PREDICTIONS OF URBAN QUALITIES IN THE CITY OF ZURICH

Jakub Lichman, Philippe Schlattner — *Florian Koch*

Department of Computer Science — Department of Physics

ETH Zurich
Zurich, Switzerland
Email: {lichmanj, pschlatt, flkoch}@student.ethz.ch

ABSTRACT

We should do as the last one.

1. INTRODUCTION

Huge predicted increase in the number of world urban area residents from 54 to 66 percent in 2050 brings many challenges which every successful city has to overcome. Cities of the future has to ensure contiguous improvements of their services despite growing population. Since it is infeasible for a city to excellent in all services, authorities have to reach compliance about the subset of prioritized ones. For instance, this year's Mercer's Quality of Living index takes into account economic and political environment, infrastructure, public transport, health, recreation and housing to decide which cities are the most desirable. Based on these factors, the list of most livable cities was constructed. However these types of rankings usually generalize the whole city and take average of the all districts which can over or underrate them. In our paper we tried to distinguish urban qualities within a city, concretely in the city of Zrich. Its public data availability enabled us to make interesting predictions about urban qualities of the individual city districts which we lately verified by examining selected places in person.

Our experiment tries to exploit data that are publicly available for the city of Zrich as well as data that we tried to obtain by ourselves. The goal was to create a map of the city with multiple layers where each layer will represent one factor of urban quality. By summing all factors and visualizing them we can obtain an interactive map which visualizes livability of each part of the city. Later we evaluated selected areas with the smart Agora platform which enabled us to collect data for validation by walking of these areas.

The main contribution of this paper is to automate reasoning about urban qualities. Nowdays it is done by real estate agents which charge big amounts of money for such services. In the future their jobs can be completely automated and so computers will monitore urban qualities and suggest accommodation prices as well as provide sets of possible actions to improve quality of life in poorly rated locations.

Related work. This paper is mainly based on the previous work of Evangelos Pournaras et al. [1] where internet of things approach is used to obtain citizen's perception of the selected areas of the city of Zrich. Paolo Neirotti et al. [2] in their paper provide comprehensive understanding of the notion of smart city through the elaboration of natural resources and energy, transport and mobility, buildings, living, government, economy and people. Kunwar P.Singh et al. [3] used principal components analysis (PCA) analysis to identify sources of pollution and tree based learning models to predict the urban air quality of Lucknow (India) using the air quality and meteorological databases.

This paper is organized as follows: Section II explains what a smart city is, its history, tools that are used in smart cities and its practical applications in the praxis. Section III describes the data that were used for our experiment and their sources. Section IV defines the criteria by which we will rate the locations inside the Zrich. Section V explains greenery detection process in our approach and section VI describes method of creating quality index map of the city. Section VII evaluates preditions through Smart Agora platform. Finally, Section VIII concludes this paper and outlines future work.

2. BACKGROUND

The term "Smart Cities" has attracted some attention recently [4, 5, 6] usually referring to the real time analysis of whole cities and their population [7]. It usually also implies economical innovation which is usually equated with entrepreneurship created by smart people.

In the process of establishing this expression, two main sources of data have been identified [7]. On one hand this compromises the increasing number of stationary installed computer connected hardware, such as traffic cameras, telecommunication networks and building management systems. On the other side the population carries an even faster increasing number of mobile computers like smart phones, navigation systems or even fully connected vehicles.

By combining these data sources and analysing them in real-time, one is able to not only steer the flow within a city but also predict the evolution. The Chinese central government has recently been in the press [8] for not only using the public surveillance system to track down criminals but also equip the police officers with glasses that have an integrated cameras [9]. According to the same article, they plan to have a public credit system in place for all its 1.4 billion citizens by 2020, which will be crucial for obtaining jobs or getting permission to travel abroad.

The real time analysis of their currently over 170 million CCTV cameras [10] enabled them to arrest a suspect, who was identified by facial recognition amongst 60'000 visitors to a concert. These are just a few examples of how the connection of various data sources enables the people having access to the data to widely control public live.

