Automatic Image Capturing and Processing for PetrolWatch

Yi Fei Dong[1] Salil Kanhere[1] Chun Tung Chou[1] Ren Ping Liu[2]

[1] University of New South Wales, Australia
{ydon,salilk,ctchou}@cse.unsw.edu.au
[2] ICT Centre, CSIRO
ren.liu@csiro.au

Technical Report
UNSW-CSE-TR-1109
May 2011

THE UNIVERSITY OF
NEW SOUTH WALES

School of Computer Science and Engineering
The University of New South Wales
Sydney 2052, Australia

Abstract

In our previous work [5], we proposed a Participatory Sensing (PS) architecture called PetrolWatch to collect fuel prices from camera images of road-side price board (billboard) of service (or gas) stations. A key part of the PetrolWatch architecture, and the main focus of this paper, is the automatic billboard image capture from a moving car without user intervention. Capturing a clear image by an unassisted mobile phone from a moving car is proved to be a challenge by our street driving experiments. In this paper, we design the camera control and image pre-selection schemes to address this challenge. In particular, we leverage the GPS (Global Positioning System) and GIS (Geographic Information System) capabilities of modern mobile phones to design an acceptable camera triggering range and set the camera focus accordingly. Experiment results show that our design improve fuel price extraction rate by more than 40%. To deal with blurred images caused by vehicle vibrations, we design a set of pre-selection thresholds based on the measures from embedded accelerometer of the mobile phone. Our experiments show that our pre-selection improves the system efficiency by eliminating 78.57% of the blurred images.

1 Introduction

The technology of Participatory Sensing (PS)[4] is changing the way information is collected and shared. PS takes advantage of the modern mobile devices such as smart phones and PDAs to form a participatory sensor networks, where the mobile devices work as sensors to gather and share ambient information.

We proposed a PS architecture called PetrolWatch in [5] that allows volunteers to automatically collect, contribute and share fuel price information, thus lowering the cost or barrier for sharing. In PetrolWatch, camera is utilized as vision sensor to capture fuel price billboard images. These images are transported to a central server where the computer vision algorithm is implemented for further fuel price billboard detection and fuel price extraction. In the PetrolWatch prototype of [5], pictures were captured manually by a person focusing on the fuel price billboard from a stationary position under ideal environment.

In this paper, we focus on the challenging task of automatically capturing images from a moving car without user intervention. The success of such task highly relies on the quality of the collected images and the design of the computer vision algorithms. However, capturing a clear image automatically in a moving vehicle is a challenging task: Firstly, the default auto focus function does not work due to the insufficient focus time. Secondly, the distance from the camera to the billboard plays an important role for the collected image's quality. Thirdly, vibration may damage an image even when all other conditions are perfect.

In this paper, we design the camera control and image pre-selection schemes to address the above challenges. Extensive experiments were conducted in suburban areas of Sydney to validate and optimize our designs. In particular, we leverage the GPS (Global Positioning System) and GIS (Geographic Information System) capabilities of modern mobile phones to design an acceptable camera triggering range and set the camera focus accordingly. Experiment results show that our design improve fuel price extraction rate by more than 40%. To deal with blurred images caused by vehicle vibrations, we design a set of pre-selection thresholds based on the measures from embedded accelerometer of the mobile phone. Our experiments show that our pre-selection can eliminate 78.57% of the blurred images with minimum processing. This improves the system efficiency by saving communication cost and local resources.

The rest of the paper is organized as follows. In the next section, we briefly review the PetrolWatch system design. In Section 3, we describe the automatic image collection designs. We investigate the camera triggering distance in Section 3.1, followed by a discussion on camera focusing methods in Section 3.2. In Section 3.3 we describe the image pre-selection method to eliminate blurred images. The performance evaluations are presented in Section 4. The related work is discussed in Section 5. We conclude our work in Section 6.

