Parallel Cities

Yuichiro Takeuchi*
Sony Computer Science Laboratories Inc.

Abstract

In this paper we introduce the concept of *Parallel Cities*—live, 3D simulations of actual cities kept "true to life" using real-time information from their real-world counterparts. A key enabler of the concept is the ever-growing tide of SNS (e.g., Twitter) updates, that can be aggregated and analyzed to obtain a rough, but *macroscopically accurate* picture of the current state of a city. Metaphorically, Parallel Cities will allow users to instantly "open windows" to any place in the world, providing an attractive enhancement to existing online mapping services. The paper will offer a concise discussion of the technical details of the concept, centered around descriptions of a working prototype that targets the city of Tokyo, Japan.

CR Categories: H.5.1 [Information interfaces and presentation (e.g., HCI)]: Multimedia Information Systems—Artificial, augmented, and virtual realities

Keywords: Parallel City; city simulation; Twitter; social networking services; urban sensing

1 Introduction

As per Mark Weiser's predictions [Weiser 1991], our world is now saturated with myriad types of digital devices, some carried by individuals (e.g., smartphones) and others embedded within the environment (e.g., interactive public displays). Such devices, combined with the increasingly wide availability of wireless connectivity, can potentially be appropriated as an army of smart, networked sensor nodes that continually capture and publish information about cities and their inhabitants. This prospect of a massive, widely distributed sensing mechanism has naturally drawn attention in UbiComp/HCI circles, spawning the neologism "urban sensing" [Cuff et al. 2008]. Although urban sensing has already produced several useful applications (traffic data overlay on Google Maps is a popular example), much of the concept's promise still lies in the future, since information that can be obtained via urban sensing is expected to grow—in both quantity and type—as new digital devices/services continue to be introduced to the market.

If we adopt a slightly speculative stance, we can extrapolate current trends and imagine that at some point in the future, the range of data that can be collected through urban sensing will be such that practically all information about cities "that are of interest" will be available to be used in applications. And combining this wealth of urban data with photorealistic, 3D city simulation (which is already technically feasible), we will be able to create lifelike, and consistently up-to-date, simulations of actual cities that could be virtually indistinguishable from live video feeds. In effect, these simulations will give users a power comparable to being able to instantly open windows to any place on earth. We call these hypothetical city simulations *Parallel Cities*.

*e-mail:yutak@acm.org

VRST'13, October 06 - 09 2013, Singapore, Singapore

http://dx.doi.org/10.1145/2503713.2503742

Figure 1: *Parallel Cities prototype.*

This paper will provide a succinct description of the concept of Parallel Cities. Discussions will center around our implemented prototype (Figure 1) targeting central Tokyo, which uses Twitter feeds as its primary data source—we believe this to be a rational choice, as the growing usage of SNSs makes them one of the most promising "urban sensors" in the near- to mid-term future.

2 Parallel Cities

Figure 2 shows the basic technical components that make up a "Parallel City". Visually, the simulation consists of an assortment of 3D models, that can be categorized into *environment* and *objects*. Here, environment is a large, mostly static model that contains the overall city landscape including buildings (essentially a 3D map, similar to Google Earth or Apple's iOS Maps), whereas objects are models of humans, vehicles, etc. placed and animated in large numbers on top of the environment. Both the appearances and behaviors of the 3D models are continuously adjusted using real-time urban data, which are collected by dynamically aggregating and analyzing raw input from a diverse range of urban sensors. For example, lighting effects on the environment model may be modified according to up-to-date weather data, and humans can be added to or subtracted from areas based on real-time cellular network data. The list of potential urban sensors is virtually endless; for near-future implementations, some of the most useful sensors would be cellular networks, microblogging/SNS feeds, surveillance camera systems, and wireless sensor networks. The recent introductions of a new generation of wearable devices (e.g., Google Glass, Nike+ Fuelband) also present exciting new opportunities for urban sensing. For example, it might become common for people to continually publish live video captured using wearable cameras, or to tag SNS updates with biometric data. Such information can further improve the fidelity of Parallel Cities.

