Chapter XVII
QyoroView: Creating a Large-Scale Street View as User-Generated Content

Daisuke Tamada
Osaka University, Japan

Hideyuki Nakanishi
Osaka University, Japan

ABSTRACT

A lot of street view services, which present views of urban landscapes, have recently appeared. The conventional method for making street views requires on-vehicle cameras. We propose a new method, which relies on people who voluntarily take photos of an urban landscape. We have developed a system called QyoroView. The system receives photos from users, adjusts the photos’ position and orientation, and finally synthesizes them to generate a street view. We conducted two experiments in which the subjects generated a street view using our system. We also observed and interviewed the subjects who participated in order to learn their impression of the system.

INTRODUCTION

Satellite or aerial photos can show users only rough geographical information. Street views can provide detailed geographical information. Street views are panoramic images of urban landscapes that are made from movies or a collection of photos taken along the streets of a city. The very first interactive street view is ‘Aspen Movie Map’ (Lippman, 1980). Recently, various kinds of Websites that provide street views have appeared, such as Google Maps Street View (maps.google.com/help/maps/streetview) and Microsoft Photosynth (labs.live.com/photosynth). Street views can be used for urban navigation, real estate property purchases, land surveying.
Cameras mounted onto cars are generally used to record movies or collect photos to create these street views. This makes it necessary to drive the cars for long distances. Since special equipment and considerable labor are required to produce street views, only large enterprises are usually able to offer large-scale street views.

The amount of user-generated content (UGC), that is, media content created by end users, is rapidly increasing. UGC has become vital in today’s daily lifestyles (Wunsch-Vincent, & Vickery, 2007, p. 9). Even if each user creates only a small piece of content, the quantity of the total amount of content sometimes exceeds that created by professionals (Lih, 2004, p. 5).

We developed a system called ‘QyoroView,’ that can produce a large-scale street view as UGC. QyoroView can synthesize street views from mobile phone photos that contain position data. We assume that typical QyoroView users will be pedestrians who have a GPS camera phone. In Japan, government required “All 3G mobile phones released after 2007 must, in principle, be equipped with GPS.” When pedestrians take and upload photos of buildings along streets, QyoroView updates the street view images so they can be viewed using a Web browser. Since QyoroView requires no extra special equipment, it is able to produce street views as UGC.

**BACKGROUND**

There are several approaches to producing street views. The basic style of this content is a hand-crafted panoramic image (Chen, 1995). An enormous amount of effort is required to construct wide-area panoramic images (Hartono et al., 2006). Using this approach, it is possible to produce street views of some spots, but it is nearly impossible to produce views along streets.

Past studies proposed several approaches to produce street views automatically using on-vehicle equipment. A common technique for collecting panoramic images of many spots along streets has been to use on-vehicle omnidirectional cameras (Koizumi, & Ishiguro, 2005; maps.google.com/help/maps/streetview; preview.local.live.com). A previous study used on-vehicle laser scanners to obtain 3D models of buildings along streets (Fruh, & Zakhor, 2001). Another previous study used on-vehicle cameras mounted to the side of a car to capture panoramic images of buildings along streets (Roman, Garg, & Levoy, 2004). These approaches require special equipment and long-distance driving. Thus, it is costly to produce wide-area street views.

‘Photo Tourism’ is a system that can make street views from a collection of photos that are freely available in photo sharing sites (Snively, Seitz, & Szeliski, 2006; labs.live.com/photosynth). So, it requires no equipment or labor to produce street views. However, a sufficient number of photos recording the same location are needed to concatenate photos using image processing technology. Many sightseers take photos at tourist attractions and some of them upload their photos to photo sharing sites, so this approach is applicable to the production of street views of such places. It is, however, not a good idea to apply the approach to the production of street views of featureless urban areas due to the lack of motivation for users to take photos.

‘OpenStreetMap’ is a UGC system that makes freely editable maps. The map data in the system has mostly been built by users who upload data from their phones or PDAs equipped with GPS devices (openstreetmap.com)]. Recently, Yahoo! allowed OpenStreetMap to use their vertical aerial imagery. So, users can create vector-based maps using online editing tools that can overlay aerial images. The system produces only vector-based maps of urban streets.

