AR GIS on a Physical Map based on Map Image Retrieval using LLAH Tracking

Hideaki Uchiyama, Hideo Saito
Keio University
3-14-1 Hiyoshi, Kohoku-ku 223-8522, Japan
{uchiyama,saito}@hvrl.ics.keio.ac.jp

Myriam Servières, Guillaume Moreau
Ecole Centrale de Nantes - CERMA IRSTV
Rue de la Noë, BP 92101, 44321 Nantes Cedex 3, France
{myriam.servieres,guillaume.moreau}@ec-nantes.fr

Abstract

This paper presents a method for retrieving a corresponding map of a captured map image from a map database. Our method is inspired from LLAH based Document Image Retrieval (DIR). LLAH is a method for recognizing a point by using a LLAH feature composed of its neighbor points. Since Map Image Retrieval (MIR) is achieved by analyzing distribution of intersections, the LLAH feature is used in order to describe the distribution. In our method, registration and retrieval in LLAH based DIR are improved for reducing the computational costs of the retrieval. In addition, the LLAH features are updated while a camera is moving. Our improvements enable MIR to the case of strong camera tilting, occlusion and fewer intersections.

1 Introduction

Geographical Information Systems (GIS) have become one of essential tools for handling urban development. GIS include spatial data such as buildings with its temporal changes. Compared with a paper map, digital GIS data can be updated anytime and be used for several applications.

One of the research issues about GIS is the way of displaying the spatial and temporal GIS data. Previous works [6, 12, 13] have shown the advantages of Augmented Reality (AR) techniques to display GIS data on a physical map. For displaying the data at a precise position, these systems have to compute the geometrical relationship between a camera and a map. In these works, they have computed the relationship by using ARToolKit [6], local feature matching [12] and analysis of distribution of intersections [13]. Use of the distribution of intersections will be suitable for enhancing general use of the system because the distribution should be the same in maps from different manufactures. However, there were some limitations for the camera motion in [13] because they applied only LLAH based Document Image Retrieval (DIR) [9], which is a method for retrieving the corresponding page of a captured document image from a document database.

In this paper, we propose AR GIS based on Map Image Retrieval (MIR) inspired from LLAH based DIR. Our method allows retrieval with fewer points with strong camera tilting and faster processing against LLAH based DIR. A captured map is retrieved from a map database by analyzing a positional relationship of intersections described by a LLAH feature. Since the positional relationship changes according to the changes of a user's viewpoint, a new positional relationship is registered to enable free camera moving, which is called LLAH Tracking. For evaluating our system, the minimum number of intersections for MIR will be discussed in the experimental results. From the aspect of computational costs, we will show that our algorithm will be compatible with AR systems. In addition, we will show our system can work in the case of strong camera tilting and occlusions.

The rest of the paper is organized as follows: we will present related works of object recognition methods in Section 2. From the related works, the details of LLAH based DIR which inspired our method will be explained in Section 3. We will then provide an overview of our system in Section 4 and its algorithm including our improvement from LLAH based DIR in Section 5. Finally, experimental results for evaluating our algorithm will be presented in Section 6.

2 Related Works

The research issue for finding a match with a query object using natural feature points has been addressed in various ways. The feature points can be described as a high-dimensional vector such as SIFT [8], SURF [2], SIFT and Fern [15]. The features are robust in terms of change of illumination, scale and rotation. The search method can be addressed as a nearest neighbor searching problem by the following approaches: ANN [1], locality-sensitive-hashing [4] and a vocabulary tree [10]. Rich descriptors are well suited to the matching of feature points with few repetitive texture patterns. On the other hand, 2D maps can be presented in different ways without using descriptors. For example, there are roads, their connectivity and intersections, which should be same according to the manufacturers. Use of such topological features enhances general use of the system because it enables use of maps from different manufactures.

Regarding the object recognition by topological features, geometric hashing (GH) is a 3D object recognition method by using corners and edges [7]. Since a geometrical invariant in GH needs huge computational

costs and memory size, GH is not suitable for AR applications.

Locally likely arrangement hashing (LLAH) focuses on 2D object recognition by using positional relationships of words' centers, which was applied to DIR [9]. Compared with GH, computational costs and use of memory are improved. Since an intersection can be represented as a point, we apply LLAH to MIR. The details of LLAH will be presented in Section 3.

3 LLAH

LLAH is a method for recognizing a point by positional relationships of its neighbor points [9]. When LLAH features of x are computed in Figure 1, its n neighbor points are firstly selected as $abcdef$ ($n = 6$). For computing a LLAH feature, m points are selected from n points as $abcde$ ($m = 5$). Since the number of combination of the selections is ${}_nC_m = \frac{n!}{m!(n-m)!}$, ${}_nC_m$ features are registered for a point. From m points, 4 points are selected for computing a ratio of two triangles, which is an affine invariant. Since the number of ratios is ${}_mC_4$, the dimension of a LLAH feature is ${}_mC_4$.