2 System Overview of PetrolWatch

PetrolWatch[5] is an automatic fuel price collection system that has two modes of operations: (i) fuel price collection and (ii) user query. Fig.2.1 presents a pictorial overview of PetrolWatch. As depicted in the picture, the data collection process involves three steps: (i) capturing images of the fuel price billboards

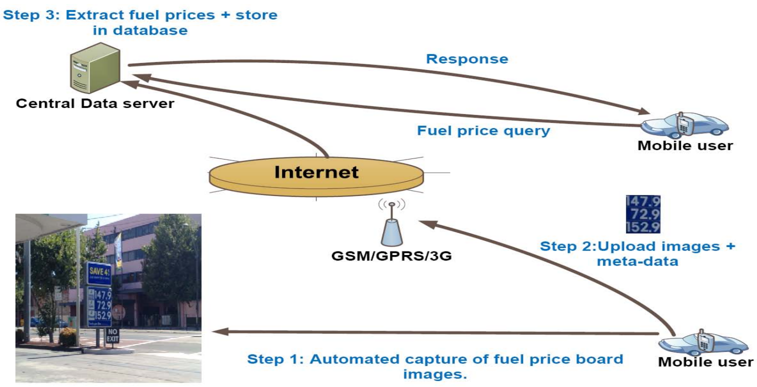


Figure 2.1: System Overview of PetrolWatch

(a) The Collected Image (b) Detected Fuel Price Board (c) Final Price

Figure 2.2: PetrolWatch's screen shots

and extract the context information (location coordinates, brand and time) as metadata (ii) uploading the images and their metadata to the central server and (iii) extracting fuel prices from the images . Each of these tasks is executed by a distinct component of the system. In the PetrolWatch prototype of [5], step (i) was conducted manually under ideal conditions. In this paper we automate this image capturing and content extracting process. Moreover, extensive experiments were conducted in practical suburban street environment in a moving vehicle.

3 Automatic Image Collection Designs

The automatic Image collection designs for PetrolWatch that include: camera's trigger distance, camera's focusing method and the blurred images' elimination mainly based on the extensive real experiments' result. We choose Noika N95 as the mobile phone hardware. The automatic image collection work flow of PetrolWatch is as the following: when the vehicle is within a proper distance to a service station, PetrolWatch automatically trigger the camera, and capture sequences of images. At the same time the accerlerometer is triggered to record the vibration information experienced by the camera when the images are captured that will be used to eliminate the blurred images detailed in Section 3.3. The time lag between each captured image is 1.3 second which is decided by the mobile phone hardware mechanism. Due to the long storage time of the inter-

Figure 3.1: Experiment Set Up

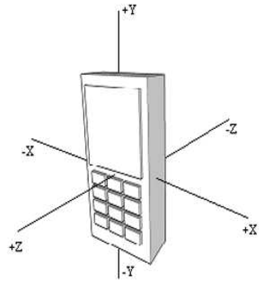


Figure 3.2: 3-axis of N95 Accelerometer

nal memory of N95 and the limited time before passing the service station, the captured images along with the context information such as: the accerlerometer readings, service station's brand, time slot and distance to the approaching service station etc... are temporally stored in the RAM of N95. After the car has passed the service station, the camera and the accerlerometer are turned off. The captured images and the context information are then saved into the internal memory for the further post-processing.

We implemented a prototype of PetrolWatch in Nokia N95 by using Symbian C++. Fig.2.2 shows the screen shots of PetrolWach implementation. Fig.2.2(a) is a collected image that includes a fuel price billboard. PetrolWatch first detect the existence of the fuel price billboard. The detected fuel price billboard is shown in Fig.2.2(b). Finally the detected board is passed through a Feedforward Back propagation Neural Network to classify the digits. Fig.2.2(c) is the extracted fuel price. The hardware installation of PetrolWatch is as shown in Fig.3.1 that N95 is held horizontally on the front dash screen of a car by a mobile phone car holder. The lens of the camera is orientated to left side due to the fact that we drive on the left side in Australia.

3.1 Trigger the Camera

At which distance should the camera be triggered? If the distance is too far, there are excessive images for processing among which many of them are possibly without fuel price billboard or the characters are too small for recognition. On the other hand, if the distance is too close, the number of captured images is limited. As a result, the camera should be triggered at the largest distance within which the captured images can be used in PetrolWatch. Whether the images are suitable to the PetrolWatch or not is decided by the system performance of the captured images after applying the computer vision algorithm of PetrolWatch. The evaluation criteria of the system performance include: the fuel price Billboard Hit Rate (BHR), fuel price Character Classification Rate (CCR) and fuel Price Classification Rate (PCR). The definition of them are as the following:

$$
\begin{aligned}
BHR &= \frac{Number\ of\ correctly\ detected\ fuel\ price\ board}{Number\ of\ total\ fuel\ price\ board} \\
CCR &= \frac{Number\ of\ correctly\ classified\ characters}{Number\ of\ total\ characters} \\
PCR &= \frac{Number\ of\ correctly\ classified\ prices}{Number\ of\ total\ prices}
\end{aligned}
\tag{3.1}
$$

Note here, PCR is more strict than CCR. This is due to the fact that only when all the characters of a price are classified correctly, the price is considered as correctly recognized. These three evaluation factors will be used through this paper as the evaluation criterions.

