The Familiar: a living diary and companion

Brian Clarkson
MIT Media Laboratory
20 Ames St.
Cambridge, MA 02139
clarkson@media.mit.edu

Kenji Mase
ATR Media Integration &
Communications Research
Laboratories, Kyoto, Japan
mase@mic.atr.co.jp

Alex Pentland
MIT Media Laboratory
20 Ames St.
Cambridge, MA 02139
sandy@media.mit.edu

ABSTRACT

We present a perceptual system, called the Familiar, that could allow a user to collect his/her memories over their lifetime into a continually growing and adapting multimedia diary. The Familiar uses the natural patterns in sensor readings from a camera, microphone, and accelerometers, to find the recurring patterns of similarity and dissimilarity in the user's activities and uses this information to intelligently structure the user's sensor data and associated memorabilia.

Keywords

Peripheral perception, clustering, data mining, sensing

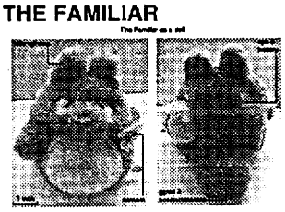
INTRODUCTION

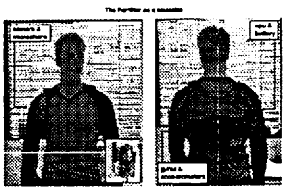
In the 70's there was a Japanese artist named On, who was in a way obsessed with time and the (usually) mundane events that mark its passage. His works such as the *I Met* and *I Went* series explored the kind of day-to-day events that tend to fall between the cracks of our memories. For years, everyday Mr. On would record the exact time he awoke on a postcard and send it to a friend or create lists of the people he met each day or trace on maps where he went each day. His work raises a few interesting questions. If we had consistent records of some aspects of our day-to-day lives over a span of a lifetime, what trends could we find? What kinds of patterns or cycles would reveal themselves? Interestingly, we wouldn't need highly detailed memories to find these trends and patterns, just a consistent sampling in time. Our dream is to build a device that can capture these life patterns automatically and render them in a diary-like structure.

THE TRADITIONAL DIARY

A diary is a hand-made record of the day-to-day emotional state and salient events in someone's life. Other than the therapeutic benefits of the actual diary writing, the diary also provides the diary-keeper with the ability to look back on the developments, trends, and patterns of activity that has led to his/her current state of being. The diary also provides a scaffolding to attach other pieces of memorabilia, which include anything from newspaper clippings and Polaroids to e-mails and home videos. However, except in the most special cases, diaries always suffer from having to many gaps and missing pieces. It is exactly those busy times where changes are rapidly occurring that diary-writing gets pushed to the side and forgotten. However those times of change and turmoil contain perhaps the most important events for a diary to include.

THE FAMILIAR


The Familiar as a doll


The Familiar as a wearable

Currently, the Familiar consists of 3 sensors:

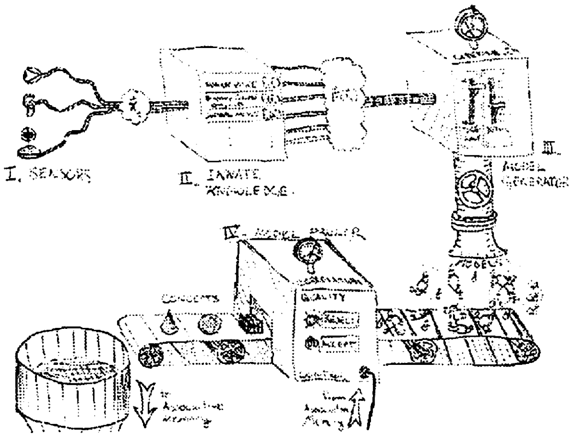
- Small video camera with extremely wide-angle lensing or parabolic mirrors
- Omnidirectional microphone
- Gyros and accelerometers

So far, we have built the sensor package into clothing (ref. Wearables) and dolls. These two types of objects have day-to-day states that are highly correlated with the user's day-to-day activities by virtue of being on the person or actually involved in the user's activities. On top of these sensors various modules have been built such as skin (from the camera) and speech detection (from the microphone), and gesture classification (from the output of the gyros and accelerometers). The purpose of these modules is to provide simple robust and salient features so that the Familiar can start learning about the structure of its user's life. Of course some of the features can be used directly to annotate a person's day (e.g. speaker identification, face detection) but more interestingly there is a chance to find more complicated and long-term patterns (for example, what is just part of your daily routine and what is a new novel occurrence).

