

# AR City Representation System Based on Map Recognition Using Topological Information

Hideaki Uchiyama<sup>1</sup>, Hideo Saito<sup>2</sup>, Myriam Servières<sup>3</sup>, and Guillaume Moreau<sup>4</sup>

<sup>1,2</sup> Keio University, 3-14-1 Hiyoshi, Kohoku-ku 223-8522, Japan  
{uchiyama, saito}@hvrl.ics.keio.ac.jp

<sup>3,4</sup> Ecole Centrale de Nantes-CERMA, 1, Rue Noë 44300 Nantes, France  
{myriam.servieres, guillaume.moreau}@ec-nantes.fr

**Abstract.** This paper presents a system for overlaying 3D GIS data information such as 3D buildings onto a 2D physical urban map. We propose a map recognition framework by analysis of distribution of local intersections in order to recognize the area of the physical map from a whole map. The retrieval of the geographical area described by the physical map is based on a hashing scheme, which is called LLAH. In the results, we will show some applications augmenting additional information on the map.

**Keywords:** GIS, Augmented Reality, LLAH.

## 1 Introduction

Geographical Information Systems (GIS) have become essential tools for studying, handling and planning urban development. GIS can superimpose layers (representing homogeneous information) that are fused together to generate maps. GIS data can be updated any time and are thus more up-to-date than traditional paper maps. They can moreover be adapted in real time to meet the user's need.

One of the research issues in GIS community is GeoVisualization, which is a way of designing an interface and displaying and handling the spatial and temporal GIS data on the interface [1, 2]. The advantages of using Augmented Reality (AR) techniques to display digital information on standard paper maps have been shown, because AR enables 3D data to be manipulated easier [3, 4, 5]. Moreover, GIS need a shift towards 3D to be compatible with sustainable development concerns. To manage increasing complexity of sustainable development requirements, spatial and temporal queries have to be handled to compute new indicators that are now being defined. For instance, a thermal comfort indicator could be 'walls that have more than 8 hours sunlight in winter and less than 2 hours in summer'. Visualizing the results of such a query requires 3D representation way because sunlight exposure is dependent on building height and neighboring buildings. 3D virtual environments are not easy to manipulate for local authorities. That is why we assume that the use of AR maps will facilitate the display of such results by letting the user manipulate both a paper map and the viewpoint in a natural way.

In this paper, we propose a framework of map recognition technique to establish a correspondence between the image of a real map captured with a camera and a GIS.

Intersections are extracted from the input image, and then matched with the GIS data, as in the problem of “Document image retrieval” [6]. We are then able to compute the camera position and orientation with respect to the map assuming that a map is flat, and display more information from the GIS. In experimental results, we will show that our framework is compatible with AR system and some applications.

The rest of the paper is organized as follows: we will first briefly present related works that can be used to match images representing the same objects, i.e. compute the geometric transformation that links the two images. We will then provide an overview of our system in Section.4 and the algorithm in Section.5. Finally, experimental results will be presented and discussed.

## 2 Related Works

The problem of finding a match for a query object using feature points has been addressed in various ways. The feature points can be described using rich descriptors such as SIFT [7] or SURF [8], that typically use image patches. These descriptors are robust in terms of change of illumination, scale and rotation and describe them with high-dimensions vectors. The search methods have then to deal with the problems of nearest neighbor search in high dimensions with approximate nearest neighbor searching [9] or locality-sensitive-hashing [10].

Rich descriptors are well suited to the retrieval of images near-identical to the ones in the database, with few repetitive texture patterns. By contrast, 2D maps can be presented in different ways, according to the manufacturer, and the retrieval method needs then to focus on the geometry of the urban environment they describe. For this reason, the feature points need to be specific to urban environments and the location of intersections are used in this paper.

It is not possible to distinguish an intersection query using only the location of a single intersection. For this reason, the essential information in retrieval is the arrangement of the features points. Such an arrangement, in our case, must be invariant to the orientation of the camera relative to the map.

One of the recognition methods by geometrical information is Geometric hashing (GH) [11]. GH is such a general model-based object recognition method widely used in computer vision as well as in other domains. The introduction of a geometric invariant yields a computational cost quite important, that is unsuitable for an augmented reality application. A probabilistic reduction of the number of feature points results in accuracy degradation and has led to the introduction of “Locally Likely Arrangement Hashing” (LLAH), which outperforms GH in both processing time and required amount of memory [6]. In this scheme, neighboring points are considered for the calculation of an affine invariant used as a key in a hashing table. A voting technique is employed for retrieval, insuring efficiency and robustness against erasure of feature points. We use a combination of this method and a more traditional tracking technique to first recognize the area in the camera field of view, then overlay 3D buildings in real-time.