For quick retrieval, a LLAH feature is stored as an index computed by

$$Index = \left(\sum_{i=0}^{{}_mC_4-1} r_{(i)} k^i \right) \bmod H_{size} \qquad (1)$$

where $r_{(i)} (i = 0, 1, ..., {}_mC_4 - 1)$ is ratios, k is quantization level and H_{size} is hash size. By numbering each point, a LLAH feature is stored as (Index, Point ID). If a collision happens at an index, the index is ignored because discrimination of the index is low.

In LLAH based DIR, a word's center is dealt as a point. The process of LLAH based DIR can be divided into off-line registration and on-line retrieval. In the registration, LLAH features of words in all documents captured from a top view are registered in a database. As described above, a word can have ${}_nC_m$ indexes. In the retrieval, a corresponding word of a word in a captured document image is retrieved from the database. The method takes into account the case where the viewpoint is rotated. In Figure 1, neighbor points are selected as $abcdef$, which order is stored in the database. If the viewpoint is rotated, the order of neighbor points may be $bcdefa$ in a captured image. For adapting the case, the method computes LLAH features of every clockwise order such as $abcdef$, $bcdefa$ and so on. Since the number of the orders is n and each order has ${}_nC_m$ indexes, $n \cdot {}_nC_m$ indexes are retrieved from the database. As a result, the corresponding word is determined by voting.

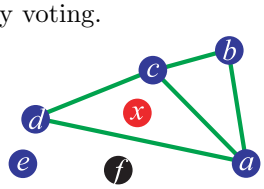


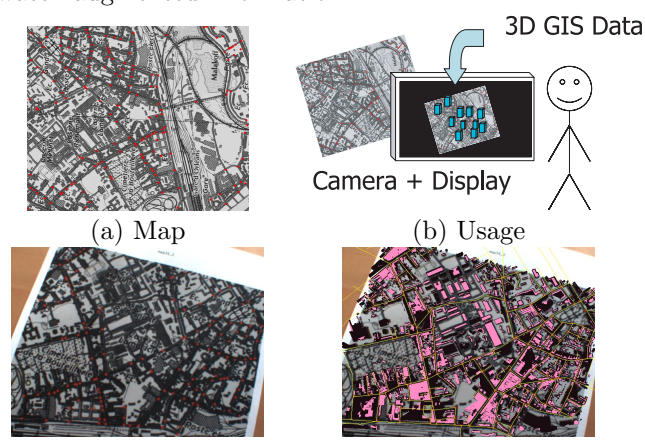
Figure 1: LLAH

4 Proposed System

In our system, we use a physical map with colored intersections for extracting intersections by simple color segmentation (Figure 2(a)). Since there are some researches for extracting intersections from a raster map [3], we would like to focus on MIR using LLAH Tracking.

The user has a hand-held device equipped with a camera and a computer (Figure 2(b)). The map should be put on any flat surface such as a table or a wall.

When the user captures a map (Figure 2(c)), its corresponding map is retrieved from the database. At the same time, a geometrical relationship between the camera and the map is computed automatically. As a result, 3D GIS data of the map is rendered (Figure 2(d)). The user can move the camera freely to watch augmented information.



(a) Map (b) Usage

(c) Input (d) Output

Figure 2: System Overview

5 Algorithm

5.1 Overview

The initial database includes LLAH features of all intersections in GIS. In the on-line process, the same process is executed at every frame (Figure 3). Intersections are extracted by using simple color segmentation from a captured image because their color was determined beforehand. For each intersection, the corresponding intersection is retrieved from the database by using the LLAH features. From 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.

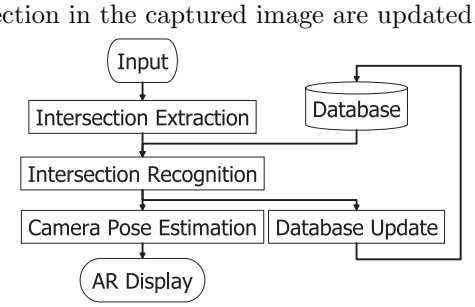


Figure 3: On-line Process

5.2 Initial Database

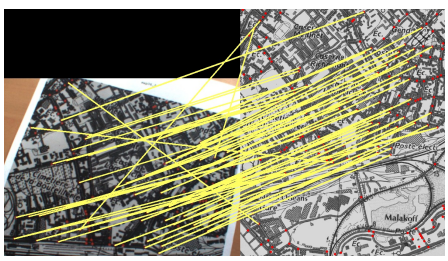
In the registration of LLAH based DIR, the LLAH features for an order of neighbor points such as $abcdef$ are registered as described in Section 3. For reducing computational costs in our retrieval, we register the

LLAH features of every clockwise order such as $abcdef$, $bcdefa$ and so on in our registration.

