

Three R's of Cyber-Physical Spaces

Vivek Menon*, Bharat Jayaraman†, Venu Govindaraju†

*Amrita Vishwa Vidyapeetham, †University at Buffalo (SUNY)

ABSTRACT

Recognition, reasoning and retrieval are three fundamental operations in a cyber-physical space that is capable of identifying and tracking the movements of people and answering questions about their whereabouts. We show that a state transition system effectively abstracts the three R's: events abstract recognition, transition function abstracts the reasoning, and states abstract the information necessary for performing retrieval. We provide quantitative metrics for the performance of a cyber-physical space by refining the standard information-theoretic measures of precision and recall. This article also presents performance improvements from integrating spatio-temporal reasoning with biometric-based recognition.

KEYWORDS

Cyber-physical spaces, State Transition System, Biometric Recognition, Spatio-temporal Reasoning, Information Retrieval, Performance Metrics.

INTRODUCTION

A *cyber-physical space* is a physical space embedded with intelligence and interfaced with humans in a natural way using vision, speech, gestures, and touch, rather than the traditional keyboard and mouse. The key to realizing this paradigm is identifying and tracking people in the space. The ability to identify and track people and answer questions about their whereabouts is critical to many applications [1]. The scenarios range from environments in which most of the individuals are known or pre-registered (health-care monitoring) to environments where most of the individuals are unknown (homeland security). We highlight below two scenarios from real-life incidents.

- *Scenario 1: Z is an elderly resident in an assisted living facility and wears an RFID badge to facilitate continuous monitoring of his presence. On one occasion, Z enters the elevator alone but gets trapped due to a power failure. The RFID signals transmitted by Z's badge are not in the range of any receiver. Only much later, when the elevator resumes its service, is Z discovered [2].*
- *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 find 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 has no other recourse but to search each room.*

The above scenarios illustrate the need for automated approaches to transforming multimedia data into a form suitable for information retrieval. This is a challenging problem spanning many research areas: video and audio processing, computer vision, spatio-temporal reasoning, and data models. They also highlight the need for unobtrusive data gathering; the idea is that people can go about their normal activities without being subject to a 'pause and declare' routine or the burden of RFID tags or badges [3]. Identifying people from their face, gait and voice is more natural and less obtrusive and hence more suited for a cyber-physical space.

We postulate three fundamental operations for a cyber-physical space: *recognition*, *reasoning*, and *retrieval*. Biometric recognition based upon modalities such as face, gait, and voice is inherently inexact and error prone, and thus the output of such a recognizer is typically represented as a probability distribution [4]. Information retrieval in a cyber-physical space is concerned with the location of people at various points in time and hence the queries will be probabilistic and spatio-temporal in nature, e.g., 'Where was X last seen?', 'What is the probability that Y and Z met in the high-security zone between 6 and 7 pm?' In order to alleviate the shortcomings of a pure recognition based approach, we show in this article that an integration of recognition with spatio-temporal reasoning enhances the overall performance of retrieval in a cyber-physical space.

At the outset, we would like to contrast our research with the extensive literature on ubiquitous or pervasive computing [5], which abounds in applications [6] where the actuation of a physical device is at the core of the paradigm. Actuation-based cyber-physical spaces tend to use simpler sensors (temperature, pressure, motion, etc.) and are more concerned with networking of sensors and computing devices. This article describes a novel paradigm where retrieval of information based on spatio-temporal queries is the main driver for identification and tracking (see fig 1). Unlike earlier approaches to the problem of tracking people [7], we will not require a complete coverage of the monitored space by massive numbers of sensing devices; rather, we will explore the more realistic scenario where only the main zones (hallways, entrances/exits) are embedded with biometric devices that capture data from a distance [8]. This is a significant departure

from current approaches that tag people based on cues such as clothing, height, etc., and establish correspondences between adjacent fields of view interspersed by blind spots - such approaches however, do not scale. This article focuses on indoor environments such as homes and offices, which 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.

