Collaborative Crowd Density Estimation with Mobile Phones

Jens Weppner, Paul Lukowicz
Embedded Systems Lab (ESL), University of Passau
{jens.weppner, paul.lukowicz}@uni-passau.de
http://wearable-computing.org/

Abstract

We present a technique for estimating crowd density by using a mobile phone to scan the environment for discoverable Bluetooth devices. The paper builds on previous work directed to using Bluetooth scans to analyse social context and extends it with more advanced features, leveraging collaboration between several close by devices, and the use of relative features that do not directly depend on the absolute number of devices in the environment. The method is evaluated on an extensive data set from a three day experiment at the Munich Octoberfest festival showing over 80% recognition accuracy (on four discrete crowd density classes) with 30% improvement over the simple method of just counting discoverable devices investigated in previous work.

1 Introduction

Knowing the density of a crowd can be relevant for a number of applications. Examples range from crowd control and emergency services through urban planning to consumer applications recommending where to go out (based where many other people also have gone to). While in some applications dedicated infrastructure such as access control gates or CCTV cameras [7] may be used, in others it would be desirable to be able to estimate crowd density without preinstalled infrastructure. One possibility is to recruit enough users to be able to estimate the density from the number of devices which report being in the relevant area. The obvious disadvantage of this method is that a significant number of users must be recruited, which is not always possible. In this paper we present an alternative method that requires only few users moving through the environment with their mobiles scanning for discoverable Bluetooth devices. The paper builds on previous work (see below in section 1.1) directed to using Bluetooth scans to analyze social context and extends it with more advanced features, leveraging collaboration between numerous devices, and the use of relative features that do not directly depend on the absolute number of devices in the environment (which may vary from venue to venue). The method is evaluated on an extensive data set from a three-day-long experiment at the Munich Oktoberfest festival.

PhoneSens'11, November 1, 2011, Seattle, WA, USA.

1.1 Related Work and Paper Contributions

The work most similar to ours is by Nicolai et al. [5] where the discovery time of Bluetooth devices as well as the relation between number of people and number of discoverable Bluetooth devices was investigated. As opposed to our approach the work relied on static Bluetooth sensing locations and only the absolute number of discovered Bluetooth devices was used. Along the same lines Morrison et al. [4] investigated crowd density estimation in stadium-based sporting events. However they did not attempt rigorous automatic classification and focused on a visualization tool for Bluetooth logs. Another use case of Bluetooth scanning is described in [3] by Kostakos. They recorded passenger journeys in public transportation by analyzing Bluetooth fingerprints. In [6] O'Neill et al. presented initial findings in Bluetooth presence and Bluetooth naming practices. Finally, slightly further away from our work, Eagle et al. showed [2] how to recognize social patterns in daily user activity, infer relationships and identify socially significant locations from using Bluetooth scans. BLIP Systems [1] exploited a stationary Bluetooth based people tracking system. Based on multiple Bluetooth zones scenarios like queue length at airports or travel times by car are indicated.

In this paper we present a Bluetooth scan based method that can distinguish between different discreet crowd densities. The main contributions beyond the above related work are as follows:

1. We rely not just on the number of devices seen by a scan, but also take into account information about average observed signal strength and the variance in both the signal strength and the number of devices.
2. We investigate the benefit of combining the information from several devices carried by different close by users.
3. We propose "relative" features based on the ratio between values observed by different devices, rather than on the absolute number of Bluetooth IDs seen by a scan. This makes the system more robust against variations in

Event	Duration	Participants	Total Bluetooth Scans	Average Devices per Scan	Median Devices per Scan	Discriminative devices
Munich Oktoberfest (DE)	3 days	2,3	2775	13.35	13.0	4454
Malta open-air festival (MT)	3 consecutive days	12,12,12	5500	8.70	8.0	1088
Wembley Stadium (UK)	1 day	6+4	4958	15.44	10.0	2509
Allianz Arena Soccer (DE)	4 days	10,16,6,12	14087	10.87	8.0	3944

Table 1: Overview of performed experiments

the number of discoverable devices that may result from the background of the people in the crowd rather than the crowd density.

We evaluate the method on a data set recorded during 3 days at the famous Munich Octoberfest Beer festival which is attended by hundreds of thousands of visitors from all over the world each day. Looking at four discrete densities that cover the range from a loosely occupied space (around 0.1 people per sqm) to dense crowd (around 0.4 people per sqm) we demonstrate recognition rates of over 80% using both relative and absolute features. This is over 30% better than the simple approach from previous work that relies on the number of devices found only.

