_{ij}) = \tfrac{1}{2}(y_{ij} - b_{ij}(y_{ij})))^2 / \sigma_{ij}^2, \quad (13)$$

is a soft penalty for each sensor, with b_{ij} being the bias that depends on the measurement y_{ij} and σ_{ij}^2 being the variance. These parameters of soft penalty functions for video count and PIR sensors were obtained by an extensive statistical analysis of the ground truth occupancy flow data.

2) Utility for Occupancy: We used smoothing procedure to obtain occupancy utility functions at zonal level by using historical data of the building usage pattern. In smoothing, at the end of each day, say at 11pm, the algorithm (11) is run over the period $[0, T] = [12am \ldots 11pm]$, with the constraint that occupancy is near zero on the boundaries, i.e. at the

beginning and end of the period. The SUN estimator returns estimates $\{\hat{\phi}^k(t : T) : t = 0, \ldots, T\}$ of occupancy and flow for each day $k = 1, \ldots, N$. Based on these estimates collected over several weeks, we obtained estimates of the mean and variance of occupancy in various zones, for each hour

$$\mu_i = N^{-1} \sum_{k=1}^{N} \widehat{x}_i^k, \qquad \sigma_i^2 = N^{-1} \sum_{k=1}^{N} (\widehat{x}_i^k - \mu_i)^2$$

where, $i = 1, \cdots, N_z$ with $N_z = 11$ being the total number of zones. This information can be translated to "utility" in the SUN estimator, as

$$\mathcal{U}_x(x) = \sum_{i=1}^{N_z} \mathcal{U}_i(x_i),$$

where,

$$\mathcal{U}_i(x_i) = -\tfrac{1}{2}(x_i - \mu_i)^2 / \sigma_i^2$$

and can be used for on-line estimation. The minus sign comes from the fact that utility is to be maximized. We applied this smoothing procedure on sensor data from 16 days to obtain the mean and variance of occupancy for the different zones. Figure 5 shows hourly occupancy estimate at building level obtained for 4 (among 16) days obtained in this process, and also the mean and variance of occupancy as a function of time for two different zones.
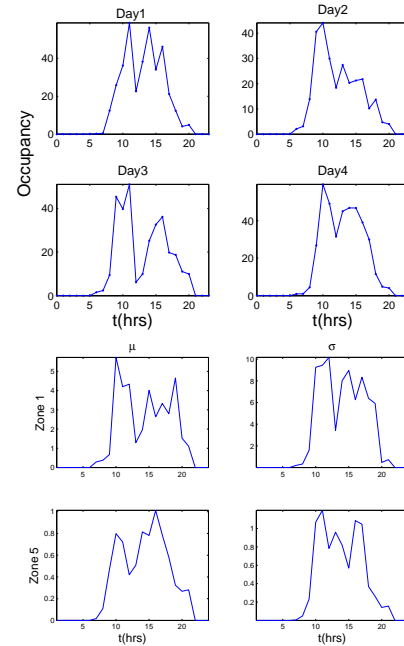

Fig. 5. Top Two Panels: Hourly occupancy estimate at building level for 4 workdays using the smoothing procedure. Bottom Two Panels: Mean and variance of occupancy as a function of time for 2 of the 11 zones.

3) Incorporating Partial Information from CO_2 Sensors: Recall that CO_2 sensors were installed only in some rooms and provide additional occupancy information for portions of each zone (see Figure 3). The simplest way to incorporate

such partial information into the existing formulation (11) is to augment the vector of decision variables with the estimated occupancies of rooms equipped with a CO_2 sensor. Consequently, the definition of vector occupancy $x(t)$ also gets modified. More precisely, the decision variable $\phi(t)$ becomes

$$\phi(t) = \begin{pmatrix} x(t) \\ x^c(t) \\ r(t) \end{pmatrix}, \qquad (14)$$

where, $x(t)$ is a vector of occupancy levels, at time t, for all zones that do not have CO_2 sensor-equipped rooms within each zone, $x^c(t)$ is a vector of occupancy levels for rooms with CO_2 sensors at time t and $r(t)$ is the vector of flow rates as before. The mass balance constraint for each zone takes the form

$$\widehat{x}_i(t+1) \;+\; \sum_{j=1}^{N_{ci}} \widehat{x}_j^c(t+1) = \widehat{x}_i(t) + \sum_{j=1}^{N_{ci}} \widehat{x}_j^c(t)$$
$$+ \; \sum_{j} \hat{r}_{ji}(t) - \sum_{l} \hat{r}_{il}(t),$$

