

Spatial Media Fusion Project

Masatoshi Arikawa, Takeshi Sagara
Center for Spatial Information Science
The University of Tokyo
{arikawa, sagara}@csis.u-tokyo.ac.jp

Koji Okamura
Computing and Communications Center
Kyushu University
oka@ec.kyushu-u.ac.jp

Abstract

Most information includes some kinds of spatial data such as the address of a restaurant and the position of a person carrying a portable phone. The spatial data are useful as meta data of multimedia data because they provide spatial connections between multimedia data. The spatial data are called *spatial keys* because they join different contents with spatial relationship. The spatial media fusion project started last year to construct a framework of multimedia contents' circulation based on the spatial keys. A geographic coordinate (x,y) is one kind of spatial data, but there are other kinds of spatial data, called *spatial referenced data*, which can be converted to geographic coordinates. We particularly focus on Japanese addresses and camera parameters as spatial referenced data. Using two kinds of spatial referenced data, we integrated text data and photo/video data in the form of spatial keys.

1. Introduction

The Web integrated all multimedia data such as texts, images, voices and movies. The Web is based on the concept of *hypertext* (Fig. 1). The idea of hypertext is simple. There are only two components: nodes and links. One piece of multimedia data is considered a node, and the transition from one node to another node is realized through a link. In the Web, Universal Resource Locator (URL) serves as a link. It is said that the Web has changed everything of the computer environment. It is called the *Web revolution*. However, there are some drawbacks in hypertexts. Losing one's way in hypertexts is a well-known problem. Since links provide only *relative* relations between pieces of information, we often lose our way in the Web. To solve the problem, search engines and directory services were innovated and became popular. Both of them provide absolute spaces as key word spaces and general con-

cept category spaces. We have researched *spatial keys* which can be another solution for the problem of losing our way in cyberspaces. The spatial keys deal with real space identifiers, while URL, key words and general concept categories are all virtual space identifiers. GIS (Geographic Information System) people dream of the *Digital Earth* which integrates most contents on Internet using spatial keys (Fig. 2).

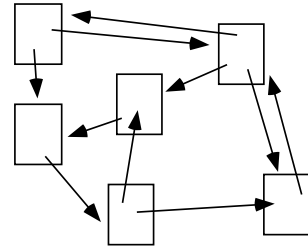


Figure 1. Hypertext integrating various data in the form of nodes and links.

Most multimedia data are related to locations in the real world. For example, we take a picture with a digital camera, then send it to a friend with the information about the location where we took the picture. Most multimedia data can be related to the locations where the data are created or their creators live. When we retrieve these multimedia data, it is useful for users to retrieve them using spatial keys. It means that we often encounter some situation in which we want some multimedia data related to our intended locations. The spatial keys are roughly defined as location data or names of places. The spatial key spaces are considered as absolute address spaces because the spatial keys are unique in our earth which is considered a real space and exists as an absolute space. In other words, all spatial keys must have their corresponding places existing in the real world. Because spatial databases are the

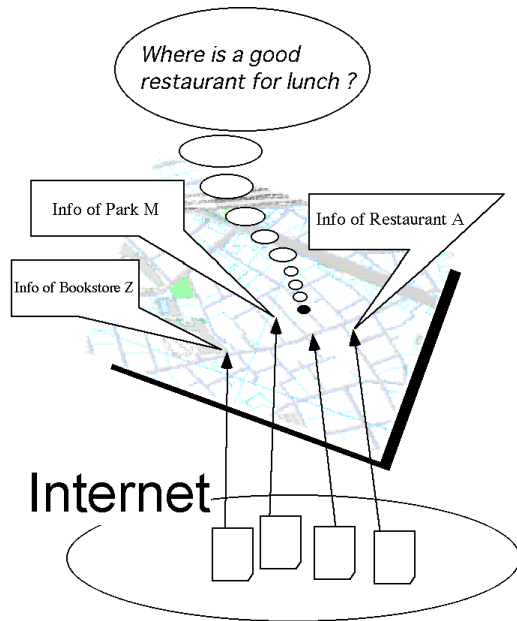


Figure 2. Spatial keys connecting/integrating various data, particularly contents on the Web.

copies of the real world, spatial data are connected to the parts in the real world. Thus, there are some links between spatial databases and the real world. Also, there are similar links between our brain and the real world. It means that spatial keys exist in our brain. Even if we communicate with other persons using natural languages, we use spatial keys such as addresses and geographic names. The spatial keys have not yet been common as access methods on Internet. For example, we cannot retrieve any Web pages about Italian restaurants near the places where we are. Spatial keys will become important components of address spaces to integrate various data including multimedia contents. We can extend this idea to time dimension, and *spatio-temporal* keys can integrate all the past, current and future contents.

