Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography
Vol. 34, No. 2, 207-215, 2016
http://dx.doi.org/10.7848/ksgpc.2016.34.2.207

ISSN 1598-4850(Print)
ISSN 2288-260X(Online)
Original article

Big Data Architecture Design for the Development of Hyper Live Map (HLM)

Moon, Sujung[1] · Pyeon, Muwook[2] · Bae, Sangwon[3] · Lee, Dorim[4] · Han, Sangwon[5]

Abstract

The demand for spatial data service technologies is increasing lately with the development of realistic 3D spatial information services and ICT (Information and Communication Technology). Research is being conducted on the real-time provision of spatial data services through a variety of mobile and Web-based contents. Big data or cloud computing can be presented as alternatives to the construction of spatial data for the effective use of large volumes of data. In this paper, the process of building HLM (Hyper Live Map) using multi-source data to acquire stereo CCTV and other various data is presented and a big data service architecture design is proposed for the use of flexible and scalable cloud computing to handle big data created by users through such media as social network services and black boxes. The provision of spatial data services in real time using big data and cloud computing will enable us to implement navigation systems, vehicle augmented reality, real-time 3D spatial information, and single picture based positioning above the single GPS level using low-cost image-based position recognition technology in the future. Furthermore, Big Data and Cloud Computing are also used for data collection and provision in U-City and Smart-City environment as well, and the big data service architecture will provide users with information in real time.

Keywords : Big Data, Cloud Computing, HLM, Multi-Source

1. Introduction

The spatial data field is rapidly diversifying and changing with the development of ICT. Recent spatial data technologies are providing realistic 3D spatial information services and other spatial data services through a variety of Web-based contents.

In the past, only experts could acquire and possess spatial data, but now the general public can easily acquire and possess spatial data. Therefore, we could say that almost everyone is encountering and using this information. If we could quickly acquire and predict spatial data, the utilization and development of spatial data could be maximized (Jang *et al*., 2013). Today, we can see the desired place in advance without going to the place through Internet maps or map applications of mobile phones. This kind of development in the spatial data field is continuing and we expect that research on the related fields will be actively conducted in the future (Cho *et al*., 2013).

Received 2016. 04. 08, Revised 2016. 04. 20, Accepted 2016. 04. 28
1) Member, Dept. of Advanced Technology Fusion, Konkuk University (E-mail: msujung@konkuk.ac.kr)
2) Corresponding Author, Member, Dept. of Civil Engineering, Konkuk University (E-mail: neptune@konkuk.ac.kr)
3) LoDiCS Co., LTD. (E-mail: swbae@lodics.com)
4) LoDiCS Co., LTD. (E-mail: drlee@lodics.com)
5) Member, Dept. of Civil Engineering, Konkuk University (E-mail: hsw721@naver.com)

However, with the acceleration of spatial data creation based on location information, spatial big data that allows the organized management of large-volume spatial data is emerging as an issue. Accordingly, the demand for technologies and services related to spatial big data is expected grow sharply, but the development of technologies and services for the collection, storage and analysis of spatial big data is still insufficient in South Korea (Lee *et al*., 2015). Until now, spatial data services have been provided by experts, but now the development of IT provided an environment in which users can offer spatial data services through images uploaded online. Multi-dimensional (spatial/non-spatial) properties are big data characterized by large volume, various forms, and fast creation speed.

To utilize the geometrically increasing information with the development of Internet, mobile, and SNS (Social Network Service) technologies and the standard/nonstandard data extracted from various multi-sensors such as CCTVs (IP cameras), digital cameras, mobile phone cameras, we need interface technologies utilizing the big data processing technology.

In this paper, the process of building a HLM using multi-source data to acquire stereo CCTV and other various data is presented and a big data service architecture design is proposed for the use of flexible and scalable cloud computing to handle big data created by users through such media as social network services and black-boxes.

2. Development of HLM and Acquisition of Multi-source Data

2.1 Development of HLM

The purpose of HLM is to provide accurate (near) real-time 3D spatial information services. As a way to build 3D spatial information for this purpose, research on 3D reconstruction is being conducted using image matching algorithms for data acquired through stereo CCTV and multi-source data (smart phone black box, digital camera, etc.).