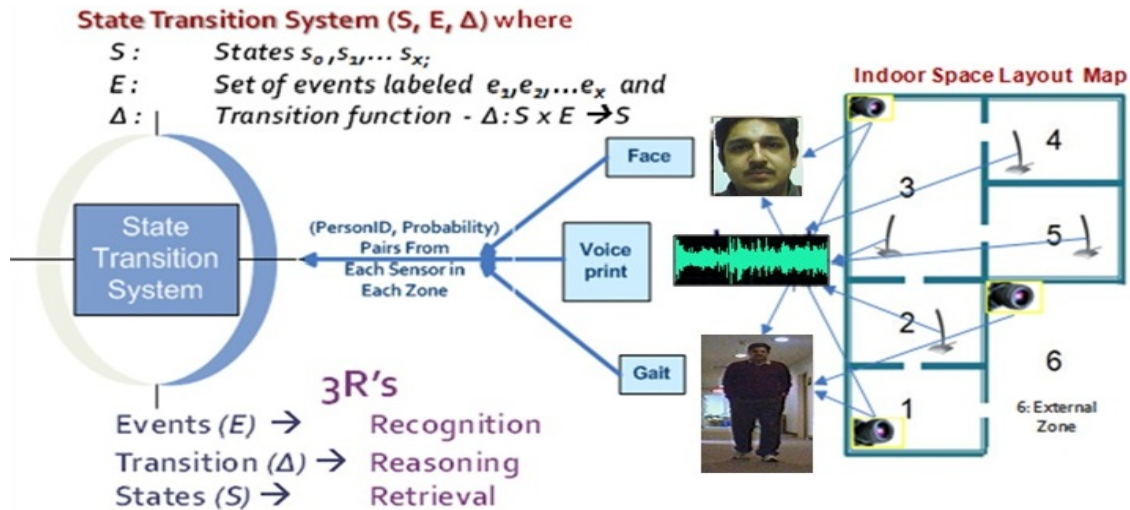


Figure 1: Identification, Tracking and Querying in Cyber-Physical Spaces

The main contribution of this article lies in providing a unified treatment of the three R's using a novel *state transition system* consisting of states, events, and a transition function. The strength of this approach is that it accommodates different approaches to recognition, reasoning, and retrieval. That is, the events abstract different approaches to recognition, the transition function abstracts different approaches to reasoning, and the states provide a basis for defining different data models to support information retrieval. This article extends our previous work [9] which focused on a pure recognition-based approach. Our state transition system is fundamentally probabilistic because the biometric recognition that underlies events is inexact in nature. We also formulate quantitative metrics for performance evaluation of identification and tracking in a cyber-physical space based on the two information-theoretic concepts: *precision* and *recall*. The concepts of precision and recall are standard performance measures in the information retrieval literature, but we adapt their definitions to suit our context.

ABSTRACT MODEL FOR CYBER-PHYSICAL SPACES

We abstract the behaviour of a cyber-physical space 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 [9]. The state transitions may be depicted as follows: $s_0 \xrightarrow{e_1} s_1 \xrightarrow{e_2} s_2 \dots \xrightarrow{e_x} s_x$

- A *state* records for each zone and each occupant o_i , $i = 1..n$, the probability of presence in that zone, $p_s(o_i)$. For each occupant, the sum of probabilities across all zones equals one. The states abstract the information necessary to support information retrieval.
- An *event* abstracts a biometric recognition step and is represented as a set of person-probability pairs, $\langle o_i, p(o_i) \rangle$, where $p(o_i)$ is the probability that occupant o_i was recognized at this event. We also have $\sum_{i=1}^n p(o_i) = 1$.
- The *transition function* abstracts the reasoning necessary to effect state transitions. In the zone of occurrence, we define $p_s(o_i) = p(o_i) + x_i * p'_s(o_i)$, where $x_i = 1 - p(o_i)$ and $p'_s(o_i)$ is probability of the occupant in the previous state. For all other zones, we define $p_s(o_i) = x_i * p'_s(o_i)$. This ensures that the sum of probabilities for an occupant across all zones in the resultant state equals one. A more detailed account of the transition function may be found in [9].