2. Spatial data acquisition system from Japanese Web pages

Spatial data are usually considered as (x,y) coordinates or collections of them. Also, spatial data should be structured well as objects, relations or tables. GISs have treated these well-structured spatial data. However, there are many non-structured spatial data, par-

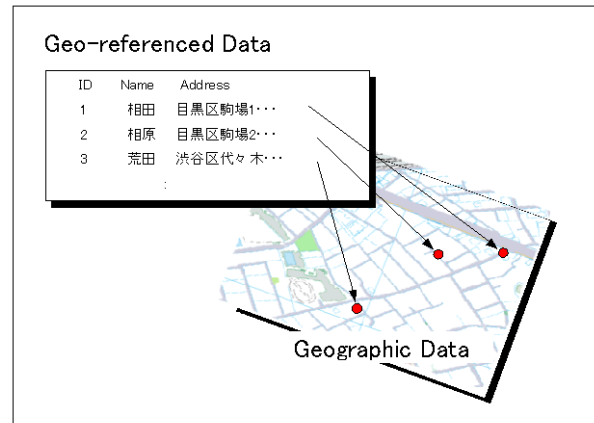


Figure 4. Geo-referenced data.

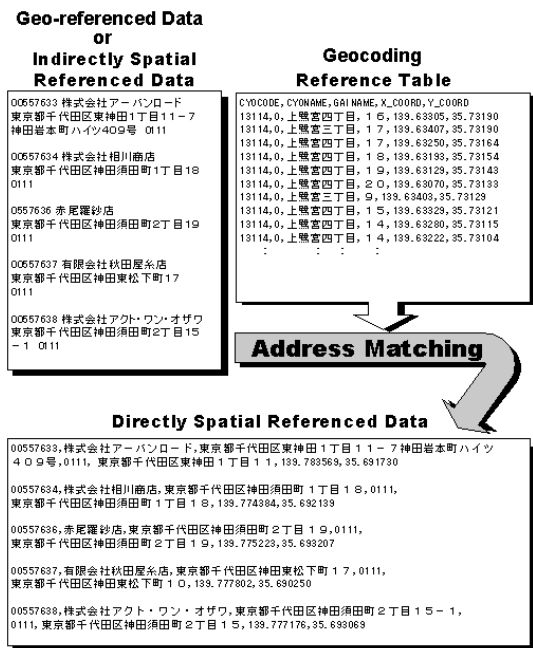


Figure 5. Address matching.

