Image Browsing, Processing, and Clustering for Participatory Sensing: Lessons From a DietSense Prototype

Sasank Reddy*, Andrew Parker*, Josh Hyman*,
Jeff Burke*†, Deborah Estrin*, Mark Hansen*
* Center for Embedded Networked Sensing,
†Center for Research in Engineering, Media and Performance
University of California, Los Angeles
{sasank, jburke}@ucla.edu, {adparker, destrin, josh}@cs.ucla.edu,
cocteau@stat.ucla.edu,

Abstract

Imagers are an increasingly significant source of sensory observations about human activity and the urban environment. ImageScape is a software tool for processing, clustering, and browsing large sets of images. Implemented as a set of web services with an Adobe Flash-based user interface, it supports clustering by both image features and context tags, as well as re-tagging of images in the user interface. Though expected to be useful in many applications, ImageScape was designed as an analysis component of DietSense, a software system under development at UCLA to support (1) the use of mobile devices for automatic multimedia documentation of dietary choices with just-in-time annotation, (2) efficient post facto review of captured media by participants and researchers, and (3) easy authoring and dissemination of the automatic data collection protocols. A pilot study, in which participants ran software that enabled their phones to autonomously capture images of their plates during mealtime, was conducted using an early prototype of the DietSense system, and the resulting image set used in the creation of ImageScape. ImageScape will support two kinds of users within the DietSense application: The participants in dietary studies will have the ability to easily audit their images, while the recipients of the images, health care professionals managing studies and performing analysis, will be able to rapidly browse and annotate large sets of images.

Categories and Subject Descriptors

C.3 [**Special-Purpose and Application Based Systems**]: Real-time and embedded systems

General Terms

Algorithms, Design, Experimentation

Keywords

Mobile Sensors, Data Auditing, Urban Sensing

1 Introduction

Imagers are an increasingly significant source of sensory observations about human activity and the urban environment. Ultimately, computer vision and image processing offer promise for automatic classification and analysis of data from such image sensors; in the near term, however, a combination of automatic processing and manual classification are required due to the complexity or ambiguity of recognition tasks. Manual human review of images can be aided by mechanisms for similarity- and/or time-based clustering during viewing.

DietSense is a software system that uses both automatic image processing techniques and manual image review. Currently under development at UCLA, DietSense will support the use of mobile devices for automatic multimedia documentation of dietary choices, efficient post facto review of captured media by users and researchers, and easy authoring and dissemination of data collection protocols. This paper describes ImageScape, a software component of DietSense for post facto processing, clustering, and browsing large sets of images. Implemented as a set of web services with a Flash-based user interface, it supports clustering by both image features and context, as well as the ability to tag images for sharing purposes. DietSense is an example of a *participatory sensing* system [1, 2], in which new network and software architecture enable individuals and communities to employ their own mobile devices as sensors in secure and credible ways.

While dietary patterns are recognized as a critical contributing factor to many chronic diseases, the reliability of existing research strategies for collecting information about an individual's daily dietary intake needs improvement [3, 4, 5]. Most existing strategies rely on self-reported information. Under- and over-reporting are common; contextual information affecting dietary consumption are often missing; and scaling to longitudinal studies of large populations is difficult.

Given the limitations of self-reporting, there has research into using personal technology, such as audio recorders and 35mm cameras, to keep electronic diaries on dietary intake, but even these practices found under-reporting due the participant's lack of confidence in operating the equipment given [6]. However, this and similar work set the stage for personal multimedia devices to play a vital role in self-reporting. Camera-equipped personal digital assistants (PDAs) and mobile phones are being used for diet and activity journaling by enabling users to upload information wirelessly and using bidirectional communication to increase adherence [7, 8, 9].

Our development of DietSense extends this body of work, framing the data collection as an exercise in participatory sensing: End users will initiate autonomous data capture on their