Figure 1 shows the HLM process. The important part of this process is to construct high-resolution color 3D data through image matching and geometric correction with stereo CCTV data and to construct (near) real-time spatial data using the data inputted from various image devices.

The HLM construction flow chart in Figure 2 consists of data acquisition, data processing and construction, and additional data processing and construction using multi-source data.

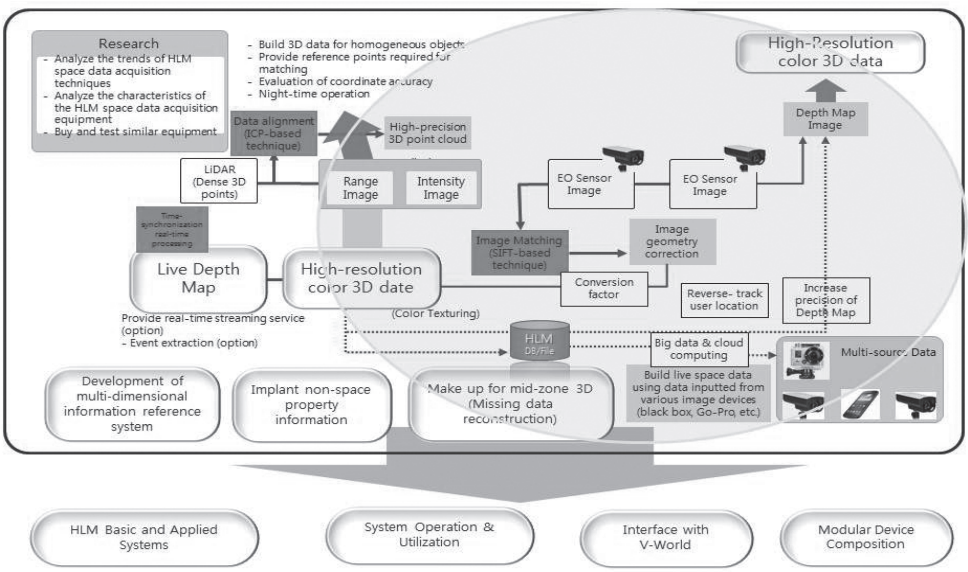

Fig. 1. HLM research process

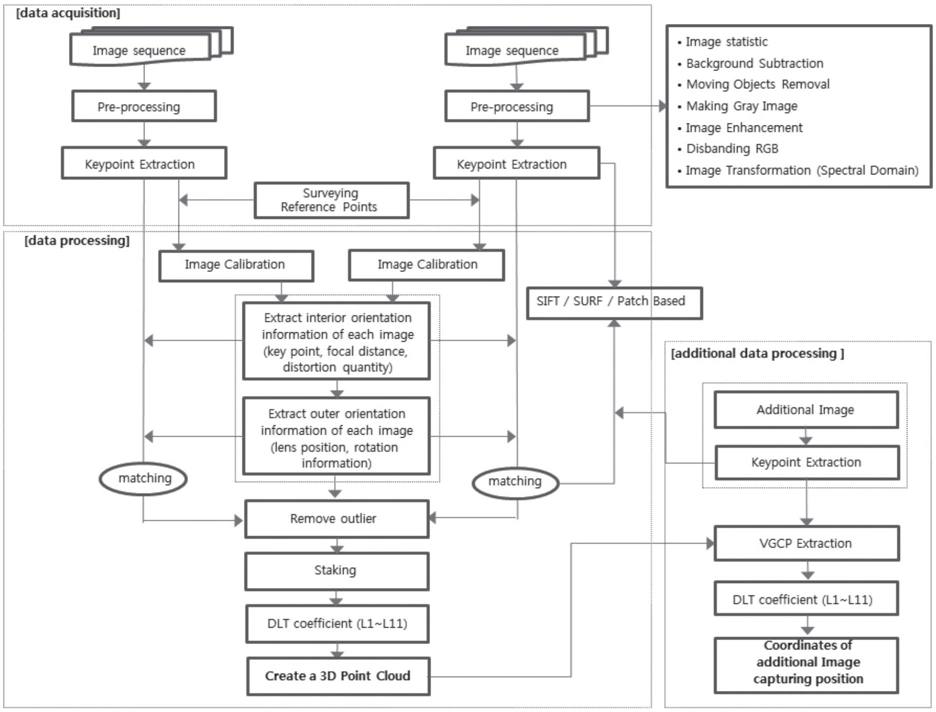

Fig. 2. HLM development process

In the data acquisition process, the stereo CCTV images that appear in pairs of left and right images are converted to images which are then corrected through camera calibration after the stereo CCTV images are converted to images. After that, they are decomposed into gray images and histogram equalization images, each of which is decomposed into RGB band again.

In the data processing and construction process, images are matched using SIFT and PATCH (Lowe, 2004), and the mismatched points are removed using RANSAC. Matching points extracted through automatic matching contain mismatching points and if they are not removed, the accuracy of the 3D point groups will be degraded (Kim and Pyeon, 2014). Finally, the 3D point cloud data is constructed using the DLT (Direct Linear Transformation) technique (Jang *et al*., 2014).

The closed area of the spatial can be represented by the stereo CCTV images alone, and additional 3D spatial data can be added by using multi-source data to the constructed 3D spatial data. In the additional image data processing and construction process, camera matrix is obtained and 3D point cloud is extracted using VGCP (Virtual Ground Control Point) from the added multi-source data (Moon *et al*., 2016).

2.2 Acquisition of multi-source data

The development of 3D spatial information using stereo CCTV is only possible in areas where CCTVs are installed. The spatial data needs to be expanded to areas beyond the visible areas of CCTVs or where CCTVs cannot be installed. Multi-source data will be used to acquire images from such closed areas.

Multi-sources include CCTV, smart phone, digital camera, black box, and action cam. The characteristics of image capturing devices that are generally used these days are outlined in the table below.

