Architecture of Image Feature DB Storage for Mobile Visual Localization

Donghyun Jeon *, Seonghyun Kim *, Sanghoon Lee *, Ju-il Jeon†

* Department of Electrical and Electronic Engineering, Yonsei University, Korea

jdh3577@yonsei.ac.kr, sh-kim@yonsei.ac.kr, slee@yonsei.ac.kr

†Intelligent Cognitive Technology Research Department, IT Convergence Technology Research Laboratory, ETRI, Korea

seventhday07@etri.re.kr

***Abstract*—Location-based services (LBS) are becoming increasingly popular. Particularly, indoor mobile localization methods that use GPS or wireless signals have problem of accuracy yet. This article presents a novel system to mobile visual localization according to a given image associated with a variety of wireless signals in indoor environment. Recently structure from motion (SfM) approaches enable to create 3D models of scenes. These reconstructed sparse 3D-point clouds can then be used for accurate image-based localization by 3D-to-2D feature matching from database. Our proposed structure of storage in database is able to efficiently handle such large amounts of image feature data.**

***Keywords*—Location-based services (LBS), visual localization, indoor localization, database, wireless fingerprint**

I. INTRODUCTION

More and more people are using their smart phones to enjoy ubiquitous location-based services (LBS). Especially, indoor localization has long been a goal of pervasive computing research. Compared with outdoor positioning, indoor localization is more challenging, since GPS signals are rarely accessible, yet room-level or even submeter precision is often required [1]. Therefore, wireless signal based on received signal strength indicator (RSSI) has been developed in indoor scenario. However, it is not enough for practical environments because Accuracy strongly depended on the number of senders in the environment.

It is worth noting that phone-captured photos contain context information about the environment, which could lead to a more precise location description, resulting in recent increased interest in research on visual (image-based) localization [2]. Visual localization technique in indoor environment is to communicate with the server using images captured by mobile devices such as smartphone. It means to estimate the user locations utilizing technique of 3D structure reconstruction through image based feature point extraction and the feature point matching with database in server. As the smartphone camera is improved, visual localization has been also treated as one of the potential positioning technique. However, large-scale visual localization has an image retrieval problem. The main challenge for visual localization is the rapid and accurate search for images related to the current scanning or capturing in a large database [3]. After finding images in a database that are most similar to the query image, the location of the query can be determined relative to it [4]. Typical visual localization approaches match the captured image against a geo-referenced image database and use the geo-tag from the retrieved images to induce the query location [5]-[7]. We focus on how to manage such a large database. We propose an architecture of efficient database storage for localization. In our proposed approach, the database is composed of wireless fingerprints and image features.

II. MOBILE VISUAL LOCALIZATION

Our proposed approach of visual localized system is shown in Figure 1. Mobile user requests location information to server sending the captured image. The high reliable features of image received from user are extracted by pre-processing and feature extraction at server. For example, the features of image are extracted after moving object detection and removing the object. Then, the server calculates the location information (camera position and direction) by matching the 2D-image features with 3D structure models in database. The user can receive the location information from server.

A. 3D Reconstruction via Structure from Motion

To achieve a higher localization accuracy, more detailed information is needed which can be obtained from a 3D-reconstruction model. Then, 3D information is stored at database. The general 3D reconstruction procedure is as follows [8]:

1) *Feature Extraction*: The common approach is to use the feature descriptors, e.g. scale-invariant feature transform descriptors (SIFT) [9] in 2D images. In our approach, the SIFT descriptor is exploited because of its strong descriptive power.

2) *3D Modeling:* 3D models are created using multiple-view vision method from features of 2D images based on structure from motion (SfM) algorithm [10]. First, for each pair of image, the feature descriptors are matched. The reconstruction started from a selected

initial image pair that has a large number of matches while containing certain viewpoint changes. The five point algorithm is used to estimate camera parameters for this initial pair [10] and matched image points are triangulated. The 3D model then incrementally is built