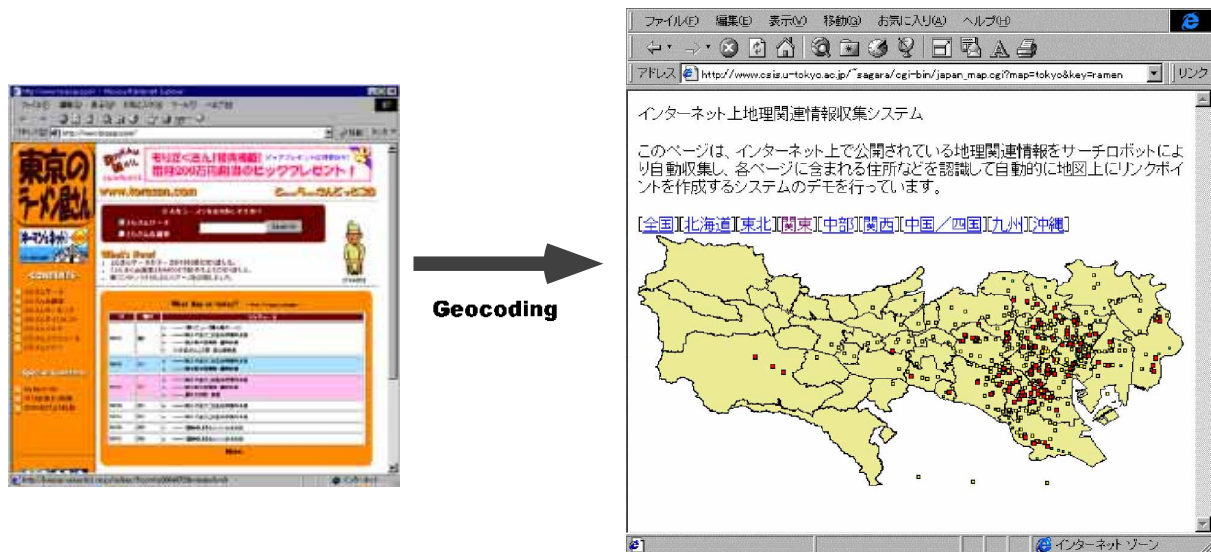


Figure 3. An automatic created map about Chinese noodles soup restaurants using both robot programs for gathering Web pages and *geocoding* programs for generating location data.

ticularly on Internet. For example, there are lots of Web pages advertising or introducing Chinese noodle soup restaurants. The Web pages are usually formed as HTML documents. It is convenient for ordinary users to browse these usual Web pages through map interfaces or geographic key words (Fig. 3). These Web pages often include *geo-referenced* data such as an address, telephone number, and zip code. The geo-referenced data are spatial data with no (x,y) coordinate data (Fig. 4). These geo-referenced data can be converted to position data using *address matching* functions (Fig. 5). It means the Web itself is a largest spatial database, and often provides fresher information than map books or map data providers. However, we cannot access most of the contents on the Web using spatial keys, because the Web is not organized as well-structured spatial databases. The conventional GIS did not focus on non-structured spatial data. We focus on making a framework to integrate the Web with a function of accessing these contents formed as non-structured spatial data. Our system was developed to realize the framework. The basic flow of the procedure of our developed system is as follows. (1)It automatically collects Web pages using robot programs. (2)It extracts segments of addresses in natural language texts or HTML documents using natural language processing technology. (3)It marks up parts of the extracted addresses in source texts or HTML documents using <spa> tags, which were proposed by us,

and will be explained in details later. (4)It converts the contents of <spa> elements indicating spatial data into (x,y) coordinates using an address matching technique. Then, the converted (x,y) coordinates are inserted in the <spa> elements as an attribute of the elements. (5)Finally, we can retrieve many Web pages including directly or indirectly geo-referenced data through map interfaces and spatial key words. The modules of (2), (3) and (4) are called “BASHO” which is the core module of our system (Fig. 7).

In treating non-structured spatial data such as Web pages including address information, a new classification of spatial data may be useful to clarify the process of non-structured spatial data (Fig. 6). Spatial data including (x,y) coordinate data are called **Directly** spatial referenced data. On the other hand, spatial data with geo-referenced data are called **Indirectly** spatial referenced data. This division is one way of classifying spatial data. Another way of distinguishing spatial data is that whether they are structured or not. **Full**-structured spatial data are formed as relations or tables. On the other hand, in natural language texts or HTML documents including spatial data, the descriptions for spatial data are not explicitly specified. The texts or HTML documents are called **Non**-structured spatial data. We also introduce **Semi**-structured spatial data which include XML elements representing spatial data. Examples of XML elements for describing spatial data are <spa>Tokyo</spa> and <spa>134,