CCTV can be used to acquire reference data through the analysis of the characteristics of multi-sources. For the areas that are not included in the CCTV capturing area, 3D spatial data is constructed using high-performance multi-source data applied to the DLT methods.

Table 1. Analysis of multi-source

Multi-source	Feature analysis	Uses of HLM construction
CCTV	CCTVs are advantageous for continuously and automatically acquiring image data from certain areas. However, they are not ideal for acquiring the reference data required for the development of 3D data because they are fixed and cannot acquire data from other areas.	- Continuously acquire image data - Easy to acquire reference data because it is fixed at one location.
Smart phone	Smart phones have excellent portability compared to other image acquisition devices and can effectively acquire image data required for the development of 3D data through interfaces with Twitter, FaceBook, Line, and M2Day where image data are actively uploaded these days.	- Effective image data acquisition by interfacing with SNS
Digital camera	Digital cameras are equipped with wireless communication features such as NFC, WiFi, and Bluetooth and can immediately transmit data after taking pictures as with smart phones.	- Can transfer the image data immediately upon capturing them as with smart phones
DSLR digital camera	DSLR digital cameras can acquire higher-quality image data compared to other image acquisition devices thanks to high-performance sensors and lenses. They can acquire more vivid image data even under little light.	- Can acquire high definition and high quality image data
Black box	The obligation of installing black boxes in vehicles is being carried forward in South Korea owing to their discernment power and economic benefits. There are many users because the black box recommendation system has been established.	- There are many users. - Appropriate for complementing the blind spots and blocked areas of the image capturing device
Action cam	Action cams use fisheye lens with a large angle of view to capture broad areas like black box and the acquired image data usually contain distortions, which need separate correction.	- Easy to acquire image data from a first-person point of view - Separate correction is required because the data contains distortions

