1 Introduction

1.1 Motivation and Significance

The vision of *ubiquitous computing* [127], also referred to as *pervasive computing*, provides the inspiration for *smart environments* saturated with computing and communication technology artifacts that are gracefully integrated with the human users. Ubiquitous computing environments offer computation and communication any time and at any place. The ultimate goal of ubiquitous computing is to improve the human experience and quality of life by obviating the need for explicit awareness of the underlying communications and computing technologies in our daily activities. This implies the technologies recede into the background of our lives, becoming more or less *invisible* thereby demanding little or no attention while serving the user - *calm technology*. The emergence and evolution of ubiquitous computing as the third wave of computing, subsuming the desktop and mobile computing domain can be attributed to significant advancements in allied areas of wireless communications and networking, mobile computing and handheld devices, embedded systems, wearable computers, sensors, RFID tags, smart spaces, middleware, software agents, etc. This phenomenon has a direct beneficial impact on indoor and outdoor environments, such as homes, offices, hospitals, shopping malls, airports and even battlefields and disaster arenas. For example, smart offices, conference/meeting rooms, lecture halls, etc. can serve the users better by anticipating their needs or adapting to them.

With increasing emphasis and attention on national and global security, there is a growing and immediate need to automatically identify and track individuals as well as answer queries about their whereabouts. These capabilities also have diverse applications, ranging from homes for the elderly or disabled to office workplaces, where the set of occupants is known in advance. Identity and location information of individuals are of paramount interest in the domain of security to determine the presence of a registered personnel or an intruder. In general, identity and location
information are key to the provisioning of various context-aware services in smart environments.

User identification based on badges and sensors, tend to be obtrusive and requires the cooperation of the user to continuously wear them. Biometric techniques for user identification based on fingerprint and iris scans require a ‘pause-and-declare’ interaction with the user [95]. They are less natural than biometric recognition based on face, gait, height and voice which are less obtrusive and hence better candidates for unobtrusive and unconstrained recognition in smart environments.

We discuss a couple of illustrative scenarios to bring out the importance of the need for unobtrusive identification and tracking with provision for querying the whereabouts of an individual.

Scenario 1: Z is an elderly individual suffering from dementia residing in an assisted living facility monitored by care givers. Every resident wears a wireless identification badge to facilitate continuous monitoring of their presence within the facility. During the course of wanderings within the facility, Z gets trapped in an elevator alone due to a sudden malfunction in the elevator. From inside the elevator, the wireless signals transmitted by Z’s badge will not be received by any of the receivers in the facility. A timeout alarm is raised by the central monitoring system after a prolonged absence of signals from Z. The care givers scour through the rooms and corridors of the facility to locate Z. Much later, when the elevator resumes its service after maintenance, Z is discovered.

Scenario 2: X is an intruder who has managed to gain illegal entry into a secure facility which is monitored by surveillance cameras. After an intruder alert has been raised, the security personnel set out to nab the intruder. The search team relies on inputs from the control room personnel monitoring the facility through multiple video feeds. The intruder no longer appears on any of the video feeds. The search team is now forced to inspect each and every room until the intruder is nabbed.
**Scenarios** 1 and 2 highlight the need for a tracking system to be able to retrieve answers to queries such as ‘Where is the individual (resident/intruder) now?’ or ‘Where was the individual (resident/intruder) last seen?’ **Scenario** 1 can further deteriorate in the absence of a genuine signal from the badge either due to a malfunction or being physically separated from the individual. Given the layout of the facility, it is possible to reason about the potential trajectory of Z based on last available location data. The need for an unobtrusive and unconstrained recognition system that can retrieve information about the whereabouts of its occupants is highlighted in **Scenario** 1.

**Scenario** 2, a homeland security scenario deals with the problem of detection and tracking of unknown, non-cooperating individuals. Unconstrained and unobtrusive mechanisms are required for classifying an individual as an intruder and tracking his/her movement through secure facilities. The efforts to trace the intruder can be more efficient if we can query the system for the intruder’s last known positions.

### 1.2 Objectives and Scope

The objective of this research is to develop smart indoor environments that can identify and track their occupants as unobtrusively as possible and answer queries about their whereabouts. We focus on small indoor environments such as homes, offices, etc., which can be partitioned into zones or blocks. Indoor environments do not suffer from the problems of power or battery-life that confront outdoor environments and thus sensors and other infrastructure can be deployed and maintained with greater ease. The sensors of interest in this work are video cameras and microphones that monitor the different zones of the environment and capture unobtrusive biometric data such as face, gait, height and ambient voice (if available) of the occupants in an unconstrained manner.