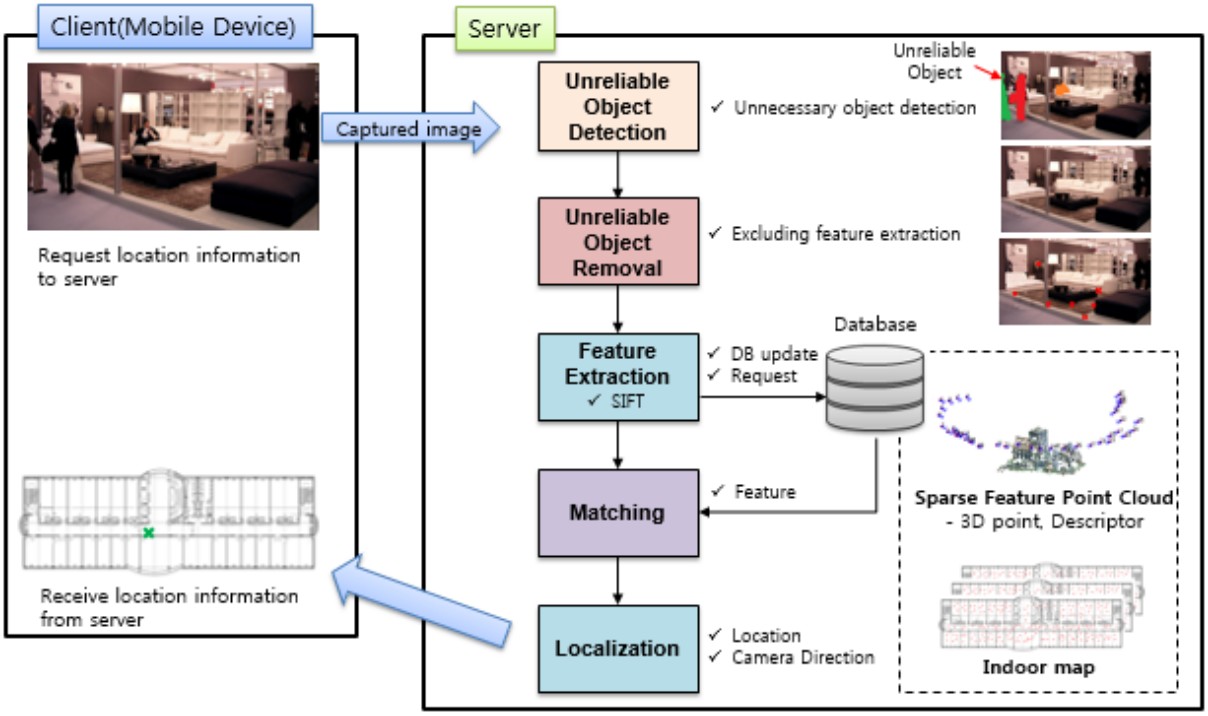

Figure 1. High-accurate mobile visual localization system

by adding a few images at a time. Finally, a bundle adjustment jointly refines the estimated camera parameters and 3D points [11].

B. 3D-to-2D Matching

3D point corresponding to a 2D feature by searching for the nearest neighbors of that feature's descriptor in the space containing the 3D point descriptors. Image retrieval technique can be used to accelerate pose estimation against large 3D models obtained from SfM methods using the 3D points belonging to the features in the retrieved images to establish 3D-to-2D correspondences [12].

When there is plenty of correspondence between image features $x = (x_{2D}, y_{2D})$ and 3D points $X = (X_{3D}, Y_{3D}, Z_{3D})$, the 3x4 projection matrix (or camera matrix) $\mathbf{P}$ is estimated.

$$\lambda \begin{bmatrix} x_{2D} \\ y_{2D} \\ 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} X_{3D} \\ Y_{3D} \\ Z_{3D} \\ 1 \end{bmatrix},$$

where $\mathbf{P} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]$, $\mathbf{K}$ is the 3x3 calibration matrix (or intrinsic matrix), $\mathbf{R}$ is the 3x3 rotation matrix, $\mathbf{t}$ is the translation vector, and λ is the homogeneous scaling factor. Generally, the calibration matrix is

$$\mathbf{K} = \begin{bmatrix} f & \gamma & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

where f is the focal length of camera, γ is the skew of camera which is often 0 (also in this paper), and (u_0, v_0) is the principal point of the image. From the 3D points that are corresponded the feature points of query image, the projection matrix can be solved by the six-point direct linear transform (DLT) algorithm. Then, the user position of image is given by $-\mathbf{R}'\mathbf{t}$ and the camera direction is $-\mathbf{R}'[0 \quad 0 \quad 1]'$.