There are many services, including ‘Panoramio’, that have a photo map function that places photos uploaded from end users onto a map (www.panoramio.com). Some projects have tried to extend this basic concept. ‘Balog’ is a Web-based
QyoroView diary system combined with a photo map function (Uematsu et al., 2004). Some researchers used multimedia projectors to display a large photo map on the floor as installation art (Nakanishi, Motoe, & Matsukawa, 2004). The ‘Degree Confluence’ project is trying to create a world-wide photo map by collecting photos of integer degree intersections for latitudes and longitudes around the world (www.confluence.org). None of the photo map systems, including those we have described here, produce panoramic images.

QyoroView can produce a large-scale street view as UGC. As discussed above, no conventional approach can do that. On-vehicle equipment enables the production of wide-area street views, but such equipment is not very affordable for end users. The Photo Tourism system can produce street views as UGC, but the system is not applicable to wide-area street views. Photo map systems can collect photos for a wide area, but they do not produce panoramic images.

SYSTEM DESIGN

QyoroView synthesizes street views from mobile phone photos. To make street views, the system needs position data and orientation data for the photos. Some recent mobile phones are equipped with GPS. GpsOne is one of the most accurate positioning services for mobile phones. Although the error rate of GpsOne is five to ten meters even when the device catches signals from more than three satellites (Wang et al., 2004, p. 342), it is still difficult to detect the side of the street on which a user is standing. Some mobile phones are also equipped with a magnetic compass. Magnetic compasses are affected by surrounding magnetic fields. Mobile phones themselves disturb magnetic compasses because the phones generate a magnetic field. Therefore, magnetic compasses are too inaccurate to orient photos. With QyoroView, users modify the positions and the system of QyoroView modifies the orientations of photos. This section explains how QyoroView collects and concatenates mobile phone photos, and how it was designed and implemented.

Collecting Mobile Phone Photos

Figure 1 shows how to take photos based on the standard user interface that is available on almost all of the GPS camera phones released in Japan. We assume that typical users of QyoroView will be pedestrians who have a GPS camera phone. In the figure, the user’s phone obtains the current position and displays a map around it. The icon, which indicates the position sent to the server side system of QyoroView, is initially located at the center of the map (Step 1 in Figure). The first task for the user is to adjust the icon’s location. The user moves the icon to a target on the other side of the street, which is usually a building (Step 2). Then, the user takes a photo of the target (Step 3).

If the user intends to upload only a single photo, the final task is to send the system an e-mail message that contains the photo and the position data. We have also prepared a more efficient way. The user can upload several photos at once, as long as the street is not very curved. As shown in Figure 1, after taking the first photo, the user walks along the street for a certain distance (Step 4) and takes a second photo (Step 5). This time the user does not need to adjust the location of the icon. The user can take a third, fourth, and more photos in the same way. Only when the user takes the last photo does he or she have to move the icon to the target on the other side of the street in the same way as when taking the first photo. Finally, the user sends the system an e-mail message that contains the photos and the data of the first and last positions. The positions of the intermediate photos between the first and last photos will be calculated by the system.

The user needs to be aware of two things to help the system smoothly concatenate photos. The first thing is that the user should not tilt the phone. Different angles of the phone result in different
altitudes of the horizon in the photos. To smoothly concatenate photos taken by different users, all users need to keep the phone at the same angle, which is perpendicular to the horizontal plane in the usual case. Another requirement is that the walking distance between each shot should not be too long. If the distance is too long, a missing area will appear between adjacent photos.

**Synthesizing Street Views**

When the server side system of QyoroView receives the e-mail message that contains the photos and the data of the first and last positions, the system calculates the positions of the intermediate photos between the first and last photos based on a linear interpolation method. As a result, every photo is accompanied by its position. At this time, the photos do not have orientation, and may not stand along the street very precisely due to the user’s inaccurate adjustment of the positions, and also due to the curving of the street. Therefore, as shown in Figure 2, the system adjusts the position and determines the orientation of each photo based on a vector map, which contains the border and halfway vectors of streets.