### 3 System

The user has a hand-held device equipped with a camera coupled with a computer, for example a cellular phone or a see-through HMD. In our experimental setup, we use a digital camera and a laptop (Fig.1 (a)).

The physical map can be displayed on a desktop, on a wall or any flat surface. At the beginning of the use, the camera needs to be in a position more or less parallel to the map, so that perspective distortion is not too important (Fig.1 (b)). After that, the user can move more freely (Fig.1 (c)) to watch the 3D buildings and the GIS data in real-time on the screen of the device (Fig.1 (d)). All physical maps should be registered in the database beforehand. The user can select a map from the registered map for watching its visual aids.

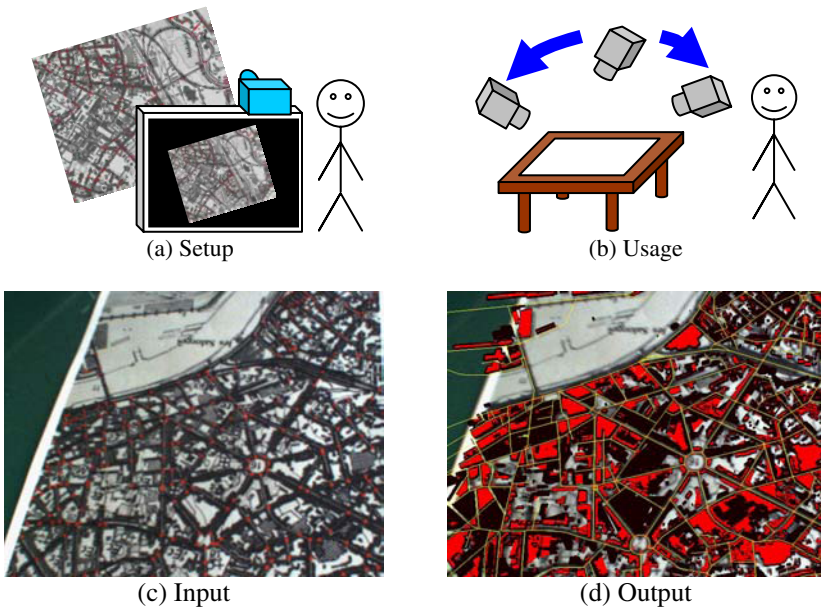


Fig. 1. System Overview

### 4 Algorithm

#### 4.1 Overview

In the off-line process, the initial database of LLAH features of all intersections in GIS is generated beforehand. In the on-line process, the same process is executed at every frame (Fig.2). From a captured image, intersections are extracted by using simple color segmentation because their color was determined beforehand. Since another automatic intersection extraction method has been proposed [12], we focus on map image retrieval by distribution of intersections. For each intersection, the

corresponding intersection is retrieved from the database by using the LLAH features. Based on the number of the retrieved intersections, the area of the map can be determined. In addition, the camera pose can be computed by using the retrieved intersections for displaying 3D GIS data of the area. At the same time, LLAH features of each intersection in the captured image are updated.