2 Approach

Background

The basic idea is based on the observation that many people have the Bluetooth transceivers of their mobile phone in the discoverable mode as default setting. This is illustrated in Table 1 and Figure 1 on data from 4 different locations and venues across Europe: (1) several soccer games from the German first and second division (collected in and around the stadium), (2) the famous Munich Octoberfest beer festival, (3) the England-France soccer game at Wembley Stadium in November 2010 and (4) a music festival in Malta. From the above only the Octoberfest data was collected explicitly for crowd density estimation (and thus contains crowd density ground truth that is used for the quantitative evaluation later in the paper). The other data sets were collected for different purposes such as inertial navigation and activity recognition. However all data sets include regular Bluetooth scans collected over periods of days by several volunteers walking through the grounds of the specific event during times of different crowd density. It can be seen that the median of the number of devices discovered per scan is between 8 and 13 with thousands of distinct devices having been seen over the course of each experiment. Figure 1 shows that only less than 10% of the scans returned no discoverable devices and up to 50 devices were seen when in dense crowd.

General Considerations

An obvious way to estimate crowd density is to perform a scan for discoverable devices and assume that the number of devices that it returns is an indication of the number of people in the vicinity defined by Bluetooth range (typically around 10m) [1]. Unfortunately, this simple approach contains a number of problems.

[1]Actual Bluetooth sensing distance is depending on the Bluetooth category, shielding objects or people and influencing jamming sources

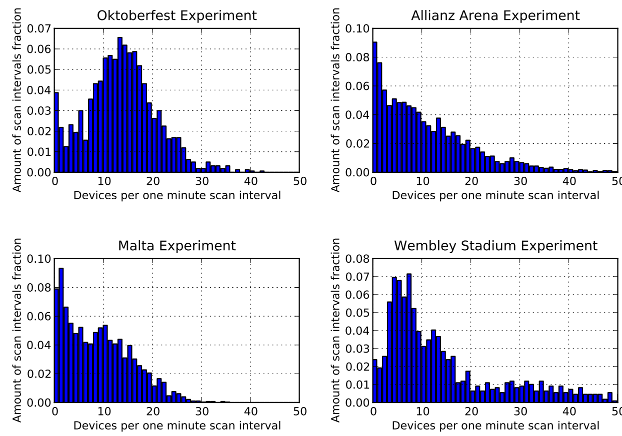

Figure 1: Distributions showing the fraction of the number of Bluetooth devices discovered per one minute scan interval at multiple experiment venues

Firstly there is the issue of sufficient statistics. With the scan limited to a radius of about 10m (approximately a circle with 300sqm area) anything between a few and a few hundred people can be within range. While in a dense crowd with a few hundred people we may get a representative sample, in less crowded areas we are likely to see very strong variations between samples. Assuming the probability of any single user having a discoverable Bluetooth device to be 10% the probability that no device is seen when 20 people are within range is $0.9^{20} = 0.12$. Thus we may sometimes be in a group of people who do not even have activated mobile phones while at other times we may be surrounded by a group where everyone has an active Bluetooth device. Secondly, there is the question of signal attenuation. At 2.4GHz (which is the transmission frequency of Bluetooth) the human body has a high absorption coefficient. This means that in a dense crowd (where we would expect to have good statistics) the effective scan range is reduced and therefore ”falsifying” the results.

Finally, we have to consider cultural factors. This means that the average number of people carrying a discoverable Bluetooth device may significantly vary depending on who the persons in the crowd are. For the same crowd density at a semester party of a technical university a different number of devices may be present than at a fifth division soccer game in a poor rural area.

To mitigate the influences above our method does not rely

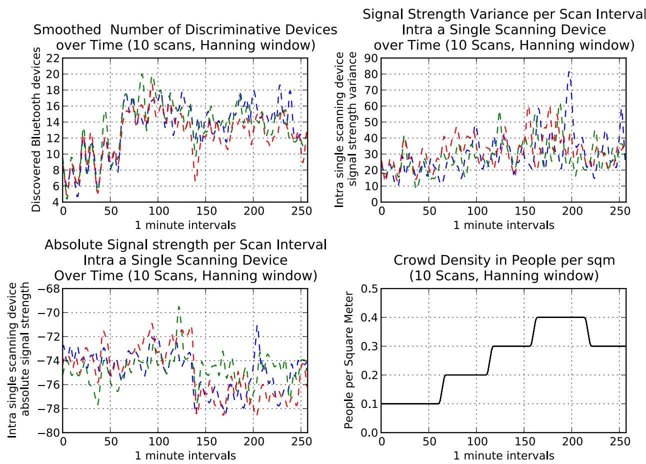

Figure 2: Visualization of individual features used for classification. Multiple lines describe multiple concurrently scanning devices. (a) Number of Bluetooth devices. (b) Signal strengths variance. (c) Absolute mean signal strength. (d) Ground truth crowd density in people per square meter.