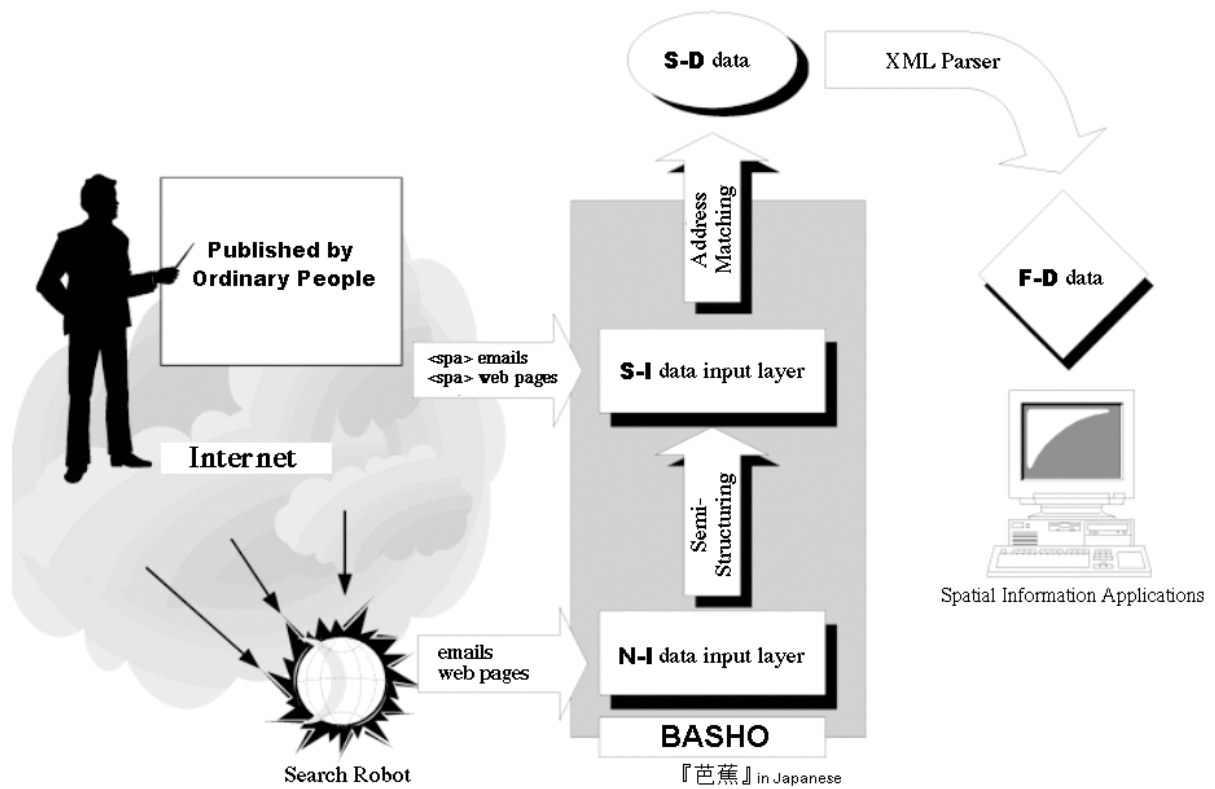


Figure 7. BASHO system. It processes non-structured and semi-structured indirectly spatial data into full-structured directly spatial data for enabling spatial access methods to text based contents on Internet.

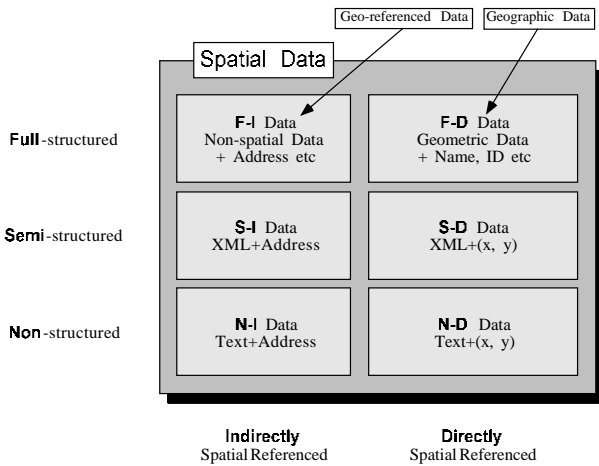


Figure 6. Category for spatial data.

45</spa>. The “spa” is an abbreviation of *spatial anchor*. There are some candidates for standards to represent spatial data as XML documents.

The conventional GIS or map data are categorized into Full-structured Directly spatial referenced data or **F-D** data. Geo-referenced data such as tables of customers’ information including their addresses are Full-structured Indirectly spatial referenced data or **F-I** data. Web pages or usual texts including address information are Non-structured Indirectly spatial referenced data or **N-I** data. Web pages or usual texts including (x,y) coordinate information, which are rare today, are Non-structured Directly spatial referenced data or **N-D** data. Examples of **S-I** data, that is, Semi-structured Indirectly spatial referenced data are XML documents including elements representing addresses. Examples of **S-D** data, that is, Semi-structured Directly spatial referenced data are XML documents including elements representing (x,y) coordinates.

Most Web pages belong to **N-I** data. The process of enabling Internet users to access Web pages through spatial keys is considered the one to convert **N-I** data to **F-D** data because only **F-D** data can be visualized as maps, and used for spatial data retrieval (Fig. 7). It is not easy to extract the parts that describe addresses from natural language texts or HTML documents. The results of extracting the parts of addresses are recorded as <spa> tags inserted in the source data. It is the process of converting **N-I** data to **S-I** data. Japanese statements have variations for representing the same Japanese addresses. Thus, it is also difficult to convert extracted addresses into (x,y) coordinates. After the process of the address matching is completed,

the generated (x,y) coordinates are inserted as the attribute values of <spa> tags in the documents. The address matching process converts **S-I** data into **S-D** data. XML parsers can convert **S-D** data to **F-D** data.