Events are assumed to be independent, but the transition function captures the dependency on the previous state, as in a Markov process. This is illustrated in table I which shows a sample state transition in a hypothetical 4-zone cyber-physical space with 5 occupants. Table Ib shows an event e_{11} occurring at zone 2 (z_2) and corresponds to the movement of occupant o_5 from z_1 to z_2 . The states s_{10} and s_{11} reflect the probability of presence of the 5 occupants in each of the 4 zones before and after the event e_{11} . The occupants listed in the last row of table Ia, Ib, Ic correspond to the ground truth (G).

Occ	z_1	z_2	z_3	z_4
o_1	0.06	0.19	0.62	0.13
o_2	0.05	0.10	0.34	0.51
o_3	0.10	0.50	0.25	0.15
o_4	0.04	0.09	0.22	0.65
o_5	0.52	0.09	0.19	0.20
G	o_5	o_3	o_1	o_2, o_4

(a) State s_{10}

Occ	z_1	z_2	z_3	z_4
0.06				
0.01				
0.18				
0.20				
0.55				
o_5				

(b) e_{11}

Occ	z_1	z_2	z_3	z_4
o_1	0.06	0.24	0.58	0.12
o_2	0.05	0.11	0.34	0.50
o_3	0.08	0.59	0.21	0.12
o_4	0.04	0.27	0.17	0.52
o_5	0.23	0.60	0.08	0.09
G	-	o_3, o_5	o_1	o_2, o_4

(c) State s_{11}

TABLE I: Sample State Transition

As we do not continuously monitor the environment, we record a discrete set of biometric recognition events corresponding to the movement of occupants from one zone of the environment to another. 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, and voice recognition might not work well in a noisy zone. Since biometric recognition based on a single modality can be error prone, fusion of multiple modalities can improve the overall accuracy of the recognition process. When recognition is based upon more than one biometric modality, the outputs from the individual recognizers are fused together in order to derive a single set of person probability pairs. Note that, when a person bearing little or no resemblance to any of the registered occupants arrives at the entry zone, the biometric recognition step would produce low probabilities for all occupants, and this is the way that one detects the presence of an ‘outsider’.

Hidden Markov Models (HMMs) *HMMs and their variants, such as Factorial HMMs and Coupled HMMs, may be regarded as examples of Dynamic Bayesian Networks (DBNs) [1]. Here, transition probabilities are to be learned from empirical data gathered about the movements of people through the space over a period of time. Since we cannot assume a predictable pattern of movement of people through various zones of the cyber-physical space, we do not adopt this approach. In our state transition system, biometric capture devices provide direct information on the occurrence of events in specific zones. Given any event in a zone, the next state is unambiguously determined, although the state information is probabilistic. Furthermore, a state with n occupants and m zones would only require $m * n$ storage, since for each of the m zones we record the probabilities of each of the n occupants being present in that zone.*

Reference

1. Z. Ghahramani, "Learning Dynamic Bayesian Networks," C.L. Giles and M. Gori, eds., *Adaptive Processing of Sequences and Data Structures*, Lecture Notes in Artificial Intelligence, Springer-Verlag, 1998, pp. 168-197.

EXPERIMENTAL TESTBED AND PERFORMANCE METRICS

We present an experimental testbed in order to validate the abstract model discussed above. Figure 2a shows the overall architecture and figure 2b shows the main interface of our simulator for a cyber-physical space. In a simulation run, a script randomly generates a set of trajectories for a given set of occupants. Each trajectory corresponds to the movements of one occupant and consists of a totally-ordered sequence of events, where an event consists of a zone visited by an occupant at a particular time, and the probabilities generated by the biometric recognizer for this event. The simulator provides options to choose a layout from a set of available layouts, specify the number of occupants who are in the space, and adjust parameters such as sensor quality (σ) and recognition threshold (θ) before a simulation run. The simulator also generates performance metrics associated with recognition before and after spatio-temporal reasoning.

