Ikarus: Large-Scale Participatory Sensing at High Altitudes

Michael von Kaenel, Philipp Sommer, and Roger Wattenhofer
Computer Engineering and Networks Laboratory
ETH Zurich, Switzerland
{vkaenemi,sommer,wattenhofer}@tik.ee.ethz.ch

ABSTRACT

Sensor networks proved to be a useful research tool in the field of environmental monitoring. While first sensor deployments consisted of a relatively small number of static nodes, mobile sensor devices have attracted growing interest for large-scale sensing applications in recent years. In this paper, we present *Ikarus*, a novel participatory sensing application having orders of magnitude more users than existing approaches. The Ikarus system exploits sensor data collected during cross-country flights by paraglider pilots to study thermal effects in the atmosphere. Based on first experiences gained from this approach, we identify three key aspects that are crucial for the success of participatory sensing applications: incentives for participation, the ability to deal with faulty data, and concise data representation.

1. INTRODUCTION

Participatory sensing lies at the intersection of two auspicious research directions, networked sensing and group intelligence. As such, participatory sensing inherits attributes from both its relatives. The general vision is that designing, deploying and operating a sensor network may not be necessary, when we want to learn about the physical environment. Instead, we may simply tap into the sensors that people carry around anyway, for instance sensors integrated into laptops or mobile phones. Unfortunately, several problems arise when trying to build a participatory sensing application.

First, and most importantly, how do we get people to participate? One possible answer to this question is simple: By paying them! And indeed, a large fraction of group intelligence projects such as Amazon's Mechanical Turk is based on monetary exchange. However, there exist some group intelligence projects that gracefully manage to motivate the crowd to participate without any financial incentives, e.g., disguised as games, or Captchas. However, there are a few reasons why browser-based applications may be easier to incentivize participation than phone-based sensing applications, essentially boiling down to form factor, power consumption, communication bandwidth and cost.

Second, in contrast to classic sensor networks, participatory sensing applications will have to deal with false data reporting. Even though this may also be an issue with traditional sensor networks, participatory sensing will be orders of magnitude more problematic. For instance, as the sensors are owned by their individual users, they will almost surely be badly calibrated. Depending on the application and the participation incentives, one may also expect a great deal of malicious (Byzantine) data, for instance in order to beat the high score of a game.

Third, depending on the parameter space of the application, data representation may be a key aspect. Sensed data may change dramatically, it may be heavily time-, location-, temperature-, or user-dependent, just to give a few examples. In order to learn most from the available data, one will have to find ways to represent the data concisely, without too many parameters. Moreover, as users will explore their own paths, data may be dense in some areas, and sparse in others. Thus, one may need mechanisms to represent the data neutrally despite these imbalances.

In this paper, we present *Ikarus*, a participatory sensing project for sensing flight conditions (Section 2). Furthermore, we address all the three points we raised, participation incentives, false data, and data representation. In terms of incentives, Ikarus features an application-data loop that truly promotes participation (Section 3). As such, we were able to use the data of 2,331 unique users, providing a total amount of raw data in the order of several Gigabytes, which makes Ikarus one of the largest existing participatory sensing projects. We will describe several problems regarding faulty data, and how we cleaned it (Section 4). And last but not least, we give a few ideas how to represent the data concisely (Section 5), despite a rich parameter space and density imbalances.

2. THE IKARUS SENSING SYSTEM

Paragliding is a popular flying sport, where a pilot uses a harness and a fabric wing, the so called *paraglider*, to gain uplift. It is a simple and relatively inexpensive way to experience the dream of flying close to nature.

Solar radiation reaching the Earth's surface warms the ground, creating a thermal column (or short: thermal), which is a vertical section of rising air. According to specific soil properties (e.g. rock surface or forest), air close to the ground heats differently [1].

Thermals are used by paragliding pilots to gain height dur-

HotMobile 2011, March 1–2, 2011, Phoenix, Arizona.