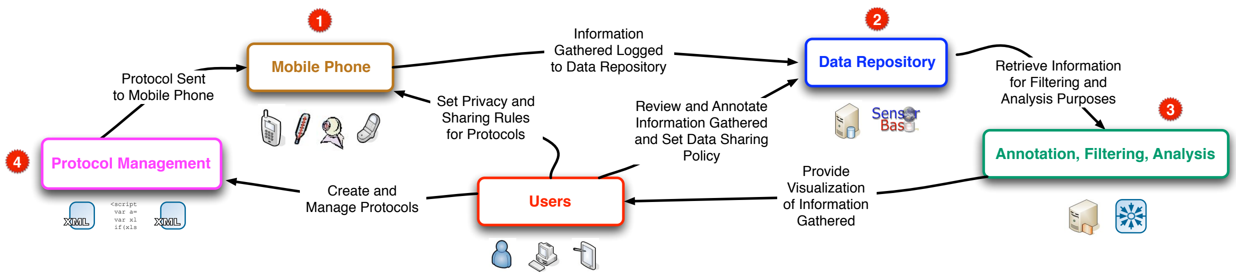

Figure 1. DietSense architecture.

mobile device and then regulate its upload (publication). We leverage digital photography for direct observation of food selection, weight estimates, and plate waste [10, 11], while eliminating most of the need for manual image capture, data entry, file upload and separate journaling techniques through the use of cell phones to collect multimedia documentation of dietary activity. Their capabilities enable automatic upload, time- and location-stamping, and gathering of contextual information that increases the usefulness of primary image data.

2 Participatory Sensing of Dietary Intake

As shown in Figure 1, there are four components involved in the DietSense architecture: (1) mobile phones, (2) a data repository, (3) data analysis tools, and (4) data collection protocol management tools. Mobile phones act as sensors, worn on a lanyard around the neck with the camera facing outwards [12, 13]. The phones run custom software to autonomously collect time-stamped images of food choice, as well as relevant context, such as audio and location. To lower the self-reporting burden and increase data integrity, devices implement adaptive data collection protocols for dietary intake, which are created by researchers. A participant data repository receives the annotated media collected by the devices and, via tools like ImageScape, allows individuals private access to their own data for auditing purposes. Participants approve which media and information to share with health care professionals. Data analysis tools, including a more feature-rich version of ImageScape, provide researchers mechanisms with which to rapidly browse the resulting large image sets and code or tag them. Such tools will support assessment of dietary intake for direct communication to the participant and to suggest modification to protocols for data collection. Easy authoring and dissemination techniques for the data collection protocols are important, but are not treated in this paper.

Rapid and intuitive exploration of captured images and associated contextual metadata (e.g., location, time and on-the-fly participant annotation) are a vital part of DietSense. Participants need to be able to quickly audit and annotate images collected of their meals and authorize them for sharing. Researchers need to rapidly cluster good images for a given mealtime, associate additional annotations, and generate reports. We thus incorporate features for fast image browsing, as well as preprocessing steps that can eliminate underexposed, overexposed and noisy images or cluster redundant images in our architecture. Our clustering techniques take advantage of time/location, image similarity and event detection [14, 15]. More information about our image browsing and filtering tools is given later in this paper.

3 Image Capture Software Pilot

In order to explore the DietSense concept and discover the project's challenges, we conducted a series of pilot studies in which students wore cell phones that captured images and other context information automatically throughout the day. The studies suggest the tools needed by a user to choose what data to share and also to help researchers or health care professionals to easily navigate contributed data.

3.1 Platform

For our pilot study, we used the Nokia N80 "smart phone" as the mobile handset platform and asked users to carry a Holux Bluetooth Global Positioning Service (GPS) receiver for location data. An application was created in Python and installed on the phone to collect data autonomously. Every ten seconds, the application captured one image and recorded one second of audio. The phone logged location information from the GPS receiver or the cell tower ID to which it was connected [1] Each image was tagged with a timestamp. At the time of capture, the user could also apply a text tag that indicated the activity and location of the user.

Recorded images, audio, and context information were automatically uploaded to a personal data store, which in this case was UCLA's SensorBase [16]. SensorBase is a sensor data archive service that enables users to have a personal repository that can only be viewed by that certain user and a shared repository that can be viewed by a larger community of users.

3.2 Study

This pilot study was performed by a group of six users for a period of two weeks, recording a maximum of twelve hours per day; the participation level varied by user. Post-deployment interviews yielded a number of system enhancements for future studies. For instance, users suggested alternative positioning of the cell phone to get a better view on the actual food being consumed and changes to the user interface (UI) to provide feedback about whether images were being taken and how they looked.

