Sensing Web Project
- How to handle privacy information in sensor data -

Michihiko Minoh **Koh Kakusho**
Academic Center for Computing and Media Studies, Kyoto University
Kyoto 606-8501, Japan
{minoh, kakusho}@media.kyoto-u.ac.jp

Noboru Babaguchi
Graduate School of Engineering,
Osaka University
Osaka 565-0871, Japan
babaguchi@comm.eng.osaka-u.ac.jp

Tsuneo Ajisaka
Graduate School of Systems Engineering,
Wakayama University
Wakayama, 640-8510, Japan
ajisaka@sys.wakayama-u.ac.jp

Abstract

This article gives an overview of the *Sensing Web* project, launched in 2007 in Japan. The project's aim is to open the data obtained by the sensors existing in our daily living environment to the public. Since the data obtained by observing the real world directly with sensors include real-world information different from the Web, a new worldwide social information infrastructure - *Sensing Web* - is realized. In this article, we discuss the research issues for arising in connection with the Sensing Web.

Keywords: Sensor data, Opening, Information Infrastructure, Privacy information

1 Introduction

Recently, many kinds of sensors, including obstacle sensors, video cameras, thermometers, and so on, have been installed in various places in our daily living environments: stations, streets, malls, etc. Some of those sensors constitute sensor networks for exchanging their data in order to attain the purpose for which they are installed, more efficiently. In this article we refer to these networks as installed in our daily environments as *Ubiquitous Sensor Networks* (USNs).

Each USN is installed by some institutions including a local government, a transit company, a security company, and so on, for specific purposes: traffic control, building management, video surveillance, etc. The sensor data obtained from the USNs are used only for the purpose by the institution exclusively. However, such sensor data can actually be used for various purposes other than their original purpose, because the data include raw real-time information of the real world. If the sensor data were opened to the public so that anyone can use the data for their own purpose, similar to the Web, the data could serve as a new worldwide social information infrastructure that supplies the information different from that supplied by the current Web. In this article, we call this new social information infrastructure the *Sensing Web*. Whereas the Web supplies organized information collected and edited manually by humans, the Sensing Web supplies raw real-time information directly acquired from the real world.

The *Sensing Web Project* is a three-year project launched in the fall of 2007, aiming to develop information technologies necessary for making the data of USNs be shared among the public.

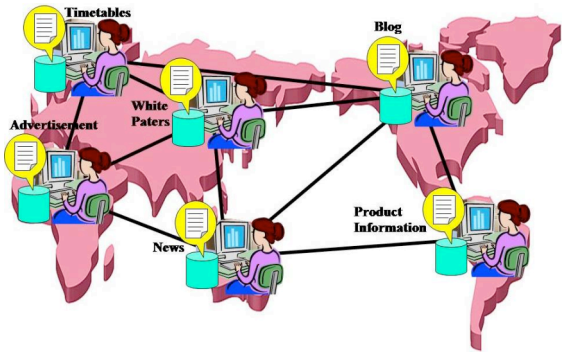
In the remainder of this article, we describe an overview of the project. First, in section 2, we compare the Web and the Sensing Web in order to make their similarity and difference clear. In section 3, we discuss research issues for realizing the Sensing Web. In section 4, we describe the general approach taken to solve the research issues. Concluding remarks are given in section 5.

2 Web versus Sensing Web

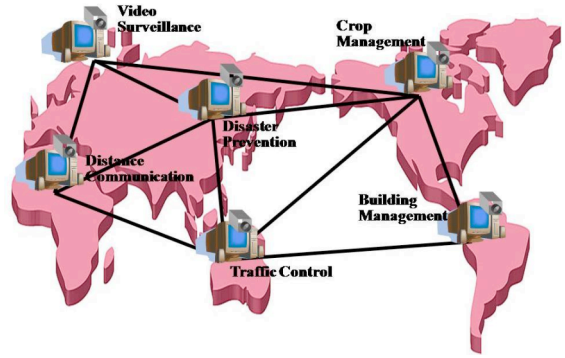
The Web is the only worldwide social information infrastructure currently open to the public. Almost all of us strongly rely on the Web as the single most important source of information for our daily life. Many businesses also rely on the Web as the preferred

L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 863–869
Torremolinos (Málaga), June 22–27, 2008

communication medium with their customers. However, all the information supplied by the Web is not raw information of the real world, but well-organized information edited manually by humans.