The data for the simulation was derived from the face images of 45 individuals using three different video cameras of varying image quality. An OpenCV [10] implementation of the eigenface algorithm was customized for this purpose. The quality of the video camera, variations in pose, illumination and expression can cause fluctuations in the overall accuracy of the recognition process, especially in unconstrained settings. Our formulation of sensor quality σ abstracts these intrinsic and extrinsic factors that can affect the recognition output. We used 10 different event templates (face images) for every individual and when the sensor quality is reduced (using the slider bar), the system chooses a lower quality image such that the event probability for the person recognized is correspondingly lower. A varying number of 'false positives' across these event templates accounts for the variability, typical of unconstrained biometric recognition.

To evaluate the performance of a cyber-physical space, we define the concepts of *precision* (π) and *recall* (ρ) for a cyber-physical space in terms of the *ground truth* G , which, for a given input event sequence, is a sequence of states wherein the presence or absence of any occupant

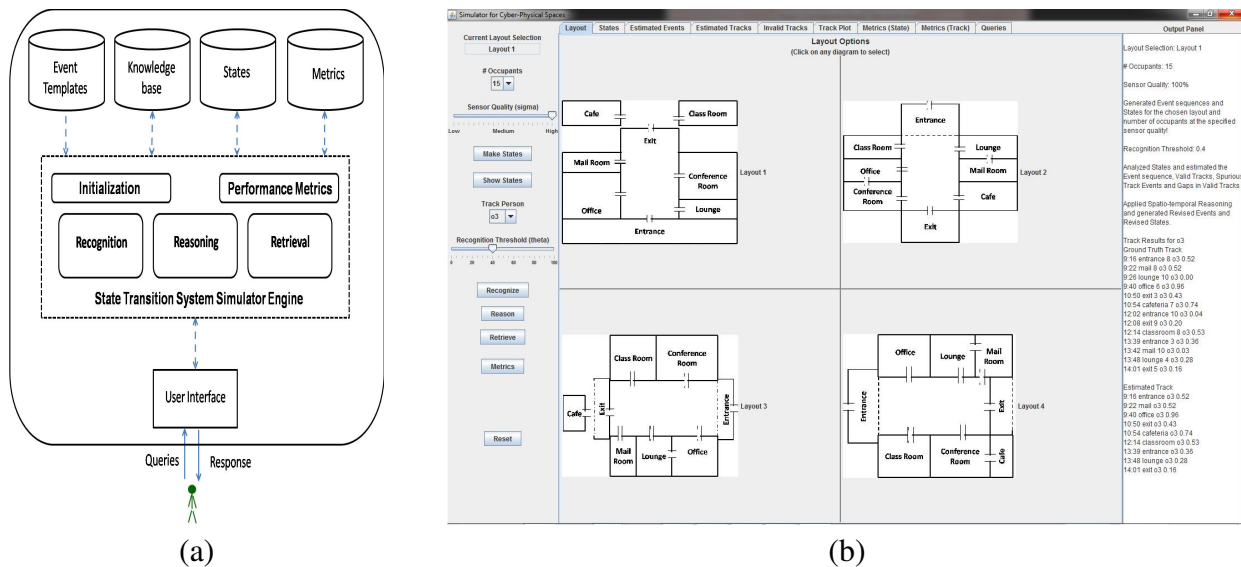


Figure 2: Simulator for Cyber-Physical Spaces

in any zone is known with certainty (0 or 1). Precision captures the extent of ‘false positives’ while recall captures the extent of ‘false negatives’. These definitions are stated in terms of a *recognition threshold* θ ; only those persons with a probability $\geq \theta$ are assumed to be present. When a person’s probability in two or more zones is $\geq \theta$, the zone with the highest probability is taken as the zone of his presence. We refer to the set of persons occurring in a ground truth G as $occ(G)$.

- 1) $\pi = tp/(tp+fp)$, where tp is the set of ‘true positives’ and fp is the set of ‘false positives’. The set $tp = \{o_i : p_s(o_i) \geq \theta \wedge o_i \in occ(G)\}$, while the set $(tp + fp) = \{o_i : p_s(o_i) \geq \theta\}$.
- 2) $\rho = tp/(tp+fn)$, where tp is defined as above, and fn is the set of ‘false negatives’. The set $(tp + fn) = \{o_i : o_i \in occ(G)\}$.