The complexity associated with identification and tracking varies across different scenarios based on the nature of the environment and the proportion of known individuals (regular occupants) to unknown individuals. For example, in assisted
homes and offices, it is quite reasonable to know in advance the set of individuals who frequent the environment on a regular basis. In such scenarios, the regular occupants can be pre-registered with their biometric templates enrolled in a database. However, there exists a possibility that a small percentage of people entering this environment are unregistered and unknown. In larger indoor environments such as hospitals, shopping malls, etc., the percentage of known individuals who are mostly likely employees, will be much less than the percentage of unknown individuals who are mostly visitors. In airports and railway stations, the percentage of unregistered people is very large and the focus might be restricted to monitoring certain zones, accessible only to a select group of registered people. Our research can be readily adapted to a wide range of scenarios discussed above, each differing in the extent to which the occupants in the environment are registered with their biometric templates stored in the enrollment database. However, in this dissertation, we focus primarily on small indoor environments where the set of regular occupants are known in advance and their biometric data such as face, gait, height and voice can be enrolled to train a recognition system for later identification.

In this research we do not continuously monitor the environment, but instead record a discrete set of recognition events corresponding to the movement of occupants from one zone of the environment to another. Conventionally, continuous tracking architectures require a very elaborate deployment of cameras and extensive data gathering, making it harder to scale. Additionally, tracking in such situations rely on ‘hand-off’ between cameras manning overlapping regions. However, our notion of tracking is discrete, with at least one of each biometric recognizer (modality) monitoring a critical segment of a zone and generating location cum identity updates of an occupant only at a zone or block level. The choice of biometric sensors for a zone may vary and depends on various factors. For example, face recognition may not be suitable in some zones for privacy reasons; voice recognition might not work well in a noisy zone; and, gait recognition may be preferable when a person is far from the camera and it is not feasible to recognize face or voice. Since biometric recognition based on a single modality can be error prone, fusion of multiple modalities, for
example, face and gait, face and voice, face, gait and voice, etc., can improve the overall accuracy of the recognition process.

Since we are dealing with a closed environment with a fixed set of occupants, we can, in general, utilize *a priori* declarative knowledge of the environment, based on the layout of zones and their connectivities, the distance between zones and whether an occupant could move between a pair of zones within a certain interval of time as well as knowledge about the occupant schedules. A simple example of reasoning would involve the case of a biometric recognizer monitoring the entry zone of an office building, which failed to resolve the identity of the first occupant to enter the building on any given day. If a subsequent biometric recognition run in an interior zone successfully established the identity of an individual, we can infer this (before the arrival of other occupants) to be the identity of the previously unrecognized occupant (at the entry zone). In practice, to avoid over-reliance on the output of a single biometric modality, we validate the track of an occupant against the spatial and temporal constraints associated with the layout of the building. Considering the shortcomings associated with unconstrained recognition, we demonstrate that an integration of recognition and spatio-temporal reasoning can provide better performance than a pure biometrics-based recognition system. As reasoning based on schedules of the occupants might not be always reliable, we focus only on declarative knowledge of the environment related to layout in this thesis.

Another objective of this research is to quantify the overall performance of a smart environment using information-theoretic concepts. Traditionally, the notion of performance in many pervasive computing scenarios has been closely tied to Quality of Service (QoS) based on speed, latency, bandwidth, etc. Given the focus of this research on identification and zone-level tracking of occupants, we are also interested in quantifying the performance of our approach in terms of how closely it models reality. Some of the key questions we are also interested include who (occupant identity), where (zone location), when (time), which (routes), and more complex queries of the form ‘What is the probability that A and B met in the high security zone of the building in the afternoon?’. Considering the classification as well as
information retrieval dimensions of our model, we explore two broad approaches to performance characterization: query-independent and query-dependent performance characterization. For this purpose, two of the standard measures from the information retrieval domain - precision and recall have been modified to suit our context. Thus, we postulate that the 3 R’s of a smart environment are recognition, reasoning, and retrieval.

1.3 Overview of Dissertation

We describe below the key areas of investigation and a brief summary of our contributions in these areas.

