Spatial Querying for Retrieval of Locomotion Patterns in Smart Environments

Gamhewage C. de Silva, Toshihiko Yamasaki, Kiyoharu Aizawa
Department of Information and Communication Engineering, the University of Tokyo
(chamds, yamasaki, aizawa)@hal.t.u-tokyo.ac.jp

ABSTRACT
A system for retrieving video sequences created by tracking humans in a smart environment, by using spatial queries, is presented. Sketches made on a graphical user interface using a pointing device are used as the means of entering multiple types of queries. After preprocessing and coordinate system conversion, the sketches are analyzed to identify the type of the query. Directional search algorithms based on the minimum distance between points is applied for finding the best matches to the sketch. The results are ranked according to the similarity and presented to the user. The results of an initial evaluation are reported. The paper concludes with an outline of possible future directions.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *video*.

General Terms
Algorithms, Experimentation, Human Factors

Keywords
Floor Sensors, Locomotion Patterns, Smart Environments, Spatial Queries, Video Retrieval

1. INTRODUCTION
Multimedia retrieval and summarization for smart environments is an active research area with several applications such as surveillance, study of human behavior, and taking care of the elderly [1]. In most of these applications, it is necessary to track the movement of persons in the environment and retrieve media showing their behavior. Depending on the environment and the application, this can be performed using only the image data, or with the help of other sensors.

However, in many of these applications, it is necessary to view the results of tracking manually to retrieve video with a particular locomotion pattern. Ability to query a collection of such results by a particular path or pattern will greatly reduce the search space in a video retrieval task. Such queries can be categorized as *spatial queries* [2]. However, facilitating spatial queries for a smart environment is a difficult task due to two main problems. First, there should be an intuitive and non-restricting way to input queries on human locomotion patterns into a computer. Second, algorithms for searching for specific movements have to be designed and developed.

MM'07, September 23-28, 2007, Augsburg, Bavaria, Germany.

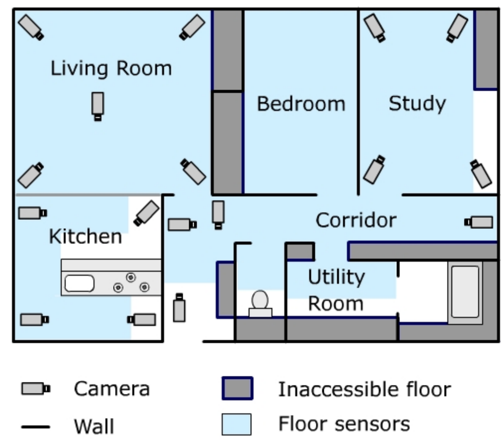

Figure 1. Ubiquitous home layout.

There has been some research towards a framework for spatial querying of locomotion patterns. Egenhofer [2] proposes a topological representation and a relational algebra for spatial queries based on sketches. Gottfried [3] uses a *locomotion base* and a set of relations to represent locomotion patterns, with emphasis on healthcare applications. However, an effective user interaction strategy for submitting spatial queries is essential for utilizing the above framework for effective retrieval of locomotion patterns. So far, there has been very little research on user interfaces for spatial querying. Ivanov and Wren [4] use simple spatial queries for video retrieval from surveillance cameras. However, the functionality of this interface is limited to specifying the direction of movement along a corridor.

In this research, we propose a novel user interaction strategy for spatial querying of locomotion patterns in smart environments. Different types of queries are designed based on a query-by-sketch approach, so that the users can specify different types of queries intuitively using a simple interface without ambiguity. We also design and implement algorithms for retrieving locomotion patterns represented by these queries.

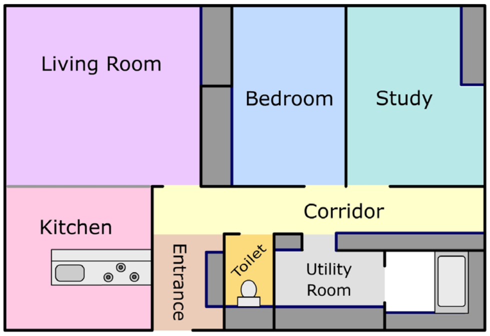

Figure 2. Partitioning the floor into regions.