Since the location coordinates of the camera and the approaching service station are available from the GPS receiver and GIS database, the distance between them can be easily calculated. In PetrolWatch this is realized by the haversine formula [12]. One of advantages of the haversin formula is particularly well-conditioned for numerical computation even at small distances which fits our mobile phone based system. If two points are represented as $P1 = (Lat1, Long1), P2 = (Lat2, Long2)$, the distance calculation by haversine method of these two points can be presented as the following:

$$
\begin{aligned}
&a = \sin^2(\tfrac{\Delta Lat}{2}) + \cos(Lat1) \times \cos(Lat2) \times \sin^2(\tfrac{\Delta Long}{2}) \\
&c = 2 \times atan2(\sqrt{a}, \sqrt{(1-a)}) \\
&d = R \times c
\end{aligned}
\tag{3.2}
$$

where $\Delta Lat = Lat1 - Lat2, \Delta Long = Long1 - Long2$ and R is the average radius of earth that is assumed to be 6,378.8 kilometers.

In respect the fact that the experiments of PetrolWatch should be conducted really in the street which involve extensive driving, we combined this experiments to decide camera trigger distance and the one to select the focusing method for the camera that details will be investigated in the next section into one experiment.

The experiment is conducted as follows: PetrolWach trigger the camera at a large distance (200 m) far away from the service station and capture sequences of images tagged with the distances to the service station. Since the sequences of images were captured while moving, it's hard for us to capture the image at an exact distance. Moreover, there exists GPS errors within the distance calculation that is approximately 5 m of true position in open sky setting[13], we sort the captured images with the value difference of the tagged distance less than 10 m (double of the GPS position error) into one group. For example, the image with a tag distance of 32.8 m and the image with a tag distance of 37.3 m are put into the image group of 30-40 meters range. We did this experiments 30 times with the camera pre focused detailed in the next section, focused by the auto focus function and un focused respectively.

The computer vision algorithms in [5] are then applied to such images to obtain the BHR, CCR, and PCR for each distance range. In order to keep the balance of avoiding excessive image processing and capturing enough images, the proper camera trigger distance is choose as the one with the largest distance value and beyond which the BHR, CCR, and PCR are smaller than an acceptable level.

From the experiments' result, we found the performance of the camera pre focused is much better than the one un focused. In the next section, we will investigate the importance of the focusing method of the camera.

3.2 Focus the Camera

When capturing a photograph manually, the photographer must first focus on the intended target object, either manually or use the auto-focus function provided in most cameras. However, in PetrolWatch, the mobile phone automatically captures images from a moving car. This makes it very difficult for the camera to focus correctly on the target object, which in our case is the fuel price billboard. Note that, even the auto-focus function requires that the camera lens is held static pointing towards the object to allow the electronics in the camera to intelligently determine the correct focal distance. Thus, the default auto-focus function does not work in this scenario due to a moving camera and the fast changing focus distance. And the camera un focused has the focus distance as infinity that is different from the real distance from the fuel price billboard to the camera. It does not work as well. To deal with such problem, we estimate a certain focus distance that is similar to the distance from the fuel price billboard to the camera and let the camera get focused by using such distance before the car started. In the experiment, we park the car in the left most lane and focus the camera by using the shops or buildings in the camera view. This works due to the fact that the offset of the shops or buildings in suburban area is similar, hereby the focus distance is similar in each run. In the system design, the estimated distance value can be set to 30 meters that is the place with the highest fuel price classification rate from the trigger distance experiment result in Section 4. As we mentioned earlier, the experiment to select the camera focus method is combined with the one to decide the camera trigger distance. In our experiments, we compare three different ways to focus the camera: (1) the default auto focus function (2) without focus at all and (3) by using an chosen distance and we will refer to this method as pre-focusing. Given that our goal is to compare different method of focusing, we need to all other variables (such as the amount of day light) are almost the same, thus making the method of focusing the main variable. We did that by conducting the experiments in the same day under similar environmental conditions.