(a) The Internet



(b) Ubiquitous Sensor Networks

Figure 1: Comparison beween the Internet and USNs as the data transmission scheme.

On the other hand, the *Sensing Web(SW)*, supplies raw real-time information obtained by observing the real world directly with sensors. In order to realize the *Sensing Web*, we first describe information processing schemes which are realized in the Web by considering how the Web has been developed before it became the worldwide social information infrastructure. We then discuss what schemes are required for realizing the Sensing Web.

2.1 Development of the Web

The development of the Web begins with the Internet, which is constituted as the result of connecting computer networks specific for military, scientific research, university LANs, etc. with one another by sharing the same communication protocol called *TCP/IP*. The Internet can be regarded as a scheme for worldwide data transmission, because it realizes data transmission between any computers in the world, regardless of their system configurations (see Figure 1).

The Internet made it possible to develop the *World Wide Web (WWW)* for sharing multimedia information transmitted over the Internet. Since the WWW makes any multimedia information in computers connected to the Internet accessible by anyone using internet browsers (as long as the information is described in the form of HTML), the WWW can be regarded as a scheme for the world wide information sharing.

With the increase of information accessible through the WWW, *search engines*[1][2][3] and *information portals*[2] appeared so that their users can find the information necessary from that stored and accessible by the WWW. These search engines and information portals can be regarded worldwide information utilization schemes.

Since this triplet of the schemes (i) for worldwide data transmission, (ii) for information sharing and (iii) for information utilization is available, the Web spread all over the world to be the social information infrastructure.

2.2 Comparison of the Sensing Web with the Web

The USNs, which have currently been constructed in various places in the real world, correspond to the computer networks constituting the Internet. By connecting the USNs all over the world with each other, a *worldwide sensor network* is realized, which makes possible data transmission between any sensors. This sensor network plays a central role in the worldwide data transmission scheme for the data obtained from various sensors installed in our daily environments. USNs are constituted of *sensor nodes*. A sensor node is a *(sensor, computer)* pair. An Internet-like structure where each node is actually a sensor node leads to the world wide sensing web, as shown in Figure 1.

However, just enabling worldwide sensory data transmission infrastructure, does not lead immediately to a social information infrastructure. Rather, as for the Web, schemes for information sharing and information utilization are required. After such schemes are developed, the existing USNs come to play a role of social information infrastructure for various kinds of applications ranging from communication, to shopping, or information

services. Figure 2 illustrates the above discussion on the comparison between the Web and the Sensing Web.