We implement the proposed spatial querying system in a two-bedroom smart home [5] equipped with a large number of stationary cameras and microphones (Fig. 1). Pressure-based sensors mounted on the floor are activated as people move inside the house. At the current status of a research project on multimedia experience retrieval from this environment, it is possible to retrieve personalized video for moving persons by changing cameras and microphones automatically as they move to different regions of the house [5]. This is performed by clustering the floor sensor data to segment the footsteps of different persons and selecting cameras and microphones automatically to ensure that the persons are seen and heard throughout the video. However, this system facilitates only temporal queries. For example, it is possible to search for video showing people walking inside the house between 6:00 pm and 7:00 pm on a given date.

We intend to enhance the capability of the above system by incorporating spatial queries to retrieve video correlated to locomotion patterns. Our objective is to facilitate queries such as: “Retrieve the video clips showing the people who walked from the living room to the study room in the morning of the 1st of May 2007”. Using the proposed user interaction strategy, we submit spatial queries by sketching paths on the floor layout of the house. The floor sensor data corresponding to the video clips are then searched for similar paths. The best matches are retrieved and shown to the user as video and footstep sequences. We evaluate the performance of the algorithm using a data set recorded during a real life experiment with actual residents.

2. USER INTERACTION STRATEGY

Our objective here is to facilitate querying by allowing the users to specify the path they want to search for, by sketching that path on a diagram of the house floor plan. In order to design an interface that is versatile, easy to use and intuitive, we took the following, user-centered approach.

When a person is standing or walking inside a house, he is always referred to as being in a “region”. Therefore, it is desirable to facilitate specification of regions during querying. For this purpose, we partitioned the house floor plan into the regions labeled in Figure 2.

We selected the following three types of possible spatial queries to be facilitated by the proposed user interaction strategy:

1. Query type 1: walking/standing within a selected region (E.g.: “inside the living room”)
2. Query type 2: walking from one region to another, irrespective of the path taken (E.g.: “from the living room to the kitchen”)
3. Query type 3: walking along a specific path.

The user interaction strategy should allow the user to submit all three types of queries in a simple and intuitive manner. There should be no ambiguity between any two types of queries. The complexity of the input is also important. Since queries 1 and 2 are less specific queries, it is desirable for them to be easier to enter.

To query for footstep sequences within a given region, the user scribbles within that region on the floor plan (Fig. 3a). Since this is the type of query with the least amount of detail, it is sensible to use a simple gesture for specifying this query.

To query for movement between any two regions, the user simply draws a line between any two of the regions, without considering the walls that partition the regions (Fig 3b). Since movement through such partitioning is impossible, only the starting and ending regions of the lines are useful as meaningful inputs to a query. The user finds the query easy to enter and intuitive, as he/she needs to be concerned about only those two regions. However, it should be noted that there can be some ambiguity when the start and end regions are not well partitioned. This is resolved later during the search algorithm.

When querying for a specific path, the user sketches the path along the house plan (Fig. 3c). The path can be either within the same region, or across several regions. This query takes more effort from the user, and returns fewer and more specific results.

We designed a graphical user interface based on the above strategy for retrieval of personalized video from the ubiquitous home. This interface is based on the concept of *hierarchical media segmentation* [5], to facilitate more accurate retrieval based on interactive querying. The inputs consist of a date input, a time line, and a drawing of the house floor layout (Fig. 4). All these inputs can be specified using only a pointing device. When the user selects a particular date, the time line shows one-minute segments where footstep sequences were present during that day. The user then draws a line segment on the time line to specify the time interval, aided by this visualization of the results. If necessary, the user can now see a summary of all footsteps within the house during this time interval (as seen by the dots on the house floor plan in Fig. 4). This helps the user to identify any unusual pattern, or directly see whether the desired pattern might be available for this time slot. Thereafter, the user sketches the path on the house floor plan. If the user does not select the date, the entire collection of data will be used as input for the search. If a time interval is not specified, the entire day is considered.

All the pixels along the sketched path and the date and time interval (if submitted) are recorded as inputs for the search algorithm described in the following section. The time to make the sketch is recorded as an additional input.

