

INTEGRATING RECOGNITION AND REASONING IN SMART ENVIRONMENTS

Vivek Menon, Bharat Jayaraman, Venu Govindaraju

Amrita University, Coimbatore 641 105, India

vivek_menon@ettimadai.amrita.edu

University at Buffalo, Buffalo, NY 14260, USA

bharat@cse.buffalo.edu, venu@cubs.buffalo.edu

Keywords: Smart environments, Biometrics, Recognition, Spatio-Temporal Reasoning, Precision, Recall.

Abstract

Our goal is to develop ‘smart indoor environments’ that are monitored unobtrusively by biometric capture devices, such as video cameras, microphones, etc. Such environments will keep track of their occupants and be capable of answering queries about the occupants’ whereabouts. In order to develop a unified model that is applicable across diverse biometric modalities, we propose an abstract state transition framework in which different recognition steps are abstracted by events, and the reasoning necessary to effect state transitions is abstracted by a transition function. We define the metrics of ‘precision’ and ‘recall’ of a smart environment to evaluate how well it tracks its occupants. We show how the overall performance of our smart environment can be improved through the use of spatio-temporal knowledge of the environment. A prototype based upon our proposed abstract framework indicates that integrating recognition and reasoning capabilities substantially improves the overall performance of the environment

1 Introduction

The goal of our research is to develop intelligent 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 are useful in homes for elderly or disabled, office workplace, department stores and shopping complexes to larger arenas such as airports, train stations, etc. While considerable research has been done in this area, the distinguishing features of our approach are:

1. the use of a unified framework for multiple biometric modalities;
2. the formulation of performance of the smart environment through the information-theoretic concepts of precision and recall; and
3. the integration of reasoning and recognition to enhance the overall performance of the environment.

For the purpose of user identification, a variety of approaches ranging from tag-based to those based on

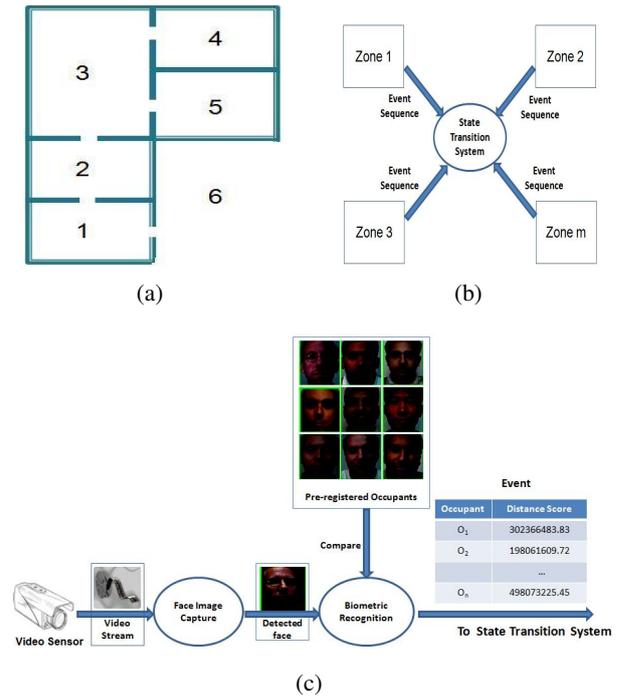


Figure 1: Biometrics Driven Smart Environment

biometrics exist. Irrespective of size, tag-based methodologies employing RFID or others means, are considered obtrusive, as the individual is expected to continuously retain them. Biometric modalities like fingerprint and iris scans, involve a ‘pause-and-declare’ interaction with the human [18]. However, biometric modalities involving face, voice, height, and gait, are more natural and less obtrusive and thus better suited for identification purposes in smart environments.

The key idea underlying our approach is to view a smart environment as a probabilistic state transition system. Each state records the set of individuals who are present in various zones of the environment. A state transition is triggered by an event, which corresponds to the detection of a subject (occupant) in some zone of the environment through biometric recognition methods such as face or voice recognition, etc. Each event thus abstracts a biometric recognition step, and the transition function abstracts the reasoning necessary to effect state transitions. The

outcome of an event is a set of distance scores indicating the degree of match between the subject and the candidates in the biometric database.

Figure 1¹ shows the overall architecture of a biometric-driven smart environment. Although figure 1(c) 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 set of person probability pairs as discussed in more detail below. 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 outputs a set of pairs $\langle o, d(o) \rangle$, where $d(o)$ is the distance score.