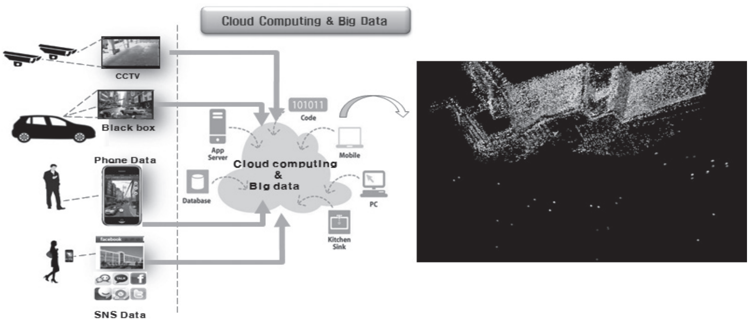

Fig. 3. Diagram of multi-source data & cloud computing

With the development of image acquisition devices, most people carry digital cameras or mobile phones embedded with a camera. Moreover, some people even use video cameras such as Go-Pro to keep or upload to web sites the videos of their activities (Kim *et al*., 2014).

Among the multi-sources, smart phone or digital camera data for which social network is possible is collected using cloud computing, and the 3D spatial data is constructed through the HLM construction process.

Figure 3 is conceptual diagram that construct multi-source to 3D spatial information by using the Big Data and Cloud Computing.

3. Cloud and Spatial Big Data

3.1 Cloud computing

Interest in cloud computing as an IT megatrend is growing among IT companies, general companies, and individual users. Cloud computing is to rent infrastructure from service providers and it is being spotlighted because there is no burden of direct investment and the cost can be saved for unused resources (Kim *et al*., 2015).

Cloud computing refers to computing services that provide ICT resources with a high level scalability (CPU, memory, storage, network, OS, applications, etc.) to a large number of customers using Internet technology. It is an on-demand outsourcing service for ICT resources on the Internet, and users pay the price for using ICT resources through the Internet as required which are provided from service providers (Lim *et al*., 2013).

Cloud service providers provide in real time ICT resource services that are needed by customers by integrating data centers distributed in many places. With this concept, users do not own ICT features but rent them. It is a new computing service model characterized by the reduction of ICT resources introduction/management costs, flexibility of computer capacity, and excellent service accessibility (Lim *et al*., 2013).

The importance of cloud is rising as the basis of new technology convergence businesses such as big data and IoT (Internet of Things). Its application areas are expanding as a new mainstream replacing the conventional implementation-type IT market (Shin, 2015).

PaaS (Platform as a Service) is service that offers an integrated environment for developing, testing, and installing applications, and provide cloud development platform, big data platform, etc. IaaS (Infrastructure as a Service) is a service that allows the use of resources whenever they are needed by virtualizing servers, storages, and networks. Integrated services include VDI (Virtual Desktop Infrastructure) service which allows the use of personal PC environment anytime anywhere through the Internet, cloud-based SECaaS (Security as a Service), and cloud-based CDN (Contents Delivery Network) (Kim *et al*., 2013).

3.2 Big-Data and spatial Big-Data

3.2.1 Big-Data

Big-Data refers to a set of large-volume data that consists of various data formats and is created very quickly and is difficult to process with existing management and analysis tools or systems. It is an information technology to extract valuable information using large-volume data and actively respond or predict changes based on the created knowledge. Big-Data is the crude oil that determines the competitiveness of the future and critical resources that open new possibilities of global development. Big-Data is expected to have the greatest impact on corporations after the Internet (Lim *et al*., 2013).

Today, the Big-Data environment has arrived in which a wide variety of information and data are created to a degree that cannot be handled easily with the rapid increase of Internet use, the propagation of smart phones, the activation of SNS, and the spread of digital economy (Kim *et al*., 2013).

Government 3.0 of the Korean Government perceives the utilization of Big-Data as a core value of the creative economy initiative. Government agencies including the Ministry of Land, Infrastructure and Transport, the Ministry of Science, ICT and Future Planning are also promoting various projects using Big-Data. In particular, spatial Big-Data is being spotlighted as a new issue which allows the organized management of large-volume spatial data with the accelerated creation of spatial data based on location information in the spatial data area (Lee *et al*., 2015).

Starting with CCTV images, the similarity of image data is defined, and the images are overlapped sequentially according to their closeness to this similarity. Then they are finally used for the development of image 3D modeling (Cho *et al*., 2014).