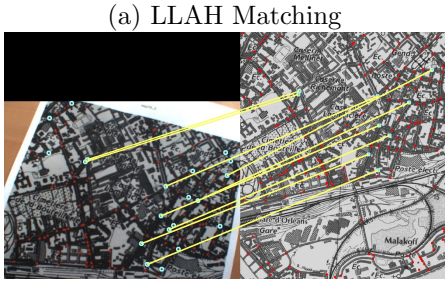
Since the number of the orders is n, a point is registered into $n \cdot {}_nC_m$ indexes. This means that memory size for the database is n times more than that of LLAH based DIR. By contrast, the computational costs of the LLAH features in our retrieval are $1/n$.

5.3 Intersection Recognition

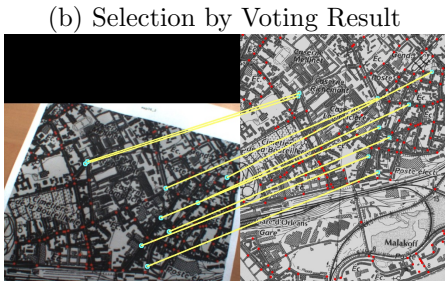
For each intersections extracted from a captured image, the LLAH features are computed in the same way as Section 3. Since ${}_nC_m$ indexes are retrieved from the database, the corresponding intersection is determined by voting. As a result, some intersections are correctly matched and other intersections are wrongly matched (Figure 4(a)). From all the matched intersection, top N intersections are selected by assuming that higher voting number is reliable (we set $N = 20$) (Figure 4(b)). Since there are still wrongly matched intersections, which are matched with an intersection in another map, we use RANSAC based homography computation [5]. For computing a camera pose, we compute a homography matrix between a captured image and a map in the database [14]. By inserting RANSAC process into the computation of the homography matrix, the wrongly matched are removed (Figure 4(c)).



(a) LLAH Matching



(b) Selection by Voting Result



(c) Selection by RANSAC

Figure 4: Intersection Recognition

5.4 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 Section 5.3, the intersections in the database can be reprojected onto the captured image (Figure 5). If a distance between the reprojected intersection and an extracted intersection in the image is within a threshold (we set 2 pixels), the extracted intersection is matched with the reprojected intersection. Compared with Figure 4(c), the number of matched intersections increases. For each intersection, the LLAH features are computed and registered in the database in the same way as Section 5.2. If a collision happens, the data is overwritten. Newly registered LLAH feature are used from next Intersection Recognition. Since the update serves as tracking, we named it LLAH Tracking.

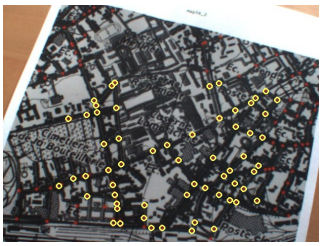


Figure 5: Matched Intersections by Reprojection

6 Experimental Results

6.1 Experimental Settings

We use real GIS data from Nantes city in France including 3760 intersections [11]. Our device is composed of a laptop (Intel Core 2 Duo 2.2GHz and 3GB RAM) with a firewire camera.

Before using LLAH, we have to determine the values of some parameters described in Section 3. Since these parameters were determined on experimental bases in [9], we set $n = 7$, $m = 5$, $k = 50$, $H_{size} = 2^{20} - 1$ from some experiments.

6.2 Minimum Number of Intersections

In this section, we confirm the minimum number of intersections for MIR with the initial database described in 5.2. Each user's map is generated by dividing the city into 100 areas (Figure 6(a)). For making the extraction of intersection easy, we use the map including only intersections. The camera is set parallel to the map to capture a top view image.

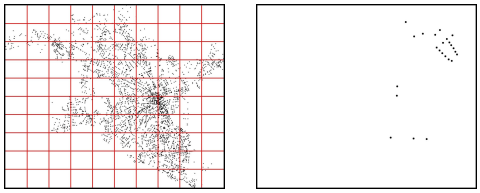
As shown in Table 1, the number of maps including less than 20 intersections is 52. All the maps couldn't be retrieved because the number of retrieved points is not enough for recognizing the area of the map. In the case where the number of points is between 21 and 40, some maps couldn't be retrieved. In one of the failure cases, many intersections were in a line (Figure 6(b)). In this case, the LLAH features couldn't be stably computed. If a map includes more than 41 points, our method works stably.