The transition function abstracts the reasoning necessary to effect the state transition for each event. It takes a state and an event as input, and first maps the distance scores to probabilities by determining a reduced set of occupants that could have participated in the event by a spatio-temporal reasoning process. The next state is then arrived by assigning revised probabilities to the occupants in the environment based upon the probabilities determined from the event. It is not necessary for us to consider nondeterministic transitions since a state itself is represented as a set of occupants and their probabilities.

We introduce the metrics 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 [21], 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; the details of our abstract framework are discussed in Section 3; the experimental prototype is described in Section 4; and conclusions and future work are presented in Section 5.

2 Related Work

There has been considerable interest in the subject of smart environments. The survey paper by Cook and Das [6] provides a good account of the state-of-the-art. A major

difference between our proposed approach and others surveyed below is our definition and use of a state transition framework for abstracting the details of diverse biometric recognition steps. Also characterized is the performance of the smart environment in terms of the concepts of precision and recall.

This paper extends our earlier paper on biometrics-driven smart environments [17] by integrating reasoning and recognition. The focus of our earlier paper was on recognition. In this paper, we show how the integration of recognition and reasoning can improve the overall precision and recall. Specifically, we show how spatial and temporal knowledge can improve overall performance. This extension has also necessitated some key technical changes in our formulation of the abstract model, especially in the definitions of an event and the transition function. We also present new experimental results in this paper.

We briefly survey closely related efforts and highlight their main features.

2.1 Location Estimation and Tracking

An extensive survey and taxonomy of location systems for a ubiquitous computing application is discussed in [12], while a more recent survey of position location techniques in mobile systems is outlined in [16]. There exist multiple approaches to tracking human movement in offices based on Abstract HMM [4], game theory [7] and Bayesian methods [9, 19]. Identity estimation and maintenance for tracking have relied on RFID tags [7], id-sensors [19] or approaches involving non-absolute, internally system generated identity of tracked persons [14].

2.2 Biometrics in Smart Environments

Pentland and Choudhury [18] 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. Hamid Aghajan et al. [1] 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. [10] 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 on a single modality to improve robustness rather than deploying a multimodal approach that fuses different biometrics. Hong et al. [13] discuss structure, operation and performance of a face verification system using Haar-like features and HMM algorithm in a ubiquitous network environment. More recently Bernardin and Stiefelhagen [2] have implemented a system for the simultaneous tracking and incremental multimodal identification of multiple users in

¹Face images blurred to preserve anonymity.

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. The improvements of combining acoustic features and 2D face images for identification of participants in a smart room environment are discussed in [15]. Techniques for score normalization, modality weighting and modality combination schemes during the fusion of the individual modalities for person identification in a smart room are featured in [8].

In terms of the focus on the use of non-obtrusive biometrics based recognition and location estimation our work is similar to [2, 8, 15]. 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 one of the key themes around which we construct our abstract framework for a biometrics driven smart environment.

2.3 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. The 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 [4]. 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 $m \times n$, meaning that for each of the m zones we record the probabilities of each of the n occupants being present in that zone. (In the Appendix, we present a state as an $n \times m$ table.) 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. The details of our state transition model are 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 $m \times n \times (x+1)$. 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 $\langle t, j, D \rangle$, where $D = \{ \langle o_i, d_{jk}(o_i) \rangle : 1 \leq i \leq n \}$ and $d_{jk}(o_i)$ is the biometric distance of an occupant o_i from the detected subject 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 $e_k = \langle t, j, D \rangle$ occurring at time t in zone j , where $D = \{ \langle o_i, d_{jk}(o_i) \rangle : 1 \leq i \leq n \}$. 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,

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

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

for $1 \leq l \leq m$ and $l \neq j$

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 $p_{lk-1}(o_i)$.

3.1 Spatio-Temporal Reasoning

The above transition function Δ is defined in terms of $p_{jk}(o_i)$, the probability of occupant o_i at zone j in state s_k . Since an event e_k only provides distance scores, we will now show how $p_{jk}(o_i)$ is defined in terms $d_{jk}(o_i)$. The key idea is to determine a *reduced database* of possible occupants who could potentially be detected in event e_k . The basic observation underlying the formation of the reduced database is that the occupants who could participate in an event at a particular zone are those that are present in the adjacent zones. Temporal information from schedules, which gives the known absences of occupants, can be used to further reduce the database. In our earlier paper [17], the mapping of scores to probabilities was carried out inside the biometric recognizer, which made use of a fixed database of possible occupants who could potentially be detected at any event. As a result, the probabilities obtained were much more conservative.