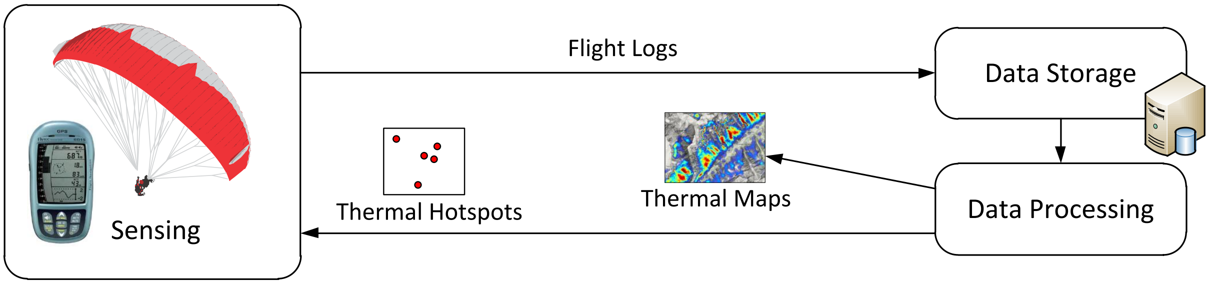

Figure 1: Overview of the Ikarus participatory sensing system. Paraglider pilots equipped with GPS-based flight navigation devices feed their flight logs into a central database in order to take part in paragliding competitions. GPS tracks are processed to generate probability maps for thermal columns. Furthermore, the coordinates of thermal hotspots can be exported to the GPS devices and used during flights.

ing cross-country flights. The paraglider soars by circling in a thermal, leaves the thermal and glides to the next thermal. While the presence of thermals is sometimes indicated by cumulus clouds at the apex of the thermal, predicting the exact location of strong thermals is difficult, even for experienced pilots.

Nowadays, atmospheric observation data are mainly gathered by sensors based on the Earth's surface or by means of single measurements using registering balloons. Therefore, wide area data sets are largely missing today. Gaining a deeper understanding of atmospheric phenomena such as the formation of thermal columns does not only help sports pilots but can be leveraged for other application domains too, e.g., increasing endurance of unmanned aerial vehicles.

In this paper, we present the Ikarus sensing system, which leverages existing flight log data collected voluntarily by paraglider pilots to generate thermal maps. The basic architecture of Ikarus is shown in Figure 1. Pilots form the mobile sensing tier by using flight navigation devices to record their current GPS position, barometric altitude and timestamp. After the flight, tracks logs are uploaded by pilots to special community websites and stored in a flight database. Records include meta information, such as the device type used for recording, and several hundred to thousands of points defining the track along with the corresponding timestamps. Tracks are stored in a well defined file format specified by the International Gliding Commission (IGC).[1]

Recorded flight logs are analyzed to gather information about thermal columns. The resulting thermal maps are then made available to the paragliding community by means of a visualization in web-based maps and Google Earth. Furthermore, thermal hotspots are extracted for visualization on the pilot's navigation device during subsequent flights.

3. PARTICIPATION INCENTIVES

The success of participatory sensing relies on an enduring commitment of volunteers towards the use of their mobile devices for data collection. While most existing initiatives in participatory sensing were initiated by research projects, it is unclear whether participatory sensing works well without continuous supervision by an organization. We argue that providing incentives to the user is crucial for the success of participatory sensing applications. Otherwise, people tend to lose their interest and quit their participation when they are not part of the feedback loop, i.e., when they have no access to the results of the measurement campaign.

Various incentive mechanisms may be successful in practice. Clearly, monetary compensation may be issued based on the quantity and/or quality of sensor data provided by the user. For example, one could pay a small amount of money to commuters participating in traffic measurement campaigns using their mobile phone's GPS. However, it is not clear if this might be feasible or beneficial for any kind of application. Paragliding, for example, is a sport that requires considerable initial investment in equipment and training. Thus, paying a small amount of cash for every uploaded track will not likely have the effect that many people start paragliding.

Instead, an alternative approach to provide incentives for participatory sensing could be based on an increased reputation in the user's social network or community. The Ikarus system shows exemplarily how this can be achieved. Cross-country paragliding pilots usually carry a flight navigation device with them during flights. These devices serve for two different purposes. First, an accurate barometric sensor and a GPS device are necessary tools for pilots to know about the current altitude and position. Second, the GPS position and altitude is recorded to prove that certain waypoints have been passed during competitions, or to analyze the flight back at home and share it with friends. Since it is not feasible to compete with a large number of pilots during a single event held at a specific location and time, the paragliding community relies on the analysis of flight logs to get a ranking of pilots. Therefore, several websites exist to collect GPS flight logs for national and international competitions.[2] Pilots upload their flight logs and get ranked in high score lists according to the distance and shape of the traveled route.