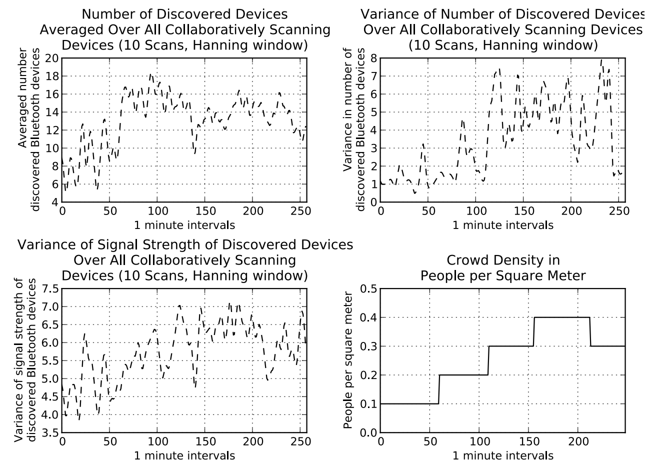

Figure 3: Visualization of collaborated features used for classification. (a) Average number of discovered Bluetooth devices. (b) Variance in number of discovered Bluetooth devices. (c) Variance of signal strengths of discovered Bluetooth devices. (d) Ground truth crowd density in people per square meter.

solely on the absolute number of discovered devices. Instead we also use the average signal strength and signal strength variations. In addition, we look at collaborative estimation from several (up to around 10) devices. In doing so we focus on differential features that are not directly dependent on the absolute number of discoverable devices in the environment or the absolute signal strength. As shown in section 3 the above measures lead to over 30% improvement in recognition rate over a method based on the absolute number of discovered devices.

Individual Device Based Estimation

To estimate the crowd density from the scans of a single mobile phone the following features were computed:

Feature 1.individual: *Number of Bluetooth devices*

Feature 2.individual: *Mean signal strength*

Feature 3.individual: *Variance of the signal strengths*

It is important to mention that the discovery of Bluetooth devices is not an infinitely small snapshot in time, but in our case a 60 seconds time window (so called Bluetooth scan interval). During this time the underlying system reports Bluetooth devices which were not found before in the given scan interval. Therefore the above features are based on a time interval of about 60sec.

The first feature (*Number of Bluetooth devices*) is the most obvious feature extracted from the raw data by counting the number of different devices per scan interval. As described above the discovered devices per 60s time window are the intersection of actual surrounding devices ($number(i) = |\cap_{x \in i}\ device(x)|$ where x is a time of occurrence and i is the current time span) in this time window independent to real duration of occurrence of a device.

The second feature (*Mean signal strength*) averages the signal strength of all devices in a scan interval ($average(i) = \frac{\sum_{x \in i}\ signal_x}{number_i}$), motivated by the assumption that the average signal strength gives a hint for the crowd density . Assuming surrounding people are shielding the signal the average of the signal strength might be lower.

The third feature (*Variance of the signal strength*) is defined by the variance of the measured signal strengths of unique devices in a scan interval ($variance(i) = \frac{\sum_{x \in i} signal_x^2}{number_i} - average(i)^2$). Assuming a shielding effect is measurable, the Bluetooth signal reception is excellent from people walking nearby and the Bluetooth signal is near the reception threshold from people walking rather further afar. At high crowd densities the signal strength variance would be higher than at low crowd densities where less people are shielding the Bluetooth signal. Figure 2 presents the extracted features from the raw data.

Collaborated Estimation

Collaborated crowd density estimation relies on multiple users walking through the same space. The analysis combines the scan results of their devices computing the following features:

Feature 1.collaborated: *Average number of devices*

Feature 2.collaborated: *Variance of the number of devices*

Feature 3.collaborated: *Variance of all signal strengths*

The first feature (*Average number of Bluetooth devices*) is computed by collecting *Feature 1.individual* of each participating sensing device and averaging the values. As opposed to a sum, this feature is independent of the number of collaborating devices.