3.2.2 Spatial Big-Data

As with the characteristics of general Big-Data, spatial Big-Data is composed of large-volume data to a degree that cannot be stored, managed, or analyzed in conventional methods. Spatial data basically contains various large-scale standard/nonstandard data such as 2D/3D numerical maps, topographical maps, satellite images, aerial photographs, and DEMs (Lim *et al*., 2013).

Expanded spatial Big-Data which can be easily interfaced with spatial data may include such standard/semi-standard/non-standard data as public facilities database, underground facilities database, property database, BIM information, public database, statistics database, sensor information, mobile object information, disaster information, u-City/

Smart City information, IoT information, location-based SNS information, user-participation-type spatial data.

We need to respond to the market demands for the efficient storage, management, and analysis services for real-time location information, existing large-volume spatial data, and external Big-Data to be interfaced, which will sharply increase in the future. We need to proactively respond to the demands for spatial Big-Data source technologies for efficient process of Big-Data that combines real-time location information, sensor information, and spatial information in addition to the general storage, management, and service technologies related to Big-Data. Furthermore, our competitiveness in domestic spatial Big-Data technologies needs to be enhanced compared to the overseas advanced technologies of market leaders such as Google, ESRI, etc (Kim, 2013).

4. Big-Data Service Architecture Design

4.1 Analysis of Big-Data solutions

AWS (Amazon Web Services) is a cloud service consisting of various remote computing services provided by Amazon.com. AWS allows you to add or remove servers within a few minutes by a few clicks. You can also easily increase 5-6,000 servers in a short period of time. The cost has been dramatically lowered.

For a domestic enterprise cloud solution, Innogrid's Cloudit3 (Cloud It, 2015) provides such services as real-time migration, template and snapshot management and usage measurement, user access management, Restful APIs, autoscaling, load balancing, adaptive IP, port forwarding, and object storage. IaaS provides hardware support such as virtual server and virtual storage as well as private cloud. Still images extracted from CCTV stream images need to be managed and serviced through such private cloud as Cloudit 3. Although additional software must be developed for interface at the data or service level between a private cloud and a public cloud (e.g., AWS), you can acquire security through the private cloud and cost saving effect through a public cloud.

4.2 Interface with web crawler

HLM services create 3D cloud points using images acquired from various devices such as CCTV, digital camera, smart phone camera, and vehicle black-box. The Big-Data images acquired from a variety of devices are widely distributed in the Internet (Web) spatial. The crawling (searching and collecting) feature for such Big-Data images crawls auxiliary data that describes the images and can determine whether or not to save the collected images in the system through content-based search. With the explosive increase of the Internet (Web), the importance of search engine as a means to obtain information on the Internet is being spotlighted. Web crawler is a program used in a search engine to collect Web pages (Shin, 2006; Shin *et al*., 2008).

When interfaced with HLM data management, 'Cloud Storage (S3)' of the data management component is commonly used for nonstandard data, and the RDS (Relational Data System), which is the RDBMS (Relational Data Base Management System) of AWS, is used for standard data. Monitor, Manager, and Agent are implemented using the service components provided by AWS (Singh, 2012),

