

Biometrics Driven Smart Environments: Abstract Framework and Evaluation

Vivek Menon^{1**}, Bharat Jayaraman², and Venu Govindaraju²

¹ Amrita University, Coimbatore 641 105, India
vivek.menon@ettimadai.amrita.edu

² University at Buffalo, Buffalo, NY 14260, USA
bharat@cse.buffalo.edu, venu@cubs.buffalo.edu

Abstract. We present an abstract framework for ‘smart indoor environments’ that are monitored unobtrusively by biometrics capture devices, such as video cameras, microphones, etc. Our interest is in developing smart environments that keep track of their occupants and are capable of answering questions about the whereabouts of the occupants. We abstract the smart environment by a state transition system: Each state records a set of individuals who are present in various zones of the environment. Since biometric recognition is inexact, state information is probabilistic in nature. An event abstracts a biometric recognition step, and the transition function abstracts the reasoning necessary to effect state transitions. In this manner, we are able to accommodate different types of biometric sensors and also different criteria for state transitions. We define the notions of ‘precision’ and ‘recall’ of a smart environment in terms of how well it is capable of identifying occupants. We have developed a prototype smart environment based upon our proposed concepts, and provide experimental results in this paper. Our conclusion is that the state transition model is an effective abstraction of a smart environment and serves as a basis for integrating various recognition and reasoning capabilities.

1 Introduction

The vision of pervasive computing [1] provides the inspiration for a smart environment saturated with sensors, computing, and communication devices that are gracefully interfaced with human users [2]. In this paper, we focus on smart indoor environments such as homes, offices, etc. Indoor environments do not suffer from the problems of power or battery-life that confront outdoor environments. The goal of our research is to develop smart indoor environments that can identify and track their occupants as unobtrusively as possible and be capable of answering queries about the occupants. Such ‘context-aware’ systems can identify and track people in environments ranging from homes for elderly or

** This work was done while the author was a Visiting Research Scientist at the *Center for Unified Biometrics and Sensors*, University at Buffalo.

disabled, office workplace, department stores and shopping complexes to larger arenas such as airports, train stations, etc.

Identification approaches vary from tag-based approaches such as those involving RFID to those based on biometrics of the user. Tag-based methodologies tend to be obtrusive, requiring the individual to continuously retain them, however small the tag maybe. Some of the biometric techniques, such as fingerprint and iris scans, require a ‘pause-and-declare’ interaction with the human [3]. They are less natural than face, voice, height, and gait, which are less obtrusive and hence are better candidates for use in our smart environments.

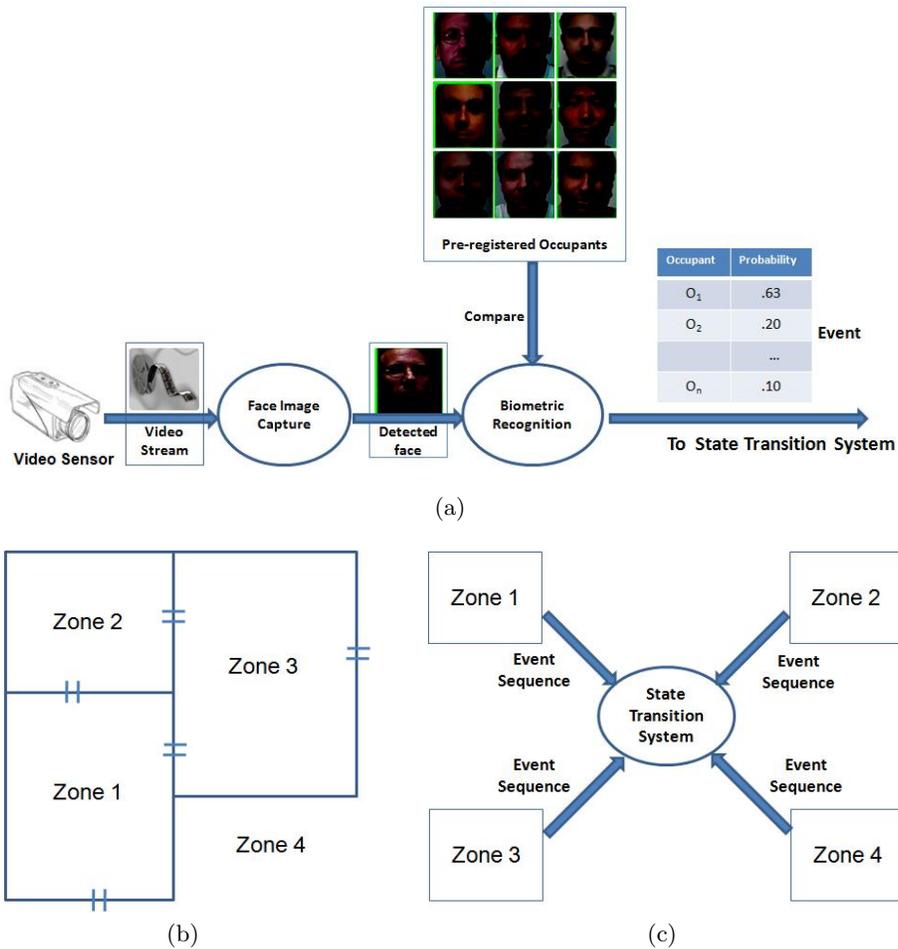


Fig. 1. Architecture of a Biometrics Driven Smart Environment

Figure 1¹ shows the overall architecture of a biometrics-driven smart environment. Although the figure illustrates a single biometric modality of face recognition, the architecture is also applicable to other biometric modalities. For example, zone 1 might use voice recognition, zone 2 might use face recognition, and zone 3 might use height estimation. However, in all cases the output of a biometric recognition is the set of person probability pairs as discussed in more detail below.