The use of reduced databases to compute probabilities substantially improves the performance of the smart environment. This point will be illustrated in section 4 through our experimental results. For example, consider an empty facility and suppose an occupant moves from the external zone to the entry zone and then on to an internal zone. While all occupants in the database are candidates for the entry event, the candidates for the internal event are only those that are detected with nontrivial probability at the entry event (say, greater than some threshold).

We incorporate the spatio-temporal reasoning in a function $f : State \times Event \rightarrow Database$, and show how f is defined. The function f can be defined in terms of two auxiliary functions f^{sp} and f^{tem} which encapsulate the spatial and temporal reasoning respectively:

$$f(s, e) = f^{sp}(s, e) - f^{tem}(e)$$

For a state s_{k-1} and event e_k , the spatial function $f^{sp}(s_{k-1}, e_k)$ returns a subset, db_k^{sp} , of occupants who could be participate in event e_k based upon the knowledge of occupants in adjacent zones. The temporal function $f^{tem}(e_k)$ returns another subset, db_k^{tem} of the occupants who cannot participate in event e_k based upon the prior knowledge of occupants' absences as given in the daily schedule. Thus the difference of these two subsets yields a

$$\text{reduced database } db_k = db_k^{sp} - db_k^{tem}.$$

Spatial Reasoning In general, spatial knowledge may include the zones of occurrence of events, room adjacencies, layouts, distances, etc. For example, with reference to figure 1(a), the set of candidate occupants corresponding to a new entry event in zone 1 (entry/exit zone) is equal to the current set of occupants in zone 6 (external zone) or zone 2. Similarly the set of candidate occupants corresponding to a new exit event to zone 6 (external zone) is equal to the current set of occupants in zone 1 (exit/entry zone). We assume a thresholding function, $thresh(s, j)$, which takes a state and a zone and returns the set of occupants with a probability above a certain threshold, say 0.5, in a particular zone. (This value could also be a parameter if need be.)

In the case of internal zones (2, 3, 4, 5), their candidate set is the union of candidate sets from their respective adjacent zones. For this purpose we define a function $spatial(s, j)$, for a given layout of rooms, as follows:

$$spatial(s, 2) = thresh(s, 1) \cup thresh(s, 3)$$

$$spatial(s, 3) = thresh(s, 2) \cup thresh(s, 4) \cup thresh(s, 5)$$

Now we can summarize the spatial reasoning as follows:

$$f_{sp}(s, e) = let e = \langle t, j, D \rangle \text{ in } spatial(s, j) \text{ end}$$

Temporal Reasoning Temporal knowledge of the presence or absence of occupants in the facility based upon *a priori* information, as given, for example, in their daily schedules. This information can be applied to the framework to narrow the candidate set of occupants. We illustrate for the case when the schedule of absence of occupants for each day is given; the case for the presence of occupants can be defined in a similar manner.

Assume that we have a schedule of absences given as a relation $absent(Start, End, Occupants)$. For example, $absent(9:00, 13:00, \{o_1, o_7\})$ could denote the absence of occupants o_1 and o_7 from 9 am to 1 pm. Then temporal function can be defined by:

$$f^{tem}(e) = S \leftarrow e = \langle t, j, D \rangle \wedge$$

$$absent(start, end, S) \wedge start \leq t \leq end$$

$$f^{tem}(e) = \{ \} \leftarrow e = \langle t, j, D \rangle \wedge$$

$$\forall (s, e) [absent(start, end, S) \rightarrow (t < s \vee t > e)]$$

Although we have given the absent relation with start and end times taking on definite values, we can generalize this approach so that these timings follow a distribution. We expect that this distribution could be learnt by the smart environment from its observation of the behaviour of the occupants over a period of time.

Thus, given a state s_{k-1} and event e_k , the reduced database of occupants is given by the $f(s_{k-1}, e_k)$. We postulate a function $g : Scores \times Database \rightarrow Probabilities$, which determines the probabilities corresponding to the reduced

database. For this purpose, we make use a standard definition from Govindaraju's earlier work [3, 5] in order to define the function g .

3.2 Precision and Recall

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(a) illustrates a 6-zone smart environment with 25 occupants who are monitored by video cameras. 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 [11]. Our prototype is based upon OpenCV's [22] implementation of the eigenface algorithm [20], which provides a basis for a simple though not robust implementation of face recognition.