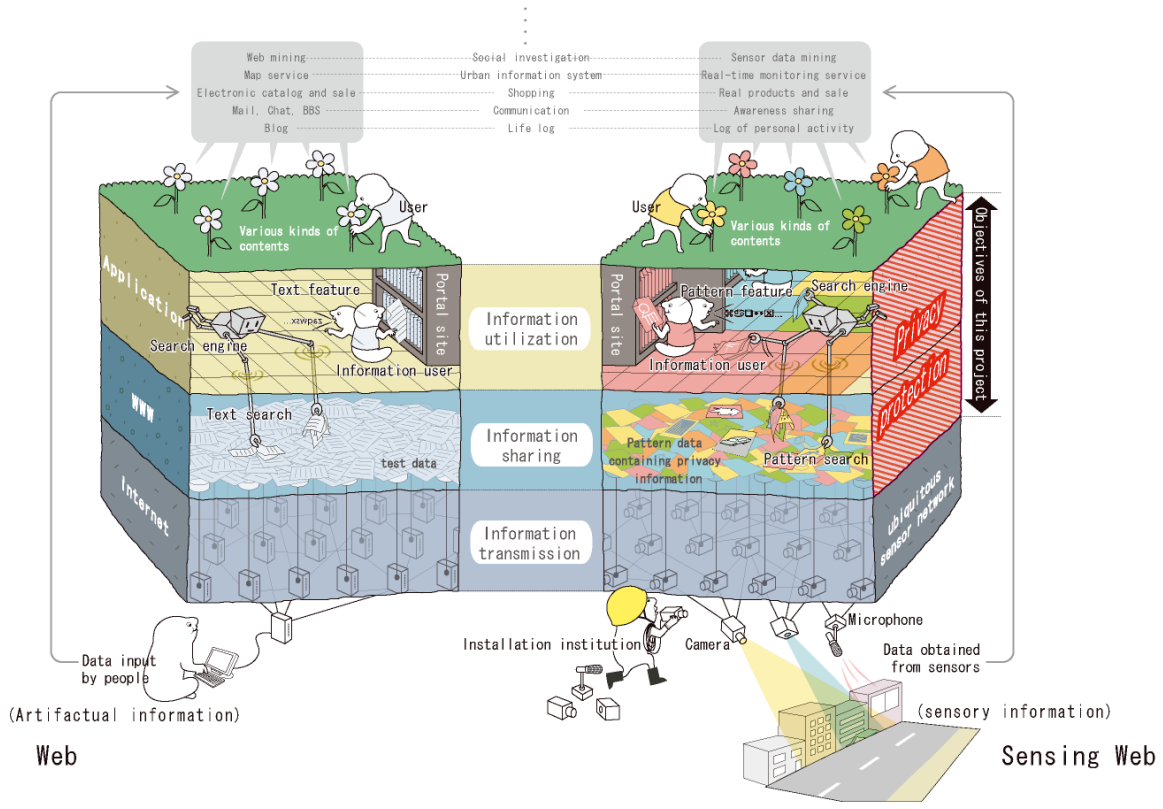


Figure 2: Comparison between the Web and the Sensing Web.

2.3 Privacy Information Management

The Sensing Web concept is similar to the Web, but exhibits important differences from it. The main difference is related to privacy issues: the information of the Sensing Web includes various kinds of privacy information related to the people observed by the sensors. The privacy issue does not arise in the Web where most of the information is entered by humans such that privacy information can be removed or otherwise protected. Unlike the Web, all the information of Sensing Web is obtained directly from sensors, and therefore there is no chance to evaluate whether the sensor information violates privacy. Therefore, a scheme for managing privacy information must be included in the sensor data, in addition to the schemes realized already in the Web (see Figure 3).

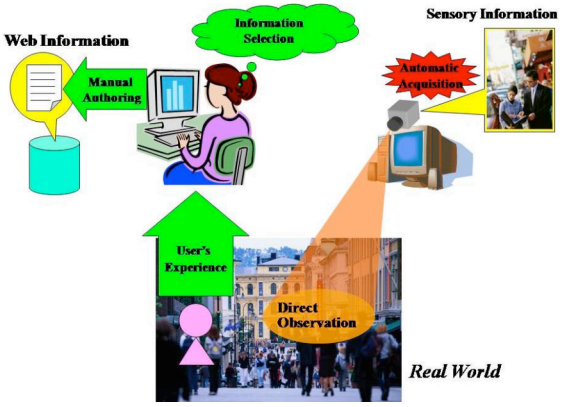


Figure 3: Difference between the Information in the Web and the Sensing Web

3 Research Issues for the Sensing Web

Characteristics of information systems can be classified by their data type and openness as shown in Table 1. Conventional information processing systems, mainly used for business purposes in office environments, deal with text-based data stored by humans. Control systems, which are used in factory environments, deal with the data acquired by various sensors automatically in the form of data streams. Both of these systems are used exclusively within a

limited institutional environment. On the other hand, the data in the Web are open to the public. However, there have not been any systems that are open and deal with pattern-based streaming data automatically acquired by various sensors. We define such systems as the Sensing Web.

Table 1: Classification of InformationSystems.

	Text-based, Manually-authored, Stored Data	Pattern-based, Automatically acquired, Streaming Data
Closed system	Information Processing System	Control System
Open system	Web	*Sensing Web*

In order for the USNs to evolve into the Sensing Web, we need to realize schemes for *information sharing, information utilization* as well as *privacy information management* for the information system characterized above. In our Sensing Web project, we will develop information technologies to realize such schemes including;

- privacy information management at sensor nodes,
- information sharing by matching information requests from the users with information services provided by sensors, and
- information integration for presenting the information obtained by various sensors in various places to the users comprehensively (see Figure 4).

In this section, we will describe the general approaches for these three research issues.