We discuss the performance metrics across multiple runs for 15 occupants. Average precision and average recall for varying values of recognition threshold θ at sensor quality $\sigma = 1.0$ are plotted in figure 3a. Note that the average precision increases up to $\theta = 0.6$ and then declines. The average precision is low at low values of θ since a high proportion of false positives are present in the set of recognized occupants. As θ increases, the proportion of false positives diminishes until a point of inflexion is reached. From this point, the average precision begins to decline since the true positives also fail to get recognized. Average recall decreases with increasing θ since the proportion of false negatives steadily increases with θ .

The average precision and average recall curves for varying sensor quality σ are shown in figures 3b and 3c respectively. The dependence of precision and recall on the recognition threshold θ causes the precision curves in figure 3b to assume a bell-shaped form. The recognition

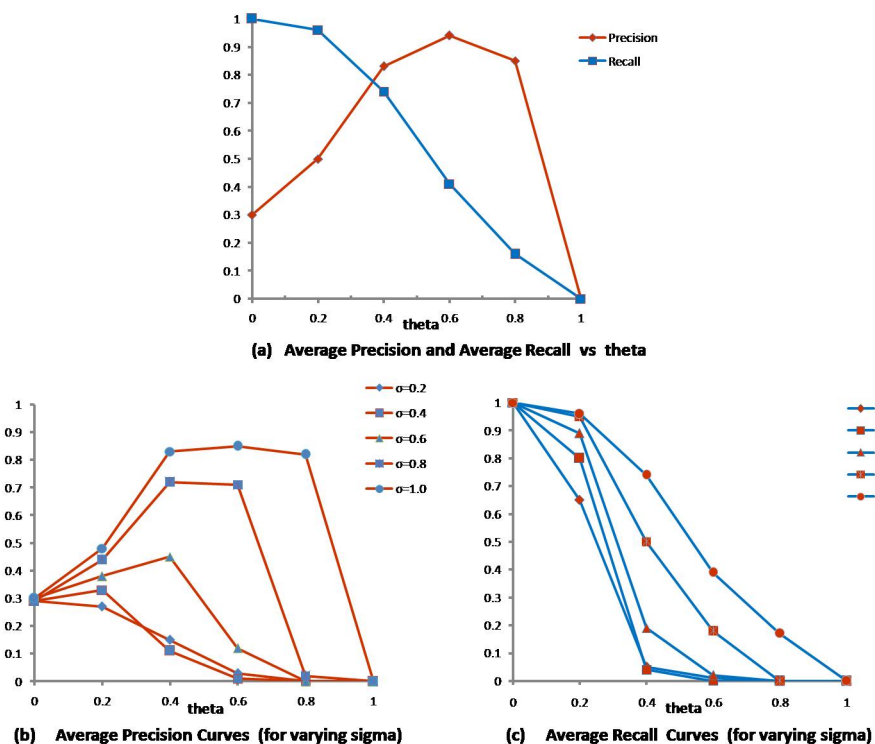


Figure 3: Precision and Recall Metrics

threshold for a cyber-physical space can be chosen by the user depending on the application at hand. For example, in security related applications, one might minimize the number of false positives, whereas in assisted living scenarios one might want to minimize the false negatives. In the absence of any additional information, a reasonable operating point would be where false positives equal false negatives. At any given σ , such a recognition threshold θ can be obtained from the intersection of the average precision and average recall curves as shown in figure 3a.

SPATIO-TEMPORAL REASONING

Our experimental results for a pure recognition based approach (indicated by the blue curve in figure 4) shows that as the number of people in the cyber-physical space increases, so does the number of spuriously identified people. In order to minimize the impact of recognition errors, the key idea is to determine the identity of a person based upon information from a track of events and corresponding states rather than a single event. The reason is that consecutive track elements of a valid track will mostly obey the zone adjacencies in the physical environment, whereas spurious tracks will mostly violate the zone adjacencies. This track-based reasoning can be captured by a transition function of the form $\Delta : \mathcal{P}(\mathcal{S}) \times E \rightarrow \mathcal{S}$. The transition function takes as input a set of previous states, computes the occupant tracks from these states, and determines the next state upon the occurrence of an event.