One critically important aspect of Parallel Cities is that they will be accurate only at the *macroscopic* level—microscopic details are not guaranteed to conform to reality. Environment models will never be exact down to each cobblestone, and a red Prius that is just about to enter an intersection in the actual city will most likely not appear in its *Parallel* counterpart. What are preserved, instead, are the higher-level characteristics: the overall city landscape, level of traffic, size/character of human crowds, etc. As a result certain usage scenarios, that would be possible if we were actually looking at a live video of the city, are by design out of reach for Parallel Cities; for example, a user cannot monitor the movement of an actual friend by looking at

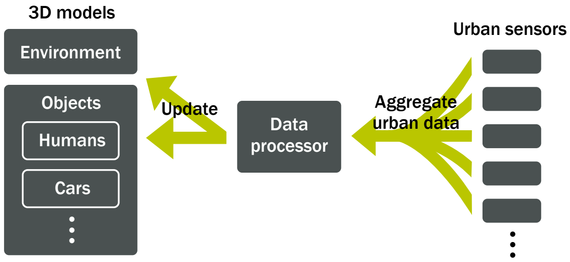

Figure 2: *System components.*

the simulated city. However, this is not to say that the truthfulness of Parallel Cities is static. Technical advances, and introductions of new urban sensors will increase the fidelity, and some details of the city that are now considered microscopic (and hence not replicable in Parallel Cities) may become replicable in the future.

If our goal is to obtain a grasp of the general atmosphere or state of the city, the lack of microscopic accuracy should not necessarily be a hindrance. In fact, for many of the minute details of the city, users should not even be able to discern if they faithfully reflect reality or not. For example, as long as the higher-level characteristics of the crowd are maintained, it is impossible to know if each of the human models in the simulation is an accurate replica of an actual person living/working there, or was instead simply sampled from a generic library of 3D human models. Also, studies in urban design [Lynch 1960] show that outside of few select areas where one has the highest levels of familiarity, small details in the environment do not feature in a person's cognitive understanding of a city. Thus, except in cases where they knowingly look at their homes, offices, etc., users may well be blind to minor inaccuracies in Parallel Cities—whether in humans, traffic or the environment.

3 Related Work

In the past decade, urban sensing [Cuff et al. 2008] has emerged as a prominent new research topic in the UbiComp/HCI communities, triggered by the rapid proliferation of mobile devices. While these works have their roots in prior research on wireless sensor networks [Akyildiz et al. 2002], urban sensing is characterized by its focus on citizen participation [Burke et al. 2006], and in its noted reluctance toward introducing proprietary sensing hardware, instead favoring clever appropriations of existing (and ideally widespread) devices such as cell phones [Eagle and Pentland 2006]. Types of information that can be collected via urban sensing include locations of humans [Reades et al. 2007] and vehicles [Leduc 2008], images/sound [Reddy et al. 2007], air quality [Dutta et al. 2009], etc.

Growing usages of social networking services have sparked a wave of studies targeting them as sources for data mining. Attempts have been made to extract data such as latent user profile (e.g., age, race, political affiliation) [Rao et al. 2010], emotions/sentiments [Kramer 2010], and occurrences of major events [Sasaki et al. 2010]. While the mining techniques themselves are typically straightforward adaptations of existing methods, the real-time nature and sheer size of the data set open previously unattainable possibilities.

Perhaps the closest precedent to Parallel Cities is AVE (Augmented Virtual Environment) [Neumann et al. 2003], where real-time footage from surveillance cameras is seamlessly integrated into 3D environment models. While AVE systems offer *microscopic* accuracy, due to high cost and privacy concerns the idea of installing legions of surveillance cameras to cover entire cities (and also making their feeds publicly available) is unrealistic, and thus unlike our approach is not scalable to city-wide (or world-wide) levels.

4 Prototype

To study the validity of the concept of Parallel Cities, we developed a preliminary prototype that targets central Tokyo. Compared to our idea of a *true* Parallel City system, the prototype is limited in many aspects—e.g., location, number and types of urban sensors, degree of photorealism, etc. Our goal in creating the prototype was to realize an implementation that is "just good enough" to perform a general evaluation of the potential and feasibility of the concept. Also, in building the prototype we faced some restrictions (related to budget and lab policy) that forced us to only use publicly available data; thus some highly useful urban data such as cellular network data are not incorporated in this prototype. While this takes a toll on fidelity, our reliance on public data also means that readers can easily follow our descriptions and build their own versions of the prototype. The prototype is written in Objective-C, and runs on Mac OS X.