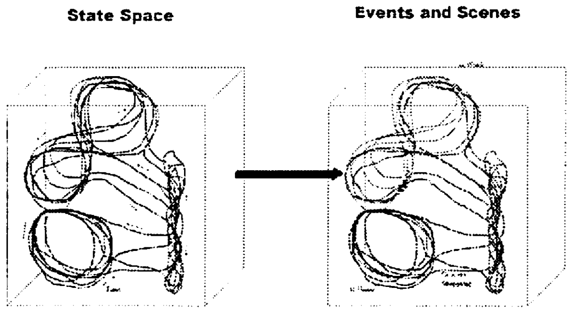
These sensors were all integrated with a small computer and HDD for the data collection and data mining algorithms.

LEARNING PIPELINE

Referring to figure below, the basic idea is that at each time step the sensors are sampled at a maximum rate of



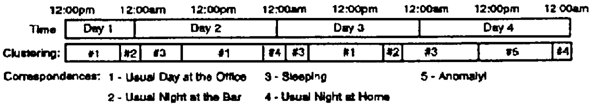
5Hz and these values are collected into a single N-dimensional state vector (step I.). These sensor values are collated and features (speech detection, image moments, skin classification, accelerometer gesture, movement) representing the bias we have designed into the system are extracted (step II.). The result of this step is a state vector in some N-dimensional space. As time progresses these state vectors trace a noisy path through N-dimensional space (see figure below this paragraph). Previous experiments have shown that there are patterns and cycles in this high-dimensional path that correspond to certain situtations such as sitting in the office, walking down a sidewalk, browsing in a video store, and so on. We have developed clustering methods allow us extract these patterns automatically as Hidden Markov Models (HMM). This is step III. However, this step will always in addition to the useful and salient patterns in the data, yield some models that are not modeling anything discernable, so we need to prune these (step IV.).



PRELIMINARY RESULTS

We have collected hours of sensor data from both the doll and wearable platforms in various conditions. Using the pipleline described above we have evaluated the system for its ability to find clusters of salient events when there is no human labeling and also its raw ability to classify situations when there are labels available. We encourage the reader who is interested in the full details of these experiments to refer to [1] [2]. The conclusion from these experiments was that we could classify most locations (both indoor, room-based locations, and outdoor locations) from our simple sensors with accuracies from 85% to 99%. Conversations could be detected quite easily with the exception that nearby music frequently confused the system (but we believe we can add an additional innate recognizer for music). For the clustering experiments (on a user walking around a campus and urban environment) we noticed that our algorithms from only a few hours of data were able to pick out short-time patterns (or events) such as walking through doorways, crossing the street, going up stairs, sitting in the office. As we constructed hierarchies of clusters we were able to extract models for more complicated scenes such as shopping for groceries, being at home, and going to work.

THE AUTOMATIC DIARY



While these results (represent the learning pipeline from step I to III) are promising and continue to be improved, it is step IV (Model Pruning and Association) that requires interesting interaction with a user in the form of a diary. We are currently organizing the fruits of the clustering and classification systems into a diary-like system which could provide semi-automatic construction of a multimedia and nonlinear diary. The figure above shows a sneak peek of how we hope the basics of the system will work. The user is provided with a timeline of his day-to-day activities. Our learning system has already been able to segment the user's history at any level of detail (coarse to fine) that is desired with the boundaries actually corresponding to scene changes. Also, the learning system has by nature of the clustering inter-related the similar segments so that similar activities are already grouped together. Now the user can associate the clusters (if necessary merge and split activities) with his own memorabilia (text, pictures, audio). The resulting structure is scrapbook of memories that is not only organized by time, but also by what events were routine or special, their period of recurrence, how perceptually similar they were, and most importantly a growing network of associations gotten automatically from correlations amongst events and manually from the links the user adds.

REFERENCE

1. Clarkson, B. and A. Pentland, *Unsupervised Clustering of Ambulatory Audio and Video*, in *ICASSP'99*.http://www.media.mit.edu/~clarkson/icassp99/icassp99.html.

2. Clarkson, B., K. Mase, and A. Pentland, *Recognizing User's Context from Wearable Sensor's: Baseline System*. Vismod Technical Report #519, 2000.