The main contribution of this paper is a framework for abstracting the behavior of a biometrics-driven smart environment in terms of a state transition system. Our proposed framework supports multiple biometric modalities in a uniform manner and facilitates a precise statement of the performance aspects of a smart environment.

The state of an environment is expressed in terms of the probabilities of the occupants being present in the different zones of the environment. The state information is probabilistic because a biometric recognizer typically provides a set of scores indicating the degree of match between the subject and the candidates in the database. Therefore, in our approach an event abstracts a biometric recognition step - whether it is face recognition, voice recognition, etc. - and is represented as a set of pairs $\langle o, p(o) \rangle$ where $p(o)$ is the probability² that occupant o has been recognized at this event.

The transition function abstracts the reasoning necessary to effect state transitions. Effectively, the transition function takes as input, a state and an event, and determines the next state by assigning revised probabilities to the occupants in the environment based upon the probabilities in the event. In this manner, we are able to accommodate different types of biometric sensors and also different criteria for state transitions, including those that incorporate declarative knowledge of the individuals and the environment. It is not necessary for us to consider non-deterministic transitions, since a state itself is represented as a set of occupants and their probabilities.

We introduce the concepts of precision and recall in order to provide a quantitative measure of the performance of a smart environment. Precision captures how well an occupant is recognized, while recall captures whether an occupant is recognized at all. These are complementary concepts and together capture the overall performance of a smart environment. The concepts of precision and recall are standard performance measures in the information retrieval literature [6], but we have adapted the definitions to suit our context.

The rest of this paper is organized as follows. The related work is presented in Section 2 while the details of our model are discussed in Section 3. The experimental prototype is described in Section 4 and the conclusions and the future work are presented in Section 5.

¹ Face images blurred to preserve anonymity.

² Such scores can be mapped to probabilities by running the recognizer on a large number of samples, as shown in [4, 5].

2 Related Work

There has been considerable interest in the subject of smart environments. A major difference between our proposed approach and several of the approaches surveyed below is our use of a framework in which the details of biometric recognition are abstracted simply as events in a state transition system. We also characterize the performance of the smart environment in terms of the concepts of precision and recall. We briefly survey closely related efforts below and highlight their main features.

The survey paper by Cook and Das [7] provides a good account of the state-of-the-art in smart environments. There are several projects that focus on smart homes (MavHome [8], the Intelligent Home [9], the House_n [10] and development of adaptive control of home environments by also anticipating the location, routes and activities of the inhabitants.

2.1 Location Estimation

An extensive survey and taxonomy of location systems for a ubiquitous computing application is discussed in [11] while [12] provides a more recent survey of position location techniques in mobile systems and draws a comparison between their parameters. Fox et al. [13] highlight the relevance of Bayesian filtering techniques in dealing with the uncertainty characteristic of sensors in pervasive computing environments and apply them in the context of estimation of an object's location. Data Association problem [14] is a pertinent issue in scenarios which involve tracking of multiple people using anonymous sensors. The ambiguity in identity estimation arising due to the absence of ID sensors and the corresponding tags on people is difficult to resolve. Thus identity estimation, absolute or relative is a desirable and integral subsystem of any object tracking mechanism.

2.2 Tracking

Bui et al. [15] propose an Abstract Hidden Markov Model(AHMM) based approach for tracking human movement in an office like spatial layout and to predict the object trajectories at different layers of detail. For location tracking in a single inhabitant smart space, Roy et al. [16] propose an optimal algorithm based on compressed dictionary management and online learning of the inhabitant's mobility profile. In a related work [17], they highlight the complexity of optimal location prediction across multiple inhabitants in a smart home and propose a stochastic game-theoretic approach, which learns and estimates the inhabitants' most likely location profiles. This study uses RFID tags to track the inhabitants. A combination of particle filters with Kalman filters to track multiple objects with accuracy of anonymous sensors and identification certainty of id-sensors is discussed in [13, 18]. Krumm et al. [19] investigate the nuances of visual person tracking in intelligent environments in the context of the EasyLiving project [20] by deploying multiple cameras for tracking multiple people.

However this tracking experiment was within the confines of a single room and the identity estimation and maintenance for facilitating the tracking only dealt with non-absolute, internally system generated identity of tracked persons.