4.1 Environment

Instead of a continuous, large-scale 3D model of the city, our environment model consists of multiple small, discrete models—i.e., a library of photographs shot at different parts of central Tokyo, onto which 3D models have been mapped manually. This difference has a major effect on the user experience; unlike in our vision of a *true* Parallel City system, where users can navigate through a continuous simulated world (à la Google Earth), our prototype merely permits users to switch between a set of disconnected, predetermined areas, and does not support camera movements outside of simple panning/zooming operations. The environment is static, aside from weather (obtained from the Yahoo! Japan website) and lighting changes depicted using image filters and OpenGL effects.

4.2 Crowd

Crowds are depicted by placing and animating (according to a simple algorithm) human models on the environment. As it has proved difficult for us to obtain free 3D human models that fit naturally into our context of Tokyo, we instead created our own (admittedly non-photorealistic) models, as shown in Figure 3. We created eight sets of models in total; one for each age/gender class used in our Naive Bayes classifier (described later).

We rely on Twitter analysis to estimate both the sizes and characters (i.e., age/gender distributions) of human crowds in cities. For crowd size estimation, cellular network data is known to give more reliable results with significantly less effort, and thus our technique is meant as a quick "hack" for cases like ours where such data is unavailable. On the other hand, we found Twitter analysis to be highly effective for estimating crowd character.

We attempt to estimate crowd size by looking at the frequency with which tweets are sent from our area of interest (*area_name*). How-

Figure 3: *Human models (subset).*

ever, since the practice of geotagging (i.e., attaching GPS locations to) tweets is not as widespread in Japan as in the US or Europe, we cannot directly obtain tweets uploaded from a particular area. Thus, we instead look for tweets that contain *area_name* as keywords; this straightforward approach works surprisingly well for Tokyo, since areas are generally called by the names of their nearest train/subway stations that are densely laid out throughout the city—in effect, central Tokyo can be divided into hundreds of tiny areas, each with its own distinct name. This simple approach may fail in other regions, however. For example, in New York City it is much more common for people to use street names (e.g., Mercer St., 3rd Ave., etc.) than region names (e.g., SoHo, Greenwich Village, etc.) to describe their whereabouts, which is problematic due to the frequent existence of identical street names, and their often coarse granularity (Broadway in New York actually runs more than 30 miles, from lower Manhattan to Sleepy Hollow). Looking for geotagged tweets may prove to be a better approach in such contexts.

Note that our strategy fails to differentiate between tweets sent *from* an area, and tweets *about* an area. Here, we simply rely on the real-time nature of Twitter (its principal usage as an outlet of real-time activity updates [Naaman et al. 2010]) to ensure that the number of former tweets will be significantly larger than that of the latter.

We assume the crowd size as being roughly proportional to the frequency of tweets—i.e., crowd size can be calculated as $c * f$, where f denotes the tweet frequency, and c is a *region-specific* coefficient whose value differs for each street, each plaza, each block, etc. For example, the crowd size for a particular street in Harajuku is calculated by multiplying the c for that street with f for the entire Harajuku area. Note that city-wide values of c must somehow be determined beforehand. While there may be ways to automatically compute them from zoning maps, statistical data, etc., for our prototype we manually set the values (through trial and error) for each of our discrete environment models.

To estimate crowd character, we similarly search for tweets recently sent out from our area of interest (again substituted with tweets that contain *area_name* as keywords). We then look at each of the users who had sent out these tweets, and determine both their gender and approximate age. This process consists of two steps. First, we compare the user name to a list of common Japanese names, to see if it can be regarded as an unambiguously male or female name. Next, we query the user's past tweets from Twitter and break them down into words, and feed them as input to a Naive Bayes classifier, that had been trained using timelines of 2574 Twitter users whose approximate age and gender are both explicitly stated in their profiles. The classifier determines each user as belonging to one of 8 classes, as shown in Table 1. (The classes are skewed to the younger side—Twitter in Japan tends to be more popular among the young, and we struggled to find enough older users to include in our training set.) After classification is complete for all users, we can easily calculate what percentage of users lies in each of the 8 classes; the values are then adjusted using survey data regarding the ratio of Twitter users in each age/gender class, to compensate for the demographic differences between Twitter users and the general population.