The second feature (*Variance in the number of devices*) is defined by the variance of the individual *'Feature1.individual'* values across all devices ($variance(i)! = var(number_x(i))$ where i is the time frame and device $x \in S$ given S as the set of sensing devices).

(a) 0.1 $\frac{people}{m^2}$ in average (b) 0.2 $\frac{people}{m^2}$ in average (c) 0.3 $\frac{people}{m^2}$ in average (d) 0.4 $\frac{people}{m^2}$ in average

Figure 4: Ground truth pictures taken during the experiment. Each picture is a sample for the corresponding class determined by the crowd density average over a 500m path.

The third feature (*Variance of all signal strengths*) is defined by the variance of the signal strengths aggregated from all participating sensing devices during a given scan interval. Potential multiple occurrences of the same Bluetooth device found by different sensing devices are not removed from the feature computation. Figure 3 presents the extracted features from the raw data.

While *Feature 1.collaborated* is an absolute value, *Feature 2.collaborated* and *Feature 3.collaborated* represent differences in the values measured by different, spatial distributed devices. Thus, they are more related to the properties of the crowd than to the absolute number of discoverable devices in the crowd (although they are not fully independent of the number of devices). The effects involved are complex and driven by a number of factors. For one, the spatial variance is likely to be reduced as the crowd density increases since each scan is likely to be based on a larger (= more representative) sample of people. On the other hand, with increased crowd density occlusions, reflections and other propagation effects are likely to play a bigger role. These depend on the specific configuration of people at scan location (where are the devices worn, how are they occluded etc.) which means that variance will increase.

In this paper we do not aim to investigate and explain such effects. We are content with the fact that in our evaluation relative features perform well and increase classification accuracy (see next section).

3 Evaluation

Experimental Setup

For our experiment three participants were equipped with three Android HTC Desire smartphones with enabled Bluetooth unit. Two phones were placed in the pant's front pockets and another phone in the back pocket. Collected raw data for each discovered Bluetooth device consisted of the following attributes:

1. Timestamp
2. Bluetooth scan interval number
3. Bluetooth device name
4. Unique Bluetooth device ID
5. Bluetooth signal strength in dBm

The data was was written to the SD card of each phone and evaluated later on.

The experiment was performed on three different days at the Oktoberfest 2010 in Munich (Germany). The event's main pedestrian zone is approximately 500m long. The width was 20m divided sparsely by merchandise stands in the center and bounded by food stands and beer tents with side street crossings where other visitors entered and left the area. Participants of the experiment were told to move continuously in a group formation with a distance of roughly 5m to each other as far as this was possible in large crowd densities while walking back and forth on the same side of the zone with mixed walking directions.

For crowd density ground truth one group member took digital camera pictures at an interval of 100m along the zone. All pictures have an embedded timestamp and were made with the same focal length of 5.8 (equivalent to 35mm) and consistent angle of vision while holding the camera above the head in forward direction. The area, covering $100m^2$ ($1.076 sq.ft.$), regarded for counting the people per picture is visualized in Figure 5 bounded in the distance by head widths of 10% of the picture width. With this technique some

Figure 5: Ground truth image showing the section (highlighted) which is used for labeling the crowd density. The distant boundary is defined by the head width of 10% of the picture width.

heads might have been covered by other heads, but compared to birds-eye-perspective-pictures this approach was feasible in this situation. All pictures were evaluated manually according to the defined boundaries and averaged over pictures taken per walk on the 500m zone resulting in a crowd density label for each segment of the experiment.
Classes used for labeling the sections of the experiment are expressed with the $\frac{people}{m^2}$ unit. Four different crowd densities were extracted out of the data set: 0.1, 0.2, 0.3, and 0.4 $\frac{people}{m^2}$. Examples of each class are shown in figure 4.

Results

We analyzed both *individual* and *collaborated* features for crowd density classification on single scans and on averages over a sliding window of 10 scan intervals. Together with the previously defined crowd density labels we trained and evaluated a decision tree classifier with 10-fold cross-validation. Figure 6 shows the classification results for non-windowed feature sets (dark grey) and windowed sequential feature sets (light grey).