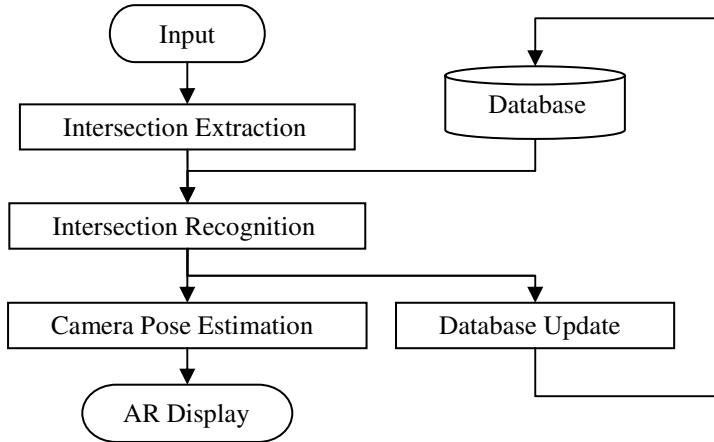


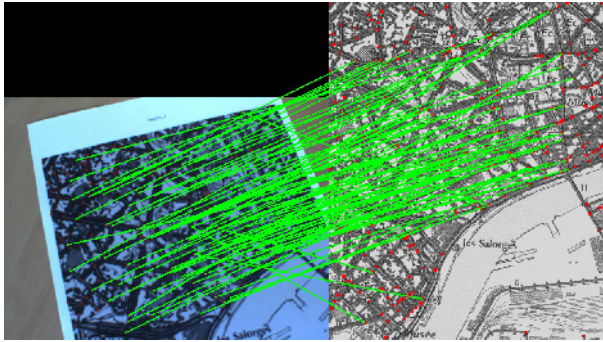
Fig. 2. Algorithm

## 4.2 GIS Data

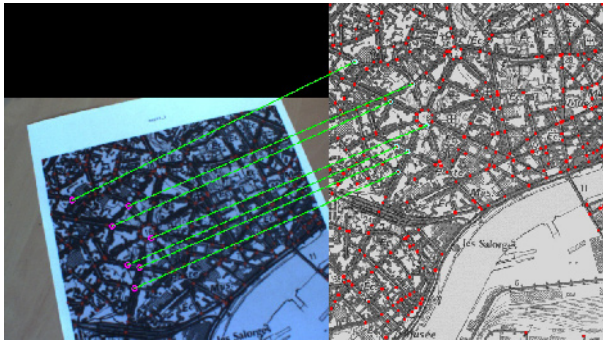
Real GIS data of a large French city is used. GDMS [13] is used to process the data in two ways:

- with a simple query, all intersections are extracted from the road network to build the features points that are used in the method.
- following Neubauer and Zipf's idea [14], we have built an XML style file that describes how the GIS database will be rendered in the virtual environment, i.e. whether a polygon layer should be rendered with flat surfaces or extruded polygons, and additional information such as the color to use. We have thus built a VRML builder above GDMS that transforms GIS data according to the XML file and generates a VRML file.

The area described by the data can in theory be very large, and must be sub-divided in sub-areas that correspond to the size of the physical maps used as queries. These sub-areas are defined by a specific ID with intersection's IDs such as (Area ID, Intersection ID1, Intersection ID2 ...). Each Intersection is stored with its belonging area such as (Intersection ID, Area ID). Additional information of the map such as 3D models of buildings is also tagged with its belonging area ID to be able to retrieve the information from area ID.



**Fig. 3.** Matching by LLAH



**Fig. 4.** Outlier Removal by RANSAC based Homography Computation

### 4.3 Intersection Recognition

From a captured image, red intersections are extracted by finding red region. For each intersection's region, the center is computed. Based on LLAH [6], the corresponding intersection of each extracted intersection is retrieved from the database. As a result, some intersections are correctly matched and other intersections are wrongly matched (Fig.3). Since there are similar arrangements of intersections, the result of LLAH sometimes includes wrongly matched intersections. For removing these wrongly matched intersections, we use RANSAC based homography computation [15].

Since the map is 2D, the correspondence between the map in the database and the map in a captured image can be described by homography. For computing a homography, several intersections are randomly selected and evaluated in the RANSAC process. After that, high confidential intersections are selected (Fig.4).

After the homography is computed, the homography can be converted into a camera position and orientation [16], which is equivalent to camera pose estimation.

### 4.4 Database Update

Since the initial database is generated by using intersections in GIS, we can say that the LLAH features in the initial database are generated by using a top view image. If

we use only the initial database, the retrieval of intersections will succeed in case of near top view including many points. By adding new LLAH features according to the changes of the user's viewpoint, the retrieval will still work when the captured image is not close to the top view image.

When the homography is computed in Intersection Recognition, the intersections in the database can be reprojected onto the captured image by using the homography. If a distance between the reprojected intersection and an extracted intersection in the image is within a threshold, the extracted intersection is matched with the reprojected intersection. Thanks to the reprojection, many intersections which are not matched with corresponding intersections in GIS by LLAH can be matched.

## 5 Experimental Results

### 5.1 Computational Costs

For evaluating computational costs of AR display, 100 frames are captured in order to compute the average computational costs. Our device is composed of a laptop (Intel Core 2 Duo 2.2GHz and 3GB RAM) with a firewire camera.

The computational costs of Intersection Recognition and Database Update depend on the number of extracted intersections, which can be represented by  $O(N)$  in the case that the number of extracted intersections is  $N$ . In our algorithm, total computational costs are 46 msec (more than 20fps). However, 3D Model Rendering took most computational costs because GIS data includes detailed polygons. The content should be appropriately selected depending on computational costs.