3.1 Privacy Information Management in Sensor Data Acquisition

As mentioned in section 2.3, the major difference of the Sensing Web from the Web is that the data obtained by the sensors include privacy information of the person. This point is heightened by the use of camera and microphone as sensors. In order to cope with the privacy information included in sensor data , we introduce the following two approaches:

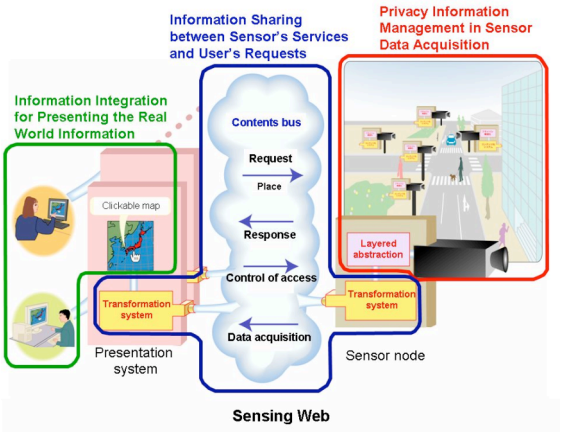


Figure 4: Research Issues for the Sensing Web.

(1) Symbolization of sensor data

The first approach is to describe the real world information by symbols free from privacy information. Although sensor data include various kinds of information, most of the real world information useful for our daily life does not related to privacy information, and can be described by a set of symbols from situation-specific vocabularies. For example, when we are interested in traffic conditions from camera images, the necessary information is not about the identity of the people walking or driving on the streets, but rather, about how many pedestrians / cars are on the street. The nature of such information is statistical, it includes no personal information, and thus it is free from privacy problems. Such information can be described by symbols from category or object dependent dictionaries (e.g. for people and cars) Although it is still very difficult to recognize in real-time people and cars in camera images robustly, approximations of number of people and cars on the streets are sufficient. The same discussion will be hold in case of audio data.

(2) Layered abstraction of privacy information

In the case of image data, the other approach is to abstract them hierarchically as shown in Figure 5. By extracting human regions from the camera images and abstracting the descriptions of those regions gradually, we can obtain layered image description with gradually reduced privacy information. Once such a layered description is obtained, sensors can be *requested to disclose their image data at the level of abstraction appropriate for the relationship between the people observed by the sensors and the user of the image data*.

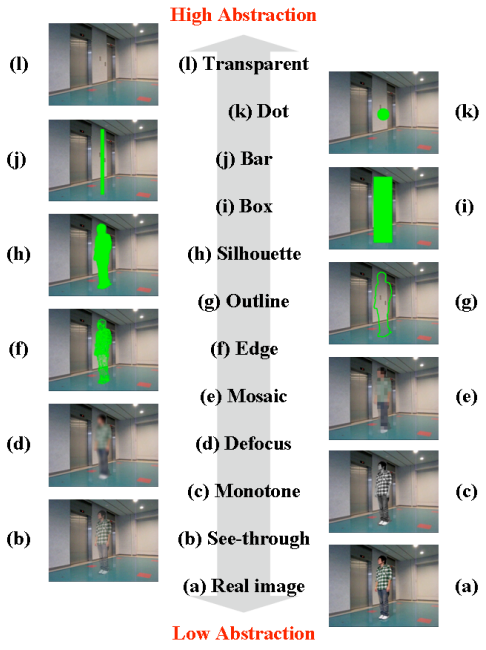


Figure 5: Layered abstraction of image data.

Extraction of human regions and degradation of the regions to cope with privacy problem were first proposed by one of the members of this project under the name of *stealth vision*[4]. We plan to introduce and extend this technique for obtaining image description based on layered abstraction of human regions. For the audio data, similar discussion is possible.

3.2 Sharing Sensory Data through Service/Request Matching

An open system requires a mechanism for matching service suppliers to service requests. Such a mechanism is in general implemented by using a *broker architecture* including its core module, the *software (or contents) bus*, which negotiates various requests and services. This approach is quite sound on architectural design, but all the difficulties are packed in the negotiation process. In the case of business applications, such as those using the Object Request Broker or the Web Services, it is not an easy task to specify requests and services on the basis of particular domain dependent business knowledge, even if a standard interface description language (IDL) is provided.

The Sensing Web will use the matching mechanism shown in Figure 6. Real world information served by the Sensing Web is however semantically much simpler than logical information of business applications. In fact, the real world information is in a single domain that involves the vocabulary obtained by symbolization through the image recognition and privacy management processes.

Several matching approaches have been used for brokering requests and services and for filtering information from enormous sources. *Lexical approaches* are used in conventional Web search engines, *syntactic approaches* are in business applications described above, and *flow model-based approaches* are often used for mining software repositories. These approaches tend to give the results of limited satisfaction, because they cannot capture precise meanings of information to match. The Sensing Web can employ a direct semantic approach identifying the names and values of each information items such as the number of people or cars, directions of movement, weather metrics, locations, etc. One of research issues in this context is a *smoothing technique* to fill in the semantic gap between user request specifications (that should be reflected in the form of final presentation) and raw (but symbolized) sensor data.