It also points out some of the threats. With the public credit and shame system [9] everyone has to constantly adhere to the accepted code of conduct in order to not loose points or be publicly denounced. This ubiquitous surveillance is feared by many privacy policy activists, slowing down similar efforts in western countries. In the end we will have to decide in which kind of world we all want to live in.

But the thus obtained data can also be used for the good. Responsive traffic management systems are a neccessity in current megacities, where an increasing number of cars tries to reach their destination on roads that cannot be extended any further. And also the public transport systems reach their limits when an ever increasing number of people tries to get from A to B all at the same time. Wisely employed camera systems not only allows to track the flow of people but the correct analysis promises to resolve some of the problems.

In our project we tried to use data from various sources to rate the living quality for people within the districts of Zurich. As most of the real time data in Zurich is not publicly available, we set out to analyse less fast paced data first. The predictions should then be checked with data obtained from study participants, who would walk those ares to collect data with their smart phones. In a future expansion this could allow to assess the living quality not only in general but also for different days of the week or even different times of the day.

3. DATA

To begin with, we evaluated data downloaded from the open data initiative of the city of Zurich [11]. Furthermore, we used maps from the Geographic Information System of the canton of Zurich GISZH [12] to get an idea of potentially interesting areas. These two sources were used, since we believed that data published by institutions of the government is in general more reliable than that published by individuals. In addition, the data is considered to be of better quality and cover more of the area. This initial trust has been partially disappointed by finding quite a few data sets with only few points. As we wanted to predict the best living areas within Zurich, we also need data that covers more or less the complete area of Zurich in order to avoid bias towards certain regions, which are better covered by the data sources. Therefore, we ommited data sets with subjectively too few entries, which we initially wanted to use due to their potential. These omitted data sets include the one on air quality, flat prices and traffic counts. The data we finally used and analysed is composed of data for public street illumination, pedestrian zones, sighing points, zones where driving is prohibited as well as restaurants. These are all sets where the available data is spread across a large part of Zurich and that we figured might be indicative for how much we as students would appreciate living there. This is a highly subjective measure that took us some time to agree amongst ourselves. The analysis of the data is described in Sec. 6. In addition we inferred the amount of greenery from satellite images as outlined in Sec. 5 to evaluate how urban the areas are. Finally we tried to evaluate the predictions by having people walk some of the areas and gather their impression by utilising Smart Agora. The details of the last bit are outlined in Sec. 7.

4. MEASURE OF ATTRACTIVENESS

As we are all students at ETH Zurich, we wanted to map the attractiveness to live in a certain area from the point of view students have. Therefore, access to public transport is important, however walking for up to 15 minutes does not bother us that much. Hence, we did not include stops of public transportation in our analysis, since they are well anough distributed that we found a stop is in wlaking distance to each and every location within Zurich. On the other hand we did not have access to detailed data on diverse measures. Accordingly we defined spheres of influence for all types of points we used for the analysis, that should represent how far we believe that item will influence how we perceive the area. This also means that accessibility for example is not that relevant, compared to bars and restaurants. Another factor that is very important to us was the proximity to places we visit frequently. However, this is even more subjective and even in our small group of people we could not identify several points of interest, that are common to all of us except for study related locations.