**Table 1.** Computational Costs

| Process                  | Time (msec) |
|--------------------------|-------------|
| Intersection Extraction  | 11          |
| Intersection Recognition | 22          |
| Camera Pose Estimation   | 3           |
| Database Update          | 10          |
| 3D Model Rendering       | 102         |

### 5.2 Application

Since the camera pose against the map is estimated, any virtual object can be overlaid at an appropriate position. In this section, we will introduce one application for AR geovisualization.

Fig.5 shows a system for displaying a picture at the captured place on the map. If a user takes a picture with its information of the captured place and input it into the application, the application displays the picture at the captured place on the map (Fig.5 (b)). The user can recognize the places and their relationships where the user captured pictured.



**Fig. 5.** Display of tagged images on a map

## 6 Conclusions

In this paper, we have presented an AR representation system for 3D GIS that are based on the augmentation of a physical map including intersections. It provides a natural device for 3D GIS information representation and manipulation. Intersection recognition is based on LLAH framework by using geometrical relationship with neighbor intersections. For free camera moving, update of LLAH features is adapted.

Our future work will be centered on two main topics. First, we will be using a real physical map, easier to manipulate, but requiring more image processing to recover the features needed in the initialization phase. Second, a map contains more information than just intersections, and this could be used to extract other features such as connectivity.

## Acknowledgement

This work is supported in part by a Grant-in-Aid for the Global Center of Excellence for high-Level Global Cooperation for Leading-Edge Platform on Access Spaces from the Ministry of Education, Culture, Sport, Science, and Technology in Japan.

## References

1. Wilson, D.C., Lipford, H.R., Carroll, E., Karr, P., Najjar, N.: Charting new ground: modeling user behavior in interactive geovisualization. In: Proc. the 16th ACM GIS (2008)
2. Wood, J., Dykes, J., Slingsby, A., Clarke, K.: Interactive visual exploration of a large spatio-temporal dataset: Reflections on a geovisualization mashup. *IEEE Trans. VCG* 13, 1176–1183 (2007)
3. Romao, T., Dias, E., Danado, J., Correia, N., Trabuco, A., Santos, C., Santos, R., Nobre, E., Camara, A., Romero, L.: Augmenting reality with geo-referenced information for environmental management. In: Proc. the 10th ACM GIS (2002)
4. Hedley, N.R., Billingham, M., Postner, L., May, R., Kato, H.: Explorations in the use of augmented reality for geographic visualization. *Teleoperators and Virtual Environments* 11, 119–133 (2002)

5. Reitmayr, G., Eade, E., Drummond, T.: Localisation and interaction for augmented maps. In: Proc. ISMAR, pp. 120–129 (2005)
6. Iwamura, M., Nakai, T., Kise, K.: Improvement of retrieval speed and required amount of memory for geometric hashing by combining local invariants. In: Proc. BMVC, pp. 1010–1019 (2007)
7. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *IJCV* 60, 91–110 (2004)
8. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
9. Arya, S., Mount, D.M., Netanyahu, N.S., Silverman, R., Wu, A.: An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *J. of the ACM* 45, 891–923 (1998)
10. Datar, M., Indyk, P., Immorlica, N., Mirrokni, V.S.: Locality-sensitive hashing scheme based on p-stable distributions. In: Proc. SCG, pp. 253–262 (2004)
11. Lamdan, Y., Wolfson, H.: Geometric hashing: A general and efficient model-based recognition scheme. In: Proc. ICCV, pp. 238–249 (1988)
12. Chiang, Y.Y., Knoblock, C.A.: Automatic extraction of road intersection position, connectivity, and orientations from raster maps. In: Proc. ACM GIS (2008)
13. Bocher, E., Leduc, T., Moreau, G., Cortés, F.G.: Gdms: An abstraction layer to enhance spatial data infrastructures usability. In: *Agile 2008* (2008)
14. Neubauer, S., Zipf, A.: Suggestions for extending the OGC styled layer descriptor (SLD) specification into 3D – towards visualization rules for 3D city models. In: Proc. UDMS, Stuttgart, Germany (2007)
15. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *C. of the ACM* 24, 381–395 (1981)
16. Uematsu, Y., Saito, H.: Vision based registration for augmented reality using multi-planes in arbitrary position and pose by moving uncalibrated camera. In: Proc. MIRAGE, pp. 99–1019 (2005)