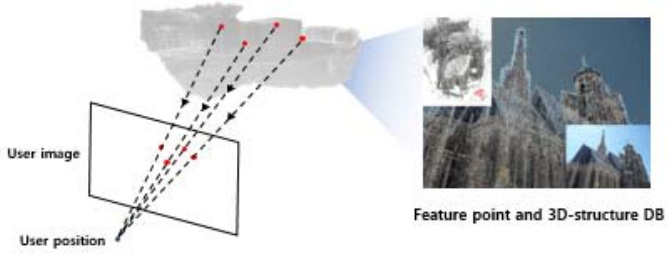

Figure 2. Visual localization based on 3D-to-2D back projection

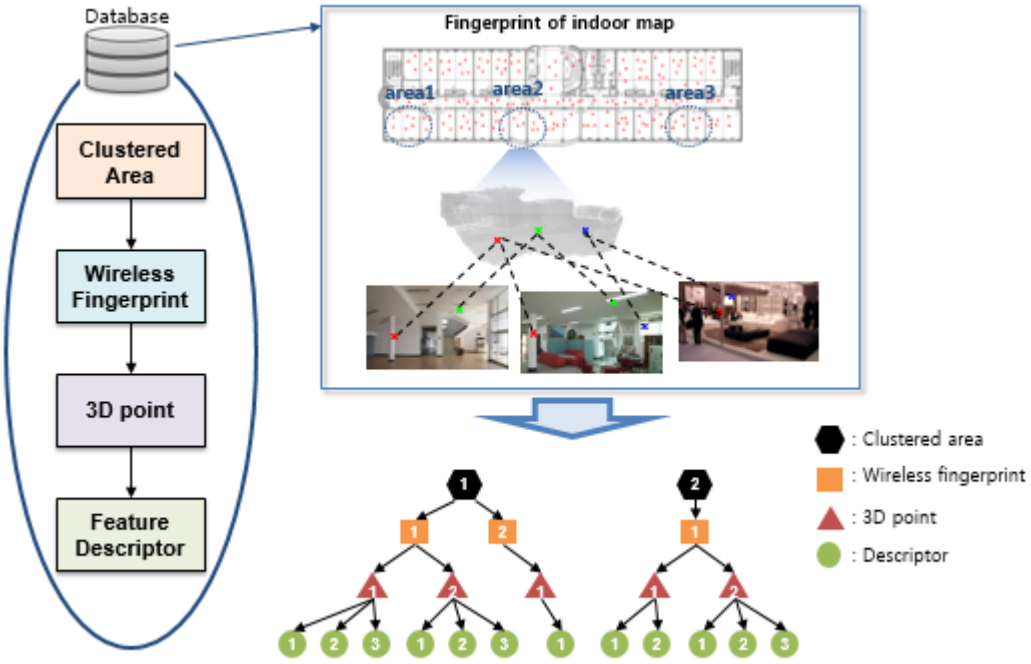

Figure 3. Hierarchical DB storage for indoor localization

III. Architecture of Database for Indoor Localization

As database size increases, the amount of memory used to index the database features can become very large. Thus, developing a memory-efficient structure is a problem of increasing interest [13]. In [14], the authors proposed to compute a representative set of 3D points that cover a 3D scene from arbitrary view points and utilize a vocabulary tree data structure for fast feature indexing. A subsequent matching approach and geometric verification directly delivers the pose of the query image. The authors in [15] retrieve the related 3D models in the database using the Bag-of-Word (BoW) representation indexed with a vocabulary tree structure. Our proposed architecture is to manage the database using hierarchical tree structure.

A. Design of image feature DB storage structure

For high performance of indoor localization system, hierarchical DB storage architecture is proposed as shown in Figure 3. The database contains indoor map, wireless finger prints, and image related data. Therefore, our localization system operates efficiently tree-based storage and search.

Based on the user geo-information such as name of building and floor of building, areas are roughly clustered first. Since wireless signal fingerprints can acknowledge the user location range in a clustered area, the wireless fingerprints are split and stored by clustered areas. Because the database contains wireless signal fingerprints such as radio map, wireless signals received from user match with database using pattern matching algorithm [16]. We expect that the range based on wireless fingerprint is about 5 meter and then make more accurate by using image features. Given wireless signal range, using the camera of mobile device, background can be scanned around. 3D scene model can be made with respect to the background and the 3D points are contained for each wireless fingerprint. Then, each 3D point include the descriptors of the feature points of each 2D-images captured from a variety of views.

Therefore, in case of new DB generation and update for visual localization, the feature descriptors are added under structure like Figure 3. For instance, many of new feature descriptor can be added at same 3D point. This structure for DB can improve the rapid and accurate for searching matched data. Moreover, our mobile visual localization system can support DB update for not only specialized company but also users directly just high-reliable DB.

IV. Conclusions

In this paper we have described compositive localization technique based on both wireless signal and image. We have presented an architecture of hierarchical database storage system for image features. Our approach system can be achieved high-speed and accurate searching DB and matching. Therefore, it can show the good performance for localization technique.

Acknowledgment

This work was supported by the ICT R&D program of MSIP/IITP. [R0116-14-3004, Development of Autonomous Location Infrastructure DB Update Technology based on User Crowd-sourcing]

References

[1] Z. Yang, Z. Zhou, and Y. Liu, “From RSSI to CSI: Indoor Localization via Channel Response,” ACM Computing Surveys, vol. 46, Nov. 2014.

[2] D. Robertson and R. Cipolla. “An image-based system for urban navigation,” in BMVC, 2004.