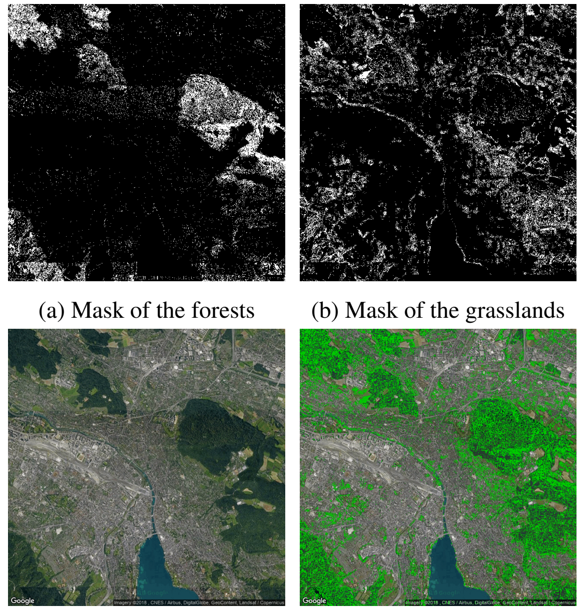
5. GREENERY DETECTION

Greenery has always played crucial part in the construction of cities. The need for green spaces has been present since ancient times. City parks are traditional place of recreation and relax for all kinds of people in their spare time. Current trends in the architecture brought new ways of connecting buildings and greenery such as roofs with green surfaces or terraces with tree pots. These trends are consequence of natural human inclination towards nature. Therefore we have decided to take greenery as a significant indicator of urban quality.

In order to predict user preferences partially based on a greenery, we needed to acquire data about it. There are two basic ways of detecting trees, parks or grass areas on a map. First approach relies on a snapshot of a satellite view [1]. It is simple to implement and yields accurate results. Detection of green areas is via pixels with RGB values that lie within a specified intervals. The only crucial requirement in order to obtain precise detection is to find high quality satellite view of the selected area which in our case was taken from the Google maps satellite view.

Second approach is to detect greenery from the Google Street View (GSV) [13]. Xiaojiang Li et. al. detected greenery by examining street pictures taken from GSV. With this method they were able to explore greenery inside the cities in a great detail. Idea behind the second method is in general more accurate since it can detect trees shadowed by taller buildings or shelters. However, its precision highly depends on availability of street view and algorithms, which are in our view still not accurate enough. That is the reason why we have decided to use the first method.

We have developed the python script which does greenery detection similarly to the first approach mentioned above. Firstly, we did the transformation of RBGA pixel format into HSV one where detection of color spectrum is more accurate. However during searching for the accurate green space range we encountered few issues. Detection of grasslands together with forests was infeasible within a single interval. While grasslands are pixels of a bright green, forests are the ones with a very dark one together with a dark grey in the locations where one tree shadows the other. Therefore we had to distinguish between them and thus do two separate detections. Combination of them then produced desired result. Detection of forests, grasslands and their combination can be seen in the Figure 1. Detected locations are marked with the RGB value (0, 200, 0) which is solid green. Furthermore our method computes intensity matrix which indicates influence of the detected greenery on every



(a) Mask of the forests (b) Mask of the grasslands

(c) Unmodified Zrich area (d) Greenery in the Zrich area

Fig. 1: Visualization of the greenery detection phases. In the original image we firstly detected forests then grasslands and finally we combined them and overlay above the original image.

place within the city of Zrich.

Our approch firstly produces binary matrix B of the image P with greenery locations G such that

$$B_{i,j} = \begin{cases} 1 & \text{if } P_{i,j} \in G \\ 0 & \text{if } P_{i,j} \notin G \end{cases}$$

In the second phase we perform so called "smoothing" of the greenery edges. After the first phase where we have obtained matrix B, we now have to compute influence of the greenery on the pixels without it i.e. the ones with the value 0. The reason why it is needed is because inhabitans do not live in the forests or on the meadows but rather takes profit from living close to them. Therefore we have to "influence" pixels that are in the close neighbourhood with the greenery ones. Our algorithm is iterative and so in every iteration visits every entry in the matrix and checks all neighbouring pixels, finds maximum of them and assigns its half to the current one. The method can be more formally defined as in Algorithm 1.

The effect of the smoothing Algorithm 1 after 2 iterations on the Matrix 1 can be seen in the Matrix 2. Remaining 0 entries would be replaced by 0.125 after the third iteration.

