Increasing Interactivity in Street View Web Navigation Systems

Alexandre Devaux
Université Paris-Est, IGN, MATIS
73 avenue de Paris
94160 Saint-Mandé
alexandre.devaux@ign.fr

Nicolas Paparoditis
Université Paris-Est, IGN, MATIS
73 avenue de Paris
94160 Saint-Mandé
nicolas.paparoditis@ign.fr

ABSTRACT

This paper presents some interactive features we have added on our street-view web navigation application. Our system allows to navigate through a huge amount of data (panoramas and laser clouds) and also to interact with it. We will detail 4 aspects of this interactivity. First, the labelling, displaying of features directly into the images in the 3D space, useful for general public but also researchers in image processing and computer vision. Secondly we propose a crowd sourcing mode for blurring people not automatically detected. Thirdly we offer the possibility for the web user to localize and measure in 3D all objects visible in the images by plotting only in one image. Finally we developed a multimedia editor that allows public administrations (like town halls, museums, operas, theaters, etc.) to add interactive content like video or images at the exact 3D position/orientation/size they chose with an easy manipulating editor to augment with realism the static scenes with dynamic or fresher elements.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: [Artificial, augmented, and virtual realities]; I.6 [**Simulation and Modeling**]: [Applications]

General Terms

Design, Experimentation

Keywords

3D measure, Annotation, City modeling, Immersive Environments, Street-View, panoramas, web-application

1. INTRODUCTION

Multimedia web-based applications using street-level high resolution images acquired by mobile mapping have been developing fast in the last years. While the quality of images is increasing, the multimedia aspects of such an approach are still to develop. All actors like Google Street-View, Microsoft Live Street-Side offer the possibility to walk through cities in panoramas with click-and-go functions but with a limited degree of interactivity. Nevertheless, users all agree that these systems are very useful to visualise the city from a pedestrian level with an excellent image quality.

In this paper, we will address the topic of generation of 3D textured city models and the interactive possibilities. [5],[2] have developed very interesting algorithms offering appealing 3D textured models who could be used over the internet for city visualization and 3D manipulations. And adding lasers to cameras, already a great amount of research has been carried out for photorealistic modeling of urban environments [1]. Nevertheless the focus of these works is on the 3D modeling of the buildings and environment, rather than the user interaction and web navigation. We have chosen a representation which is well adapted for fast immersive navigation in streaming but which above all allows intuitive 3D measurement tools that of course are the key issues for 3D updating by crowd sourcing means of these city models. We use multi-planar image panoramas that do not distort the representation of different objects within the scene (straight lines in the images remain straight in the panorama). This eases of course the manual annotation and the 3D measurement of objects. We will not detail our mobile acquisition system which provides us an excellent georeferencing quality of all data (images and laser point clouds) acquired by the mobile mapping system. This paper will concentrate on all interactive aspects around the multimedia applications of such systems especially because the State-of-the-Art on

Figure 1: A screenshot of the web application while measuring the height of a porch with the laser cloud visualization activated.