2.3 Biometrics in Smart Environments

Pentland and Choudhury [3] highlight the importance of deploying audio-and-video based recognition systems in smart environments as these are modalities similar to those used by humans for recognition. They summarize the face recognition efforts and discuss various commercial systems and applications as well as its novel applications in smart environments and wearable computing. Chen and Gellersen [19] propose a new method to support awareness based on fusion of context information from different sources in a work environment which included integration of audio and video sources with more specific environment sensors and with logical sensors that capture formal context. Reasoning over the context information generated by applying different perception techniques on the raw data collected is used to generate a refined context.

Hamid Aghajan et al. [22] propose a vision-based technology coupled with AI-based algorithms for assisting vulnerable people and their care givers in a smart home monitoring scenario. However, users are expected to wear a wireless identification badge that broadcasts a packet upon sensing a significant signal by one of the accelerometers. Gao et al. [23] propose a new distance measure for authentication in their face recognition system for a ubiquitous computing environment which relies on a fusion of multiple views of each person. Their work focuses on optimizing a single modality to improve robustness rather than deploying a multimodal approach that fuses different biometrics. Hong et al. [24] discuss structure, operation and performance of a face verification system using Haar-like features and HMM algorithm in a ubiquitous network environment. Zhang et al. [25] propose a distributed and extensible architecture of a continuous verification system that verifies the presence of logged-in user. A Bayesian framework that combines temporal and modality information holistically is used to integrate multimodal passive biometrics which includes face and fingerprint.

Driven by the proliferation of commercially available hand-held computers, Hazen et al. [26] research the improvements in identification performance by adopting a bimodal biometric approach to user identification, for use on mobile devices by integrating audio and visual biometric information in the form of voice and face. They also report significant improvements in identification performance that can stem from using dynamic video information instead of static image snapshots on a database of 35 different people. Highlighting the privacy concerns of video recording discussed in Bohn et al [27] and reliability issues of face recognition techniques for user authentication, Vildjiounaite et al. [28], deploy a fusion of accelerometer based gait recognition and speaker recognition as an unobtrusive and marginally privacy-threatening means of user authentication with personal mobile devices. They have reported improved performance in the combination mode as opposed to individual modalities for a user base of 31. More recently Bernardin and Stiefelhagen [29] have implemented a system for

the simultaneous tracking and incremental multimodal identification of multiple users in a smart environment which fuses person track information, localized speaker ID and high definition visual ID cues opportunistically to gradually refine the global scene model and thus increase the system’s confidence in the set of recognized identities.

In terms of the focus on the use of non-obtrusive biometrics based recognition and location estimation, our work is similar to [29]. However, in our research, we propose an abstract framework where in a variety of biometric modalities can be incorporated in a uniform manner. Our approach to identity estimation deals with the absolute identity of people across multiple zones of a facility. However, we attempt to highlight the inherent uncertainty of automated face recognition by recasting the eigen distances generated by eigenface algorithm into a probability distribution of the registered faces, instead of the conventional approach of assigning the value with the least eigen distance as the matching face. This probabilistic approach to biometric recognition is a key theme around which we construct our abstract framework for a biometrics driven smart environment.

2.4 State Space Representation

It might appear that a Hidden Markov Model(HMM) would serve as an elegant basis for representing the state space. From a HMM perspective, a smart environment with n occupants and m zones can have m^n distinct possible states. Thus probabilities are not associated with the states but with the transitions between them; these transition probabilities are to be learnt from past behavior or by simulation [15]. Thus an HMM approach is computationally more complex due to a state space explosion and the requirement of *a priori* probabilities of trajectories. In our approach, the size of a state is $m * n$, meaning that for each of the m zones we record the probabilities of each of the n occupants being present in that zone. Therefore, in Section 4 (Evaluation), we depict a typical state as a table with n rows and m columns. The transitions from one state to another are deterministic. Therefore, given any event in a zone, the next state is unambiguously determined. In contrast with the HMM approach, we do not need to learn the transition probabilities in order to determine the next state because biometric recognition (or event) provides a direct means for effecting state transitions. Our state transition model is discussed in the next section.

3 Framework

Definition (Smart Environment): *An n -person smart environment is abstracted as a state transition system (S, E, Δ) where S is the set of states labeled $s_0, s_1, \dots s_x$; E is the set of events labeled $e_1, e_2, \dots e_x$ and $\Delta : S \times E \rightarrow S$ is a function that models the state transition on the occurrence of an event. The state transitions may be depicted as follows:*

$$s_0 \xrightarrow{e_1} s_1 \xrightarrow{e_2} s_2 \dots \xrightarrow{e_x} s_x$$

We shall consider a smart environment as being divided into a number of zones, each of which may be a region (or a set of rooms). We include two special zones, an external zone and a transit zone, for the sake of convenience.

Definition (State): Given n occupants, $o_1 \dots o_n$ and m zones labeled $1 \dots m$, a state s_k of the environment is represented by an m -tuple $\langle Z_{1k} \dots Z_{mk} \rangle$ where for $1 \leq j \leq m$, $Z_{jk} = \{ \langle o_i, p_{jk}(o_i) \rangle : 1 \leq i \leq n \}$. Also, in each state s_k and for each occupant o_i , $\sum_{i=1}^m p_{jk}(o_i) = 1$.