3.3 Results

Figure 2 contains examples of images that were collected while running the pilot study, which raised several issues with

[1] Global System for Mobile Communications (GSM) cell id.

the collection and analysis of images. Capturing an image every 15 seconds quickly creates a large collection of images. In a twelve hour period, a user could amass 2,880 images, the vast majority of which are unrelated to dietary habits. Furthermore, images that are related to diet have other objects or attributes that reveal information about the user or his surroundings. For instance, faces of individuals that surround the user while s/he is consuming food, as well as personal data such as material that is being read or snapshots of the computer screen, are often present in images of food. Many images are repetitive when the carrier is not moving and the environment is relatively static; images are also often too blurry or dark to be useful.

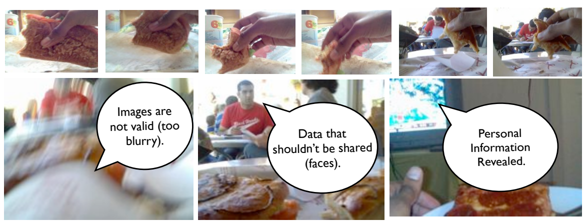

Figure 2. DietSense example images.

4 Design of ImageScape

Our pilot data collection "campaign" confirmed what we had proposed in the original DietSense architecture: New tools are necessary, for both users providing data and for the health professionals receiving the data, to interactively manage the large number of images that were being collected. To address the issues related to image management, we created ImageScape, the first of a suite of application modules to support the DietSense architecture. ImageScape works as a thin-client platform that can be initiated on a local image store or on a larger set of remotely-stored images that needs to be navigated quickly. It has an Adobe Flash front-end to a set of web services that interact with a database backend. In the DietSense application, ImageScape will be used for images related to food consumption, but we designed the tool to be able to work with a wider set of image domains. Figure 3 contains the UI for the ImageScape application. Currently it supports browsing information using tags, image features and other context information, inspired by work in [17, 18, 19, 20].

To help the user browse and tag photos more effectively, ImageScape organizes images by both linked metadata (as context) and features calculated by image processing techniques. Our approach draws from existing research in image processing, detailed below, as well as other picture browsing research, such as time- and color-based clustering in [14], photo library visualization methods and boolean search across many attributes [21], and time-based clustering of personal photos [15]. Furthermore, ImageScape uses information regarding the physical context (decibel level of captured audio from the environment) to provide an additional parameter to browse the image data.

4.1 Context

Reviewing the image collections generated by the DietSense campaign confirmed that organizing captured images using context information would make navigation more effective and help enable the images to be used as "data". Thus, in addition to automatically capturing images at a certain rate, context information was also logged. Currently, the Python client software on the phone records three types of context information:

- *location*, including GSM cell identifier, and GPS information via a Bluetooth GPS receiver;
- *time*, including time stamps from GPS receiver and also network time from the phone's cell service
- *physical environment*, including audio samples by using the microphone

Using location, time, and audio samples helped in organizing images that have been gathered. Clusters based on location and time were created for the images. This enabled us to create UI components that let users to browse by particular locations ("home" and "work") and search between time intervals. The decibel level was calculated from the audio samples so images can be organized based on varying loudness. Thus, there are clusters for loud and quiet situations. In the case of DietSense, typically users remember the location and time associated with food consumption and preparation. These context clues help reduce the number of images that need to be browsed for sharing.

4.2 Image Processing

In addition to browsing based on context, ImageScape provides a platform to create image filter components that can be linked together to enable various views on the image data. A set of filters can be combined together, in a stack, to promote smart browsing of the images for sharing and tagging purposes. Currently, this functionality exists in the backend but we plan to enable users to build, manipulate, and save stacks using the UI. We will explore three types of image filters: features, comparisons, and classifiers. Features are attributes that represent an image; comparisons enable us to compare features; classifiers help in clustering or grouping of images using comparison metrics. Table 1 contains filter components that we plan to have.

Features	Comparisons	Classifiers
Histogram	K-L Divergence	Clustering
Edge	Jetson Shannon	Bayesian
Average	Euc. Distance	Decision Tree
Wavelet/FFT	Max Likelihood	Markov Model
Haar Features		Neural Network
Mix. Models		Linear Classifier

Table 1. Image filter components.

We define a standard interface for the different components, which includes the data types that can be used as the outputs and inputs. In the future, the UI will support specifying the underlying "wiring" between blocks for the application. Users will be able to get snapshots of results at the various end points of the components to see the images associated with the filter stacks.

The first version of ImageScape explores standard image processing techniques, including dominant color, edge detection, and histogram Kullback-Leibler (K-L) divergence, to fil-

Figure 3. ImageScape view with query interfaces, feature graph, and feature thumbnails.

ter images and enable easier browsing of the image data. In order to quantify how well these filters work in helping the user audit images, we analyzed a days worth of collected images of a particular user, which was approximately 2500 photos.

