Visualizing Urban Sports Movement

Ozgun Balaban\textsuperscript{1}, Bige Tuncer\textsuperscript{2}
\textsuperscript{1}\textsuperscript{,2}Singapore University of Technology and Design; ETH Zurich, Future Cities Laboratory, Singapore-ETH Center (SEC), Singapore
\textsuperscript{1}balaban_ozgun@mymail.sutd.edu.sg \textsuperscript{2}bige_tuncer@sutd.edu.sg

In this study, a visualization tool that maps outdoor physical activity such as runs on a map by specifying time, location, activity, gender, age group, etc. is created. This tool reveals the usage patterns of streets within a city for outdoor physical activity. This tool is created within a larger research project that investigates the influence of streets on the leisure walking activity within cities. For this purpose, the tool is capable of presenting the collected multi-modal data that includes personal fitness data, weather data, spatial data, and crime data. Moreover, the tool creates new analysis capabilities such as displaying usage of streets by urban joggers. The research project in which this tool will be used is aimed for designers/planners to improve streets for 'runnability'.

Keywords: Sports Activity, Big Data, Urban Visualization, Fitness Applications

INTRODUCTION
Urban planning and design requires detailed analysis that relies also on collections of data concerning the site and various issues such as utilization of urban spaces, accessibility, sustainability, etc. Often this data comes from traditional sources such as surveys or interviews (Reades et al 2007). Surveys and interviews are labor intensive tasks, as it is costly to collect data personally. Moreover, these are done a limited number of times, and it is hard to update the data. Additionally, for running and biking, as these involve time and path data, people may not remember the exact time and location of their activities.

There are other methods that are available for an urban planner especially when planning for leisure. Behavior observation is observations of people using urban spaces. It was the important tool for Whyte’s study of utilization of plazas in New York (Whyte and Underhill 2009). For sports it includes recording of people using parks or other spaces for sports. It can be done using several different techniques such as taking videos and photos, or watching and recording of users manually. This provides detailed information about how residents use urban spaces, such as parks for running. However, this is resource intensive work, requires workforce and time for recording.

With the rising popularity of mobile phones and applications running on 'smart devices', there is another possible method of collecting data about how residents use cities. Especially, fitness applications which keep track of location of their users' physical activities are valuable in understanding sports participation within a city. As these applications record users location and time of activities, it is possible to know a person's usage patterns of spaces during a sports activity. Furthermore, by acquiring data from more than one user, it is possible to record location specific queries such as who is using a park for running and where they start their runs, which will in turn indicate a practical catchment area of a park.
The main aim of this research, therefore, is to use personal physical activity data generated from fitness applications to visualize how an urban setting is used for running within a selected neighborhood. The questions to be answered in this study are:

- How to use fitness application data to indicate running behavior of residents of a neighborhood?
- How to improve the data collection methods mentioned above with the help of this new data?
- How to analyze the influence of streets to urban running activity with the help of implemented tool?

**BACKGROUND**

With the recent developments of sensors and their availability for the masses, it is possible to study behaviors of urban residents. Tools that are available for planners, designers and policy makers create a new notion of cities, which is generally referred to as "smart cities". Smart cities are the cities that can adapt to the behavior of their residents and in a way their citizens are constantly involved in the design and planning processes without even realizing it. Traditional way of understanding public utilization of urban spaces is having some questionnaires and polls. However, these methods are costly and often cannot provide the whole picture.

Use of traces of urban residents, left by the sensors they carry with them while using the urban space (smart phones) is a recent way of acquiring knowledge instead of doing surveys and polls. One example is Mobile Landscapes project, (Ratti et al 2006) which uses mobile phone location data to understand city residents' movement and their activities in time. Another example comes from Lee and Kam's work (2014) on the study of Singapore commute patterns using the data from smart cards used for train commutes. This provides an overview of travel patterns during the day which can be used to plan new lines to relieve congestion during peak hours.

Visualizations of fitness data collected from different fitness applications are popular within data visualization communities. These generate artistic visualizations that show the frequency of runs within cities [1,2]. Strava, one of the fitness application companies, has even created an interface to show heat maps of runs in a city [3]. While these visualizations provide an overall understanding of which paths are popular within a city, they lack filtering capabilities and are limited in effectively informing urban designers and planners.

Collection of fitness data and its usage in urban planning is discussed in several studies. Clarke and Steele (2011) suggest combining fitness data with social media data and discuss how the data should be collected. Cortes et al take (2014) Endomondo as an exemplary fitness application, collect the data and provide some statistical data about overall usage of the application. Both of these studies focus more on collection of the data and they suggest how to use this data for urban planning. However, none of these are implemented.