4. DATA INTEGRITY AND QUALITY

Researchers using data from Ikarus can rely on the integrity of flight data since each track is signed by the device using a private key. Thus, invalid flights can be sorted out easily. Furthermore, it is nearly impossible to upload a track log containing bogus waypoints. However, measurement errors are directly reflected in the recorded data. Each device introduces new specialties and shortcomings. Furthermore, incorrect handling of the device by the user could possibly

---

[1] http://www.fai.org/gliding/

[2] E.g. http://www.xcontest.org

introduce further measurement errors. In this section we describe our approach to detect and correct these errors in the raw sensor data.

A recorded flight is transfered to a computer by the pilot. The upload to the local hard drive is performed by aid of various tools. The users are able to trim a flight record such that it covers only the time period between launch and landing, but further transformations of the waypoints are not possible. This subset of the track is signed on the device using a private key distributed by the IGC to various device manufacturers. Nearly all observed track logs are recorded on such trusted devices, which assures data integrity. In the first step of our analysis, flight logs are checked for data consistency and integrity before they are stored in a central database. Data sets containing errors that can not be corrected are discarded. Out of more than 30,000 flight tracks collected during the years 2003 to 2010 within Switzerland, we were able to use 24,169 tracks (81.1%). Table 1 lists different type of errors due to which tracks had to be removed from further analysis.

None or invalid altitude information	7.0 %
Flight outside area of interest	4.0 %
Too many successive GPS spikes	2.5 %
Tracks far below terrain after correction	5.0 %
Other errors (night flight, altitude outliers)	0.4 %
Total removed tracks	18.9 %

Table 1: Percentage of non-recoverable errors.

4.1 Timestamping Errors

About a third of all tracklogs define an erroneous start date. In this case, the date specified manually by the pilot while uploading the file to the competition web-page was used instead. Furthermore, it is difficult to check whether the recording time specified in the file is correct. However, some obvious errors are detectable. Sometimes the time was reset to 00:00 UTC shortly before launch, while in other cases the time shift was bigger. Even though flights at night are generally possible, they are not of interest for thermal evaluation.

4.2 Positioning Errors

The most obvious and frequently encountered problems are short errors, so called spikes, in the position or altitude. Most of the time they happen at the beginning of the track during device startup (5 % of flights affected), but they can happen at any time during flight, too. In free air during flight, the device is generally well exposed to the GPS satellites. Nevertheless, the signal gets lost frequently, which is sometimes recognized by the sensor and marked accordingly in the track log. Often, strongly depending on the recorder model, the spike just appears as a valid waypoint. Spikes in elevation might be a result of a short hit to the barometric sensor, which is not uncommon during launch or landing.

Position spikes are determined according to speed. Paragliders have a very limited speed window between 30 and 60 km/h in calm air. Because the GPS sensor only covers the speed over ground, the relative speed might sink to zero, or it can rise over 100 km/h with tailwind. Spikes, on the other hand, have a much higher speed between two waypoints. Often only a single or a few of the waypoints need to be removed. We observed position spikes in 2 % of all flights, each affecting on average 1.9 track points.

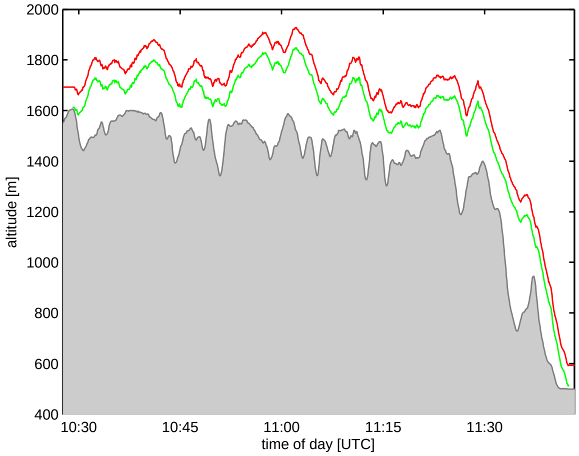

Figure 2: Start and end point of the original trace (red) do not match with the digital elevation model (gray). The corrected trace (green) is pinned to the altitude of the launch spot.