The state of an environment is expressed in terms of the probabilities of the occupants being present in the different zones of the environment. The constraint $\sum_{i=1}^m p_{jk}(o_i) = 1$ indicates that sum of probabilities of any occupant being present across all the zones in any state equals one. In the initial state s_0 , we may assume without loss of generality that all occupants are in the external zone with probability 1. Given a smart environment with n occupants, m zones, and x number of events, the total size of the state space is $m * n * (x + 1)$. Thus, the size of the state space is quadratic in m and n rather than exponential, as in HMMs. In this paper we model all exit events as entry events into a transit zone. Hence it suffices in our model to only consider entry events. An event is essentially an abstraction of a biometric or feature recognition step performed in the environment.

Definition (Event): Given n occupants $o_1 \dots o_n$, an (entry) event e_k occurring at zone j ($1 \leq j \leq m$) at time t is represented as $\langle t, j, P \rangle$, where $P = \{ \langle o_i, p_{jk}(o_i) \rangle : 1 \leq i \leq n \}$ and $p_{jk}(o_i)$ is the probability that an occupant o_i was recognized at zone j in event e_k .

As noted earlier, an event is an abstraction of a recognition step. For simplicity, we assume that events happen sequentially in time, i.e., simultaneous events across different zones are ordered arbitrarily in time. That is, the entry of an occupant o_i into zone z_i and occupant o_j to zone z_j at the same time t can be modeled as o_i before o_j or o_j before o_i .

Definition (Transition Function): $\Delta : S \times E \rightarrow S$, maps state s_{k-1} into state s_k upon an event $e_k = \langle t, j, P \rangle$ occurring at time t in zone j , where $P = \{ \langle o_i, p_{jk}(o_i) \rangle : 1 \leq i \leq n \}$. Let $s_{k-1} = \langle Z_{1k-1} \dots Z_{jk-1} \dots Z_{mk-1} \rangle$ and $Z_{jk-1} = \{ \langle o_i, p_{jk-1}(o_i) \rangle : 1 \leq i \leq n \}$. Then Δ determines state $s_k = \langle Z_{1k} \dots Z_{jk} \dots Z_{mk} \rangle$ as follows: Let $x_i = 1 - p_{jk}(o_i)$. Then,

$$Z_{jk} = \{ \langle o_i, p_{jk}(o_i) + x_i * p_{jk-1}(o_i) \rangle : 1 \leq i \leq n \} \quad (1)$$

$$Z_{lk} = \{ \langle o_i, x_i * p_{lk-1}(o_i) \rangle : 1 \leq i \leq n \}, \text{ for } 1 \leq l \leq m \text{ and } l \neq j \quad (2)$$

The transition function maps a state s_{k-1} to a state s_k upon an event e_k occurring at zone j . For zone j , we sum the new probability $p_{jk}(o_i)$ for an occupant generated by event e_k with the complement of the new probability value $1 - p_{jk}(o_i)$, apportioned by a factor of the existing probability $p_{jk-1}(o_i)$. In the event of a revision, there might be a violation of the constraint that the sum of probabilities for any occupant across all zones equals one ($\sum_{i=1}^m p(o_i) = 1$). To

restore adherence to this constraint, for each occupant o_i , we apportion to the probability of o_i being present in each zone $l \neq j$ by redistributing the complement of the new probability value, $1 - p_{jk}(o_i)$, in the ratio of the probability value in existing state $p_{lk-1}(o_i)$. This transition function ensures that the probability values associated with a new event as well as the current state figure in the determination of the new state as in a Markov process.

Since we are dealing with a closed environment with a fixed set of occupants, $o_1 \dots o_n$, we can, in general, utilize *a priori* declarative knowledge regarding the occupants, such as their schedules or knowledge of the environment, such as the distance between rooms and whether an occupant could move between a pair of rooms within a certain interval of time. However, the transition function presented in the above definition does not make any use of such declarative knowledge of the occupants. The nature of the declarative knowledge can also be fuzzy to factor in the probabilistic nature of the environment. The reasoning component can alleviate some of the weaknesses in the recognition component. However it should ensure that its results are not totally contradictory to the recognition system thereby generating inconsistencies.

We now define the concepts of precision and recall for a smart environment. These are defined in terms of the ground truth, which, for a given input event sequence, is a sequence of states of the environment wherein the presence or absence of any occupant in any zone is known with certainty (0 or 1). Precision captures how well an occupant is recognized, while recall captures whether an occupant is recognized at all.

Definition (Ground Truth): *Given n occupants $O = \{o_1 \dots o_n\}$ and an event sequence $e_1 \dots e_x$, then the ground truth is the sequence of states $g_1 \dots g_x$ where each $g_k = \langle T_{1k} \dots T_{jk} \dots T_{mk} \rangle$ and $T_{jk} = \{ \langle o_i, q_{jk}(o_i) \rangle : 1 \leq i \leq n \wedge q_{jk}(o_i) \in \{0, 1\} \}$. Also, $\langle o_i, 1 \rangle \in T_{jk} \rightarrow \langle o_i, 1 \rangle \notin T_{lk}$, for all $l \neq j$ in state g_k .*

Given a zone of a state, the precision for that zone of the state is defined as the average probability of those occupants that are present in that zone of the state as given in the ground truth. The average precision across all zones (where at least one occupant is present as per the ground truth) is the precision for the state, and the average precision across all states is the precision for a given ground truth. Finally, the average across multiple ground truths is the precision of the smart environment.