4.2.1 Dominant Color Analysis

The dominant color analysis [22] of the images was very affective when places have a distinct color associated with them. For instance, images that were primarily blue or green corresponded to periods where the user was outdoors during the daytime. Also, we were able to find pictures related to food consumption by using the fact that the user was in a area that had a distinct shade of white when he was eating. Finally, dominant color could be used to find images outdoors at night easily.

4.2.2 Edge Detection

All 2500 images were run through the Roberts cross edge detection algorithm [23]. The algorithm performs a simple, quick to compute, 2 dimensional spatial gradient measurement on the images based on a specified threshold value. The threshold value varies based on the image set, so we plan to expose setting this value through a UI component. The output of the algorithm is a black and white image where black pixels represent the regions of high spatial gradient (edges). The "edginess" of an image was calculated by counting the number of black pixels in each image. Using the algorithm, over 350 images were classified as being images that were considered not of interest. These images contained homogeneous environments, such as walls, empty desks, pictures of the floor, and were too blurry or too dark. In terms of DietSense, images with low edge structure were typically not images related to diet.

4.2.3 Color Histogram

Finally, the K-L divergence of the RGB histograms of the images was used to bin related photos into clusters, a technique inspired by [14]. The algorithm sequentially processes images based on their timestamp, and clusters are identified by grouping images based on similarity in terms of a bin's average KL divergence. Clusters will contain images within a window of 200 images to take advantage of the temporal nature of the image time series. Small buckets are discarded and images which do no end up in a bucket are placed in one during a second pass. Overall, the set of 2500 images were binned into 150 sets of clusters. Each cluster contained at least two images. There were over thirty clusters that contained 20 or more images. Currently, the image with the most edginess is chosen to represent each cluster in the UI, based on the idea that an image with more structure will likely be more memorable to the user. We plan to explore other techniques to find the image, or set of images, that represent a particular bin and also explore the use of hierarchical clustering techniques.

4.3 User Interface

ImageScape was conceived as a set of common web services supporting various front end UIs for different users, including technology researchers, domain specific researchers or practitioners, and end-users. Thus, we try to make the user interface fairly intuitive while still enabling users to have advanced views and capabilities. We employ linked graphics, image statistics and graphs that are updated based on user input, as well as various query methods, such as the ability to search on time, locations, and activities. Also, our drag-and-drop interface simplifies selection of which images to share.

5 Conclusions and Future work

Dietary patterns have been recognized as contributing factors to many chronic diseases. Current methods of logging dietary habits, which typically involve keeping daily journals, have problems with accuracy, missing context data, and scalability. With DietSense, we suggest that mobile phones may provide a means to more accurately and easily document dietary choices over larger populations. In addition to their broad availability, mobile phones can collect images with context information, such as location, time, and audio and autonomously upload the data to a repository.

Based on our initial prototype and pilot of DietSense, we recognized that tools to manage, share, and analyze data are essential. These tools need to be available during all phases of data collection and review. As a first step towards this goal, we focused on creating a tool for users to review and audit gathered image data. We created ImageScape for this task and enabled concepts such as clustering based on context, organizing similar images, and being able to filter based on image features.

Our current prototype is a simplified version of the overall DietSense architecture. We mainly focused on making diet image capture and auditing easy for the user. In the near term, we plan to investigate the use of more context information, such as the motion or physiological state of the user, to enable more intelligent views to browse the image data. Furthermore, there is some refining needed to our current image processing techniques to incorporate items in Table 1 and more advanced techniques such as text and face recognition. Finally, we plan to add the ability to annotate the diet image capture with voice recordings from the user to easily provide context tags which can be used to organize the information further. Overall, with ImageScape as a prototype model, we hope to build tools to enable users to quantify sensor information in an effective manner. We believe that these types of tools are not only necessary in the DietSense architecture but are essential for sensor systems more widely.

6 Acknowledgments

Dr. William McCarthy from the School of Public Health and Drs. Karol Watson, Preethi Srikanthan, and Tony Kuo from School of Medicine at UCLA provided input in understanding current problems in dietary monitoring. Also, Dr. Mani Srivastava from Electrical Engineering at UCLA helped in refining the architecture of DietSense. Mobile phones were provided by the Nokia SensorPlanet program.