where N_{ci} is the number of rooms equipped with CO_2 sensors in ith zone. The mean and variances required for the CO_2 utility functions are computed empirically using 16 days of CO_2 sensor measurement data, similar to as described in the section III-B.2. However, before extracting the mean and variance for CO_2 utility functions, significant preprocessing of CO_2 measurements is required. First, in order to eliminate the day to day variations, the CO_2 readings of a typical day is subtracted by a day of CO_2 data without any occupancy (such as a weekend or a holiday). The top and middle panels in Figure 6 shows CO_2 profile in an unoccupied and occupied room, respectively. The CO_2 measurements are then denoised using a 4th order low-pass FIR digital filter. The model used to relate the filtered CO_2 measurements $y_i(t)$ (with background CO_2 level subtraction) to the underlying occupancy estimates $x_i^c(t)$ (accounting for a single step 10 minute time lag) is given by

$$y_i(t+1) = \alpha y_i(t) + \beta \widehat{x}_i^c(t), \qquad (15)$$

where the coefficients α and β are fitted using the ground truth occupancy data and corresponding CO_2 measurement data available, as shown as a red curve in bottom panel in Figure 6. The modeled occupancy is set to zero when the CO_2 level is below a "tolerance" of 50ppm; the observed variability in CO_2 measurements for zero occupancy is approximately 50ppm. The bottom panel in Figure 6 shows the correspondence between the actual and modeled occupancy using this approach.

4) Bounds on Occupancy and Flow Rates: The lower bound on occupancy level in any zone is set to be zero, while the zone level upper bound is taken to be summation of the upper bounds imposed on each rooms in the zone. For a room, we used the room level upper bounds given in Table I, which depends on the type of room and its expected usage pattern over the course of a day. Currently, the flow
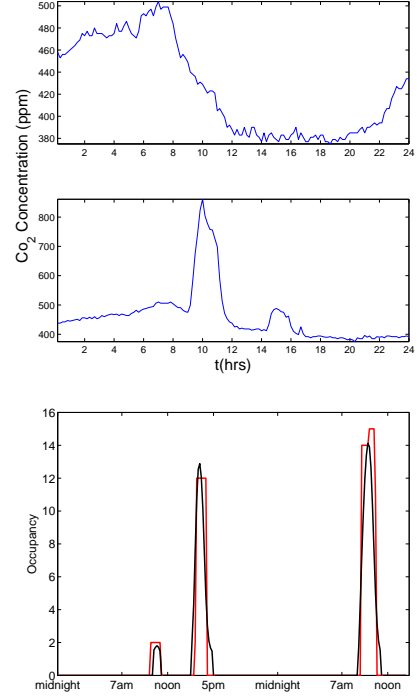

Fig. 6. Top panel shows CO_2 concentration with no occupancy, while the middle panel shows a typical profile on a weekday. Bottom panel shows the modeled occupancy (black line) and actual occupancy (red).

	Office	Small CR	Big CR
12 - 6am	1	0	0
6am - 9am	1	0	0
9am - 12pm	3	10	20
12pm - 1pm	2	5	10
1pm - 5pm	3	10	20
5pm - 6pm	2	5	10
6pm - 9pm	2	0	0
9pm - 12am	1	0	0

TABLE I

OCCUPANCY BOUNDS FOR DIFFERENT TYPES OF ROOMS (CR STANDS FOR CONFERENCE ROOMS).

rate bounds are set high enough such that it never constrains the problem.