3. Networked spatial video hypermedia

It is natural for users to use real-time videos to interact with spatial databases. This kind of application of realizing the real world itself as an user interface is called *augmented reality* or AR because it adds more information to the real world. Thus, users can appreciate more information in the real world and interact with databases naturally. The basic principle of realizing augmented reality is overlapping and synchronizing video data with spatial data (Fig. 8). In Fig. 8, there are several rectangles, called spatial anchors, which correspond to the spatial data. In order to select a spatial object in the video, we only click its part in the video. Clicking the part of a spatial object is interpreted as clicking a rectangle representing the spatial object and traversing its spatial anchor so that we can obtain some information related to its selected object. This is the same procedure for clickable images in Web pages. We only extend this idea to video data. It means that if a camera moves, the view of the video also changes depending on the parameters of the camera at the same time (Fig. 9). Even if the view of the video is changed, these rectangles representing spatial anchors are dynamically generated and synchronized with the camera’s movement on the video. Thus, we can obtain a clickable live video. We implemented such spatial live video system called *Name-at*. We also incorporate functions of levels of detail (LOD) into *Name-at*. When we zoom in, the new anchors appear to provide more information with users, and when we zoom out, less important or small anchors disappear to reduce the complexity of display objects in the video. Fig. 10 shows the control panel of *Name-at*. Users can control the direction of a camera and its zoom ratio. Also, users can select geographic names from a list of them like selecting URLs. The system converts its selected geographic names into identifiers of cameras and parameters of the cameras. We can thus obtain our intended video scenes viewing real objects with users’ selected geographic names. By clicking a part of a video, we can move to a related Web page. Also, from the Web pages, we can select addresses, then change the view of the live videos we want to watch. Thus, *Name-at* achieves interactions among the Web spaces, geographic spaces and video spaces. Users can move to the contents of their interests using spatial keys which are used to connect the Web spaces, geographic spaces

and video spaces.

We also incorporate map interfaces into video interfaces. When we click a part in a map, a window shows a video showing a scene at our specified position. Fig. 11 and 12 show the demonstrations of our system integrating live video and map data. A gray area on the map represents the extent of a camera's views.

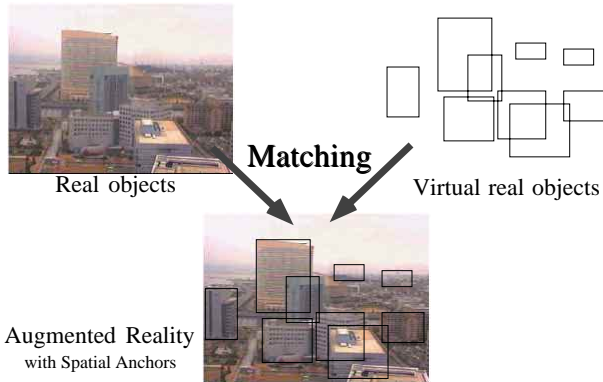


Figure 8. Augmented reality as an overlap of a video including real objects, with the graphics representing the real objects stored in spatial databases.

4. Spatial media fusion project

GIS researchers and developers are trying to extend the conventional GIS to Web GIS. Using the Web GIS, we can appreciate geographic data through usual Web browsers. Their approaches are considered map-enhanced Web applications. The difference between our approach SMFP, that is, spatial media fusion project, and Web GIS is that we use map data as a basis of integrating various kinds of multimedia data. In other words, we do not use maps as visual interfaces, but we use geographic data as spatial dictionaries to connect and convert spatial data, and to make links between various multimedia contents. For example, when we walk in the real world using augmented reality interfaces, we retrieve some information of the place we are now through spatial keys. Furthermore, we can retrieve video data from the spatial keys which are geographic addresses as the parts of Web pages and ordinary texts. Thus all media data are related to the position, lines or areas in the real world. It is convenient for ordinary people to access all Web contents through the spatial keys. This concept is generally called *digital earth* in the field of GIS. We believe