1. **Adjusting the Position**

   The system finds the border vector that is closest to the photo’s original position (Step 1 in Figure 2). Then, the system changes the photo’s position to a new position that is on the border vector and that is also closest to the original position (Step 2).

2. **Determining the Orientation**

   Since the orientation of the photo must be perpendicular to the border vector on which the photo is placed, there are only two choices for the orienta-
The system finds the halfway vector that is closest to the photo's new position (Step 3). Then, the system picks up the orientation that aims the photo toward the halfway vector (Step 4).

Figure 3 illustrates how uploaded photos are concatenated and pasted along the street. Currently, the street view produced by the system is just a collection of photos pasted on a monochromatic street map that is drawn based only on the vector map.

Figure 4 shows the snapshots of the client side system of QyoroView, which is an interactive map running on a Web browser. You can scroll the map by drag operation, and zoom in and out by rotating the mouse wheel. The figure shows the map in three zoom levels. These functions are similar to conventional two-dimensional map services. You can also rotate the map. Clicking the right mouse button rotates the map ninety degrees counterclockwise. Usually, only three-dimensional map services have this kind of function. QyoroView is basically a two-dimensional map service, but the rotating function is essential, because it is not easy to recognize scenes of upside-down photos.

**System Architecture**

Figure 5 depicts the architecture of QyoroView. The server side system of QyoroView consists of four modules, three databases, and two data folders. Each module works as follows.

1. **Receiving Module**

The receiving module parses the e-mail messages sent from the user's mobile phone to retrieve the photos and position data. Then, the module stores the photo images in the photo data folder and stores the indices to the photo images, the position data, the sender's name, and a timestamp in the e-mail database. The e-mail software of mobile phones usually has some limitations. Examples include a maximum number of files and position data that can be attached to a single message. So, the system allows users to send a series of data, which are the photos and the data of the first and last positions, via multiple e-mail messages. To enable the reconstruction of this series of data, the module stores the sender's name and a timestamp for each e-mail message.

2. **Positioning Module**

The positioning module reads the position data and the indices to the photo images of the same sender from the e-mail database. Then the module reconstructs the series of data, which are the ordered photo indices and the data of the first and last positions, according to the timestamps. After the reconstruction, as explained in the previous subsection, the module calculates the positions of the intermediate photos between the first and last photos based on a linear interpolation method. Finally, in the position database, the module stores the photo indices, each of which is accompanied by its position data.
3. Placing Module

The placing module reads the position data which have been added to the position database by the positioning module. The module also reads a vector map around the positions from the vector map database. Then, as shown in Figure 2 and detailed in the previous section, the module adjusts the position and determines the orientation of each photo based on the vector map. Finally, the module stores the adjusted position data and the orientation data in the position database.

4. Drawing Module

In the street view data folder, the street view image is divided into blocks of a fixed size (e.g., 500 pixels square) and stored. When the drawing module updates the street view, the module again produces the blocks that are changed by the new photos. First, the module reads the vector map from the vector map database and redraws the monochromatic street map of the blocks based on the vector map. Next, the module reads the photo images from the photo data folder, and reads the photos’ adjusted position data and orientation data from the position database. Then, as shown in Figure 3, the module pastes the photo images on the blocks according to the position and orientation data. Finally, the module overwrites the old blocks in the street view data folder with the produced blocks.

As explained above, the module does not just paste the new photos onto the old blocks, but produces new blocks, because the images of the blocks are stored in JPEG format to keep the file size small. If the module were to read the blocks from the data folder, paste new photos on the blocks, and store the blocks in the data folder again, the quality of the blocks’ images would be degraded.

The client side system of QyoroView requests the street view blocks that are necessary to display the map from the Web server of the QyoroView service. The parameters sent to the Web server contain the location, zoom ratio, and rotation degree of the map. Mouse operation on the map determines these parameters and causes a new request to be sent to the Web server.