4.3 Traffic

Traffic simulation in our prototype is quite straightforward; models of cars basically follow a simple algorithm of running straight along a road, and randomly making turns at intersections. Since traffic is in general orderly (though it can be heavy at times) in central Tokyo, a modest set of rules can realistically model traffic behavior in most situations. Numbers of cars on roads are constantly adjusted using real-time traffic data from Google Maps. To maintain a consistent aesthetic style, we created our own 3D models for the automobiles,

	Age			
	$<= 24$	20–34	30–44	$>= 40$
Female	Class 1	Class 2	Class 3	Class 4
Male	Class 5	Class 6	Class 7	Class 8

Table 1: *Age/gender classes.*

in the same way as the human models. Cars displayed in the simulation are randomly chosen from a single library—i.e., the prototype does not take into account possible regional variations in car styles, such as color, shape or brand. (It may be possible to infer these from income statistics or auto-market data in future implementations.) In addition, the prototype currently does not display any vehicles other than cars, such as bicycles, motorcycles, trams, etc.

4.4 Special Events

In theory, we can incorporate any type of special events, so long as they can be detected through analyses of urban data and we prepare specialized graphical effects to visualize them in the simulation. In our prototype, we have incorporated the blooming of cherries in the spring (again detected via Twitter) as a special event. Other events that may be helpful if simulated include festivals, bazaars, car accidents, queues for restaurants/concert halls, etc. Many of such events can be monitored (albeit with varying accuracy) through Twitter—nowadays we can expect any noteworthy event to eventually make its way into the collective consciousness of tweets.

4.5 Performance

We conducted several tests, to quickly evaluate the performances of our Twitter analysis techniques.

To test the accuracy of crowd size estimation, we calculated crowd sizes for eight areas in central Tokyo, at 10-minute intervals for an entire 24-hour day. We then compared these results with our prior expectations, derived from general area usage (residential, commercial, etc.) and train/subway station usage statistics. As a result, we found estimated crowd sizes to be roughly consistent with expectations for commercial and business districts; for areas known to be nightlife hubs (such as Shibuya and Roppongi) crowd sizes show a clear peak in the evening, and for areas known as corporate districts (such as Shinagawa and Osaki) the values are more stable but with small peaks in the morning and evening (likely caused by Tweeting commuters). On the other hand, in residential neighborhoods crowd sizes barely fluctuate throughout the day—the number of tweets is evidently too small in these areas to yield reliable results.

We conducted two tests to evaluate the accuracy of crowd character estimation. In the first test, we assessed the performance of our age/gender classification method (consisting of user name checking and Naive Bayes classification), using a set of 150 Twitter users whose age/gender were both already known. As a result, the system could classify users into their correct age/gender class 75.3% of the time, and could determine either the age or gender correctly 98.0% of the time. While these are good results, remember that since we are only interested in *macroscopic* accuracy, what really matters to us is not classification accuracy per se, but whether our method can correctly estimate the higher-level demographic properties of an entire group of users. We thus calculated the mean age & percentage of females for the entire group of 150 users—the mean age was off only by 1.5 years, and the percentage of females was off only by 2.0%.

In the second test, we calculated macroscopic demographic properties (mean age & percentage of females) for six areas within central Tokyo (we took 10 separate measurements throughout a single day,

Figure 4: *Prototype screenshots juxtaposed with photographs.*

and averaged the results), and compared them with our prior expectations. While we cannot check the accuracy of the actual numbers themselves, the results do seem to reflect the general characteristics of the different areas; for example, there is a clear contrast between the estimated female ratios in the male-dominated business districts (Shinagawa and Shinbashi, 21.7% and 24.5% respectively) and the teenage fashion center of Harajuku (65.6%). Surprisingly, even for residential neighborhoods such as Shirokanedai or Nezu, the results match our impressions of these areas quite nicely; this presumably is because crowd characters generally do not fluctuate as rapidly as crowd sizes, and hence the small number of tweets in these areas is less of a handicap compared to estimation of crowd sizes.