Definition (Precision): *Given an environment with m zones, n occupants $O = \{o_1 \dots o_n\}$, an event sequence $E = e_1 \dots e_x$, a ground truth $G = g_0, g_1, \dots g_x$, and state transitions $S = s_0, s_1, \dots s_x$. We define the precision, π , with respect to G as follows:*

Let $\pi_{jk} = a_{jk}/b_{jk}$, where

$$a_{jk} = \sum \{ p_{jk}(o_i) : 1 \leq i \leq n \wedge q_{jk}(o_i) = 1 \}$$

$$b_{jk} = |\sum\{o_i : 1 \leq i \leq n \wedge q_{jk}(o_i) = 1\}|$$

Then $\pi_k = \sum_{j=1}^m \pi_{jk}/m$, and we define $\pi = \sum_{k=1}^x \pi_k/x$.

Now, given a set of ground truths $\{G^1, G^2, \dots, G^t\}$ with the corresponding precisions $\{\pi^1, \pi^2, \dots, \pi^t\}$, the precision of the smart environment, $\Pi = \sum_{l=1}^t \pi^l/t$.

For a given ground truth, state and zone, we define recall with respect to a threshold θ as the ratio a/b , where a is the number of occupants of that zone with probabilities greater than θ and who are present in the ground truth, and b is the number of occupants who are present in the ground truth for that zone. The recall for a state is the average of the probabilities across all zones where at least one occupant is present as per the ground truth. The average recall across all states is the recall for a given ground truth, and the average across multiple ground truths is the recall of the smart environment.

Definition (Recall): Given an environment with m zones, n occupants $O = \{o_1 \dots o_n\}$, an event sequence $E = e_1 \dots e_x$, a ground truth $G = g_0, g_1 \dots g_x$, and state transitions $S = s_0, s_1, \dots, s_x$. We define the recall, ρ , with respect to a threshold θ as follows:

Let $\rho_{jk} = a_{jk}/b_{jk}$, where

$$a_{jk} = |\sum\{o_i : 1 \leq i \leq n \wedge q_{jk}(o_i) = 1 \wedge p_{jk}(o_i) > \theta\}|$$

$$b_{jk} = |\sum\{o_i : 1 \leq i \leq n \wedge q_{jk}(o_i) = 1\}|$$

Then $\rho_k = \sum_{j=1}^m \rho_{jk}/m$, and we define $\rho = \sum_{k=1}^x \rho_k/x$.

Now, given a set of ground truths $\{G^1, G^2, \dots, G^t\}$ with the corresponding precisions $\{\rho^1, \rho^2, \dots, \rho^t\}$, the recall of the smart environment, $R = \sum_{l=1}^t \rho^l/t$.

As it is clear, the recall is inversely proportional to the threshold, θ , since lowering the threshold will result in more occupants being identified. This figure is generally arrived at experimentally for a smart environment. A reasonable choice of θ is 0.5, and this is also the value that we adopt in our experiments. In the above definition, the recall was defined zone-wise. An alternative approach is to disregard the zones while taking the ratio; doing so would increase the overall recall. Our definition gives due importance to zones, and hence is a relatively more conservative.

4 Evaluation

We have developed an experimental prototype embodying the ideas presented in this paper. Figure 1(b) illustrates a 4-zone smart environment (fourth zone

representing the external zone) with 25 occupants who are monitored by video cameras installed in each of the zones. Our experimental prototype collects sample face images of the 25 occupants of an office facility and pre-registers them in a training database. The image sensors deployed in each zone detects the presence of occupants as they move through the zones and verify the face images extracted from the video against the database. The distance scores generated by eigenface is recast into a probability value [4, 5] which denotes a posterior probability of the detected face matching the pre-registered occupants. This set of person-probability pairs generated essentially constitutes an event as defined in Section 3.

Although our abstract framework is independent of the details of any particular biometric modality, we illustrate our concepts in terms of face recognition. Automated face recognition is yet to attain any comparable levels of robustness as that of humans. Factors such as viewing angle, distance, background clutter, lighting spectrum, intensity, angle and diffuseness of lighting, differences between posed photographs and spontaneous expression can cause fluctuations in the performance of computer vision based on statistical classifiers [30]. Our prototype is based upon OpenCV's [31] implementation of the eigenface algorithm [32], which provides a basis for a simple though not robust implementation of face recognition.

This prototype can be extended to incorporate other biometric recognizers in a similar manner. For example, for voice recognition, the voice samples of occupants are pre-registered in a database instead of face image samples. The ambient voice that is picked up by the voice sensors installed in different zones can be matched against the voice database to generate a set of person-probability pairs. In this manner the different biometric recognizers are interfaced in a uniform way with the rest of the system.