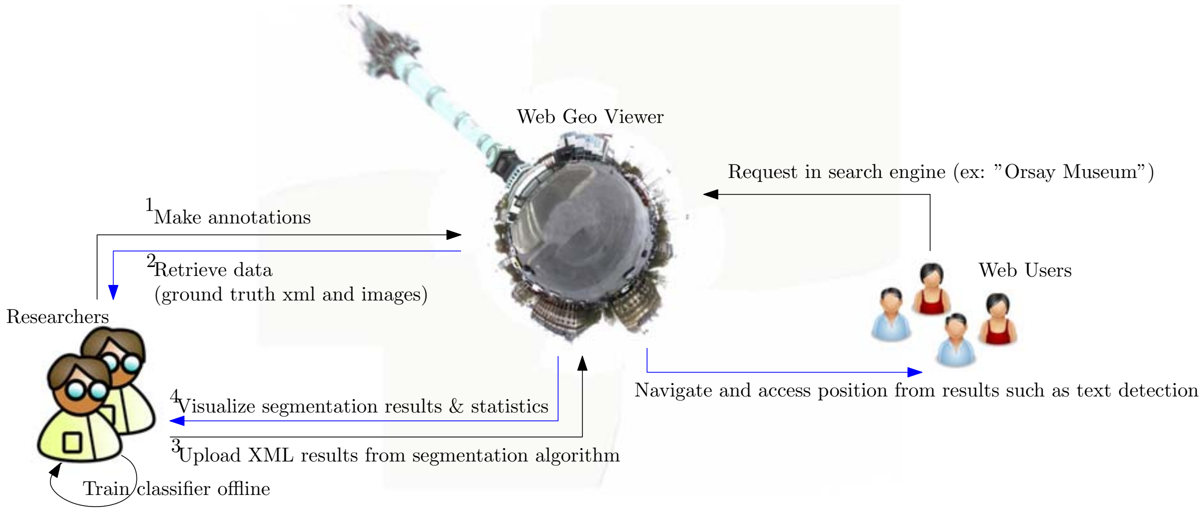

Figure 2: Schema of the features annotation functions and interrogation of the system using automatic extraction results.

interactivity in web-based immersive geoviewers is rather light.

In the following, we first present the scientific collaboration aspects we offer in the application with the labelling of features, then we show the crowd sourcing possibilities for missed face blurring. In a third part we present the 3D measuring function with the laser clouds visualization in the internet navigator and finally we present a multimedia integration editor to augment the interactive contents of these city models in order to provide updated and useful local information (e.g. cultural, functional, etc.) on the different services or properties of the different city units.

2. A TOOL FOR BENCHMARKING

One of our key features is to provide researchers with huge datasets of high quality urban images which they can label and use to train and measure their automatic segmentation and annotation algorithms efficiency. Indeed, augmenting the interactivity with the content of the images is a key issue to interact with cities and find a spatial location from partial cues.

2.1 Basic principles

Very few web applications for image labeling for scientific purpose exist. A remarkable one is labelMe [3] which allows annotation on thousands of images in an internet navigator. In our application, in the same spirit, you can navigate through the city, annotate whatever you want but you can also view all the annotations results, manual or automatic, provided by the users of the system and researchers that have uploaded results of different segmentation algorithms like face detection (decisive for privacy aspects on web diffusion), text detection, road mark and road sign detection, etc.

All the annotations are stored in a database on the server with xml as exchange format for uploading and downloading results. We defined a structure for every feature studied in the urban scene. There is no limitation on the number of classes and in a near future it will be possible for users to create new object classes. For the time we describe the world with a very simple tree-like ontology: we have classes (vehicle, text, road mark, pedestrian, etc) in which we can find several sub-types or specialisations. For example, in the text class there are three types: line, block and word. Then every object has an attribute value which can be a list of details like color, material... Table 1 shows the structure of the object as it is stored in the database. It is possible to import or export these objects in xml format direclty over the street-view web-application.

Table 1: Visual object description

filename	class	type
attribute	confidence	verified
polygon	date	user

2.2 A scenario

One typical scenario would be the following: A group of scientists is working on face detection. They need a ground truth to train their algorithm (for example to train an adaboost classifier), they can create the ground truth with the tool online (labeling feature in just one click with bounding boxes, circles) and then download the xml files with the photos from http or ftp. They can then train their classifier with those files and after that test and benchmark their algorithm with another ground truth and visualize the performance statistics (distance between the truth and their results) online. Every scientist has the right to edit his own data online but has a read-only access to the data of other scientists. It is possible to display different classes, like a layer system with a special menu. Fig 2 presents the global schema of the labeling/detection functions and Fig 3 shows a screenshot of displaying results with the layer menu in the internet navigator. We are encouraging the use of our system in an infrastructure for feature extraction challenges and research.

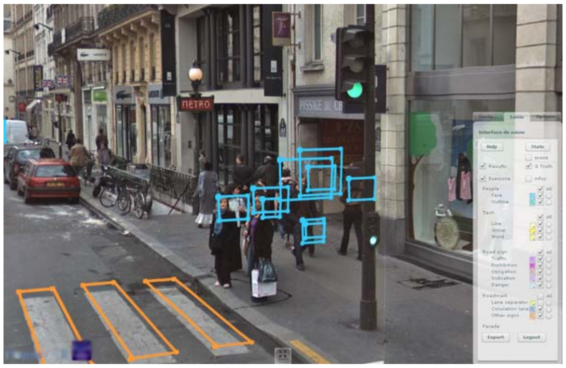

Figure 3: Displaying automatic detection results in the web-viewer.