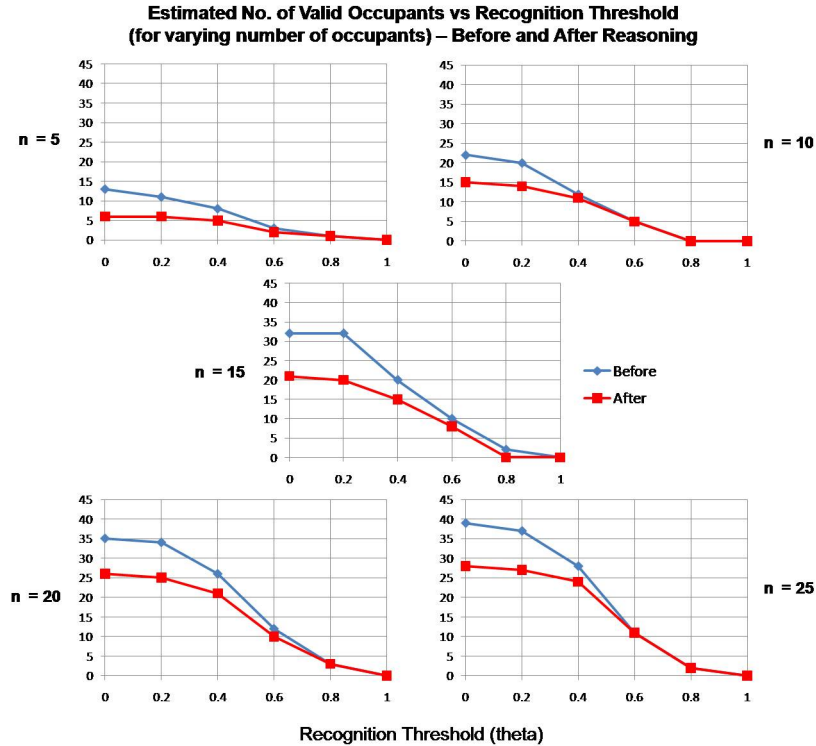


Figure 4: Estimated Number of Occupants - Before and After Reasoning

In order to determine the tracks from a set of states, the transition function determines for each event occurrence, the individual who moved between two zones of the cyber-physical space. From this information, the set of all tracks (both valid and spurious) is immediately obtained. The criterion for determining which occupant moved is defined in terms of the maximum difference in occupant probabilities between two consecutive states in the zone of occurrence of the event. It might appear that the person with the highest probability in an event is the one who moved. However we do not adopt this approach for two reasons: event information could be erroneous; and comparing consecutive states gives due importance to both event and historical information which is captured and retained only in the states.

The initial states of a cyber-physical space are likely to have more errors because track-based reasoning on shorter tracks is less effective in mitigating the errors due to recognition. Over a period of time, as longer tracks are formed, the reasoning process not only determines subsequent states with less error, it can also correct the errors in the initial states. Such a transition function would have the form $\Delta : \mathcal{P}(\mathcal{S}) \times E \rightarrow \mathcal{P}(\mathcal{S})$. That is, it takes a set of states as input, computes the tracks from these states, and determines as output the next state along with a revised set of previous states. Figure 4 shows the benefits of integrating recognition and track-based reasoning (indicated by the red curve) so as to reduce the extent of spuriously identified occupants. The benefits of reasoning are more pronounced at lower value of θ where the number of false positives are higher.

Spatio-Temporal Reasoning *Spatio-temporal reasoning over occupant tracks is similar to a higher-order Markov process, since the next state depends upon multiple previous states. When the transition function also updates the information in previous states, the resulting inference is closer to that of a Markov Random Field (MRF) analysis. In the MRF approach, the operation of a cyber-physical space may be modeled by an undirected graph whose nodes correspond to space-time (or zone-event) points and edges capture space-time adjacency. Spatio-temporal reasoning with MRF is based upon a neighborhood analysis around the zone of occurrence of an event. While it is more general in principle, it is also computationally more complex than track-based reasoning, which is more specialized and hence can more efficiently incorporate a global view of the system. Spatio-temporal reasoning has also been investigated from a logic and constraint perspective, with applications in geographical information systems (GIS), computer vision, planning, etc. [1].*

Reference

1. A. Gerevini, B. Nebel, "Qualitative Spatio-Temporal Reasoning with RCC-8 and Allen's Interval Calculus: Computational Complexity," *Proc. European Conf. on AI (ECAI 2002)*, IOS Press, 2002, pp. 312-316.

