M-075

# Spatial Querying for Video Retrieval from Smart Environments

Gamhewage C. de Silva† Toshihiko Yamasaki‡ Kiyoharu Aizawa‡

## 1. Introduction

Multimedia retrieval and summarization for ubiquitous 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.

In many of these applications, it is desirable to be able to search for video showing a particular pattern of movement. However, this is a challenging task that has been less researched at the current state. There are two main issues that have to be tackled in performing this task. First, there should be an intuitive and non-restricting way to submit queries on movement patterns. Second, algorithms for searching for specific movements have to be designed and developed. There has been some recent work in this direction [2], with limited functionality.

This research is based on *Ubiquitous Home* [3], a two-room house equipped with a large number of stationary cameras and microphones (Fig. 1). The place is built as a test-bed for recording, retrieval and awareness of human behavior. Pressure-based sensors mounted on the floor are activated as people move inside the house. The amount of data recorded in a single day is approximately 500 GB.

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 [4]. 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 those persons are seen and heard throughout the video.

However, at the current state, it is possible to search among such sequences only based on time. For example, it is possible to search for videos 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 this system by facilitating search based on *spatial queries*. For example, it should be possible to query the system and retrieve videos showing people walking from the study room to the kitchen.

We propose to implement the above system using the following approach. A user interaction strategy is designed to 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.

† Intelligent Media Laboratory, The University of Tokyo

‡ Dept. of Information and Communication Engineering, The University of Tokyo

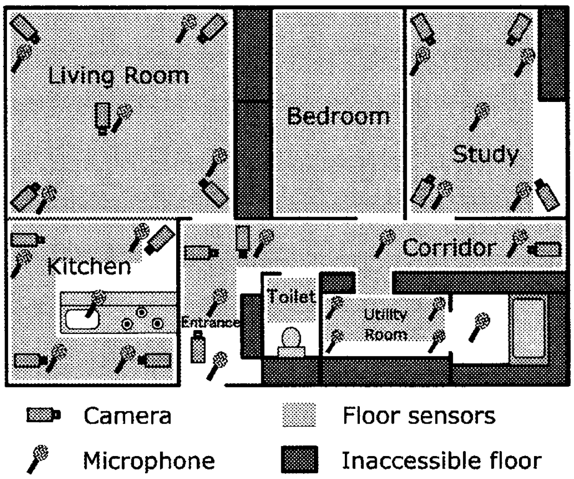


**Figure 1. Ubiquitous home layout.**

## 2. Submitting Queries

The user interface for submitting spatial queries consists of a date input, a time line, and a drawing of the house floor layout. After selecting the date and the time interval (if known), the user specifies a path for searching the collection of video sequences, by sketching the desired path on the house floor plan. All the pixels along the sketched path are recorded as input for search.

## 3. Search Algorithm

Each footstep sequence, segmented using the algorithm in [4], is an ordered set of 4 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 Fig. 1. These data are sufficient for automatic handover of video and audio data and creation of video sequences. The proposed search algorithm uses only the coordinates and the time order of the elements for retrieval of these sequences.

The array of pixel coordinates contained in the sketch has to be preprocessed before searching. First, the points are transformed to 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 paths were also corrected to remove instances of crossing the corners of walls due to less precise sketches. After preprocessing, the ordered set of points $P=\{P_1, P_2, \ldots, P_n\}$ , is submitted as input to the search algorithm.

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 minimum deviation from the sketched path, while preserving direction. The threshold value of 360 corresponds to an average deviation of two floor sensors. The paths with different starting or ending rooms/regions (as labeled in Fig. 1) from those of $P$ are removed from the 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. The user can browse the results to find the exact sequence/s that he/she was looking for. Figure 2 shows a sketch made by a user (the curved path) and the retrieved path that matches the best (the piecewise linear path ).

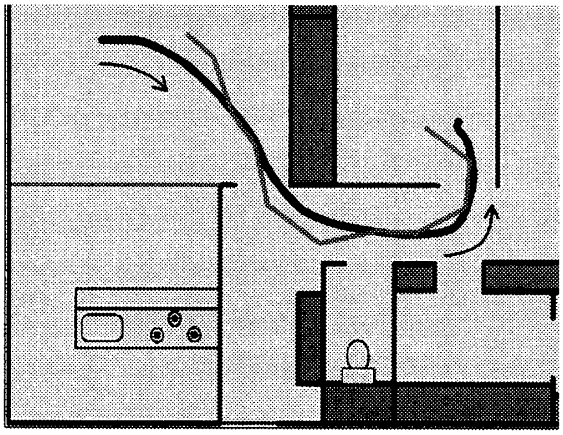

**Figure 2. A query and its best-matching result**

## 4. Evaluation and Results

In order to evaluate the system, the house floor was partitioned into small regions with an area of 90x90 cm. This size corresponds to 5 floor sensors along each side. We decided to carry an initial evaluation of the system by searching for only those paths between two of these regions. Retrieval of sequences with the correct starting and ending 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.

During evaluation, five paths between each pair of 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 and recall R for retrieval were calculated as

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

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

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

where $N_C$ is the number of correctly retrieved video clips, $N_O$ is the number of clips that were not retrieved, and $N_M$ s 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 $P$ but match well with the corresponding segment of $P$ .

## 5. Conclusion and Future Work

We have implemented 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 work and the results reported in this paper are late-breaking, with a fair amount of research in progress. The user interaction strategy is being developed further to facilitate more types of queries using simple gestures. A comprehensive user study for evaluation will be necessary, to evaluate the effectiveness and usability and find directions for improvement. The current evaluation for the search algorithm is simple, and considers only the movement between predefined regions of the house. An evaluation with finer levels of details will be designed and conducted later, to evaluate the accuracy of both retrieval and ranking.

## Acknowledgments

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

## References

[1] T. Mori et al., "Sensing Room: Distributed Sensor Environment for Measurement of Human Daily Behavior", Proc. INSS2004, pp.40-43, 6 (2004).

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

[3] T. Yamazaki, "Ubiquitous Home: Real-life Testbed for Home Context-Aware Service", Proc. *Tridentcom2005*, pp.54-59, 2005.

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