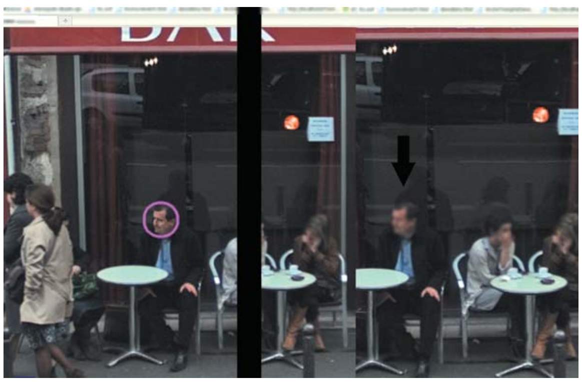

Figure 4: Example of a miss-detected face reported by two web-users, blurred after automatic verification process.

3. CROWD SOURCING

Using the same system as described in the last part we offer the possibility for users to improve the system on some aspects. The automatic face detection system has a detection rate of 86%, more than one person in ten will not be blurred. Web users can blur faces that are miss-detected directly in the images using the same drawing system as for labelling ground truth. The main problem then is a moderation process. As we want to make the whole application the most autonomous possible, we created a verification process that has an automatic mode and a manual one. Using the same idea as [4] we analyse the redundancy of information to verify the authenticity of a user's action. If three different users indicate the same face for blurring it means very probably that the information is right so it will be blurred without any administrator's action. But as there are many images, missed faces can be hard to see and might be seen by nobody or only one web user. Hence, if it remains too much time in the task list, the administrator will receive an alarm and see the picture to decide if he has to blur it. We could imagine to launch automatically the face detection algorithm just on the part the web-user asked for blur, it is more efficient to scan a very small zone well delimited. Fig 4 shows an example of a face that the algorithm did not detect but different web users reported so the face get blurred without the need of an operator's action.

4. MEASURING POSSIBILITIES

The acquisition vehicle while taking panoramic pictures geo-localized also acquires a laser cloud very precise around its position. As the system is fully calibrated it is possible to associate for each laser point a pixel in the panoramic pictures. We have for every position of a panoramic acquisition a laser cloud, in fact the laser acquisition is continuous, it never stops scanning. The acquisition vehicle is equipped with different lasers, the one we use for 3D measuring in the viewer is a 10 000 measurements/sec Riegl with a very good precision oriented perpendicular to the façade resulting in a cloud of 3D laser points every 10 cm on the façade. The images and laser points are geo-referenced in a global reference frame with the help of an Inertial Navigation Systems (integrating 2 GPS, an Inertial Measurement Unit and an odometer) providing overall a submetric absolute localization. The relative measurement allows to measure distances, surfaces and volumes with a centimetric precision.

The interesting aspect of using lasers is that we do not need to do technique such as Structure From Motion, we have instantly a very precise 3D information even if the resolution is not as high as the panoramic image resolution. Remains the difficulty of accessing the laser data online so we created a structure where you can access locally to small part of the laser acquisition already geo-referenced. We divided thousands of square meters of laser data into blocks of one meter square and used a geo hierarchy on the web server, so the user actually just loads the laser data around its position quite fastly because everything is already preprocessed. We are currently working on another architecture subdividing the global laser cloud into façade packet which is more efficient when measuring functions are enabled.

For visualization aspects it is possible to show the laser cloud in the viewer, superposed to the image or alone. While activating only the laser cloud you can evolve fluidly in the city through 3D points which are streamed dynamically around your position. Passing the mouse over an image and you get instantly the 3D position on the earth of what the mouse is on. But dislike the laser cloud acquisition is continuous, the panoramic pictures are taken every 3 meters so moving objects or hidden objects won't have a good projection between laser data and images. An optimal measure would be on the 3D points near the camera axis. Different tools allow the user to make measurement like distance, surface (Fig 5) and also volume. A great use of Papervision [6] and special tuned library from it were used to offer the possibility for any internet user (having flash) to visualize all that 3D without downloading a heavy plug-in. Actually this was also an important choice to make for us, many different way of displaying 3D exists on internet but it is not already generalized. We took the choice of a light plug-in present on 99% of computers but limited in 3D graphics as hardware rendering is not really supported. We could imagine another version using a heavy plug-in allowing to use GPU to accelerate a lot the rendering if we need to integrate large 3D models but for now this was the best compromise, it forced us to optimize rendering algorithms.