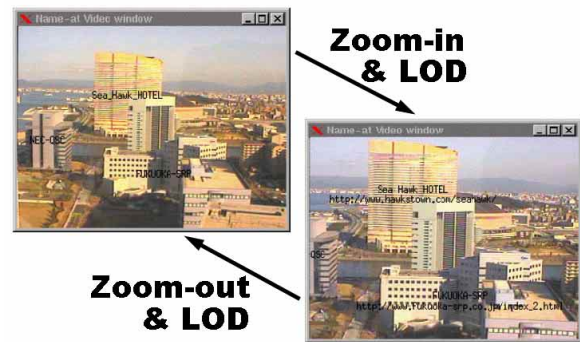


Figure 9. A prototype *Name-at* which provides spatial anchors connected to the Web space through the video space, and controls the number of spatial anchors using the rule of levels of detail (LOD).

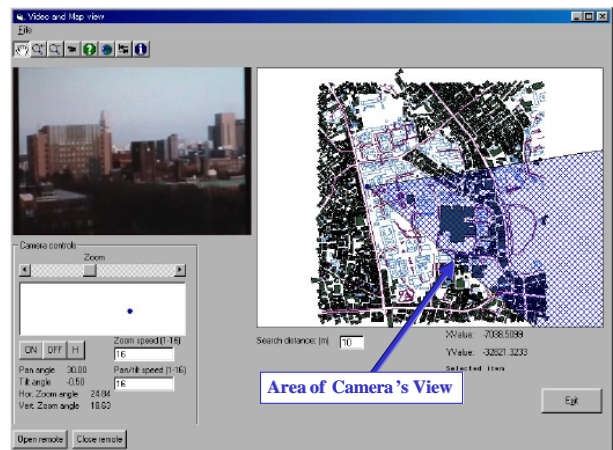


Figure 11. Integration of a real video and a large scale map(1). The area of the camera's view is displayed on the map.