INFORMATION RETRIEVAL

The state transition system provides a natural basis for retrieval of answers to queries about the whereabouts of occupants in the cyber-physical space. We show how to formulate spatio-temporal queries using the well-known SQL database query language, focusing on the computation of probabilities [11], an aspect that is novel to our model. We have also developed more complex queries in a constraint-based extension of logic programs, called CLP(R), which permits general recursive queries and reasoning over real-valued variables and arithmetic operations [12].

We define an occupancy relation, *occupancy(start, end, person, zone, probability)*, where *start* and *end* define a time interval consisting of a discrete totally ordered set of points (since events are also discrete). The attribute *probability* $\in \mathcal{R}$, the set of real numbers, and is functionally dependent on the other four attributes. This relation captures the state information after recognition and reasoning have been performed and a set of valid occupants have been determined. The basic syntax of SQL queries is as follows: `SELECT attributes FROM relations WHERE condition`. The *condition* is typically a conjunction of simpler tests that serve as a basis for tuple selection. There are numerous extensions to the basic syntax outlined above, in order to perform aggregate operations, grouping, ordering, etc.

Query: What is the probability that o_7 was in the lounge during 10:00 AM to 11:00 AM?

Answer: Since there could be multiple sub-intervals within 10:00 AM to 11:00 AM during which o_7 was in the lounge (with different probabilities), the answer to the query is 1 minus the product of the probabilities that he was not in the lounge during every such sub-interval. The probability that an occupant was not in the lounge at a given time is sum of the probabilities that he was

in one of the other zones at this time (since the sum of the probabilities across all zones = 1 at any given time).

```
(1 - PROD (
  SELECT SUM(prob) as sumprob
  FROM occupancy
  WHERE person = o7 and
         zone ≠ lounge and
         10:00 ≤ start and end ≤ 11:00
  GROUP BY start)
)
```

While query-independent performance characterization is holistic from a system level, characterizing the performance of a cyber-physical space with respect to queries may be more meaningful to a user and cater better to his interest. The performance metrics of any given query are defined in terms of the ground truth, which is a set of true answers associated with the query. The nature of the response set may vary depending on the type of query posed and may comprise of entities such as occupants and zones, attributes such as probabilities of presence, time of occurrence, or derived attributes such as duration of presence, tracks, etc., based on relations defined as part of the data model. From an information retrieval perspective, precision is the fraction of retrieved answers that are relevant to the query while recall is the fraction of the answers relevant to the query that were successfully retrieved.

Spatio-Temporal Databases *The data in cyber-physical spaces is fundamentally probabilistic and spatio-temporal in nature since people are moving between different zones over a period of time and we are interested in their trajectories. Hence the data models and query languages of interest in a cyber-physical space are also probabilistic and spatio-temporal. There has been considerable research on spatio-temporal databases over the past two decades. For example location-based systems have been a major driver for the interest in moving object databases (MOD), and their associated data models, query languages, indexing, and uncertainty [1, 2]. In addition to the challenges involved in spatio-temporal databases, research into probabilistic databases [3] has gained momentum over the years due to the emergence of a broad range of applications that need to manage large and imprecise data sets in domains such as sensor networks, information extraction and business intelligence. Our research in cyber-physical spaces makes crucial use of probabilistic and temporal concepts, while the spatial issues are treated more in a qualitative (symbolic) than a quantitative (geometric) manner.*

References

1. R. H. Guting, M. Schneider, *Moving Object Databases*, Morgan Kaufmann, 2005.
2. N. Pelekis et al., “HERMES: Aggregative LBS via a Trajectory DB Engine”, *Proc. ACM Intl. Conf. Management of Data*, 2008, pp. 1255-1258.
3. N. Dalvi, C. Re, D. Suciu, “Probabilistic Databases: Diamonds in the Dirt,” *Comm. of the ACM*, Jul. 2009, pp. 86-94.