Our experimental prototype collects sample face images of the 25 occupants of an office facility and pre-registers them in a training database. The fixed cameras deployed in each zone detect the presence of occupants as they move through the zones and verify the face images extracted

from the video against the database. Simultaneously, a height estimation module determines the height of the moving contour of an occupant (which corresponds to the actual height of the occupant) by calibrating it against the known height of the topmost point in the view of the corresponding camera.

In order to demonstrate the benefits of applying spatio-temporal reasoning, we performed four separate experiments on the same underlying database of 25 occupants and ground truths. In the first experiment, no spatio-temporal reasoning is performed and the distance scores generated by eigenface are cast into probability values directly. In the second experiment, spatial reasoning is performed so that the database of occupants that is used in determining probabilities varies for each event. In the third experiment we observe the database reduction achieved by incorporating reasoning based on height estimation. The effect of using height estimation is similar to spatial reasoning and it helps to further reduce the database of possible occupants who could be detected at any event. In the final experiment, spatial reasoning in conjunction with height estimation is performed for achieving optimal database reduction.

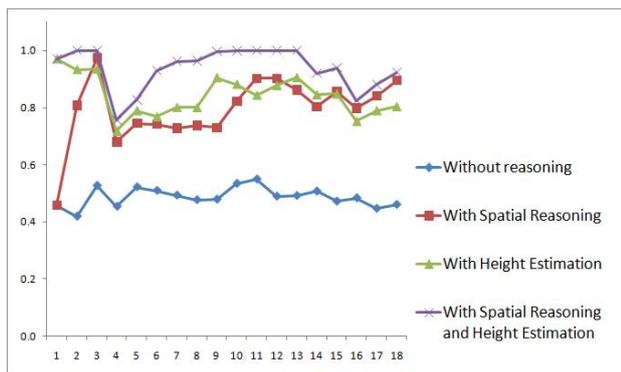


Figure 2: Precision

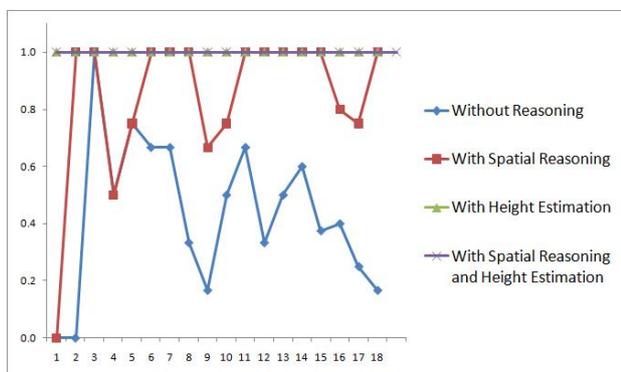


Figure 3: Recall

Figures 2 and 3 illustrate the performance improvements in precision and recall as a result of using spatial reasoning and height estimation separately and in conjunction. The two graphs show the precision and recall results

for a sequence of 18 events. Figure 4 shows the overall improvement in performance, by presenting the average figures across all states. Although we have not incorporated temporal reasoning in these experiments, it should be clear that it would further enhance the precision and recall, as temporal reasoning would help to further narrow down the database of possible occupants.

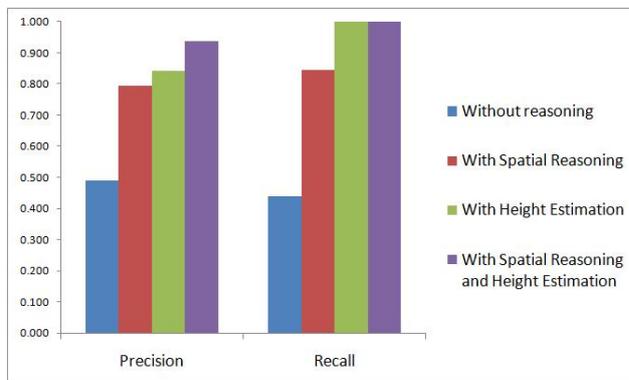


Figure 4: Overall Improvement

The prototype we have developed can be extended to incorporate other biometric recognizers in a similar manner. For example for voice recognition, voice samples of occupants are pre-registered in a database instead of face image samples. The output of the voice recognizer is again 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.