We illustrate with figures 2, 3, and 4, the computation carried out by the state transition system that is embodied in the 4-zone smart environment. We have presented the observations and results obtained in a tabular form for ease of understanding. Figure 2 presents 10 events involving four of the occupants of the smart environment. Each event is presented as a column of probability values - these form the output of the face recognition module as it matches a given face against the training database. Shown in boldface are the probability values corresponding to the actual occupants who were involved in the corresponding events, as per the ground truth. Italicized values indicate the probabilities corresponding to the occupants who were not involved in the corresponding events, but falsely assigned values by the recognizer. This ambiguity may arise due to any of the reasons already discussed above.

Figure 3 illustrates a sample transition from state s_8 to s_9 upon the occurrence of event e_9 in zone 1. The probabilities of the occupants in zone 1 in state s_9 are obtained as per equation (1) defined under the transition function definition. For the remaining zones of s_9 , the probability values are obtained as per equation (2) defined under the transition function.

	Events and corresponding Zones									
	Event 1	Event 2	Event 3	Event 4	Event 5	Event 6	Event 7	Event 8	Event 9	Event 10
Occupants	Zone 1	Zone 2	Zone 3	Zone 1	Zone 2	Zone 1	Zone 2	Zone 2	Zone 1	Zone 3
O1	0.66	0.81	0.47	0.00	0.00	0.00	0.71	0.00	0.03	0.00
O2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.63	0.50
O3	0.01	0.00	0.00	0.00	0.00	0.79	0.00	0.72	0.00	0.01
O4	0.02	0.07	0.00	0.00	0.00	0.00	0.11	0.01	0.00	0.00
O5	0.02	0.12	0.00	0.00	0.00	0.00	0.13	0.02	0.04	0.44
O6	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.02	0.00
O7	0.06	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.03	0.00
O8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
O9	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.14	0.01	0.01
O10	0.00	0.00	0.51	0.72	0.52	0.01	0.00	0.00	0.20	0.00
O11	0.00	0.00	0.00	0.00	0.35	0.00	0.00	0.00	0.00	0.00
O12	0.10	0.00	0.00	0.23	0.00	0.00	0.00	0.00	0.00	0.00
O13	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.02	0.00	0.00
O14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
O15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
O16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
O17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
O18	0.00	0.00	0.00	0.01	0.03	0.00	0.00	0.00	0.00	0.00
O19	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
O20	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
O21	0.04	0.00	0.00	0.00	0.00	0.07	0.01	0.04	0.00	0.00
O22	0.00	0.00	0.02	0.04	0.06	0.01	0.00	0.00	0.00	0.00
O23	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
O24	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.02	0.00
O25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.02
Ground Truth	O1	O1	O1	O10	O10	O3	O1	O3	O2	O2

Fig. 2. Event Sequence

Figure 4 illustrates the results of the precision and recall for the ground truth corresponding to the event sequence of figure 2. The low values for precision at zone 3, corresponding to states s_3 and s_{10} in particular, can be traced to the ambiguity arising in the face recognition step at events e_3 and e_{10} , both occurring at zone 3 which results in a low probability of recognition of occupants o_1 and o_2 at these events respectively. For the same reason, the values for recall at zone 3 also suffers, thereby affecting the average recall of the states $s_3 \dots s_6$.

5 Conclusions

We have presented a novel framework for non-obtrusive biometric based indoor smart environments for identification and tracking.

1. A state transition framework in which events abstract different biometric recognition steps and transitions abstract different reasoning steps.

State	S8				Event 9 at	S9			
Occupant	Zone 1	Zone 2	Zone 3	Zone 4	Zone 1	Zone 1	Zone 2	Zone 3	Zone 4
O1	0.02	0.84	0.13	0.01	0.03	0.05	0.82	0.13	0.01
O2	0.00	0.00	0.00	1.00	0.63	0.63	0.00	0.00	0.37
O3	0.22	0.72	0.00	0.06	0.00	0.22	0.72	0.00	0.06
O4	0.02	0.17	0.00	0.81	0.00	0.02	0.17	0.00	0.81
O5	0.01	0.25	0.00	0.74	0.04	0.05	0.24	0.00	0.71
O6	0.00	0.04	0.00	0.96	0.02	0.02	0.04	0.00	0.95
O7	0.06	0.02	0.00	0.93	0.03	0.08	0.01	0.00	0.90
O8	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
O9	0.06	0.14	0.00	0.81	0.01	0.06	0.14	0.00	0.80
O10	0.35	0.51	0.07	0.06	0.20	0.49	0.41	0.05	0.05
O11	0.00	0.35	0.00	0.65	0.00	0.00	0.35	0.00	0.65
O12	0.30	0.00	0.00	0.70	0.00	0.30	0.00	0.00	0.70
O13	0.03	0.02	0.00	0.95	0.00	0.03	0.02	0.00	0.95
O14	0.00	0.00	0.00	0.99	0.00	0.01	0.00	0.00	0.99
O15	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
O16	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
O17	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
O18	0.01	0.03	0.00	0.97	0.00	0.01	0.03	0.00	0.97
O19	0.00	0.02	0.00	0.98	0.00	0.00	0.02	0.00	0.98
O20	0.08	0.00	0.00	0.92	0.00	0.08	0.00	0.00	0.91
O21	0.10	0.04	0.00	0.86	0.00	0.10	0.04	0.00	0.86
O22	0.05	0.07	0.02	0.86	0.00	0.05	0.07	0.02	0.86
O23	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.99
O24	0.02	0.04	0.00	0.94	0.02	0.03	0.04	0.00	0.92
O25	0.00	0.00	0.00	1.00	0.03	0.03	0.00	0.00	0.97
Ground Truth	O1, O3, O10				O2	O1, O2, O3, O10			