7 References

[1] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and MB Srivastava. Participatory Sensing. *WSW at Sensys*, 2006.

[2] M. Srivastava, M. Hansen, J. Burke, A. Parker, S. Reddy, G. Saurabh, M. Allman, V. Paxson, and D. Estrin. Wireless Urban Sensing Systems. Technical report, Report 65, CENS, UCLA, Apr. 2006., 2006.

[3] A. Schatzkin, V. Kipnis, R.J. Carroll, D. Midthune, A.F. Subar, S. Bingham, D.A. Schoeller, R.P. Troiano, and L.S. Freedman. A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study. *J of Epidemiology*, 32:1054–1062, 2003.

[4] KS Kubena. Accuracy in dietary assessment: on the road to good science. *J Am Diet Assoc*, 100(7):777–83, 2000.

[5] AA Stone, S. Shiffman, JE Schwartz, JE Broderick, and MR Hufford. Patient compliance with paper and electronic diaries. *Control Clin Trials*, 24(2):182–99, 2003.

[6] C.H. Kaczkowski, P.J. Jones, and J. Feng. Four day multimedia diet records underestimate energy needs in middle aged and elderly women as determined by doubly-labeled water. *Journal of Nutrition*, pages 802–805, 2000.

[7] D.H. Wang, M. Kogashiwa, and S. Kira. Development of a New Instrument for Evaluating Individuals' Dietary Intakes. *JADA*, pages 1588–1593, 2006.

[8] A. Farmer, O. Gibson, P. Hayton, K. Bryden, C. Dudley, A. Neil, and L. Tarassenko. A real-time, mobile phone-based telemedicine system to support young adults with type 1 diabetes. *Informatics in Primary Care*, 13(3):171–178, 2005.

[9] H.S.P.D. Kim, N.C.P.D. Kim, and S.H.P.D. Ahn. Impact of a Nurse Short Message Service Intervention for Patients With Diabetes. *Journal of Nursing Care Quality*, 21(3):266–271, 2006.

[10] DA Williamson, HR Allen, PD Martin, AJ Alfonso, B. Gerald, and A. Hunt. Comparison of digital photography to weighed and visual estimation of portion sizes. *J Am Diet Assoc*, 103(9):1139–45, 2003.

[11] I.T.L. Lillegaard, N.C. Øverby, and L.F. Andersen. Can children and adolescents use photographs of food to estimate portion sizes? *European Journal of Clinical Nutrition*, 59:611–617, 2005.

[12] S. Hodges, L. Williams, E. Berry, S. Izadi, J. Srinivasan, A. Butler, G. Smyth, N. Kapur, and K. Wood. SenseCam: a Retrospective Memory Aid. *Ubicomp*, 2006.

[13] HP Laboratories. The Casual Capture Prototype. http://hpl.hp.com/, 2004.

[14] J.C. Platt, M. Czerwinski, and A. Brent. Phototoc: Automatic clustering for browsing personal photographs. *IEEE PCM*, 1:6–10, 2003.

[15] A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd. Time as essence for photo browsing through personal digital libraries. *ACM Digital libraries*, pages 326–335, 2002.

[16] K. Chang, N. Yau, M. Hansen, and D. Estrin. Sensorbase–a centralized repository to slog sensor data. In *DCOSS/EAWMS*, June 17 2006.

[17] J.C. Platt, M. Czerwinski, and B.A. Field. PhotoTOC: Automatic Clustering for Browsing Personal Photographs. *IEEE PCM*, 2003.

[18] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood. Does organisation by similarity assist image browsing? *SIGCHI*, pages 190–197, 2001.

[19] D. Zhong, H.J. Zhang, and S.F. Chang. Clustering methods for video browsing and annotation. *Proc. SPIE*, 2670:239–246, 1996.

[20] Yahoo Inc. FlickR. http://flickr.com/, 2007.

[21] H. Kang and B. Shneiderman. *Visualization Methods for Personal Photo Collections: Browsing and Searching in the PhotoFinder*. Human-Computer Interaction Laboratory, University of Maryland, 2000.

[22] BS Manjunath, J.R. Ohm, VV Vasudevan, and A. Yamada. Color and texture descriptors. *Circuits and Systems for Video Technology, IEEE Transactions on*, 11(6):703–715, 2001.

[23] J.R. Parker. *Algorithms for Image Processing and Computer Vision*. Wiley. New York, NY, USA, 1996.