In order to detect altitude errors we proceed accordingly. The mean climb rate for paragliders in a thermal is 1.3 m/s in the observed dataset. In general, the variance is low, but climb rates might rise over 6 m/s for a short period of time. However, climb rates exceeding 20 m/s are certainly outliers and such tracks are removed.

About 20 % of all flights were not cropped properly in time to launch and landing. Right before the launch, the barometric sensors have to be calibrated, which results in position spikes. After removing trackpoints corresponding to the time before launch and after landing, only about 2.5 % of the flights showed spikes that had to be removed in order not to be interpreted as thermals.

4.3 Sensor Calibration Errors

Barometric sensors are the preferred method to measure small changes in elevation of a paraglider since they are much more accurate than the height information provided by the GPS device. However, the altitude must be calibrated before use. Some flight recorders correct the barometric height by the measured GPS height over a longer time range, but it is more common that the pilot himself must adjust the height properly. However, this is frequently forgotten by the pilot, resulting in a constant shift in altitude over the whole flight, assuming the air pressure remains constant during the flight. Since launch and landing are always on the ground, the correct altitude can be computed using a digital elevation model, as shown in Figure 2. About 20 % of all flights had an altitude that differs significantly from the height provided by the digital elevation model. On average, altitude information had to be adjusted by 59 meters.

5. DATA REPRESENTATION

Processing measurement data and drawing the right conclusions from it is an inherently difficult problem in participatory sensing. While sensor data is collected at fixed locations and at a constant rate in traditional sensor networks, participatory sensing can only provide a snapshot of

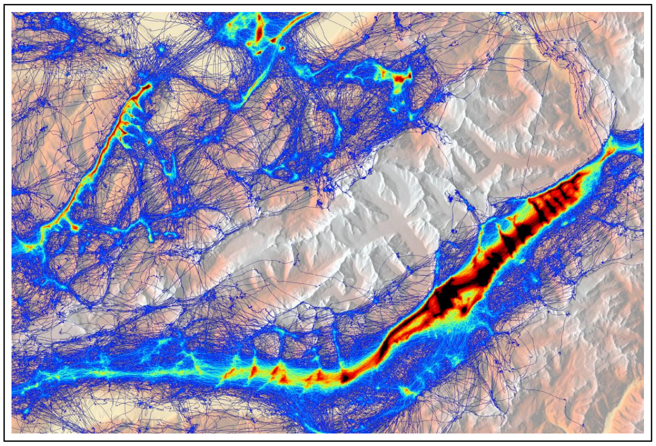

Figure 3: GPS flight tracks collected during the years 2003 to 2010 in the Goms Valley, a popular region for paragliding in the Swiss alps. The map has a dimension of 80 x 50 km.

the situation at a certain location and time. We argue that representing sensor data in a concise manner is a crucial aspect to reach acceptance by the users of the system.

In the following, we describe our first experiences with data representation in the Ikarus system. Having a large dataset of more than 24,000 valid flights at hand, we implemented different heuristics to identify geographical regions with high thermal uplift. The basic idea is to identify thermal columns based on the first order derivate of the recorded altitude. However, pilots may lose accidentally the area with stable uplift and have to start circling until they can find it again. Finally, a flight phase with thermal uplift is reduced to two points, an entry point and an exit point of the uplift. Then, the thermal uplift is assigned to a point of origin located on the Earth's surface. By the use of trigger points on the surface we can reduce the effect of wind direction and strength on thermal uplifts. A first analysis revealed that both the time of year and the time of day have a large influence on the probability to find a thermal column in a certain area.

Analyzing the geographical distribution of sensor measurements revealed that flights are very unequally distributed, as shown in Figure 3. Not so surprisingly, there are certain well known routes with thousands of flights while almost no flights have been carried out in other areas. It is not clear whether these areas do not offer good flight conditions, or if they are too far off from civil infrastructure. Furthermore, since pilots choose a launch pad based on the current weather conditions, all flights in a region are performed under similar conditions.