3.3 Eliminate the Blurred Images

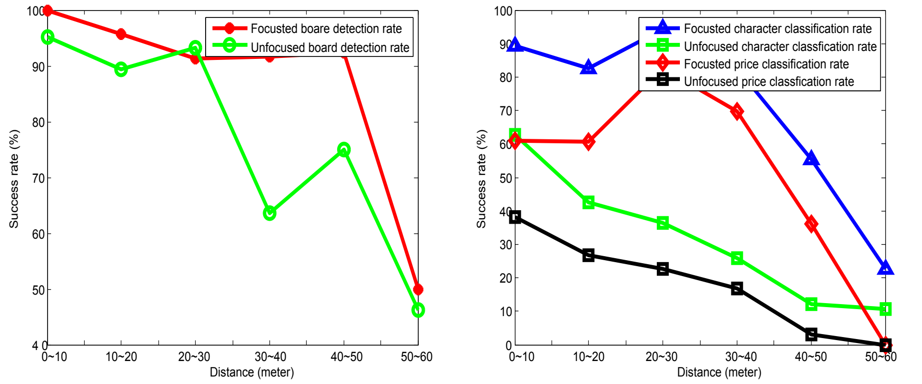
While we have investigated the trigger distance and the focus of the camera to capture a clear image, there still exists some images possibly blurred due to the automatic image collection schedule and the motion of the camera. Obviously, such blurred images should be eliminated immediately rather than let go through the next stage of transmission, and processing. By eliminating these blurred images through a simple pre-selection process, resources such as memory space, transmission cost, or local image processing can be avoided.

In this section, we discuss the blurred images' elimination of PetrolWatch that is based on the vibrations measured by the accelerations in X-Y-Z directions. Through experiments, we find the thresholds values in the X-Y-Z accelerations beyond which the image is deemed to be blurred. These thresholds are then built into PetroWatch for the pre-selection process. The ground truth of the images blurred or not is classified by human vision instead of by the computer vision algorithm, this is because of the dependance between the computer vision algorithm and the quality of the captured images each other. In PetrolWatch, we try to capture a clear image recognizable by human vision firstly, and based on the captured images, we design the respective computer vision algorithm to mimic human's recognition behavior and try the best to catch the performance of human beings. The computer vision algorithms used in the previous sections are designed based on the clear images classified by human.

We assume the users participates in PetrolWatch by holding the mobile phone as we mentioned in the experiment set up that the lens oriented to the left side. PetrolWatch utilizes Nokia N95 8G as the prototype hardware which embedded with a 3-axis accelerometer from STMicroelectronics (type LIS302DL). The gather of raw sensor data use an interrupt-event-driven sampling method with the sampling frequency of 20Hz. The official explanation of 3-axis (X, Y, Z) of Nokia N95 8G's accellerometer from Nokia Forum [2] is presented in Fig.3.2. Recall, Fig.3.1 is the mobile phone's mount method of PetrolWatch that N95 8G is mounted on the windscreen horizontally. As a result, in PetrolWatch, X axis is up and down fluctuation; Y is the left and right acceleration; and Z is the forward and backward acceleration. Since the accelerometer reading is quite sensitive, we smoothed the raw readings by using moving average method with the average buffer size of 20 samples. By checking the accelerometer readings tagged in the captured images, we find there is not much fluctuation in the X-axis (the up and down direction) and it's hard to find a direct relationship between the image quality and the value of X-axis. This is due to the good quality of Australia's road. As a result, due to the page limitation, we don't show the result of X-axis in this paper. Fig.4.2 is the captured images ' acceleromerter readings in Y-axis and Z-axis. For Y-axis, the positive value is right side acceleration and the negative value is left side acceleration. For Z-axis, the positive value is forward acceleration and the negative value is backward acceleration. In Fig. 4.2, there are few samples with negative values in Y-axis and Z-axis. This is due to the fact we drove at the most left lane when we did the experiment, and when we try to change lane, we always turn right and accelerate. From Fig. 4.2, We can find the blurred images arise in the area of $Y > 6$ and $Z > 55$. This means when the car turns right and increase the speed, the captured images are probably blurred. However, as we mentioned we drove in the most left lane, and when we try to change lane we always accelerate to overtake others. As a result, the blurred images are mainly caused when we change lane. We can get the conclusion that if $|Y| > 6$, or $|Z| > 55$ (it doesn't matter turn right or left, accelerate or decelerate), the captured images are possibly blurred.