C. Occupancy Estimation Results

In this section we compare the SUN estimator with a conventional estimator that relies solely on flow measurements. The accuracy of the two methods is assessed with respect to the ground truth occupancy distribution which was obtained manually by analyzing individual video frames and correcting them for under counts, over counts, and misdetects. In order to obtain building level occupancy ground truth, five peripheral video cameras (labeled as 1,2,3,5, and 9, see Figure 3) which monitor people traffic at entrance and exit locations were chosen. In addition two video people count sensors (7 and 8, see Figure 3) were used in the interior of building to assess the accuracy of occupancy estimates at
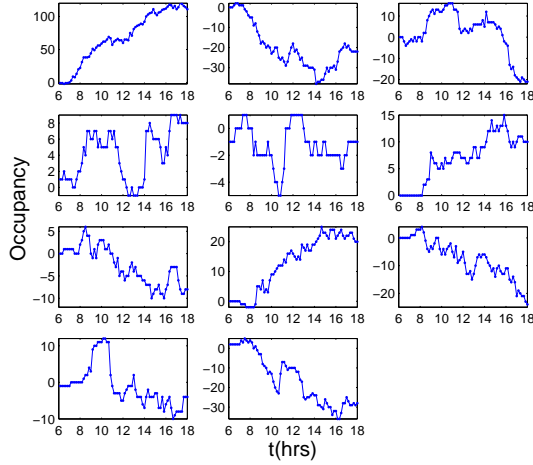
Fig. 7. Zone level occupancy estimates obtained from conventional people count estimator.
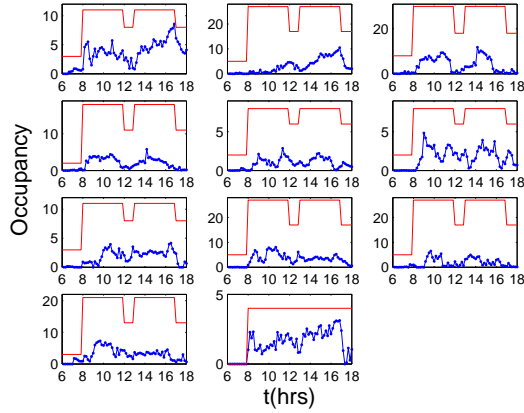

Fig. 8. Zonal level occupancy estimates obtained from SUN (blue). Also shown in red are the zonal occupancy bounds used during SUN estimation.

the zonal level. This zone was chosen, since it is one of the high traffic areas in the building and shows wide variability in occupancy levels during the course of the day. The estimation error was computed as the average % error, using

$$E = \frac{1}{T_f - T_0} \sum_{t=T_0}^{T_f} \frac{|x(t) - \widehat{x}(t)|}{x(t)}, \qquad (16)$$

where, $x(t)$ is the ground truth occupancy and, $\widehat{x}(t)$ is its estimate, with $T_0 = 6$ hrs corresponding to 6am and $T_f = 18$hrs corresponding to 6pm.

The conventional estimator is based on simple counting,

$$\widehat{x}_i(t+1) - \widehat{x}_i(t) = \sum_j \hat{r}_{ji}(t) - \sum_l \hat{r}_{il}(t), \qquad (17)$$

where, $\hat{r}_{ji}(t)$ are the estimates of people count from video cameras. The estimator is initialized to a zero occupancy level at 12am. Figure 7 shows occupancy estimates in the 11 zones based on this estimator over a typical weekday. The occupancy levels are either unexpectedly high at the
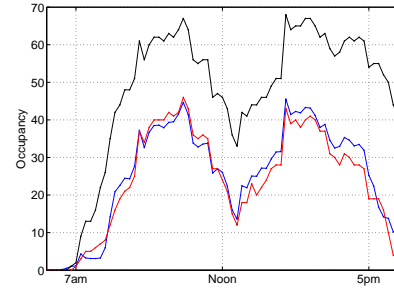

Fig. 9. Occupancy estimation at building level using, conventional people count estimator(black), SUN estimator (blue) and ground truth (red).
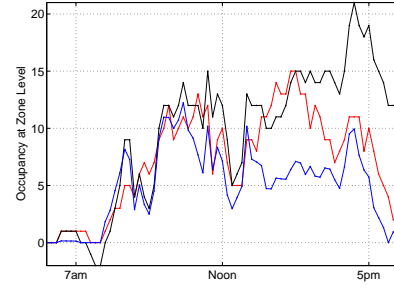

Fig. 10. Similar plot as above but at zonal level.

end of the workday in many zones (notably among them being the zone 1) or have non physical negative values. The lack of observability is evident from these plots: the sensors providing flow measurements that have a small positive bias that is integrated over the course of the day results in a massive error at 5:00 p.m. At building level (see Figure 9), the occupancy variation is more reasonable due to cancelation of mutual errors, although high occupancy at the end of the day is again unrealistic. The average estimation error based on this simplistic people count estimator is around 70%. Figure 10 shows a similar plot, but at zonal level (this zone is made up of zones 7 and 8, see Figure 3) with an estimation error of 30%.