System Implementation

The client side system of QyoroView is an Ajax-based application. The server side system of QyoroView runs on Debian GNU/Linux 4.0. The modules were implemented in Java 1.5. The databases were implemented in PostgreSQL 7.4. Apache 2.2 is used as the Web server.

Digital Map 2500 (Spatial Data Framework) published by the Geographical Survey Institute of Japan is used as the vector map. The ‘2500’ means that the vector data are in scale of 1/2500.

Currently, the receiving module can deal with e-mail messages sent from KDDI’s ‘au’ mobile phones with the ‘EZ Navi Walk’ service, since KDDI’s phones make up the majority of GPS phones in Japan. KDDI’s ‘au’ is the second largest mobile phone network in Japan and about ninety percent of ‘au’ users have mobile phones that are equipped with GPS. We are also working on adapting the module to other mobile phone networks, since GPS phones for other networks are appearing.

EXPERIMENT 1

We compared the two methods: the standard method of uploading only a single photo (one-shot capture) and the continuous capture method of uploading several photos at once. We asked three students of our research group to collect photos from around a residential area close to our university campus. The subjects collected photos by the one-shot and continuous methods.
Results 1

Table 1 shows how long it took to take photos with both methods. The table shows that the continuous capture method is more than twice as fast as the one-shot capture method. Figure 6 shows a part of street views produced by the continuous capture method. It took about three and a half hours to collect the photos included in the figure’s area.

It is apparently not common to take pictures of buildings from the other side of the street. The subjects reported that it was especially awkward to take pictures of personal houses, elementary schools, and kindergartens. Interestingly, they reported that it was not awkward to take pictures of strange-looking buildings that did not fit with the surrounding buildings.

In this experiment, subjects discussed at laboratory to divide areas where they should take photos. They seemed to think it would be better to cooperate with each other. In the real case of creating UGC, users may need interaction with others. So, we conducted second experiment.

EXPERIMENT 2

We conducted the second experiment to observe how subjects participated in making street views. In this observation we tried to determine how we should support interaction among users to facilitate the production of street views. As listed in Table 2, the subjects consisted of eight undergraduate students (four male students and four female students). They did not begin taking photos on the same day. Two subjects began in August. One of them retired from the experiment on August 5. The left map of Figure 7 shows the location of the two campuses of our university. Five students lived near the west campus, and the other three students lived near the east campus. The subjects constructed street views of a loop road that includes the two campuses. As a simple incentive, we paid a fixed amount of money for each uploaded photo. We also paid them for transportation expenses to visit the area to take photos, and communication expenses to upload photos. We told them that we could not pay for duplicated photos, meaning that they could not earn money if they took photos at a part of the street where other subjects had already taken photos.

Table 1. Comparing photo-taking times

<table>
<thead>
<tr>
<th></th>
<th>One-shot</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of pictures</td>
<td>29</td>
<td>428</td>
</tr>
<tr>
<td>Total time</td>
<td>50 min</td>
<td>5 hour 31 min</td>
</tr>
<tr>
<td>Average time per image</td>
<td>1 min 44 sec</td>
<td>46.5 sec</td>
</tr>
</tbody>
</table>