Open Data approach, which means sharing of different kinds of data with the public by organizations or governments, leads to new possibilities in urban planning. Availability of data for planners makes it easier to use this data as sometimes collection of this data might be tedious. As there are data covering different aspects, it is possible to bring them together to form multi-modal data. An example of multi-modal data is collecting crime data, running data, and GIS data such as street width in a database and correlate street choices of the users to get informed about perception of crime. Availability of many different sensors makes multi-modal approach possible.

**TECHNICAL OUTLINE**

In this study, multi-modal data approach is used to collect different data from different sources. The data collected is weather data, personal fitness data, spatial data regarding streets of a city such as width of a street, width of a pedestrian walk, etc. But the pri-
mary mode of the data comes from fitness applications, the other data sources complement fitness applications.

There are many different personal fitness applications on the market. While their interfaces differ, what they do is basically same: that is to record users' location within a timeframe and report it back. The most important personal fitness applications are Google Fit, Strava, Endomondo, Nike+, and Runkeeper. For our study, initially we choose Endomondo as data source. The reason for this selection is that, in Endomondo every workout is presented in a webpage which is listed sequentially which makes data collection easier.

Table 1
An example of a workout data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workout ID</td>
<td>23007236</td>
</tr>
<tr>
<td>Start date and time</td>
<td>6/12/2015 12:22</td>
</tr>
<tr>
<td>Distance</td>
<td>10.12 km</td>
</tr>
<tr>
<td>Duration</td>
<td>1h 1m 23s</td>
</tr>
<tr>
<td>Average Speed</td>
<td>6.45 km/h</td>
</tr>
<tr>
<td>Calories</td>
<td>950 kcal</td>
</tr>
</tbody>
</table>

Endomondo is one of the most used personal fitness applications. It has over 20 million users worldwide [4]. In Endomondo every user has a unique ID and every workout such as a run is stored with a workout ID as presented in Table 1. In every workout there are certain parameters related to that workout that is stored, these are workout ID, type of the sport, the date and time of the workout, distance covered, duration of the workout, calories spent during the workout, average speed, minimum and maximum elevations, and ascent and descent during the workout. These are the basic parameters that are available when users' have basic subscription. If a user has premium subscription, other features such as heart rate data are also available.

An Endomondo activity webpage is created by the ID of the workout such as: endomondo.com/workouts/23007256/. This ID is sequential. As of June 2016, there are over 741 million workouts however some of these workouts are deleted or private so it is not possible to access those workouts. Endomondo presents workouts in its interface as shown in Figure 1.

The same approach is used for users as well, all Endomondo users have user ID and these IDs are also sequential. One can list every workout of a unique user if the user has not selected to be a private user. Some users also associate their social media accounts with their identity. Profiles of users are also presented in a webpage and an example of a URL of a profile page is endomondo.com/profile/4356/. In the profile
parameters of users are displayed such as username, country, workouts that are done by the user, birthday, gender, and total distance covered.

Workouts that include GPS data in Endomondo are divided into parts by time (for running it is around 9 secs) (Cortes et al 2014). Accordingly, every 9 seconds the application locates the user by a GPS point and uploads this data. These are called GPS tuples and they include longitude, latitude, altitude, distance, duration, and pace.

The first step to collect the data from Endomondo is to download all public data. As all the workouts in Endomondo are sequential one can browse through these workouts by increasing the number of ID and saving the data. To automatize this process, a script that can visit every webpage that is available to public is created. The only task of this application is to save the html code to a text document. Runs are structured as JSON objects by Endomondo, which is helpful in text processing.

After this phase, a huge document containing all public workouts by the whole Endomondo community is created. This document is ever expanding as the application gets more and more popular and there are new workouts created by the community every moment. This document needs to be text processed such that every workout data is separated into new entries that contains information such as the user ID, username, the time of the workout, the type of the workout, duration of the workout, and distance of the workout. If the application collecting the data is used on a GPS enabled hardware, then there will be GPS tuples available. For the purpose of this study workouts without GPS data are not useful. After creating these entries, this data is recorded in a mySQL database with the respective tables and fields as shown in Figure 2.