3. ALGORITHMS FOR SEARCHING

The personalized video from the ubiquitous home are created by automatically selecting cameras and microphones automatically for footstep sequences of different persons [5]. The attributes

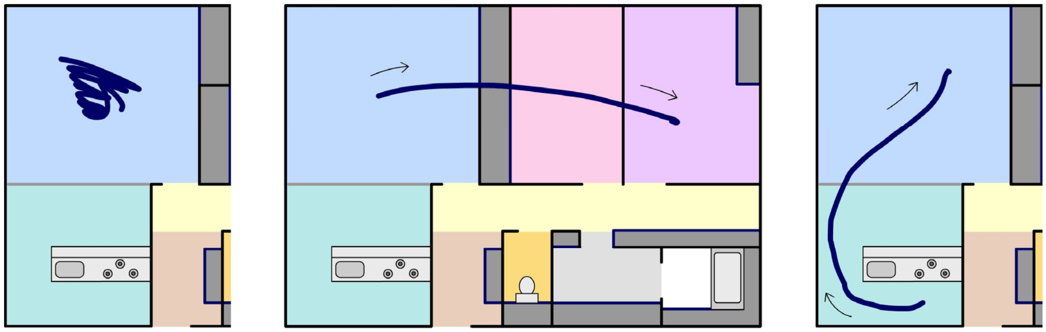

Figure 3. Different types of spatial queries.

contained in the footstep sequence data are sufficient to fully represent the video clip. Therefore, it is possible to retrieve video by querying only the collection of footstep sequences using the parameters obtained from the inputs, namely the date, time interval and the sketch.

Each footstep sequence is an ordered set of four-dimensional data elements with the following variables:

- X coordinate of the position of the sensor
- Y coordinate of the position of the sensor
- Time stamp indicating when the sensor was activated
- Duration that the sensor was active

The X and Y coordinates are specified in millimeters, starting from the bottom left corner of the house floor as seen in Figure 1. The proposed search algorithm uses only the coordinates and the time order of the elements for retrieval of these sequences. The duration of the activations apparently has no correlation with the time taken to input the sketch.

The array of pixel coordinates contained in the sketch has to be preprocessed before searching. First, the points are transformed to the house floor coordinates. Ideally, this should be possible using a linear transform if the house floor layout on the interface is drawn to scale. However, due to calibration errors in floor sensor data, minor adjustments are needed. The ordered set of points $P = \{P_1, P_2, \ldots, P_n\}$ is submitted as input to the search algorithms.

The next step is to identify the type of the query by analyzing the distribution of points in P and the time consumed for making the sketch. Type 1 queries are identified using the distribution of the points. Type 2 queries are distinguished by identifying the crossing of partitions between regions. When there is an ambiguity between type 2 and type 3, it is resolved using the time consumed to draw the sketch since type 2 queries are faster to sketch.

3.1 Query Type 1

To perform this query type, we retrieve all segments footstep sequences present in the specified region. If a time interval is specified, the results are filtered using that interval. The results are ordered by the starting time of the sequences.

3.2 Query Type 2

First, we retrieve all the footstep sequences that include floor sensor data from both regions, and filter them by the specified time interval. Thereafter, the following algorithm is applied for each of the candidate paths $C=\{C_1, C_2, \ldots, C_m\}$:

1. Starting from C_1, scan C until the first point in the starting region, C_s is found
2. within the sub-sequence $C' = \{C_s, C_{s+1}, \ldots, C_m\}$, find the first point in the ending region, C_e
3. select $C'' = \{C_s, C_{s+1}, \ldots, C_e\}$ as a match

Again, the matches are ordered by the starting time stamp.

3.3 Query Type 3

We start the search by selecting all the footstep sequences recorded during the time interval that the user specified. For each of the candidate paths $C = \{C_1, C_2, \ldots, C_m\}$, selected as above, the following algorithm is applied:

1. Set overall mean distance $D = 0$
2. For the first point C_1 in the selected candidate path C, find the closest point P_a in P
3. Add the Euclidean distance between C_1 and P_a to D
4. Repeat the steps 1 and 2 for the next point in C and $P' = \{P_a, P_{a+1}, \ldots, P_n\}$, until all points in C are used for calculation.
5. Divide D by m
6. If $D < 360$, select C as a match.