[3] G. Schroth, R. Huitl, D. Chen, M. Abu-Alqumsan, A. Al-Nuaimi, and E. Steinbach, “Mobile visual location recognition,” IEEE Signal Processing Magazine, vol. 28, no. 4, pp. 77–89, July 2011.
[4] W. Zhang and J. Kosecka. “Image based localization in urban environments,” in 3DPVT, 2006.
[5] D. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg. “Mapping the world’s photos,” In Proceedings of the 18th International Conference on World Wide Web. ACM, 761–770, 2009.
[6] Y. Avrithis, Y. Kalantidis, G. Tolias, and E. Spyrou. “Retrieving landmark and non-landmark images from community photo collections,” In ACM Multimedia, 2010.
[7] A. R. Zamir and M. Shah. “Accurate image localization based on google maps street view,” In ECCV, 2010.
[8] R. I. Hartley and A. Zisserman. “Multiple View Geometry in Computer Vision,” Cambridge Univ. Press, 2nd edition, 2004.
[9] D. Lowe. “Distinctive image features from scale-invariant keypoints.” IJCV, 60(2):91–110, 2004.
[10] F. Dellaert, S. Seitz, C. Thorpe, and S. Thrun, "Structure from Motion without Correspondence" IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2000.
[11] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, “Bundle Adjustment-A Modern Synthesis,” Lecture Notes in Computer Science, vol. 1883, pp. 298-375, 2000.
[12] T. Sattler, B. Leibe, and L. Kobbelt, “Fast Image-Based Localization using Direct 2D-to-3D Matching,” In ICCV, 2011
[13] B. Girod et al., ‘‘Mobile Visual Search,’’ IEEE Signal Processing Magazine, vol. 28, no. 4, pp. 61-76, July 2011
[14] A. Irschara, C. Zach, J.-M. Frahm, and H. Bischof., “From structure-from-motion point clouds to fast location recognition,” In CVPR, 2009.
[15] G. Schindler, M. Brown, and R. Szeliski. “City-scale location recognition,” In CVPR, June 2007.
[16] S.-H. Fang, T.-N. Lin, and K.-C. Lee, “A Novel Algorithm for Multipath Fingerprinting in Indoor WLAN Environments,” IEEE Trans. Wireless Comm., vol. 7, no. 9, pp. 3579-3588, Sept. 2008

Donghyun Jeon received the B.S. degree in electrical engineering in 2014 from Yonsei University, Seoul, Korea, where he is currently working toward the M.S. and Ph.D. degrees with the Multidimensional Insight Laboratory. His research interests include indoor localization, Beyond Fourth- and Fifth-Generation (B4G/5G) systems, wireless multimedia communications.

Seonghyun Kim received the B.S. degree in electrical engineering in 2009 from Yonsei University, Seoul, Korea, where he is currently working toward the M.S. and Ph.D. degrees with the Multidimensional Insight Laboratory. His research interests include deep learning, cloud computing, indoor localization, ad hoc networks, Beyond Fourth- and Fifth-Generation (B4G/5G) systems, wireless multimedia communications, and cross-layer optimization.

Sanghoon Lee (M’05–SM’12) received the B.S. in E.E. from Yonsei University in 1989 and the M.S. in E.E. from Korea Advanced Institute of Science and Technology (KAIST) in 1991. From 1991 to 1996, he worked for Korea Telecom. He received his Ph.D. in E.E. from the University of Texas at Austin in 20 00. From 1999 to 2002, he worked for Lucent Technologies on 3G wireless and multimedia networks. In March 2003, he joined the faculty of the Department of Electrical and Electronics Engineering, Yonsei University, Seoul, Korea, where he is a Full Professor. He was an Associate Editor of the IEEE Trans. Image Processing (2010-2014). He has been an Associate Editor of IEEE Signal Processing Letters (2014-) and Chair of the IEEE P3333.1 Quality Assessment Working Group (2011-). He currently serves on the IEEE IVMSP Technical Committee (2014-), was Technical Program Co-chair of the International Conference on Information Networking (ICOIN) 2014, and of the Global 3D Forum 2012 and 2013, and was General Chair of the 2013 IEEE IVMSP Workshop. He also served as a special issue Guest Editor of IEEE Trans. Image Processing in 2013, and as Editor of the Journal of Communications and Networks (JCN) (2009-2015). He received a 2012 Special Service Award from the IEEE Broadcast Technology Society and a 2013 Special Service Award from the IEEE Signal Processing Society. His research interests include image/video quality assessment, medical image processing, cloud computing, sensors and sensor networks, wireless multimedia communications and wireless networks.

Ju-il Jeon received the B.S and M.S. degree in information and communication engineering in 2009 from Chungbuk National University. He is currently researching at positioning & navigation technology research section in Korea Electronics and Telecommunications Research Institute. His research interests include video coding, video streaming technology, image recognition and indoor localization.