5. MULTIMEDIA EDITOR

What would be an application where you could walk virtually through cities if the pictures are old and do not take into account recent changes?

Figure 5: One click surface measurement over a large advertising panel shows the square meters of the surface selected.

5.1 Concept

A key-aspect of city modelling is that it has to be recent, up-to-date. But no actual acquisition technique allows to make many geo-localized pictures simply at a very low-cost. In our case the vehicle will pass approximately every year through one city. In a year there are not so many changes but still, building get transformed, decoration changes... Systems like Google Street-View, Microsoft StreetSide have the possibility to add users picture from panoramio or flickr. They are then displayed in the good position and geometry with PhotoSynth for Microsoft and a bit more simply for Google. It is a very interesting feature they offer but this system does not allow regular interactive panoramic updating. We created a function in the application where identified people could add the information they like. A simple example: the administrator of the Operas of the city can put the current programme of the season, or a video presentation of the season with a redirection to their website when clicked. He can do that easily from his computer and manipulate the data in every aspect, size, scale, position, 3D orientation. It becomes possible to replace an actual real picture with a new one without noticing it is not original (Fig 6).

5.2 Technique

As for the annotation part we get the 3D point of the panorama plane touched by the mouse by projecting the mouse from the camera to the 3D object in the scene. That means we get the point in the 3D plane where the new image should be attached. To ensure there are no bad intersections with the others 3D planes (there are 10 planes composing the panorama, one for each camera) we put the new image closer to the camera. Then the user can manipulate it completely free and save after editing to the interactive data base. The same approach could be used for any users to indicate a change using their own picture. A typical example would be a commerce who changed his owner and façade and would like that to be seen in the web viewer. The only problem here is the verification process which cannot be done automatically. As we talked in the second part (Crowd Sourcing for face blurring) about the moderation process, allowing people to add data online on a public service points out a lot of complex verification processes. The search for automatic moderation is in that case quite difficult as probability law

Figure 6: A screenshot of the multimedia editor, modeling the new Opera season program.

on multiple sources for the same action is not represented here.

6. CONCLUSIONS

We presented some aspects where interactivity could be improved in web immersive geo-viewers by integrating them in our navigation system. It is an area moving very fast these years helped by the high speed connection everyone has at home and the nearly unlimited storage capacity of big companies. Imagery acquisition systems are also getting lighter, more accurate and cheaper. We showed that with a precise system (image geometry and orientation/position) it is possible to add high-level functions like detection algorithm displaying, labelling, crowd-sourcing, 3D measures and multimedia editing. Remains infinity of services to add and challenges. We will continue to work on those aspects to discover new effective visualizations.

Acknowlegments

This work is partly funded by the ANR (French National Research Agency), iTowns-MDCO project. www.itowns.fr

7. REFERENCES

[1] C. Frueh, S. Jain, and A. Zakhor. Data processing algorithms for generating textured 3d building facade meshes from laser scans and camera images. *IJCV*, 61(2), 2005.

[2] B. Micusik and J. Kosecka. Piecewise planar city 3d modeling from street view panoramic sequences.

[3] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. Labelme: a database and web-based tool for image annotation. *IJCV*, 77:157–173, May 2008.

[4] L. von Ahn, B. Maurer, C. Mcmillen, D. Abraham, and M. Blum. recaptcha: Human-based character recognition via web security measures. *Science*, 321(5895):1465–1468, August 2008.

[5] J. Xiao, T. Fang, P. Zhao, M. Lhuillier, and L. Quan. Image-based street-side city modeling. *ACM Trans. Graph.*, 28:114:1–114:12, December 2009.

[6] A. Zupko, C. Ulloa, J. Grden, and T. Knip. Papervision: An open source realtime 3d engine for flash.