In the SUN implementation we studied in detail the importance of different terms in the objective function (11). Not surprisingly, it was found that the best estimates are obtained by a suitable combination of the different terms, each of which model different source of information: we only report the final results. The zone level SUN estimates are shown in Figure 8, while the blue curve in Figure 9 shows the estimates at the building level. These estimates were obtained for 10 minute intervals by using a backward receding horizon of 1 hour, i.e. $T = 60$ minutes in objective (11). The quadratic program solver in MATLAB was used for numerical implementation of SUN algorithm. The initial estimates at 12am were taken to be zero both for the occupancy and the flow rate. The estimation error at building level is found to be approximately 11%, while that at zonal level is around 21% (see Figure 10). Inspite of the large bias and variance in the occupancy and flow measurements, the

SUN estimator yields occupancy estimates that are in good agreement with ground truth, especially at the building level.

IV. CONCLUSIONS

We introduced the sensor-utility-network (SUN) method to accurately estimate occupancy levels and distribution in buildings. Based on inputs from a variety of distributed sensor measurements (video, PIR, access control, and CO_2), along with historical data regarding building utilization, the SUN estimator produces occupancy estimates through the solution of a receding-horizon convex optimization problem. In our experiments we found that this approach reduced average occupancy estimation errors at building level from 70% obtained using conventional estimator, to 11%.

There are many challenges that remain to be addressed. These include evaluation of performance of SUN estimator for predictive use, adaptive techniques for learning building usage and associated utility functions, sensitivity of the impact of sensor placement and sensor types on occupancy estimation performance. Finally, to make SUN estimator scalable, decentralized algorithms should be developed.

V. ACKNOWLEDGEMENTS

REFERENCES

[1] H. Chen. Similarity-based analysis for large networks of ultra-low resolution sensors. Technical Report, 2006.

[2] Kun Deng, Wei Chen, Prashant G. Mehta, and Sean Meyn. Resource pooling for optimal evacuation of a large building. In *47th IEEE Conference on Decision and Control*, Cancun, Mexico, December 9-11 2008.

[3] T. M. Garwin, N. A. Pollard, and R. V. Tuohy, editors. *Project Responder: National Technology Plan for Emergency Response to Catastrophic Terrorism*. The National Memorial Institute for the Prevention of Terrorism / United States Department of Homeland Security, 2004.

[4] G. C. Goodwin, M. M. Seron, and J. A. De Doná. *Constrained control and estimation. An optimisation approach*. Communications and Control Engineering. Springer-Verlag, London, 2005.

[5] Xiaoshan Pan Kincho Law, Kenneth Dauber. Energy impact of commercial building controls and performance diagnostics: Market characterization, energy impact of building faults and energy savings potential. Technical Report TIAX Reference D0180, TIAX LCC for DOE, November 2005.

[6] D. J. C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003. available from http://www.inference.phy.cam.ac.uk/mackay/itila/.

[7] S. Meyn, Y. Lin, M. S. Oggianu, A. Surana, and I. Fedchenia. System and method for occupancy estimation and monitoring. *USPTO patent application under review*, 2008.

[8] S. P. Meyn. *Control Techniques for Complex Networks*. Cambridge University Press, Cambridge, 2007.

[9] National Fallen Firefighters Foundation. *Report of the National Fire Service Research Agenda Symposium*, Emmitsburg, MD, 2005.

[10] J.S. Niedbalski, P. G. Mehta, and S. P. Meyn. Model reduction for reduced order estimation in traffic models. In *IEEE American Control conference, Seattle, USA*, pages 914–919, June 11-13 2008.

[11] M. Schreckenberg and S.D. Sharma, editors. *Pedestrian and Evacuation Dynamics*. Springer, Berlin, 2002.

[12] J. M. Smith and D. Towsley. The use of queuing networks in the evaluation of egress from buildings. *Environment and Planning B: Planning and Design*, 8(2):125–139, 1981.

[13] Robert Tomastik, Satish Narayanan, Andrzej Banaszuk, and Sean Meyn. Model-based Real-Time Estimation of Building Occupancy During Emergency Egress. 4th International Conference on Pedestrian and Evacuation Dynamics. http://www.ped2008.com/, February 27–29 2008.

[14] Peng Wang, Peter B. Luh, Shi-Chung Chang, and Jin Sun. Modeling and optimization of crowd guidance in building emergency evacuation. In *Proceedings of the 2008 IEEE Conference on Automation Science and Engineering*, 2008.