The uneven distribution of flights has to be taken into account in order to generate thermal maps that are comparable over the whole area of interest. Thus, we assign an uncertainty value to each thermal trigger point. This uncertainty value is high for thermals that are only based on a single flight track. On the other hand, a low uncertainty value is assigned to a thermal uplift that has been confirmed through multiple independent flights. Finally, the resulting map allows us to compare the probabilities to find good thermal conditions in different geographical areas, as shown in Figure 4.

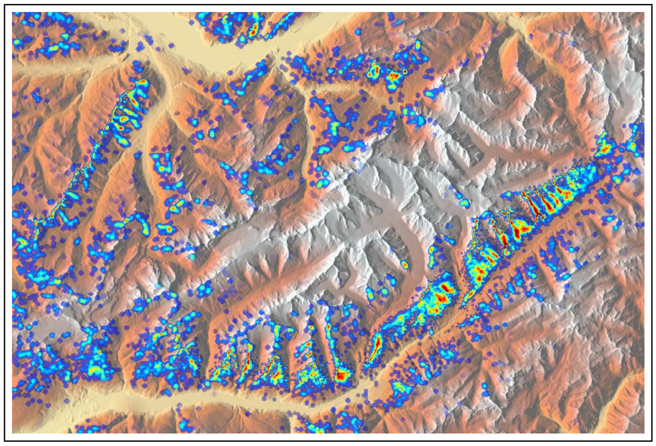

Figure 4: Resulting thermal map with a relief layer in the background. Red color indicates a high probability to spot thermal columns in that area.

6. CLOSING THE FEEDBACK LOOP

Thermal maps serve as a useful tool for pilots to prepare flights and inquire information about unfamiliar areas. As part of the Ikarus system we developed thermal maps for different data formats, as shown in Figure 5. A public web application enables to visualize thermal maps in the browser during flight preparation.[3] The application is targeted towards paraglider pilots and allows to explore maps for different seasons (spring and summer) and time of days (morning/noon/evening). The maps can be examined along several base layers from Google Maps or aside pure elevation data. Furthermore, thermal maps can be visualized in 3D using KML files for Google Earth.

While map views are useful in the preparation phase, they are of limited use during the actual flight phase. Therefore, we decided to export the coordinates of single points with high probability for air uplifts to a XML file, which can be read by the GPS devices carried by the pilots. This allows for a quick lookup of nearby thermal hotspots during flights. We implemented two different algorithms to extract hotspots from thermal maps. In the greedy algorithm we select the geographic location with the highest thermal probability as a new hotspot in each round. After each round, the thermal probability in the proximity of a newly found hotspot is reduced linearly with the distance. However, the greedy algorithm marks each thermal area at its peak point only. Therefore, no additional information about the actual shape of the area is available to the user. Based on user feedback, an advanced algorithm has been implemented, which represents wide or strong thermal areas by the use of multiple hotspots. Modeling this problem as an instance of the well-known facility location problem [5] provided good results in our experiments.

Future generations of GPS-based flight devices will likely include wireless communication capabilities, e.g., by tethering with a mobile phone. This will allow for novel applications, such as providing pilots with real-time information about nearby thermal hotspots observed by other pilots in the area.