Fig. 3. Sample State Transition

Precision				
State	Zone 1	Zone 2	Zone 3	Average
S1	0.66	-	-	0.66
S2	-	0.81	-	0.81
S3	-	-	0.47	0.47
S4	0.72	-	0.47	0.59
S5	-	0.52	0.47	0.49
S6	0.79	0.51	0.47	0.59
S7	0.79	0.67	-	0.73
S8	-	0.69	-	0.69
S9	0.63	0.65	-	0.64
S10	-	0.65	0.50	0.57
			Overall	0.62

(a)

Recall				
State	Zone 1	Zone 2	Zone 3	Average
S1	1	-	-	1.00
S2	-	1	-	1.00
S3	-	-	0	0.00
S4	1	-	0	0.50
S5	-	1	0	0.50
S6	1	1	0	0.67
S7	1	1	-	1.00
S8	-	1	-	1.00
S9	1	0.67	-	0.83
S10	-	0.67	1	0.83
			Overall	0.73

(b)

Fig. 4. Illustrating Precision and Recall

2. A characterization of the performance of the smart environment in terms of the concepts of precision and recall.

Our state transition system is fundamentally probabilistic because the biometric recognition that underlies events is inexact in nature. Our formulation of precision and recall succinctly characterizes the performance of a smart environment. We believe that our state transition model is an effective abstraction of a smart environment and serves as a basis for integrating different recognition and reasoning capabilities.

In our model, a state provides location information at a zone level and a sequence of consecutive states implicitly contains zone level tracking information for all occupants. While it is possible to define precision and recall with respect to any query of interest, we have formulated them in a query independent manner which we believe is more general. Our model is capable of supporting spatial and temporal queries, such as: the location of an occupant in the facility; time of entry/exit of an occupant; the first/last person to enter/leave the facility; the current occupants present in the facility; etc. A query of special interest is tracking of individuals over time and space.

We plan to enhance our current prototype by incorporating a variety of biometric recognizers, such as for height, gait, voice, etc. It is possible to fuse two or more biometric modalities to enhance the overall performance of recognition. We also plan to incorporate spatio-temporal reasoning based upon declarative knowledge of the environment as well as the occupants. Through such enhancements in recognition and reasoning, we can improve the overall precision and recall of the smart environment. We plan to test this approach on larger environments and support speech-based information retrieval queries about the environment.

Acknowledgments

Thanks to Philip Kilinskas for his help in developing the experimental prototype; Dr. Jason J. Corso for discussions on Markov models; and members of the *Center for Unified Biometrics and Sensors* for their comments and suggestions.

References

1. Weiser, M.: The Computer for the 21st Century. *Scientific American* 265(3), 66–75 (1991)
2. Satyanarayanan, M.: Pervasive Computing: Vision and Challenges. *IEEE Personal Communications* 8(4), 10–17 (2001)
3. Pentland, A., Choudhury, T.: Face Recognition for Smart Environments. *IEEE Computer* 33(2), 50–55 (2000)
4. Cao, H., Govindaraju, V.: Vector Model Based Indexing and Retrieval of Handwritten Medical Forms, In: Proc. of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), pp. 88–92. IEEE Computer Society, Washington, DC (2007)