Abstract Model

The main contribution of this thesis is a model for abstracting the behavior of a multimodal smart environment in terms of a state transition system: states, events, and a transition function. The state of a smart environment captures the probabilities of presence of all the registered occupants in the different regions, or zones, of an environment. The state changes upon an event, i.e., the movement of an occupant from one zone to another. In this approach, an event abstracts a biometric recognition step - be it face recognition, gait recognition, etc., and is represented as a set of pairs person-probability pairs, \((o, p(o))\), where \(p(o)\) is the probability that occupant \(o\) was recognized at this event. Due to the inexact nature of biometric recognition, the associated state information is also probabilistic in nature. 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 of the event. In this manner, we are able to accommodate different types of biometric sensors as well as different criteria for state transitions, including those that incorporate declarative knowledge of the individuals and the environment, wherever available. Our presentation here expands upon the publications [84, 85].
Recognition
Our approach to recognition deals with the absolute identity of people across multiple zones of a facility. The state transition model facilitates multimodal recognition output to be uniformly abstracted as events. In conventional biometric recognition, the smallest distance score generated by the identification algorithm over the set of enrolled people corresponds to the identity of the person in question. However, we highlight the inherent uncertainty of unconstrained, automated biometric recognition by recasting the distances scores generated by the recognition algorithm (for example, eigenface algorithm for face recognition) into a probability distribution of the registered occupants. This probabilistic approach to biometric recognition is one of the key themes around which we construct our unified approach to biometrics driven smart environments. Our model is transparent to the presence of multiple biometric modalities such that the outputs from the individual recognizers can be fused together in order to derive a single set of person-probability pairs. The presentation here expands upon the publication [87].

Reasoning
Unconstrained biometric recognition using the proposed modalities can lead to errors in the recognition process leading to misidentification. In this work, we discuss how spatio-temporal reasoning can help alleviate some of the limitations in the underlying recognition technology. Our key idea is to determine which tracks (of the occupants) are spurious and which are valid. In order to determine spurious tracks, we observe that consecutive track elements of a valid track will mostly obey the zone adjacencies in the environment, whereas spurious tracks will mostly violate the zone adjacencies. The events in a spurious track are candidates for event rectification: some of them can be reassigned to the valid tracks at those places where there is a missing event (or events). For the remaining events, their probabilities are re-determined using knowledge of the occupants in adjacent zones given the time and location of the event as well as the valid tracks. The revised probabilities determine new events which in turn are fed back to the state transition system to determine new states and an improved set of tracks. The presentation here expands upon the publication [86].
Retrieval

The state transition system model provides a natural basis for retrieval of answers in response to various queries about the presence and movements of occupants within the smart environment. We formulate a data model based upon an occupancy relation which captures the state information after recognition and reasoning have been performed and a set of valid occupants have been determined. A set of spatio-temporal queries are formulated in SQL, focusing on the computation of probabilities, an aspect that is novel to our framework. Given the shortcomings of SQL, the use of constraint-based extension of logic programs, CLP(R), is explored in the formulation of more complex queries. We discuss techniques for the formulation of various probabilistic spatio-temporal queries of the form such as: ‘What is the probability that an occupant A was in a specific zone Z during a given time interval?’; ‘What is the probability that two occupants A and B met in a zone Z of the building in the afternoon?'; ‘If A was in the environment at noon, what is the probability that B was also present?’; etc.

Performance Metrics

We provide quantitative metrics for performance evaluation of identification and tracking in a smart environment as well as the queries, by refining the basic information-theoretic concepts of precision and recall. From a classification perspective, precision captures the ‘false positives’, or the extent to which the smart environment over-estimates its occupants. On the other hand, recall captures the ‘false negatives’, or the extent to which the smart environment under-estimates its occupants. These are complementary concepts and together capture the overall performance of a smart environment. Though we have formulated precision and recall metrics in a generic and query independent manner to estimate the overall performance of a smart environment, these metrics can also be applied in the conventional manner with respect to a query of interest. Our definition of query-independent precision and recall refines the standard definitions by incorporating a recognition threshold; thereby leading to bell-shaped precision curves instead of monotonically increasing or decreasing behavior. This point of intersection between
the precision and recall curves provides an optimal operating point for the smart environment.

*Simulation Environment*

We discuss a novel simulation environment that realizes the implementation of the abstract model of the smart environment. The simulator is used as a means to gain insights into the functioning of our model for occupant tracking and querying in a biometrics driven smart environment and serves as a platform for integration of the 3 R’s of our smart environment: recognition, reasoning, and retrieval. The results from the simulation run can be used as a pointer for determining the choice of modalities of the biometric recognizers, the quality of the corresponding sensors, optimal recognition thresholds, and the type of reasoning needed in order to achieve a specified level of overall performance of the smart environment.

The rest of this dissertation is organized as follows; Chapter 2 provides an overview of smart environments, biometric recognition, performance metrics, location estimation and tracking. Chapter 3 describes our state transition system model of a smart environment and defines quantitative metrics for query-independent performance evaluation. Chapter 4 discusses our approach to spatio-temporal reasoning to enhance the overall performance of recognition in smart environments. Chapter 5 provides a detailed description of the simulation based experimental system and a discussion of the results from multiple simulation runs. Chapter 6 discusses a data model for our system and formulation of a basic set of probabilistic spatio-temporal queries and corresponding query-based metrics for evaluation of query performance. Chapter 7 describes our conclusions and suggests the scope for future work.