Algorithm 1 Smoothing

procedure SMOOTHING(iterations, B)
    $S \leftarrow B$
    $i \leftarrow 0$
    while $i < 0$:
        forall p **in** S :
            $p := max_{neighbours}(p)/2$
        $i \leftarrow i + 1$

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{1}$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0 & 0.25 & 0.5 & 0.5 & 0.5 \\ 0 & 0.25 & 0.5 & 1 & 0.5 \\ 0 & 0.25 & 0.5 & 0.5 & 0.5 \end{pmatrix} \tag{2}$$

6. URBAN QUALITY PREDICTION

TODO: This should be "core" chapter of our paper. Philippe here you can describe whole procedure of data visualization + correlation matrix etc.

7. PREDICTIONS EVALUATION

For evaluating the predictions we derived in the previous section, we chose to let people rate that areas based on their visual impression after walking through the respective part of the city. For this we initially wanted to use Smart Agora, as it allows to pose the questions at various locations.

Instead we generated a map with markers in various positions and asked people to answer the same questions at those locations. Whereas this does not allow to gather additional information, it is more comfortable to use and easier to collect data from different people. All particpants were asked to answer the following five questions, where the first two are single choice and the last three are multiple choice questions. The order of the questions was the same for each path but randomly varied between different locations. The quesitons and answer options can be found in App. 11.1.

8. DISCUSSION

TODO: Discuss possible drawback and ways how to improve them + results and whether we are satisfied with our predictions.

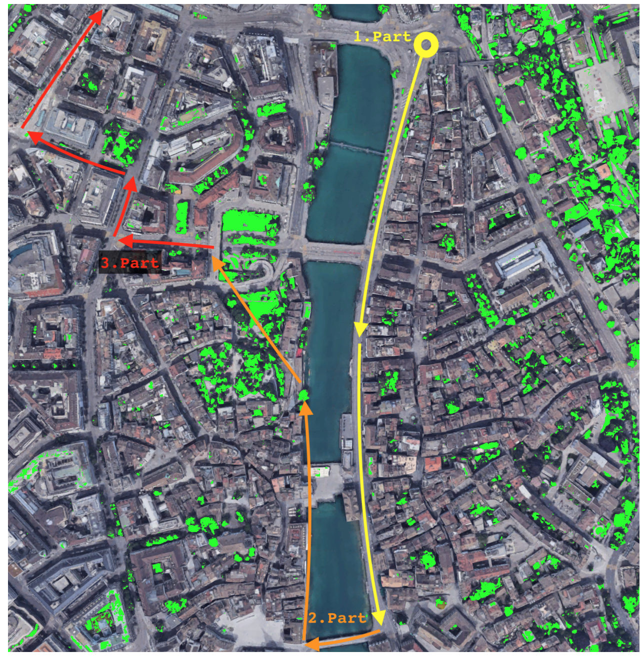


Fig. 2: Greenery detection in the area of our path.

9. CONCLUSION

We should do as the last part together with abstract.

10. REFERENCES

[1] Danielle Griego, Varin Buff, Eric Hayoz, Izabela Moise, and Evangelos Pournaras, "Sensing and Mining Urban Qualities in Smart Cities," in *2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*. mar 2017, pp. 1004–1011, IEEE.

[2] Paolo Neirotti, Alberto De Marco, Anna Corinna Cagliano, Giulio Mangano, and Francesco Scorrano, "Current trends in Smart City initiatives: Some stylised facts," *Cities*, vol. 38, pp. 25–36, jun 2014.

[3] Kunwar P. Singh, Shikha Gupta, and Premanjali Rai, "Identifying pollution sources and predicting urban air quality using ensemble learning methods," *Atmospheric Environment*, vol. 80, pp. 426–437, dec 2013.

[4] Aidin Ferdowsi, Ursula Challita, Walid Saad, and Narayan B. Mandayam, "Robust Deep Reinforcement Learning for Security and Safety in Autonomous Vehicle Systems," may 2018.

[5] Mohammad Abu-Matar and John Davies, "Data Driven Reference Architecture for Smart City Ecosystems," may 2018.