---

[3] http://thermik.kk7.ch/

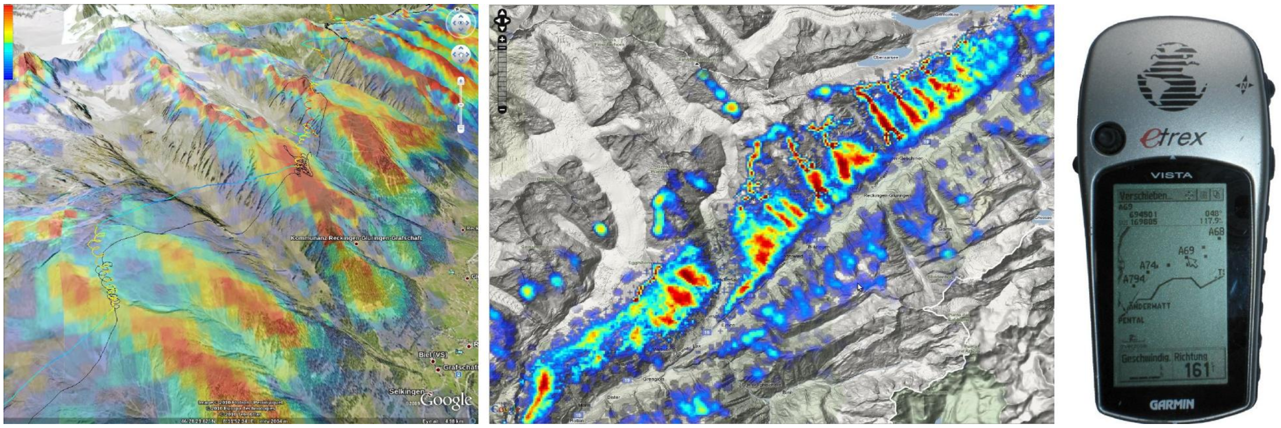

Figure 5: Flight analysis with a thermal probability overlay in Google Earth (left). Thermal columns can be shown as an overlay layer in browser-based maps (middle). Visualization of thermal hotspots on a GPS device for pilots (right).

7. EVALUATION

Since this work is the first of its kind, no existing thermal maps based on real sensor data are available for a direct comparison. Therefore, we evaluated the accuracy of our thermal prediction by a two-fold approach. First, we compared thermal maps based on data from the years 2003 to 2009 with measurement data collected during spring 2010. Second, we compare our measurement data with thermal maps based on simulations of a physical model for thermals.

Thermal maps are evaluated against flights from spring 2010 that were not used for the initial creation of the maps. First evaluation results indicate that thermal hotspots from the training years can also be found at the same place in the reference year. In areas with low densities of previous flight data, thermal prediction is less accurate than in regions with higher flight density. Because of this fact one can expect even better maps, as soon as more flights are added covering weakly known areas.

TherMap [21] is a raster model of thermal convection, which is based on physical properties such as seasonal soil characteristics. A direct comparison of thermal maps created by Ikarus and TherMap is shown in Figure 6. Ikarus maps show the probability to find a usable thermal for paragliders while TherMap shows the thermal pressure representing approximately the climb rate for a glider. Since Ikarus relies on flight data to predict thermals in a certain area, a comparison only makes sense in areas with high coverage. In these areas, the thermal prediction of TherMap and the measurement data of Ikarus show a very good match. However, there exists also a set of small differences, mainly due to the effect of local winds, which are not taken into account in the model, but play an important role for the position of thermal columns.

8. RELATED WORK

In the last years, mobile phones have attracted a growing interest as a platform for people-centric sensing applications [4]. Thanks to the large penetration of mobile phones in all demographic groups and due to the increasing capabilities of their integrated sensors, they provide the opportunity to gather data at unprecedented fidelity and scale. While modern devices already offer built-in imaging capabilities, microphone, position information and accelerometers/gyroscopes, future advances in miniaturization of micro-electromechanical sensor systems will bring further sensing possibilities to smart phones. Participatory sensing [2] enables the use of privately owned and controlled mobile devices to perform a common sensing task. Individuals collect sensor readings during work or leisure activities, which allows to gather high fidelity data about the surroundings of participants leveraging the local knowledge of the user [20]. It is even possible that participatory sensing applications are initiated by members of the community itself without the need for a supervising organization. However, the lack of centralized storage and control requires special efforts to ensure sensor data integrity and user privacy [9, 10]. Recently, Lane et al. [12] proposed the concept of opportunistic sensing where users may not be aware what is being sensed by their mobile phone.