6.3 AR Display

Our method can still work in the cases of strong camera tilting and occlusions (Figure 7). This is because we register new LLAH features according to the changes of the user's viewpoint.

Table 1: Number of Failures in Map Retrieval

Intersections	Maps	Failures
0-20	52	52
21-40	13	4
41-	35	0



(a) 100 Maps (b) Failure

Figure 6: Number of Failures in Map Retrieval

For evaluating computational costs of AR display, 100 frames are captured in order to compute the average computational costs as shown in Table 2. 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, GIS Data Rendering took most computational costs which are more than twice that of our algorithm. Since GIS data includes detailed polygons, the data has to be simplified in order to reduce computational costs.

Table 2: Computation Cost

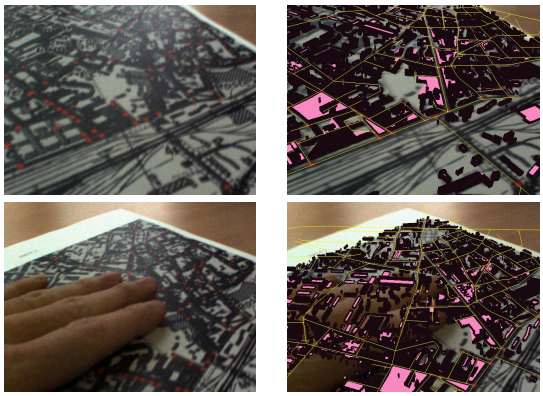
Process	Time (msec)
Intersection Extraction	11
Intersection Recognition	22
Camera Pose Estimation	3
Update Database	10
GIS Data Rendering	102

7 Conclusions

In this paper, we proposed a method for retrieving a corresponding map of a map image from a map database. We extended LLAH based DIR in order to enable free camera moving. In our method, registration and retrieval in LLAH based DIR were improved for reducing computational costs in the retrieval process. LLAH features of rotated map images were registered in the registration. In addition, LLAH feature was updated when the user's viewpoint changed. Our improvements can enable MIR to the case of strong camera tilting, occlusion and fewer intersections. To enables use of maps from different manufactures, a method for intersection extraction will be implemented.

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.



(a) Input (b) Output

Figure 7: GIS Rendering

References

[1] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Wu. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *J. of the ACM*, 45:891–923, 1998.

[2] H. Bay, T. Tuytelaars, and L. V. Gool. SURF: Speeded up robust features. In *Proc. ECCV*, pages 404–417, 2006.

[3] Y. Chiang, C. Knoblock, C. Shahabi, and C. Chen. Automatic and accurate extraction of road intersections from raster maps. *GeoInformatica*, 2008.

[4] M. Datar, P. Indyk, N. Immorlica, and V. S. Mirrokni. Locality-sensitive hashing scheme based on p-stable distributions. In *Proc. SCG*, pages 253–262, 2004.

[5] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *C. of the ACM*, 24:381–395, 1981.

[6] N. R. Hedley, M. Billinghurst, L. Postner, R. May, and H. Kato. Explorations in the use of augmented reality for geographic visualization. *Teleoperators and Virtual Environments*, 11:119–133, 2002.

[7] Y. Lamdan and H. Wolfson. Geometric hashing: A general and efficient model-based recognition scheme. In *Proc. ICCV*, pages 238–249, 1988.

[8] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60:91–110, 2004.

[9] T. Nakai, K. Kise, and M. Iwamura. Use of affine invariants in locally likely arrangement hashing for camera-based document image retrieval. In *Proc. DAS*, pages 541–552, 2006.

[10] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *Proc. CVPR*, pages 2161–2168, 2006.

[11] OrbisGIS. http://orbisgis.cerma.archi.fr/.

[12] G. Reitmayr, E. Eade, and T. Drummond. Localisation and interaction for augmented maps. In *Proc. ISMAR*, pages 120–129, 2005.

[13] H. Uchiyama, H. Saito, V. Nivesse, M. Servières, and G. Moreau. AR representation system for 3D GIS based on camera pose estimation using distribution of intersections. In *Proc. ICAT*, pages 218–225, 2008.

[14] Y. Uematsu and H. Saito. Vision based registration for augmented reality using multi-planes in arbitrary position and pose by moving uncalibrated camera. In *Proc. MIRAGE*, pages 99–107, 2005.

[15] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond, and D. Schmalstieg. Pose tracking from natural features on mobile phones. In *Proc. ISMAR*, pages 125–134, 2008.