Figure 6. A part of street views

Table 2. Subjects of the second experiment

<table>
<thead>
<tr>
<th>Subject</th>
<th>Gender</th>
<th>Campus</th>
<th>Photos</th>
<th>Begin</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Male</td>
<td>West</td>
<td>861</td>
<td>7/24</td>
<td>8/5</td>
</tr>
<tr>
<td>B</td>
<td>Male</td>
<td>West</td>
<td>1898</td>
<td>7/25</td>
<td>8/30</td>
</tr>
<tr>
<td>C</td>
<td>Male</td>
<td>East</td>
<td>320</td>
<td>8/12</td>
<td>8/30</td>
</tr>
<tr>
<td>D</td>
<td>Female</td>
<td>East</td>
<td>153</td>
<td>7/24</td>
<td>8/30</td>
</tr>
<tr>
<td>E</td>
<td>Female</td>
<td>East</td>
<td>139</td>
<td>8/13</td>
<td>8/30</td>
</tr>
<tr>
<td>F</td>
<td>Female</td>
<td>West</td>
<td>116</td>
<td>7/27</td>
<td>8/30</td>
</tr>
<tr>
<td>G</td>
<td>Female</td>
<td>West</td>
<td>718</td>
<td>7/23</td>
<td>8/30</td>
</tr>
<tr>
<td>H</td>
<td>Male</td>
<td>West</td>
<td>241</td>
<td>7/25</td>
<td>8/30</td>
</tr>
</tbody>
</table>
The experiment consisted of two phases. In the first phase the subjects were asked to construct street views of only the north part of the loop road (See the left map of Figure 7). The length of this part was about 5.5 kilometers. In the second phase they were asked to construct street views of the entire loop road, the length of which is about 32 kilometers. The first phase began on July 23, 2007. The first phase ended and the second phase began on July 27, 2007. The second phase ended on August 30, 2007. Since the experiment was conducted in summer, it was very hot and bright around the loop road. So, the subjects tended to take photos during the morning.

We prepared a mailing list for communicating with other subjects. The message log of the mailing list was a resource to tell us what kind of information sharing was needed to construct street views.

We interviewed the subjects on August 30 when the experiment finished. The purpose of this interview was to explore ways to improve QyoroView.

Results 2

This subsection describes the results of analyzing the three kinds of data described above.

System Log

In Figure 7 you can see how the photo taking of the loop road proceeded during the experiment. The figure shows the street view and the number of uploaded photos on July 27 and August 30. July 27 was the day when the message log (described later) was obtained. August 30 was the day when the experiment finished.

The experiment was successful. A total of 4,446 photos were taken by all of the subjects, which covered about three-fourths of the approximately 32-kilometer loop road. However, we learned that we should have been more aware of the distances between the subjects’ living areas and the targeted area, i.e., the loop road. It seemed much better to involve subjects whose living areas were spread over the targeted areas.
Message Log

The following exchange of messages occurred on July 27, which was the first day of the second phase. The names of streets, intersections, and stations have been replaced with alphabetical letters.

Message 1 sent from subject G at 9:37 a.m.

I am now at intersection P. I will walk down street Q toward the south. I’m going to take the photos of the east side of street Q. Let me know if you are taking the photos of that part.

Message 2 sent from subject F at 9:40 a.m.

I joined the experiment today. I’m going to begin taking photos near station R and end up taking photos around the city office. I’ll proceed toward the south more if I have time.

Message 3 sent from subject H at 11:09 a.m.

I’m going to take photos of the part of street S between station T and intersection U. Let’s everybody take many photos, although it’s hot again today.

Message 4 sent from subject G at 11:14 a.m.

I finished taking photos of the part of street Q’s east side between intersection P and intersection V. Traffic is very congested, so I’m giving up taking photos now. Good luck to everybody who’s beginning now!

Message 5 sent from subject F at 11:21 a.m.

I finished taking photos of the part of street Q’s east side between station R and the post office. I have to stop because I need to prepare for term-end exams. Good luck to everybody who’s beginning now!

Message 6 sent from subject H at 15:03 p.m.

I finished taking photos of the part of street S between intersection U and intersection W. That part includes a crossing with an overpass where I did not take photos.

As shown in messages 1, 2, and 3, the mailing list was used to notify the part of the road where the subjects were going to take photos. After the experiment we asked subjects F, G, and H about how they decided the area to take photos on July 27. In the interview we found that the notifications were able to successfully coordinate the subjects’ work. Subject F had planned to begin at intersection P, proceed to the south, and end around station X. But she decided to take photos at the southern area due to message 1, which she received on her mobile phone just after leaving her home. She moved to station R, which was south of station X, and began taking photos. Subject H found that subjects F and G were taking photos of the area around his home when he read messages 1 and 2. So, he decided to take photos at an area that was a little distant from his home.

As shown in messages 4, 5, and 6, the mailing list was also used to announce the part of the road where the subjects had finished taking photos. In the interview they all reported that they carefully chose the spot to end their photo taking. They thought it would be easy for others to take photos after their work if the area they finished was clearly recognizable, e.g., the area between the two large intersections.