4 Performance Evaluation

Extensive experiments were conducted by driving around suburban areas in Sydney with a mobile phone fixed to the windscreen of the test car. The mobile phone is installed with our automatic image capturing embedded software. During the experiments, the software control the target parameters, such as triggering distance, focus, and pre-selection thresholds, to validate and optimize our designs. Fig.4.1 is the experiment result of BHR, CCR, PCR at different distance based on 30 images within each distance range. In Fig.4.1(a), the BHR



(a) Board Detection Rate Vs. Distance

(b) Character and Price Classification Rate Vs. Distance

Figure 4.1: The fuel price billboard detection rate and character, price classification rate obtained at different distance. At each distance, the result is based on 30 automatic collected images.

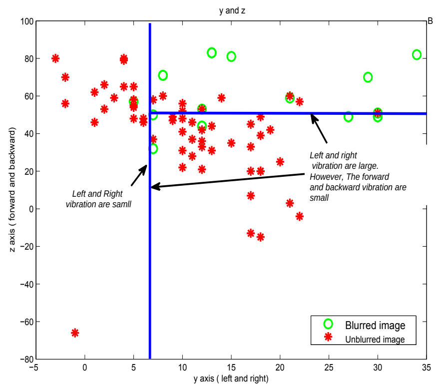


Figure 4.2: Y-axis and Z-axis of N95 accelerometer readings

within 50 meters keeps above 90%. However, it drops to around 50% from 60 meters. As a result, the largest distance for an acceptable BHR is 50 meters that means the images captured beyond 50 meters are hard to detect the board. Applying the same analysis on Fig.4.1(b), we find the largest distance for an acceptable CCR is 40 meters that means the characters in the images captured beyond 40 meters are hard to recognized. By combining these two results, the distance to capture clear and enough number images is at 40 meters. As a result, the camera should be triggered when it's within the area away from the approaching service station 40 meters. In Fig.4.1(b), the highest CCR and PCR are at the distance of 30 meters. As a result, 30 meters will be set as the estimated focus distance for the pre-focusing of the camera.

Now, for PetrolWatch, the focus methods for the camera to select includes: (1) the default auto focus function (2) without focus at all (3) by using our proposed pre-focusing by setting an estimated focus distance. Fig.4.1 is the experiment result of camera focus method comparison as well. The result is based on 30 images captured by the camera auto focused, un focused and pre-focused respectively at different distance range. Since the performance of the images captured by the camera with auto focus function is too poor to compare with the other two focus methods. In Fig.4.1, we only showed the result of camera pre-focused and un focused. As a result, in Fig. 4.1, focused camera represents the camera focused by the pre-focusing. From Fig. 4.1, we can find, BHR of pre-focused camera is higher than the one of unfocused camera 5% averagely at all distances ranges. Compared with unfocused camera, CCR of pre-focused camera averagely increased from 40% to 80% and PCR of focused camera averagely increased from 20% to 70%. Hereby, experiment results show that our design improve fuel price extraction rate by more than 40%.

We tested 138 images among which 28 blurred images and 110 unblurred images by using our image pre-selection method. 22 blurred images were successful eliminated. The success rate is 78.57%. 5 un blurred images were wrongly eliminated. The false positive rate is 4.5%. As a result, our image pre-selection method effectively eliminate the blurred images by sacrificing a relative low miss eliminated unblurred images.

5 Related Work

There already arose many GPS related PS applications. NoiseTube[7] and Earphone[8] utilize the GPS-equipped mobile phones to measure participating users' personal exposure to noise in their everyday environment. Each user contributes by sharing their geo-localised measurements and further personal annotation to produce a collective noise map. MobiSense[11] and [15] estimate the traffic profile and users' mobility by analyzing the mobile phone embedded GPS receiver readings. In the above work, GPS is utilized as the annotation information or as the analysis data. Different from the usage of them, PetrolWatch utilizes GPS both as the geo-annotation information of the extracted fuel price and the coordination information to calculate the camera trigger distance. Accerolermeter is widely used in PS applications as well. [9] and [14] use mobile phone embedded accerolermeter with GPS to recognize physical motion modes. And accerolermeter is used to monitor road surface condition in [6]. The only difference between the usage of accerolermeter in PetrolWatch and its usage as mentioned above is the application scenario. Accerolermeter is used to eliminate the blurred images in PetrolWatch. A particular sensing modality of PS is image since it can provide a visual representation of the urban landscape at specific location and time slot which can be considered as the most intuitionists information. Google Streetview [1] allows users both browse the street view of a specific location and contribute images with buildings, natural sight or even their personal activities near the location to share with others; SnapScout[3] is a crime prevention application. It allows and encourages users to snap any suspicious people or item in their neighborhood and send the snapped images to a central server where the trained security professionals review the submitted photos for possible illegal activity; Dietsense [10] provides participants uploading images of their daily meals annotated with the description information such as voice or text messages onto a server. Thereby, the dietary specialists will provide further analysis and give some comments on their diet. However, most of existing image related PS systems rely on significant amount of users' intervention to complete data collection. They either requires user take the images, input the annotation for the image or identify objects or patterns from collected images. PetrolWatch is a unmanned image based PS system which sensors/collectors are automatically triggered to take samples/pictures, and these pictures are processed automatically to extract relevant information.