It can be seen that the combination of all features from a single device *Feature x.individual* leads to a 39% accuracy over individual scans and can be improved to 63% by a windows over 10 scans ("window case"). Individual features have an even lower accuracy. The results can be significantly improved with the collaborative approach. All collaborative features (including the average absolute number of discovered devices) leads to a single scan accuracy of 64% and a windowed accuracy of 81%. Without considering *Feature 1.collaborated* which is directly influenced by the percentage of people carrying a discoverable Bluetooth device, it is possible to obtain a 40% accuracy on single scan basis and 82% windowed accuracy.

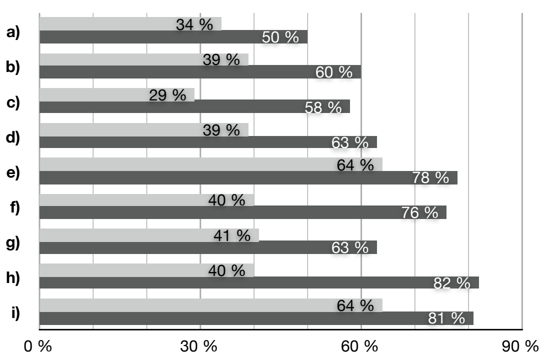

Figure 6: Classification accuracy (light grey: non-windowed, dark grey: windowed feature set). INDIVIDUAL: (a) Number devices (b) Mean signal strength (c) Variance signal strength (d) All features. COLLABORATIVE: (e) Overall number (f) Variance number devices (g) Overall signal strength variance (h) Combined collaborative (relative) features (i) All collaborative features

4 Conclusion

We have shown how Bluetooth scan data from just a few users equipped with standard mobile phones can be used to estimate crowd density. The core of the method is the comparison and fusion of data from different devices which leads to over 30% improvement in accuracy over a simple single device approach. The just over 80% accuracy on four classes must be seen in the context of noisy ground truth resulting from arbitrary class definition, extrapolation between photos taken every 500m, and inaccuracies in the counting process. In addition, confusions occur nearly exclusively between neighboring classes (see figure 7). Note that the experimental data did not include the "nearly empty space" class which can be trivially recognized from the near absence of

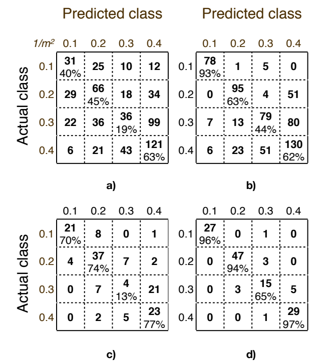

Figure 7: Confusion matrices (a) Individual - all features, not windowed (b) Individual - number and number variance, windowed (c) Collaborated - all features, not windowed (d) Collaborated - all features, windowed (10-folds cross validation / Decision tree min 20 elements per leaf)

Bluetooth devices and could be easily integrated into the system.

In summary, we believe that the method presented in this paper is potentially useful for many applications. To further improve it future work will focus on better understanding and modeling the relative features. We will also collect data from other events and countries to verify the hypothesis that the relative features are robust against culture related differences in the percentage of people carrying discoverable Bluetooth devices.

5 Acknowledgments

This work is supported by the FP7 ICT Future Enabling Technologies programme of the European Commission under the grant agreement No. 231288 (SOCIONICAL).

6 References

[1] BlipSystems marketing and tracking solutions. `http://www.blipsystems.com/`. Accessed: 10/10/2011.

[2] N. Eagle and A. (Sandy) Pentland. Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.*, 10:255–268, March 2006.

[3] V. Kostakos. Using bluetooth to capture passenger trips on public transport buses. *arXiv*, 806, 2008.

[4] A. Morrison, M. Bell, and M. Chalmers. Visualisation of spectator activity at stadium events. In *2009 13th International Conference Information Visualisation*, pages 219–226. IEEE, 2009.

[5] T. Nicolai and H. Kenn. About the relationship between people and discoverable bluetooth devices in urban environments. In *Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology*, pages 72–78. ACM, 2007.

[6] E. O'Neill, V. Kostakos, T. Kindberg, A. Schiek, A. Penn, D. Fraser, and T. Jones. Instrumenting the city: Developing methods for observing and understanding the digital cityscape. *UbiComp 2006: Ubiquitous Computing*, pages 315–332, 2006.

[7] H. Su, H. Yang, and S. Zheng. The large-scale crowd density estimation based on effective region feature extraction method. In *Proceedings of the 10th Asian conference on Computer vision - Volume Part III*, ACCV'10, pages 302–313, Berlin, Heidelberg, 2011. Springer-Verlag.