There were many exchanges similar to those described above. The mailing list was used to share the plans and results of the subjects’ work. The information about the plans was useful for parallel work being done simultaneously. The information about the results was helpful for the work that occurred later.
Interview Results

We selected subject answers that could be categorized as navigation and situations, each of which represents an aspect that may lead to a more efficient production of street views.

1. Navigation

Q. “What should the experimenter have done to facilitate taking photos?”
A. “I think I would have taken more photos if the experimenter let us know which part of the street view of the road had not yet been constructed.”
A. “I think I would have been more eager to take photos if the experimenter taught us techniques to take them efficiently.”

These answers indicate that the experimenter should have guided the subjects more, i.e., by recommending areas to take photos, and by training the subjects in photo-taking.

2. Situations

Q. “What kind of area was easy to take photos in?”
A. “Areas with a lot of shade made by buildings.”
A. “Areas that were less congested with traffic.”
Q. “Which periods of time were easy to take photos in?”
A. “Periods when it was cool and there were few pedestrians and cars.”

These answers implied that photo-taking could have been facilitated by sharing the knowledge of the area and time when it was easy to take photos.

DISCUSSION

The message log suggests that collaborative support tailored to geographical content creation is a key to the successful creation of geographical UGC such as street views. Collaborative support functions are already common in some UGC systems [8]. Collaborative support peculiar to geographical content creation has, however, not yet been clarified. There are several previous studies on geographical UGC (Cheverst et al., 2001; Espinoza et al., 2001; Riva, & Toivonen, 2006), but few of them proposed functions to support collaborative content creation. Below, we discuss functions that should be provided by systems for geographical UGC.

The most basic and important collaborative support seems to be a function to avoid conflicts in the area where each participant creates media content. The system should inform participants of areas that have been filled with content, areas that are now under construction or that will enter construction soon, and areas that are still untouched. Furthermore, as described in the navigation category of the interview results, a function to actively allocate areas for participants to work in may be effective. As the system log showed, there should be a function to take the location of each participant’s living area into consideration.

As described in the situations category of the interview results, we think that sharing knowledge of the areas and times that are easy to work in would facilitate geographical content creation. We think conventional geographical UGC systems (Cheverst et al., 2001; Espinoza et al., 2001; Riva, & Toivonen, 2006) can be used for such knowledge sharing.

CONCLUSION

We have proposed and implemented a system that can produce large-scale street views as UGC.
system uses vector maps to concatenate photos sent from pedestrians’ GPS phones.

We also presented the results of two experiments. In the first experiment, we compared two methods. The bottleneck in our ‘one-shot capture’ method was the difficulty of users having to take a lot of pictures. We suggested a ‘continuous capture’ method and confirmed that a ‘continuous capture’ method could reduce the time needed to collect photos along streets. In this experiment, subjects seemed to cooperate with others. So we conducted the second experiment to determine how we should support interaction among users to facilitate the production of street views. We presented the results of this experiment that suggests collaborative support functions for geographical content creation. Sharing the plans, current status, and achievements of users who are moving around seemed to be effective for coordinating users.

REFERENCES


Degree Confluence Project: http://www.confluence.org/


Google Maps Street View: http://maps.google.com/help/maps/streetview/


Microsoft Photosynth: http://labs.live.com/Photosynth/


OpenStreetMap: http://openstreetmap.com/

Panoramio: http://www.panoramio.com/


Windows Live Technology Preview: http://preview.local.live.com/


**KEY TERMS**

*Street View:* Panoramic images of urban landscapes that are made from movies or a collection of photos taken along the streets of a city

*User-generated content:* Media content created by end users

*Vector map:* Digital data of maps that is consisted of X-Y coordinates. If you draw lines and polygons with these coordinates, you can create maps

*Border vector:* Digital data of Border line between a road and a City block

*Halfway vector:* Digital data of Center line of a road

*Photo map:* A map onto which are placed photos uploaded from end users

*One-shot capture method:* A method of uploading only a single photo

*Continuous capture method:* A method of uploading several photos at once