This algorithm looks for paths with less deviation from the sketched path, while preserving direction. The threshold value of 360 corresponds to an average deviation of two floor sensors, and was selected empirically. The paths with different starting or ending regions from those from the sketch are removed from the

Figure 4. Date and time input for filtering results.

set of matches, to prevent false retrievals of much shorter paths with good overlap.

The matched paths are presented to the user in ascending order of the overall mean distance. Figure 5 shows a sketch made by a user (the curved path) and the retrieved path that matches the best (the piecewise linear path).

4. EVALUATION AND RESULTS

The system needs two types of evaluations; firstly, a user study on the usability of the interaction strategy, and secondly, an evaluation of the accuracy and proper ranking of the results retrieved. At the time of writing, such detailed evaluations are not available since the results are late-breaking. However, we conducted the following experiment to evaluate the accuracy of the search algorithms.

The results retrieved by the first two types of queries are straightforward, since they are direct queries on a relational database followed by a linear search. In order to evaluate the performance of the search algorithm for query type 3, the house floor was partitioned into *sub-regions* with an area of 90x90 cm (corresponding to 5x5 floor sensors). We decided to evaluate the system by searching for only those paths between a pair of these sub-regions. Retrieval of sequences with the correct starting and ending sub-regions was used as an objective measure of the accuracy of retrieval.

The system was tested on a selected set of 94 footstep sequences obtained from 12 hours of data gathered during a “real-life experiment”, where a family of three members stayed in ubiquitous home for 10 days. A set of pairs of sub-regions were selected by observing these data.

During evaluation, five paths between each selected pair of sub-regions were drawn and results were retrieved. Both the instances where wrong paths were retrieved and correct paths were missed were recorded. The precision P, recall R and balanced F-measure F for retrieval were calculated as

$$P = N_C / (N_C + N_M)$$

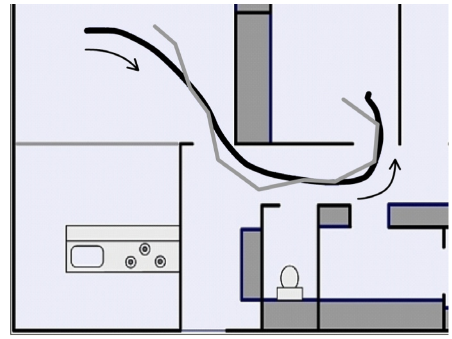

Figure 5. Date and time input for filtering results.

$$R = N_C / (N_C + N_O)$$

$$F = 2PR / (P + R)$$

where N_C is the number of correctly retrieved video clips, N_O is the number of clips that were not retrieved, and N_M is the number of mistakenly retrieved clips.

The precision of retrieval was 92.5% and the recall 98.8%. The balanced F measure was 95.2%. The low precision is mainly due to some candidate paths that are shorter than the sketch but match well with the corresponding segment of the sketch.

5. CONCLUSION AND FUTURE WORK

We have proposed a user interaction strategy and a search algorithm for querying and retrieving from a collection of video sequences based on spatial information. Both of these can be applied to the results of human tracking based on any type of sensors, and therefore are not restricted to environments with floor sensors. The accuracy of retrieval, when applied to real-life data from a home-like environment, was approximately 95%.

The user interaction strategy is being developed further to facilitate more types of queries using simple gestures. We are working on a detailed user study and an evaluation experiment, as described in Section 4.

6. ACKNOWLEDGMENTS

We thank NICT for their cooperation. This work was partially supported by CREST and JST.

REFERENCES

[1] J. C. Augusto, C. D. Nugent, “Designing Smart Homes – the Role of Artificial Intelligence”, LNCS 4008, Springer 2006.

[2] M. Egenhofer, “Spatial Query by Sketch”, Journal of Visual Languages and Computing, 1997 - 8(4), pp. 403-424.

[3] B. Gottfried, “Spatial Health Systems”, Pervasive Health Conference and Workshops, 2006, Vol., Iss., Nov. 29 2006-Dec. 1 2006, pp. 1-7.

[4] Y. A. Ivanov, C. R. Wren, "Toward Spatial Queries for Spatial Surveillance Tasks", In proc. *Pervasive PTA 2006*.

[5] Gamhewage C. de Silva et al., "An Interactive Multimedia Diary for the Home", *IEEE Computer*, 2007-05, pp. 52-59.