A rich body of different participatory sensing applications has been proposed in the literature, e.g., [6, 8, 11, 14, 17, 18, 20]. However, most previous applications consisted of a much smaller number of participants than in Ikarus. The CenceMe [14] application classifies data from the phone's internal sensors to share the context of the user through social networking applications (22 subjects), Cyclopath [17] (29 subjects) and Biketastic [19] (12 subjects) target bicycle drivers to gather data about quality of routes and driver performance. Pothole [7] (7 subjects) and Nericell [15] use smartphones for sensing traffic and road surface conditions. BikeNet [6] (5 subjects) uses custom hardware together with a mobile phone to gather information about cyclist performance and environmental conditions. Thus, every bike participating has to be equipped with additional hardware. This is different from paragliding, where almost every pilot carries a flight device anyway to determine the current position and altitude.

Large measurement campaigns with more participants have been initiated to gather traces for vehicular mobility in the cities of San Francisco [16] (500 taxis) and Shanghai [13] (4000 taxis) by recording GPS positions of taxicabs. Similar to the approach used in Ikarus, De Choudhury et al. [3] used large collections of photos uploaded to the flickr website in order to automatically generate travel itineraries for tourists by analyzing where and when a picture was taken.

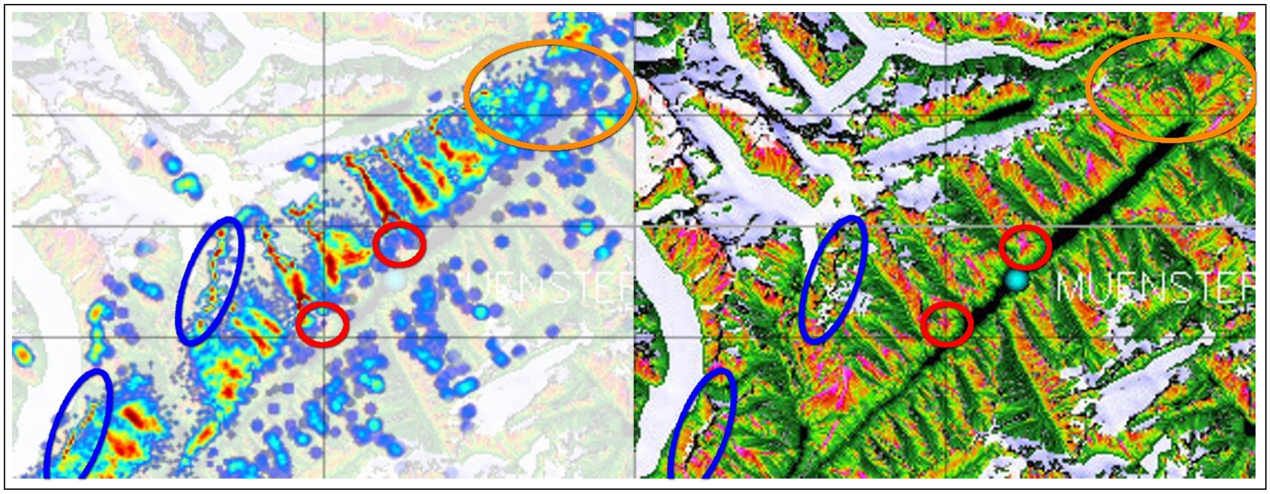

Figure 6: Comparison of the thermal maps of Ikarus for a summer noon (left) versus data from the TherMap (right) simulation. Main differences between the maps are highlighted with circles.

9. CONCLUSIONS

In this paper, we presented Ikarus, a novel large-scale participatory sensing system, which exploits sensor data collected during flights by paraglider pilots. The availability of more than 30,000 flight tracks by a total of 2,331 unique users provides a huge potential for the analysis of thermal uplift. This makes Ikarus one of the largest existing participatory sensing projects. We argue that successful participatory sensing projects are not possible without providing strong incentives for their users. Furthermore, we show that even though sensor devices are owned and controlled by the users, implementing simple mechanisms for handling sensor data of various qualities results in high data yield in practice. We proposed to use thermal probability maps, the first thermal maps of this kind, to cope with the unbalanced distribution of measurements both in time and place. Several applications were developed to assist paraglider pilots in flight preparation and analysis. The probability maps have shown their usefulness to predict thermals accurately in practice. Finally, we believe that thermal maps gathered by participatory sensing hold great potential for future applications in atmospheric research.