4.3 Big-Data service architecture design

For cloud point data, the final cloud point file, origin file,

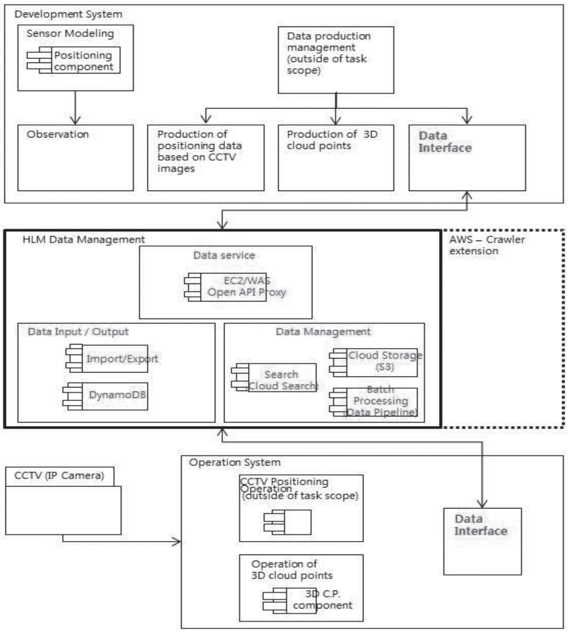

Fig. 4. Big-Data service architecture design

Table 2. Service components

Division	Description
Common properties	ID, file path
CCTV information	Photographed location (area or group name), left image, right image
CCTV still image file	left image, right image
Control point information	Photographed location (area or group name), coordinate system
Control point file	x, y, z (text)
Cloud point information	Photographed location (area or group name), number of points, implementation method (matching method, image type)
Cloud point file	Last point x, y, z / Origin x, y, z
Texture file	Texture file, texture type, texture source path
Digital camera information	Model, resolution, photographed location (area or group name)
Digital camera file	Image name

Table 3. Data management

Data Management	
Data Pipeline	Data batch process workflow
	This web service stably processes data between AWS computing and storage services as well as field data sources and moves them according to the specified cycle. You can periodically access stored data, convert the data size, and efficiently send the results to S3, DynamoDB, and other services.
S3	Cloud storage
	This is a stable cloud storage service with excellent scalability. You can store and search data of a required volume through a simple Web service interface. It also supports data management according to the life cycle policy. When a specific policy is set, data migration to the proper storage class is possible with no change of application.
CloudSearch	Data search
	You can simply set, manage, and adjust search functions for Web site or application. Highlight, auto complete, topography search, and multi-language support

Table 4. Data input/output

Data Input/Output	
Import/Export	Large transmission for local operation system
	Large transmission/reception of interface data with the development and operation clients
DynamoDB	Binary data processing, metadata processing
	Storage of target data and management of metadata

Table 5. Data service & map service

Data Service	
EC2	Web service/Web application service
	Web service and Web application service that are accessed by data interface service users and managers. Automatic scale-up is possible according to the degree of use.
Map Service	
External map service	Integrated gateway through proxy on EC2
	A gateway service for using external map services to mark on the map the locations of images and supplementary data (control point, etc.)

texturing file, and target area name are sent. For digital camera image data, the target file, area name, the left and right image file for sending stereo image data acquired with CCTV, and area name are sent. Furthermore, the access information (IP, port, etc.) for data interface are set and the data is searched. The type of data to search (cloud point, digital camera image, CCTV stereo image, or control point) is selected and resolution, camera name, photographed area, etc. are searched.

Figure 4 is block diagram of HLM service architecture that sync operation system with construction system for HLM data management (data service, data input/output, data management). Table 1,2,3,4 are component and explanation of HLM service architecture.

5. Conclusion

Recent spatial data technologies are providing realistic 3D spatial information services and other spatial data services through a variety of Web-based contents. A big data service architecture was designed to apply big data technology so as to use image data generated from the Internet and smart phone in the process of 3D city spatial data development. Many big data analysis technologies are existing now, but they have not reached the level of processing video and image data rather than text yet. The platform for using image-based spatial data required for HLM development in big data has not been constructed, but the application of big data technology will enable the extraction and processing of large-volume distributed image data which are required for the development of 3D spatial information.

The HLM development process using multi-source data for the acquisition of stereo CCTV and various other data was presented, and a big data service architecture design for using flexible and scalable cloud computing was proposed to process big data created by users such as SNS and black box. Cost savings from the reduced data collection time and the minimized use of specialty devices can be expected by developing and providing spatial information services so as to use standard/nonstandard data that exist online for HLM development through big data and cloud computing.

Acknowledgment

This research was supported by a grant (15CHUD-C061156-05) from National Spatial Information Research Program funded by Ministry of Land, Infrastructure and Transport of Korean government & This paper was written as part of Konkuk University's research support program for its faculty on sabbatical leave in 2015.

Reference

Amazon (2013), *Amazon Web services Inc*., https://aws.amazon.com (last date accessed: 5 April 2016).