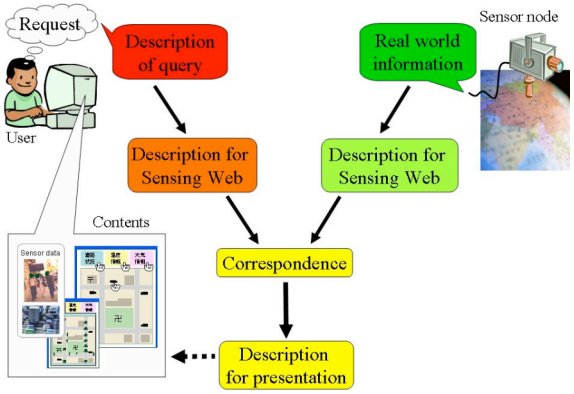


Figure 6: Matching the requests from the users with the information given by the sensors.

3.3 Information Integration for Presenting the Real World Information

As an application of the Sensing Web, we plan to employ the *Digital Diorama* technology which allows the users to view the real world from any virtual viewpoint (see Figure 7). Digital Diorama may serve a similar function to *Google Earth*[5]. Google Earth supplies the users the views of the earth from the viewpoints and scales specified by the users. However, those views do not reflect dynamic change of the real world in real time, because they are generated by synthesizing the static views from satellite images. When used in conjunction with the Sensing Web, Digital Diorama supplies the

views *synthesized from the real-time data* obtained by the sensors of the USNs.

In order to realize Digital Diorama, the sensor data obtained at various places with various kinds of sensors are integrated into a single uniform description of the real world. This process includes *interpolation of sensor data*, *association with related information obtained from other information resources* including the Web.

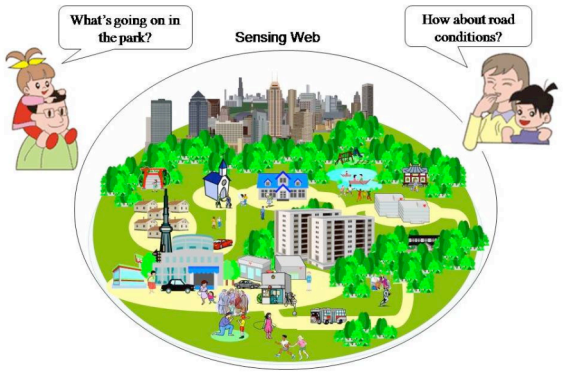


Figure 7: Presentation of integrated sensory information by Digital Diorama.

3.4 The whole Research Issues

Each sensor of the Sensing Web can only supply the information of the area in the real world. This information is represented as a point in the spatiotemporal space. On the other hand, the requirement of the real world information from the users is wide area information and/or time varying information. This information is represented as a line in the spatiotemporal space. Therefore, the whole research issue is how to convert the sensor information into the user view, i.e. how to generate lines from points in the spatiotemporal space. Since we aim to construct the Sensing Web, the architecture necessary is completely distributed processing and no communication is assumed among the sensors due to the cost of them.

In this situation, we have to consider the spatiotemporal service providers that combine the sensor information represented as a point into the line in the spatiotemporal space. This service should be realized in the form of Web services as shown in Figure 8.

When a user requests to show a view of a certain area, the system searches and finds the sensors located in the area and integrates their information into a space view. To do this process in real time, a spatiotemporal server should be attached to the area and have collected the information in advance, because it may take time to integrate the information of sensors in space due to the existence of the area where no sensors can observe. Hence, the user site can request the necessary information to the server, while the server continuously collects the information from the sensors in the area to meet the request of the user. At the user site, it is also desirable to relate the sensor information with that obtained from other information resources including the Web.

The information technologies required for solving these research issues are related to the field of pattern processing, software engineering and media processing. Thus, the Sensing Web project needs contributions by researchers in those research fields.

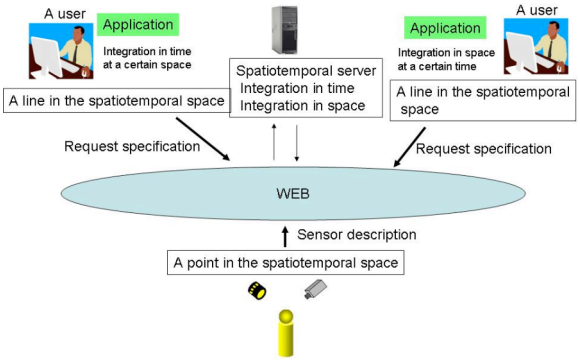


Figure 8: System architecture of the Sensing Web system.