10. ACKNOWLEDGMENTS

We thank the anonymous reviewers and our shepherd Gaetano Borriello for their helpful comments. Furthermore, we would like to thank Heini Wernli and Michael Sprenger for their useful advice on thermal maps, and Beda Sigrist for providing us with TherMap data. Last but not least, our thanks go to the paragliding community for providing their flight data and the valuable feedback on the thermal maps.

11. REFERENCES

[1] T. Bradbury. *Meteorology and Flight: Pilot's Guide to Weather (Flying and Gliding)*. A&C Black Publishers, 2000.

[2] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava. Participatory Sensing. In *WSW*, 2006.

[3] M. De Choudhury, M. Feldman, S. Amer-Yahia, N. Golbandi, R. Lempel, and C. Yu. Automatic Construction of Travel Itineraries using Social Breadcrumbs. In *HT*, 2010.

[4] N. Eagle and A. (Sandy) Pentland. Reality Mining: Sensing Complex Social Systems. *Personal Ubiquitous Comput.*, 10(4):255–268, 2006.

[5] M. A. Efroymson and T. L. Ray. A branch-bound algorithm for plant location. *Operations Research*, 14(3):pp. 361–368, 1966.

[6] S. B. Eisenman, E. Miluzzo, N. D. Lane, R. A. Peterson, G.-S. Ahn, and A. T. Campbell. The BikeNet Mobile Sensing System for Cyclist Experience Mapping. In *SenSys*, 2007.

[7] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan. The Pothole Patrol: Using a Mobile Sensor Network for Road Surface Monitoring. In *MobiSys*, 2008.

[8] R. K. Ganti, N. Pham, H. Ahmadi, S. Nangia, and T. F. Abdelzaher. GreenGPS: A Participatory Sensing Fuel-Efficient Maps Application. In *MobiSys*, 2010.

[9] R. K. Ganti, N. Pham, Y.-E. Tsai, and T. F. Abdelzaher. PoolView: Stream Privacy for Grassroots Participatory Sensing. In *SenSys*, 2008.

[10] P. Gilbert, L. P. Cox, J. Jung, and D. Wetherall. Toward Trustworthy Mobile Sensing. In *HotMobile*, 2010.

[11] J.-H. Huang, S. Amjad, and S. Mishra. CenWits: A Sensor-Based Loosely Coupled Search and Rescue System using Witnesses. In *SenSys*, 2005.

[12] N. D. Lane, S. B. Eisenman, M. Musolesi, E. Miluzzo, and A. T. Campbell. Urban Sensing Systems: Opportunistic or Participatory? In *HotMobile*, 2008.

[13] P. Luo, H. Huang, W. Shu, M. Li, and M.-Y. Wu. Performance Evaluation of Vehicular DTN Routing under Realistic Mobility Models. In *WCNC*, 2008.

[14] E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell. Sensing Meets Mobile Social Networks: The Design, Implementation and Evaluation of the CenceMe Application. In *SenSys*, 2008.

[15] P. Mohan, V. N. Padmanabhan, and R. Ramjee. Nericell: Rich Monitoring of Road and Traffic Conditions using Mobile Smartphones. In *SenSys*, 2008.

[16] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser. CRAWDAD data set epfl/mobility (v. 2009-02-24). `http://crawdad.cs.dartmouth.edu/epfl/mobility`, Feb. 2009.

[17] R. Priedhorsky and L. Terveen. The Computational Geowiki: What, Why, and How. In *CSCW*, 2008.

[18] R. K. Rana, C. T. Chou, S. S. Kanhere, N. Bulusu, and W. Hu. Ear-Phone: An End-to-End Participatory Urban Noise Mapping System. In *IPSN*, 2010.

[19] S. Reddy, K. Shilton, G. Denisov, C. Cenizal, D. Estrin, and M. Srivastava. Biketastic: Sensing and Mapping for Better Biking. In *CHI*, 2010.

[20] K. Shilton, N. Ramanathan, S. Reddy, V. Samanta, J. Burke, D. Estrin, M. Hansen, and M. Srivastava. Participatory Design of Sensing Networks: Strengths and Challenges. In *PDC*, 2008.

[21] B. Sigrist. TherMap - Thermal maps for mountain regions. `http://www.aerodrome-gruyere.ch/thermap/`, Jan. 2011.