SUMMARY AND CONCLUSIONS

This article describes a novel state transition system approach for modeling a cyber-physical space that can identify and track its occupants unobtrusively and answer queries about their whereabouts. Our model accommodates different approaches to each of the three R's of a cyber-physical space and effectively abstracts and unifies them: recognition (through events), reasoning (through transitions), and retrieval (through states). We also provide a succinct statement of its performance in terms of information-theoretic concepts of precision and recall. This unified model serves as an elegant basis for integration of recognition and spatio-temporal reasoning capabilities for improved performance over a pure biometrics-based recognition system and serves as an effective blueprint for developing practical deployments.

REFERENCES

- [1] J. Hightower, G. Borriello, "Location Systems for Ubiquitous Computing," *IEEE Computer*, Aug. 2001, pp. 57-66.
- [2] K. Schindler, private communication, Feb. 2009.
- [3] A. Pentland, T. Choudhury, "Face Recognition for Smart Environments," *IEEE Computer*, Feb. 2000, pp. 50-55.
- [4] D. Bouchaffra, V. Govindaraju, S. Srihari, "A Methodology for Mapping Scores to Probabilities," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Sep. 1999, pp. 923-927
- [5] M. Satyanarayanan, "Pervasive Computing: Vision and Challenges," *IEEE Personal Communications*, Aug. 2001, pp. 10-17.
- [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*, Mar. 2007, pp. 53-73.
- [7] J. Krumm et al., "Multi-Camera Multi-Person Tracking for EasyLiving," *Proc. 3rd IEEE Intl. Workshop on Visual Surveillance (VS'2000)*, IEEE Computer Society, 2000, pp. 3-10.
- [8] M. Tistarelli, S. Z. Li, R. Chellappa, eds., *Handbook of Remote Biometrics, for Surveillance and Security*, Springer, 2009.
- [9] V. Menon, B. Jayaraman, V. Govindaraju, "Multimodal Identification and Tracking in Smart Environments," *Personal and Ubiquitous Computing*, Dec. 2010, pp. 685-694.
- [10] OpenCV, <http://www.intel.com/technology/computing/opencv/index.htm>
- [11] R. Subrahmanyam, private communication, Jul. 2009.
- [12] V. Menon, "Integrating Recognition and Reasoning for Tracking and Querying in Smart Environments," *PhD Thesis*, Amrita University, Mar. 2010.

AUTHORS BIOS AND CONTACT INFORMATION

Vivek Menon is an assistant professor of information systems in the school of business at Amrita Vishwa Vidyapeetham (Amrita University), India. His research interests center around intelligent systems and smart environments. He received his PhD in computer science from Amrita Vishwa Vidyapeetham, India. He is a member of IEEE, ACM and AIS.

CONTACT

Email: vivek_menon@cb.amrita.edu, Phone: +91 (422) 268-5000

Mailing Address: Amrita School of Business, Amrita Vishwa Vidyapeetham, Ettimadai, Coimbatore 641 105, Tamil Nadu, India.

Bharat Jayaraman is a professor in the CSE department at the University at Buffalo (SUNY), and has served as its Chair from 2001-09. His research interests center around software systems and languages. He received his PhD in computer science from the University of Utah, USA. He is a member of ACM and a senior member of IEEE.

CONTACT

Email: bharat@buffalo.edu, Phone: +1 (716) 645-4755

Mailing Address: CSE Department, University at Buffalo, NY 14260-2000, U.S.A.

Venu Govindaraju is a SUNY distinguished professor and directs the Center for Unified Biometrics and Sensors (CUBS) at the University at Buffalo (SUNY). He received his PhD in computer science from the University at Buffalo. His research interests center around pattern recognition applied to biometrics and document analysis. He is a fellow of AAAS, ACM, IEEE, IAPR.

CONTACT

Email: govind@buffalo.edu, Phone: +1 (716) 645-2356

Mailing Address: CSE Department, University at Buffalo, NY 14260-2000, U.S.A.