4 Related Work

There have already been other proposals to utilize the data obtained from sensors installed in the real world. IBM constructs *Smart Surveillance System*[6], which uses many cameras to detect and track objects, for face cataloging, to index events and so on. In *Sensor Webs Project*[7] conducted by NASA JPL, many sensor nodes named *Pods*, equipped with various kind of sensors, are installed in real environments so that the sensor nodes gather the data of the environments by exchanging their data with each other. Similar research projects are also conducted by many universities such as MIT[8], Stanford[9]. However, most of these research projects use homogenous sensor nodes installed by the research organization itself for the express purpose of their research. By contrast, the Sensing Web project described here

assumes the use of the sensors in the existing USNs, which have already been installed by various institutions for various purposes, independent of the Sensing Web. This means that sensors are heterogeneous and their positions cannot be designed and controlled by us.

Microsoft Research is also running *Sense Web Project*[10], which integrates still images uploaded by the users all over the world in the description of a map. This description is similar to the Digital Diorama of the Sensing Web. However, Digital Diorama does not limit the sensor data to still images taken by humans, but it can use any kind of sensor data obtained from the existing USNs. As a result, a richer and more dynamic real world description can be obtained while privacy of information issues are also addressed.

Other projects, including *Irisnet*[11] and *HiFi*[12], are also related with the Sensing Web from a viewpoint of processing complex queries over distributed sensors. Although this infrastructure may be partially effective for the Sensing Web, they do not consider the privacy protection.

5 Conclusions

We described an overview of the Sensing Web project, which aims to open the real world information obtained by various sensors of the existing USNs installed by various institutions for various purposes. The research issues for realizing the Sensing Web are management of privacy information included in the sensor data by pattern processing, transformation from the description of information query by users into that of information services given by sensors, and information presentation by integrating fragments of the real world information obtained by sensors. Since the information technologies required to solve these research issues include pattern processing, software engineering, media processing, researchers in the fields of these technologies join the project. Although the project is just at the beginning, we will develop these technologies for opening the sensor data of the USNs in the upcoming three years. Since it is quite important to extend this research to various countries for making the scale of the Sensing Web really worldwide, we are very much interested in the collaboration with the researchers in various countries.

Acknowledgements

The Sensing Web Project, which is officially named "Social Exploitation of Sensory Information as Real-World Content," is funded by Effective and Efficient Promotion of the Coordination Program of Science and Technology Projects in the Special Coordination Funds for Promoting Science and Technology, which is conducted by Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, and Japan Science and Technology Agency (JST).

The authors acknowledge the contribution of all the members of the Sensing Web Project.

References

[1] http://www.google.com/

[2] http://www.goo.co.jp/

[3] http://www.yahoo.com/

[4] Itaru Kitahara, Kiyoshi Kogure, Norihiro Hagita, "Stealth Vision for Protecting Privacy," Proc. of 17th International Conference on Pattern Recognition (ICPR 2004), Vol.4, pp.404-407, (2004).

[5] http://www.earth.google.com/

[6] Arun Hampapur, Lisa M. Brown, Jonathan Connell, Max Lu, Hans Merkl, S. Pankanti, Andrew W. Senior, Chiao-fe Shu, and Ying-li Tian, "Multi-scale Tracking for Smart Video Surveillance," IEEE Transactions on Signal Processing, Vol. 22, No. 2, March 2005.

[7] http://sensorwebs.jpl.nasa.gov/

[8] http://sensorweb.mit.edu/

[9] http://infolab.stanford.edu/stream/

[10]http://research.microsoft.com/nec/senseweb/

[11] Phillip B. Gibbons, Brad Karp, Yan Ke, Suman Nath, Srinivasan Seshan, "IrisNet: An Architecture for a World-Wide Sensor Web," IEEE Pervasive Computing, Vol. 2, No. 4, 2003.

[12] Michael J. Franklin, Shawn R. Jeffery, Sailesh Krishnamurthy, Frederick Reiss, Shariq Rizvi, Eugene Wu, Owen Cooper, Anil Edakkunni, Wei Hong, "Design Considerations for High Fan-in Systems: The HiFi Approach," CIDR, pp.290-304, 2005.