6 Conclusion

In this paper, we mainly focus on dealing with the major challenges related to automatic best effort image quality images' collection for PetrolWatch. The proper camera triggering distance, focus method, and blurred images' elimination are investigated through real implementation and experiment result is demonstrated in this paper as well. Experiment results show that our image collection design improve fuel price extraction rate by more than 40% and our image pre-selection scheme can eliminate 78.57% of the blurred images.

Bibliography

[1] Google street view. http://www.google.com/intl/en_us/help/maps/streetview/.

[2] Nokia forum. http://library.forum.nokia.com.

[3] Snapscouts. http://snapscouts.org/.

[4] Jeffrey A. Burke, D Estrin, Mark Hansen, Andrew Parker, Nithya Ramanathan, Sasank Reddy, and Mani B. Srivastava. Participatory sensing. In *WSW06 at SenSys 06*, New York, NY, USA, October 31, 2006. ACM.

[5] Y. F. Dong, S. Kanhere, C. T. Chou, and N. Bulusu. Automatic collection of fuel prices from a network of mobile cameras. In *Distributed Computing in Sensor Systems*, pages 140–156, 2008.

[6] Jakob Eriksson, Lewis Girod, Bret Hull, Ryan Newton, Samuel Madden, and Hari Balakrishnan. The pothole patrol: using a mobile sensor network for road surface monitoring. In *Proceeding of the 6th international conference on Mobile systems, applications, and services*, MobiSys '08, pages 29–39, New York, NY, USA, 2008. ACM.

[7] Nicolas Maisonneuve, Matthias Stevens, Maria E. Niessen, and Luc Steels. Noisetube: Measuring and mapping noise pollution with mobile phones. In Ioannis N. Athanasiadis, Pericles A. Mitkas, Andrea Emilio Rizzoli, and Jorge Marx Gmez, editors, *ITEE*, pages 215–228. Springer, 2009.

[8] Rajib Kumar Rana, Chun Tung Chou, Salil S. Kanhere, Nirupama Bulusu, and Wen Hu. Ear-phone: an end-to-end participatory urban noise mapping system. In *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, IPSN '10, pages 105–116, New York, NY, USA, 2010. ACM.

[9] Sasank Reddy, Min Mun, Jeff Burke, Deborah Estrin, Mark H. Hansen, and Mani B. Srivastava. Using mobile phones to determine transportation modes. *TOSN*, 6(2), 2010.

[10] Sasank Reddy, Andrew Parker, Josh Hyman, Jeff Burke, Deborah Estrin, and Mark Hansen. Image browsing, processing, and clustering for participatory sensing: lessons from a dietsense prototype. In *in EmNets 07: Proceedings of the 4th workshop on Embedded networked sensors*, pages 13–17. ACM Press, 2007.

[11] Sasank Reddy, Vidyut Samanta, Jeff Burke, Deborah Estrin, Mark H. Hansen, and Mani B. Srivastava. Mobisense - mobile network services for coordinated participatory sensing. In *ISADS*, pages 231–236. IEEE, 2009.

[12] C. C. Robusto. The cosine-haversine formula. *The American Mathematical Monthly*, 64(1):pp. 38–40, 1957.

[13] G. Wing, Aaron Eklund, and Loren D. Kellogg. Consumer-grade global positioning system (gps) accuracy and reliability. *Journal of Forestry*, 103(4):169–173, 2005.

[14] Jun Yang. Toward physical activity diary: motion recognition using simple acceleration features with mobile phones. In *Proceedings of the 1st international workshop on Interactive multimedia for consumer electronics*, IMCE '09, pages 1–10, New York, NY, USA, 2009. ACM.

[15] Jungkeun Yoon, Brian Noble, and Mingyan Liu. Surface street traffic estimation. In *Proceedings of the 5th international conference on Mobile systems, applications and services*, MobiSys '07, pages 220–232, New York, NY, USA, 2007. ACM.