Cho, Y.W., Pyeon, M.W., and Kim, D.S. (2013), Comparison of Image Matching Algorithms in the Spatial Information System, *Journal of Korean Society for Geospatial Information System*, Vol. 2013, No. 5, pp. 139-142. (in Korean)

Cho, Y.W., Pyeon, M.W., Kim, D.S., Moon, S.J., and Jang, I.W. (2014), *Video Image Based Hyper Live Spatial Data Construction*, Springer Berlin Heidelberg, Heidelberg, Berlin, Vol. 274, pp. 371-376.

Cloud It (2015), *Innogrid corp*., http://www.cloudit.co.kr (last date accessed: 6 April 2016).

Jang, I.W., Pyeon, M.W., Eo, Y.D., Jeon, M.C., and Lim, S.B. (2013), Trend Analysis of Spatial Information Using Big Data, *Research Notes in Information Science*, Vol. 14, pp. 396-399.

Jang, I.W., Pyeon, M.W., Hong, K.H., Kim, J.H., Moon, K.I., and Kim, K.S. (2014), Realtime Digital Map by Using Multi Source Video Image, *International Conference on Innovative Engineering Technologies (ICIET)*, 28-29 Dec., Bangkok, Thailand, pp. 169-171.

Kim, Y.S. (2013), *the planning Report of Big Data Analysis and Utilization Technology Development Based in Spatial Information*, Report, R&D/2013-City Planning and Architecture-A03, Korea Aerospace Research Institute, 14p.

Kim, J.T., Jo, S.S., Oh, D.H., and Noh, J.G. (2015), An Analysis on the Technology Stack of Cloud Computing, *Journal of Korean Institute of Next Generation Computing*, Vol. 11, No. 6, pp. 79-89. (in Korean with English abstract)

Kim, J.H. and Pyeon, M.W. (2014), An Experiment of Eliminating Mismatching Points on Stereo Images for RDM, *Journal of Korean Society for Geospatial Information System*, Vol. 2014, No. 5, pp. 147-148. (in Korean)

Kim, J.H., Pyeon, M.W., Eo, Y.D., and Jang, I.W. (2014), An Experiment of Three-Dimensional Point Clouds Using GoPro, *World Academy of Science, Engineering and Technology*, Vol. 8, No. 1, pp. 82-85.

Kim, J.J., Shin, I.S., and Han, K.J. (2013), *Spatial Big Data*, Report, National IT Industry Promotion Agency, Seoul, Korea, pp. 14-25.

Lee, M.H., Park, J.M. Shin, D.B., and Ahn, J.W. (2015), A Study on the Selection of Core Services for Geo-Spatial Big Data, *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, Vol. 33, No. 5, pp. 385-396.

Lim, Y.J., Baek, S.K., Jang, S.I., and Won, H.S. (2013), *Cloud & Big Data for the Smart Internet Services*, Report, Korea Communications Agency, Seoul, Korea, pp. 1-31.

Lowe, D.G. (2004), Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91–110.

Moon, K.I., Pyeon, M.W., Kim, J.H., Kim, K.S., and Lim, Y.S. (2016), A Study of Construction of Utilization of Realtime Digital Map (RDM) Using CCTV and Multi-source Data Tools, *Proceedings of 2016 International Conference on Urban Planning, Transport and Construction Engineering (ICUPTCE'16)*, 2-3 Jan., Pattaya, Thailand, pp. 1-6.

Shin, E.H. (2015), Now Cloud's Generation, the Trend of the Cloud Computing and Prospect, *LG CNS*, Korea, http://blog.lgcns.com/770 (last date accessed: 7 April 2016).

Shin, E.J. (2006), *Design and Implementation of a Parallel Web Crawler for Large-Scale Search Engines*, Master's thesis, KAIST, Daejeon, Korea, 35p. (in Korean with English abstract)

Shin, E.J., Kim, Y.R., Heo, J.S., and Whang, K.Y. (2008), Implementation of a Parallel Web Crawler for the Odysseus Large-Scale Search Engine, *Journal of KIISE : Computing Practices and Letters*, Vol. 14, No. 6, pp. 567-581. (in Korean with English abstract)

Singh, H. (2012), Current Trends in Cloud Computing a Survey of Cloud Computing Systems, *International Journal of Electronics and Computer Science Engineering*, Uttar Pradesh, http://www.ijecse.org/ (last date accessed: 8 April 2016)