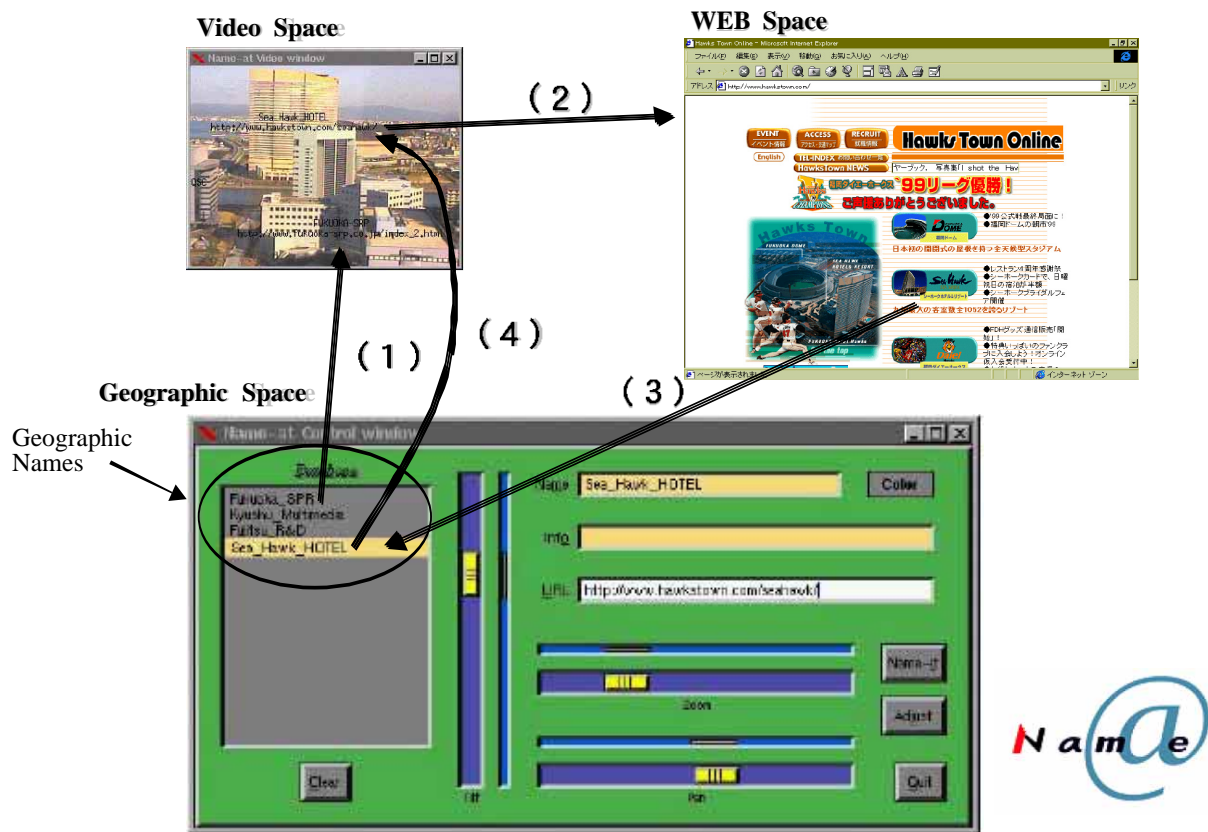


Figure 10. An example of spatial media fusion. Interactions among Web space, geographic space and video space. For instance, (1) if we select the name of a building from a list of geographic names, we can obtain the video showing the building we selected, (2) if we click a part of the video, the corresponding Web page appears, (3,4) if we click an anchor in the Web page, the corresponding video can be selected through an geographic name space extension to URL.

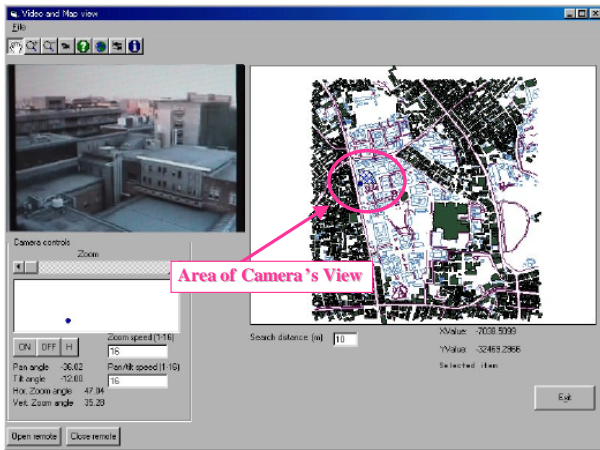


Figure 12. Integration of a real video and a large scale map(2). The camera faces to near buildings.

the digital earth is the next significant paradigm of Internet, following the Web, search engines and directory services.

5. Concluding remarks

This paper presented our project, spatial media fusion project. It enables the integration of all multimedia data on the Web using spatial keys. It is possible to provide ordinary users with a tool to access all kinds of multimedia data through the location on the earth. We are now developing a prototype system and are constructing our philosophy. We also try to clarify the process of human's understanding of information in view of cognitive map. The process of understanding things can be regarded as making some cognitive map in our brain. Thus a spatial concept is not special but rather essential. Before computers, papers, texts and natural languages were innovated, our ancestors survived the real space using our spatial ability.

Acknowledgements

This research was supported in part by "Research for the Future" Program of Japan Society for the Promotion of Science under the Project "Advanced Multimedia Contents Processing" (Project No. JSPS-RFTF97P00501). We would like to thank Mr. Michihiro Murao and Mr. Thorvald Boe for their cooperation in developing ideas and prototype systems of *Name-at*.

References

- [1] M. Arikawa, "Spatial Hypermedia as Augmented Reality Based on Spatial Information Bases," Advanced Database Research and Development Series - Vol. 9, Advanced Database Systems for Integration of Media and User Environment '98, World Scientific Publishing Co. Pte. Ltd., 1998, pp. 9-14.
- [2] R. Azuma, "A Survey of Augmented Reality," in *ACM SIGGRAPH '95 Course Notes No.9 - Developing Advanced Virtual Reality Applications*, August 1995.
- [3] M. Murao, M. Arikawa, K. Okamura, "Networked Augmented Spatial Hypermedia System on Internet," *Advances in Visual Information Management, Visual Database Systems*, Edited by Hiroshi Arisawa and Tiziana Catarci, Kluwer Academic Publishers, 2000, pp. 239-253.
- [4] K. Okamura, M. Arikawa, Y. Yoshimura, M. Murao, "Virtual Video Frameworks for Generic Video Applications on Internet," the proceedings of the 2001 Symposium on Applications and the Internet - Workshops (SAINT 2001 Workshops) to be held 8-12 January 2001 in San Diego, California, USA, (to appear).
- [5] <http://www.real.com/>
- [6] Spatial Media Fusion Project, CSIS, Univ. of Tokyo, <http://smfp.csis.u-tokyo.ac.jp/>