Figure 4 shows some screenshots of the prototype, juxtaposed with photos shot at roughly the same time/place. Ultimately, the question of whether simulations produced by Parallel Cities seem realistic or not is a highly subjective one, but to our eyes the prototype already succeeds reasonably well in reproducing the overall *vibe* of the different regions of Tokyo—the contrast between drab Shinagawa and young, colorful Harajuku is eloquently visualized, for example.

5 Conclusion

In this paper we described the concept of Parallel Cities—city simulations of the future kept “true to life” using real-time urban data. Users of Parallel Cities can freely open windows to any place in the world, taking virtual tours to foreign nations, or checking the neighborhood to see if anything interesting is going on. Though our current prototype is limited in many aspects, it already conveys general atmospheres of cities with reasonable success.

One important issue we have failed to discuss is privacy. The exact types of urban data that would be available for Parallel Cities is contingent on our constantly evolving attitudes towards privacy, which is hard to predict and also would have regional differences (already the US and EU have vastly different laws regarding privacy issues). However, through our prototype we have shown that it is possible to extract many useful information from Twitter alone—i.e., data that had been willingly shared by users, not collected surreptitiously or involuntarily without consent.

References

AKYILDIZ, I. F., SU, W., SANKARASUBRAMANIAM, Y., AND CAYIRCI, E. 2002. Wireless sensor networks: A survey. *Computer Networks 38*, 4, 393–422.

BURKE, J., ESTRIN, D., HANSEN, M., PARKER, A., RAMANATHAN, N., REDDY, S., AND SRIVASTAVA, M. B. 2006. Participatory sensing. In *Proc. of WSW at SenSys 2006*.

CUFF, D., HANSEN, M., AND KANG, J. 2008. Urban sensing: Out of the woods. *Communications of the ACM 51*, 3, 24–33.

DUTTA, P., AOKI, P. M., KUMAR, N., MAINWARING, A., MYERS, C., WILLETT, W., AND WOODRUFF, A. 2009. Common sense: Participatory urban sensing using a network of handheld air quality monitors. In *Proc. of SenSys 2009*, 349–350.

EAGLE, N., AND PENTLAND, A. 2006. Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing 20*, 4, 255–268.

KRAMER, A. D. I. 2010. An unobtrusive behavioral model of “gross national happiness”. In *Proc. of CHI 2010*, 287–290.

LEDUC, G., 2008. Road traffic data: Collection methods and applications. JRC Technical Note.

LYNCH, K. 1960. *The Image of the City*. MIT Press.

NAAMAN, M., BOASE, J., AND LAI, C. 2010. Is it really about me? message content in social awareness streams. In *Proc. of CSCW 2010*, 189–192.

NEUMANN, U., YOU, S., HU, J., JIANG, B., AND LEE, J. 2003. Augmented virtual environments (ave): Dynamic fusion of imagery and 3d models. In *Proc. of VR 2003*, 61.

RAO, D., YAROWSKY, D., SHREEVATS, A., AND GUPTA, M. 2010. Classifying latent user attributes in twitter. In *Proc. of SMUC 2010*, 37–44.

READES, J., CALABRESE, F., SEVTSUK, A., AND RATTI, C. 2007. Cellular census: Explorations in urban data collection. *IEEE Pervasive Computing 6*, 3, 30–38.

REDDY, S., PARKER, A., HYMAN, J., BURKE, J., ESTRIN, D., AND HANSEN, M. 2007. Image browsing, processing, and clustering for participatory sensing: Lessons from a dietsense prototype. In *Proc. of EmNets 2007*, 13–17.

SASAKI, T., OKAZAKI, M., AND MATSUO, Y. 2010. Earthquake shatters twitter users: Real-time event detection by social sensors. In *Proc. of WWW 2010*, 851–860.

WEISER, M. 1991. The computer for the 21st century. *Scientific American 265*, 3, 94–104.