5. Bouchaffra, D., Govindaraju, V., Srihari, S.: A Methodology for Mapping Scores to Probabilities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 21(9), 923–927 (1999)
6. van Rijsbergen, C.J.: *Information Retrieval*. London: Butterworths, Boston (1979)
7. Cook, D. J., Das, S. K.: How smart are our environments? An updated look at the state of the art. *Pervasive and Mobile Computing* 3(2), 53–73 (2007)
8. Youngblood, M., Cook, D. J., Holder, L. B.: Managing adaptive versatile environments. In: *Proc. of the Third IEEE Intl. Conf. on Pervasive Computing and Communications*, pp. 351–360. IEEE Computer Society, Washington, DC (2005)
9. Lesser, V. et al.: The Intelligent Home Testbed. In: *Proc. of the Autonomous Agents 1999 Workshop on Autonomy Control Software*. ACM, New York (1999)
10. *House n Living Laboratory Introduction* (2006)
11. Hightower, J., Borriello, G.: Location Systems for Ubiquitous Computing. *IEEE Computer* 34(8), 57–66 (2001)
12. Manesis, T., Avouris, N.: Survey of position location techniques in mobile systems. In: *Proc. of the Seventh Intl. Conf. on Human Computer interaction with Mobile Devices and Services. MobileHCI '05*, pp. 291–294, ACM, New York (2005)
13. Fox, D., Hightower, J., Liao, L., Schulz, D., Borriello, G.: Bayesian filtering for location estimation. *IEEE Pervasive Computing* 2(3), 24–33 (2003)
14. Bar-Shalom, Y., Li, X.-R.: *Multitarget-Multisensor Tracking: Principles and Techniques*, Yaakov Bar-Shalom (1995)
15. Bui, H. H., Venkatesh, S., West, G.: Tracking and surveillance in wide-area spatial environments using the Abstract Hidden Markov Model. *Intl. Journal of Pattern Recognition and Artificial Intelligence* 15(1), 177–195 (2002)
16. Roy, A., Bhaumik, S., Bhattacharya, A., Basu, K., Cook, D. J., Das, S. K.: Location aware resource management in smart homes. In: *Proc. of the First IEEE Intl. Conf. on Pervasive Computing and Communications*, pp. 481–488. IEEE Computer Society, Washington, DC (2003)
17. Das, S. K., Roy, N., Roy, A.: Context-aware resource management in multi-inhabitant smart homes: A framework based on Nash H-learning. *Pervasive and Mobile Computing* 2(4), 372–404 (2006)
18. Schulz, D., Fox, D., Hightower, J.: People tracking with anonymous and id-sensors using Rao-Blackwellised particle filters. In: *Proc. of the 18th Intl. Joint Conference on Artificial Intelligence (IJCAI)*, pp. 921–926 (2003)
19. Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M., Shafer, S.: Multi-Camera Multi-Person Tracking for EasyLiving. In: *Proc. of the 3rd IEEE Intl. Workshop on Visual Surveillance (VS'2000)*. IEEE Computer Society, Washington, DC. (2000)
20. Brumitt, B., Meyers, B., Krumm, J., Kern, A., Shafer, S. A.: EasyLiving: Technologies for Intelligent Environments. In: Thomas, P.J., Gellersen, H. (eds.) *Proc. 2nd Intl. Symposium on Handheld and Ubiquitous Computing, LNCS vol. 1927*. pp. 12–29, Springer-Verlag, London (2000)
21. Chen, D., Gellersen, H.: Recognition and reasoning in an awareness support system for generation of storyboard-like views of recent activity. In: *Proc. of the Intl. ACM SIGGROUP Conference on Supporting Group Work. GROUP '99*, ACM Press, New York (1999)
22. Aghajan, H. et al.: Distributed Vision-Based Accident Management for Assisted Living. In: Okadome, T., Yamazaki, T., Mokhtari, M. (eds.) *ICOST 2007, LNCS vol. 4541*. pp. 196–205, Springer, Heidelberg (2007)
23. Gao, Y., Hui, S. C., Fong, A. C.: A Multi-View Facial Analysis Technique for Identity Authentication. *IEEE Pervasive Computing* 2(1), 38–45 (2003)

24. Hong, K. et al.: Real Time Face Detection and Recognition System Using Haar-Like Feature/HMM in Ubiquitous Network Environments. In: Gervasi, O. et al. (eds.) ICCSA 2005, LNCS vol. 3480. pp. 1154-1161, Springer, Heidelberg (2005)
25. Zhang, S. et al.: Continuous Verification Using Multimodal Biometrics. In: Zhang, D., Jain, A.K. (eds.) ICB 2006, LNCS vol. 3832, pp. 562-570, Springer, Heidelberg (2005)
26. Hazen, T., Weinstein, E., Heisele, B., Park, A., Ming, J.: Multi-Modal Face and Speaker Identification for Mobile Devices. In: Hammoud, R. I., Abidi, B. R., Abidi, M. A. (eds.) Face Biometrics for Personal Identification: Multi-Sensory Multi-Modal Systems, pp. 123-138, Springer, Heidelberg (2006)
27. Bohn, J., Coroama, V., Langheinrich, M., Mattern, F., Rohs, M.: Social, Economic, and Ethical Implications of Ambient Intelligence and Ubiquitous Computing, In: Weber, W., Rabaey, J., Aarts, E. (eds.) Ambient Intelligence, pp. 5-29, Springer, Heidelberg (2005)
28. Vildjiounaite, E., et al.: Unobtrusive Multimodal Biometrics for Ensuring Privacy and Information security with Personal Devices. In: Fishkin, K.P., Schiele, B., Nixon, P., Quigley, A.J. (eds.) Pervasive 2006. LNCS vol. 3968. pp. 186-201. Springer, Heidelberg (2006)
29. Bernardin, K., Stiefelhagen, R.: Audio-visual multi-person tracking and identification for smart environments. In: Proc. of the 15th International Conference on Multimedia, pp. 661-670. ACM, New York (2007)
30. Hewitt, R., Belongie, S.: Active Learning in Face Recognition: Using Tracking to Build a Face Model. In: Proc. of the 2006 Computer Vision and Pattern Recognition Workshop, pp. 157-157. IEEE Computer Society, Washington, DC (2006)
31. OpenCV, <http://www.intel.com/technology/computing/opencv/index.htm>
32. Turk, M., Pentland, A.: Eigenfaces for Recognition. *J. Cognitive Neuroscience* 3(1), 71-86 (1991)