5 Conclusions

We have presented a novel framework using state transitions for non-obtrusive biometric-based indoor smart environments that are capable of identifying and tracking its occupants. We believe that our framework provides an effective abstraction of a smart environment and serves as an elegant basis for integrating various recognition and reasoning capabilities. The main contributions of this paper are as follows:

1. A unified framework based on state transitions where recognition based on different biometric modalities are abstracted as events and the reasoning necessary to effect state transitions are abstracted by a transition function.
2. A characterization of the performance of the smart environment using the information-theoretic concepts of the precision and recall.
3. A demonstration of the improvement in the performance metrics by integrating reasoning and recognition.

We plan to enhance our current prototype by incorporating a variety of other biometric recognizers involving gait,

voice, etc. and fuse two or more of such biometric modalities to enhance the overall performance of recognition. We also intend to incorporate additional spatio-temporal reasoning based upon declarative knowledge of the environment as well as the occupants to seek further improvements in the performance of the smart environment in the long run.

Another technique to enhance the overall performance is that of ‘back-propagation’ of information through previous states. Once the smart environment is able to confirm the identity of an occupant with certainty, this knowledge can be propagated to all previous states in which this occupant was detected. We believe that the combination of spatio-temporal reasoning together with back-propagation would be crucial to achieve satisfactory performance as we scale up the occupant database.

Acknowledgments

This work was done while the first author was a Visiting Research Scientist at the *Center for Unified Biometrics and Sensors*, University at Buffalo. 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] H. Aghajan et al.: Distributed Vision-Based Accident Management for Assisted Living. In: T. Okadome, T. Yamazaki, M. Mokhtari (eds.) ICOST 2007, LNCS vol. 4541. pp. 196–205, Springer, Heidelberg (2007)
- [2] K. Bernardin, R. Stiefelagen: 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)
- [3] D. Bouchaffra, V. Govindaraju, S. Srihari: A Methodology for Mapping Scores to Probabilities. IEEE Transactions on Pattern Analysis and Machine Intelligence. 21(9), 923–927 (1999)
- [4] H. H. Bui, S. Venkatesh, G. West: 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)
- [5] H. Cao, V. Govindaraju: 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)
- [6] D. J. Cook, S. K. Das: How smart are our environments? An updated look at the state of the art. Pervasive and Mobile Computing 3(2), 53–73 (2007)
- [7] S. K. Das, N. Roy, A. Roy: 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)
- [8] H. K. Ekenel, M. Fischer, Q. Jin, R. Stiefelagen: Multi-modal Person Identification in a Smart Environment. In: Proc. of 2007 CVPR Biometrics Workshop, pp. 1–8. IEEE (2007)
- [9] D. Fox, J. Hightower, L. Liao, D. Schulz, G. Borriello: Bayesian filtering for location estimation. IEEE Pervasive Computing 2(3), 24–33 (2003)
- [10] Y. Gao, S.C. Hui, A.C. Fong: A Multi-View Facial Analysis Technique for Identity Authentication. IEEE Pervasive Computing 2(1), 38–45 (2003)
- [11] R. Hewitt, S. Belongie: 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)
- [12] J. Hightower, G. Borriello: Location Systems for Ubiquitous Computing. IEEE Computer 34(8), 57–66 (2001)
- [13] K. Hong et al.: Real Time Face Detection and Recognition System Using Haar-Like Feature/HMM in Ubiquitous Network Environments. In: O. Gervasi et al. (eds.) ICCSA 2005, LNCS vol. 3480. pp. 1154–1161, Springer, Heidelberg (2005)
- [14] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, S. Shafer: 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)
- [15] J. Luque: Audio, Video and Multimodal Person Identification in a Smart Room. In: R. Stiefelagen, J. Garofolo (eds.) CLEAR 2006, pp. 258–269, Springer, Berlin (2007)
- [16] T. Manesis, N. Avouris: 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)
- [17] V. Menon, B. Jayaraman, V. Govindaraju: Biometrics-Driven Smart Environments: Abstract Framework and Evaluation. In: 5th Intl. Conference on Ubiquitous Intelligence and Computing. UIC-08 (Accepted) (2008)
- [18] A. Pentland, T. Choudhury: Face Recognition for Smart Environments. IEEE Computer 33(2), 50–55 (2000)

- [19] D. Schulz, D. Fox, J. Hightower: 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)
- [20] M. Turk, A. Pentland: Eigenfaces for Recognition. *J. Cognitive Neuroscience* 3(1), 71–86 (1991)
- [21] C. J. van Rijsbergen: *Information Retrieval*. London: Butterworths, Boston (1979)
- [22] OpenCV, <http://www.intel.com/technology/computing/opencv/index.htm>