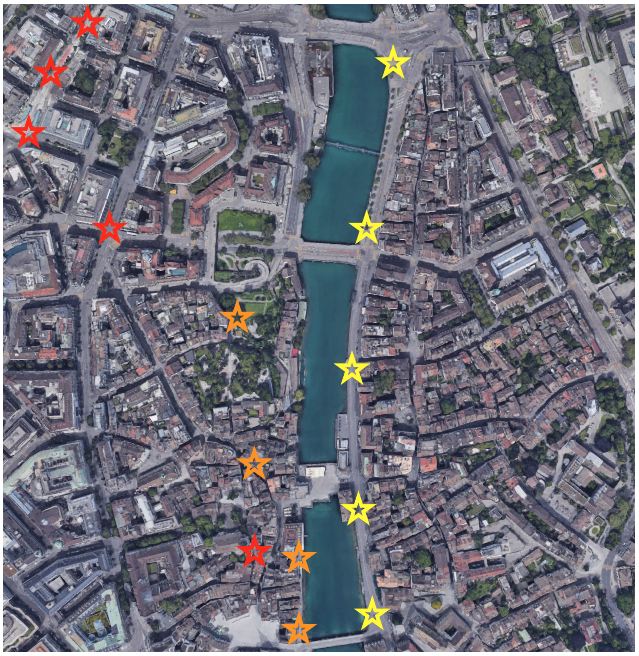


Fig. 3: Locations of question points along the path

[6] Muhammad Usman, Muhammad Rizwan Asghar, Imran Shafique Ansari, Fabrizio Granelli, and Khalid Qaraqe, "Technologies and Solutions for Location-Based Services in Smart Cities: Past, Present, and Future," *IEEE Access*, vol. 3536, no. c, pp. 1–1, 2018.

[7] Rob Kitchin, "The real-time city? Big data and smart urbanism," *GeoJournal*, vol. 79, no. 1, pp. 1–14, Feb 2014.

[8] "China's ever-expanding survaillance state," online: https://thediplomat.com/2018/04/chinas-ever-expanding-surveillance-state/, 2018, Accessed: 12/May/2018.

[9] "China turns to tech to monitor, shame and rate citizens," online: https://www.cnet.com/news/china-turns-to-tech-to-monitor-shame-and-rate-citizens/, 2018, Accessed: 12/May/2018.

[10] "Chinese man caught by facial recognition at pop concert," online: http://www.bbc.com/news/world-asia-china-43751276, 2018, Accessed: 12/May/2018.

[11] "Zurich open data," online: https://data.stadt-zuerich.ch/, Accessed: 08/May/2018.

[12] "Geographic information system of the canton of Zurich," online: http://maps.zh.ch/, Accessed: 08/May/2018.

[13] Xiaojiang Li, Chuanrong Zhang, Weidong Li, Robert Ricard, Qingyan Meng, and Weixing Zhang, "Assessing street-level urban greenery using Google Street View and a modified green view index," *Urban Forestry and Urban Greening*, vol. 14, no. 3, pp. 675–685, 2015.

11. APPENDIX

11.1. Questions for survey

The following questions were asked to evaluate our predictions. After each question one finds the possible answers. The first two questions are single choice, whereas the other three are multiple choice. The order of questions was the same for each participant, but varied for the different locations.

1. Would you like to live hear if you had the opportunity?
   - certainly yes
   - rather yes
   - rather not
   - certainly not
2. Is it likely, you could currently afford a flat in this area?
   - certainly yes
   - rather yes
   - rather not
   - certainly not
3. What do you like about this place?
   - It's close to transport.
   - It's close to shopping.
   - It's close to other palces I often visit.
   - It's calm with lots of green.
   - It's visually appealing.
   - It's bright and open.
   - It's modern and well maintained.
4. What do you dislike about this place?
   - It's far from transport.
   - There are no/insufficient shopping possibilities nearby.
   - It's too far to the places I usually visit.
   - It's very busy and there are no tree nearby.
   - It looks ugly.

- It’s very confined and poorly illuminated.
- It’s old and shaby.

5. What can you see from here?

    - A stop for public transport.
    - Parking facilities.
    - Store for buying food.
    - A restaurant or a bar.
    - Construction site.
    - A park.
    - A nice flat that I could live in.