In the last phase, this collected information in the database is used for visualizing users’ movements. A heatmap showing all the workouts that took place in a certain city is done by using ggmap and ggplot package of R, which is a language and environment for statistical computing and graphics. Ggmap along
with ggplot is used for plotting spatial data on top of static maps such as Google Maps [5]. For this purpose, first a query is made in the database for selecting runs in a city by filtering out runs that are out of the city coordinates. Then since every run has a GPS point that is recorded every 9 seconds, it is easy to plot these points in the map and draw a path. All these runs happening in a city form a heatmap which indicates the intensity of usage of that space for running. With the database capabilities it is easy to create different visualizations for an urban planner, such as runs that start from a certain location, runs within a time period, runs by gender or age groups, etc (figure 3). Also in this visualization other modes of data such as crime and weather are visible to the planner/designer as well. Therefore he/she can see the crime rate of a street to understand its effect on people’s usage of that particular street.

Crime rate data comes from open data approach which some cities has already started to participate. Boston city has a portal for publishing crime incident reports with the public [6]. Reported crimes are presented in this portal with GPS coordinate, time, and the event type such as burglary, assault, etc. Unfortunately, this data is not yet available in Singapore.

Weather data is also included in the database as it affects people’s running behavior. The source of the data comes from Weather Underground website which publishes past weather data of most of the cities worldwide [7].

In addition to heatmaps of activities, there are additional visualization capabilities. This tool is planned to be used in a study that inspects the influence of street configurations in the urban running behavior. For this purpose, there are specific visualizations such as a certain street with the graph that shows the runs and the time of the day that these runs occurred.

Another one is used to present the statistics of street typologies such as boulevard, main road etc. that users prefer running along. To achieve this, streets are classified into different street typologies and every run of each user is divided into segments of street typologies where occurred. Then the resulting statistics are show both for individual runners (see figure 4) and the overall population of runners. This is still in development as it needs the street typology data which is acquired by manually assigning the streets into street typology.

**DISCUSSION**

This study develops a promising approach to use available fitness data in order to assist urban planners. It improves traditional techniques such as surveys and behavior observation, if not replacing some of them. These traditional techniques however very valuable, are resource intensive, thus sometimes omitted. This visualization tool displays the fitness activities on a map by specifying time, location, activity, gender, age group, etc, which in turn indicates the usage patterns of certain spaces. Moreover, it creates a tool for studying the influence of streets in urban running behavior.

This study has some limitations due to using data methodologies. The first limitation is that the data is open to errors, Errors from sensors that are used should be taken into account. However, for the purpose of this study, the accuracy is enough to indicate streets used during the runs. The second limitation may be privacy issues. We handle this by making the data anonymous and removing all the information regarding personal contact such as names and emails.

Another limitation of the study comes from using Endomondo. Endomondo is just one personal fitness application on the market where there are dozens of other applications as well. It gives a sample set of runners within a city but it might be open to some bias. To overcome this issue, there will be other fitness applications integrated in the system as well. Also most of the data study is done considering this issue, so for example in the case of checking which street is more appealing for the runners it is avoided to consider the number of runs as this might be deceiving. Instead, runs that use that street are analyzed and divided into segments and the percentage of usage of that street
is presented. That shows the runners’ willingness of choosing that street when he/she has other options.

As mentioned earlier this visualization tool will be part of study that is inspecting the influence of streets in urban running behavior. Therefore, the capabilities are devised to help understand how streets affect runners but they are also instrumental in helping planners and designers understand the usage of streets by runners.

This study is precursor for a detailed analysis study about people’s urban running behavior. In this stage the analysis part is not finished, as still there is ongoing effort of collecting different modes of data. In the future the results of this study will be published with an analysis of the data that is collected.

ACKNOWLEDGEMENTS
The research was conducted at the Future Cities Laboratory at the Singapore-ETH Centre, which was established collaboratively between ETH Zurich and Singapore’s National Research Foundation (FI 370074016) under its Campus for Research Excellence and Technological Enterprise (CREATE) programme. In addition, it is part of PhD study conducted by the author at Architecture and Sustainable Design Pillar at Singapore University of Technology and Design.

REFERENCES
Clarke, A and Steele, R 2011, ‘How personal fitness data can be re-used by smart cities’, Proceedings of the 2011 7th International Conference on Intelligent Sensors, Sensor Networks and Information Processing, ISSNIP 2011, pp. 395-400
Cortes, R, Bonnaire, X, Marin, O and Sens, P 2014, ‘Sport Trackers and Big Data: Studying user traces to identify opportunities and challenges’, no title given
Whyte, WH and Underhill, P 2012, City: Rediscovering the Center, University of Pennsylvania Press, Incorporated
[2] https://flowingdata.com/2014/02/05/where-people-run/
[5] https://cran.r-project.org/