

Server-Side Web Browser with Use of Videophone

Makoto OKITA^{***}, Hideyuki TSUGANE^{*}, Norio YAMAGUCHI^{**}, Takayuki SHIGEMATSU^{*}
Osamu TAKAHASHI^{**}, and Eiichi MIYAMOTO^{**}

^{*}HEWLETT PACKARD Japan LTD.

7, Gobancho, Chiyoda-ku, Tokyo, 102-0076, Japan, okita@jpn.hp.com

^{**}FUTURE UNIVERSITY-HAKODATE

116-2, Kameda Nakanochi, Hakodate-shi, Hokkaido, 041-8655, Japan

ABSTRACT

As mobile phones have become increasingly advanced in recent years, it became more intense in users' demand for viewing web sites designed for PC's on their mobile phones. However, since a standard PC web browser requires exclusive applications, the mobile phone will need the application environment. Furthermore, a certain approach is required since mobile phones have limited system resources, network bandwidth, screen size, and operation forms.

In order to be able to view a PC web page on a mobile phone without an additional application environment, research has been done to develop a server-side web browser in which the web browsing capabilities exist on the server side, while the user operates and views the web pages via a videophone client (server conversion-type web browser). This document describes a way of converting web pages to video and streaming them to videophones, viewing and navigating the PC web sites with a small screen, operating numeric keypads of a mobile phone, and selecting hyperlink systems. It also explains the actual implementation and the evaluation of this system.

1 BACKGROUND

Almost all mobile phones in Japan become standard with a web browser, and more than 80 percent of users browse websites with mobile phones. However, most of these mobile phone browsers can only display web pages that have been specifically made for mobile phones (in CHTML[1], XHTML[2], WML[3], etc.) and cannot display the majority of PC web pages.

Many people wish to view PC web pages on their mobile phones. According to the survey[4] conducted in May 2005, more than 60 percent of the respondents answered that they would like to do so. Recently, mobile phones with preinstalled PC web browsers, and browser applications that can be additionally installed to existing phones, are starting to appear in the market. However, the mobile phones that have these capabilities are limited since those applications need to be installed on the handset devices.

Considering this point, the 'server conversion-type web browser' has emerged. This system converts web pages into video on the server and sends it to users utilizing the

videophone feature embedded on most 3G[5] mobile phones. With the fact that about a quarter of the whole mobile phones in use as well as over 70% of the shipment of the new mobile phones in Japan (as of March 2006) are embedded with a videophone feature, it allows many mobile phone users to browse PC web pages. In this document, we will discuss the mechanism of this system, the results of its implementation, and the subsequent operational testing.

2 RELATED WORKS

There are three types of web browsers for mobile phones:

- Client applications such as Opera[6] and NetFront[7].
- Server/client applications such as jig[8], SiteSneaker[9], ibisBrowser[10], Scope[11], and weview[12][13].
- Server conversion-type applications such as jig browser web[14].

Fig.1 shows examples of display screens browsing with each method.

For client applications like Opera, as few features as possible are included in the browser; an approach is taken to operate per client. Thus, a special server is not required. However, since clients need large resources and the program size up to 1MB, it is necessary to design mobile phones that accommodate these requirements.

For server/client applications, such as jig and Scope, the server takes a part of the processing work of browsing; an approach is taken to minimize the number of clients. Internal processing, such as session management and page analysis performed by the server, enables the clients to customize the users' interface in function and minimizes the application size to under 100KB on existing mobile phone devices.

However, both web browser of client applications and server/client applications require additional client applications. Mobile phones which have no additional applications to execute this function cannot be utilized.

For server conversion-type applications of jig browser web, the system converts the web contents designed for PC's to the mobile phone's standard browser format such as CHTML or WML on the server. This does not require the installation of any additional client applications, however, since the display width of a page is fixed and it is limited to vertical scrolling, the original page layout would be lost.



Fig.1 Examples of screenshots

As a way of viewing the display of the server on mobile phones, there have been studies such as ‘considering thin-client system for mobile phones’[15], iVNC[16], J2ME VNC[17], and ubiquitous viewer[18]. Those require customized applications on the client side, since it is necessary to transfer and display the differences of the bitmap information in real time when the bitmap is changed on the server-side display screen.

As one way of converting HTML into video, there is ‘multimedia provider system, multimedia conversion server, and multimedia terminal’[19]. In this study, it says that ‘the data of the HTML image can be converted to video and videophone screens through a streaming device’, however, it does not describe its processing mechanism of converting HTML to video in detail such as the way of viewing web pages larger than the screen size, the method of selecting hyperlinks on videophones, or the mechanism needed for web browsing.

3 PROPOSAL OF THE SERVER CONVERSION-TYPE WEB BROWSING MECHANISM FOR VIDEOPHONES

In this document, we propose a new method of the server conversion-type web-browsing mechanism, which does not require any customized applications from the clients, by using mobile videophone handsets.

This method allows the user of a 3GPP standard videophone to browse PC web sites without special applications. It does not depend on proprietary specs of handsets or mobile operators. Moreover, since the handsets and servers are linked through circuit-switched connections (32Kbps or 64Kbps), a packet switched infrastructure at a high bandwidth is not necessary. Therefore, a stable action speed unaffected by network congestion can be obtained.

Here is a list of issues to be surmounted to browse PC web sites from videophones:

(1) Basic Architecture

- (1-1)The server architecture that allows circuit-switched connections through the mobile

videophone functions, call controls, and video streaming

- (1-2)The method of selecting the initial web page for viewing
 (1-3)The client identification system to control handset sessions on the server
 (1-4)The method of converting web pages to video data
 (2) The method of viewing operation
 (2-1)The method of viewing web pages in a limited display area (176×144pix) and operating numeric keypads (DTMF signal [20]) from handsets
 (2-2)The method of selecting hyperlinks
 (3) The method of creating a video menu
 (3-1)The method of dynamically making a video menu screen

How to cope with these issues above will be described in the next section.

3.1 Basic Architecture

- (1-1) The server configuration system

Fig.2 shows the basic structure of the server conversion-type web-browsing method proposed in this document. The media server, used for the mobile phone media distribution, can also perform the call-control processing, the receiving processing of DTMF signals, and the video streaming processing.

With 3G-324M G/W, since mobile phones and media servers are connected through circuit-switched networks, the connections with telephone lines and the protocols (SIP[21] and RTP[22]) of media servers are converted into the 3G-324M protocols[23] for videophones.

The rendering server converts the information to video data by accessing the target web site, and creating and processing the image data of the web page.

- (1-2) The web page selection system

The media server identifies phone numbers (incoming call numbers) and sub-addresses (numbers after asterisk) and selects which contents of the web pages should be displayed first. Each sub-address should be pre-assigned

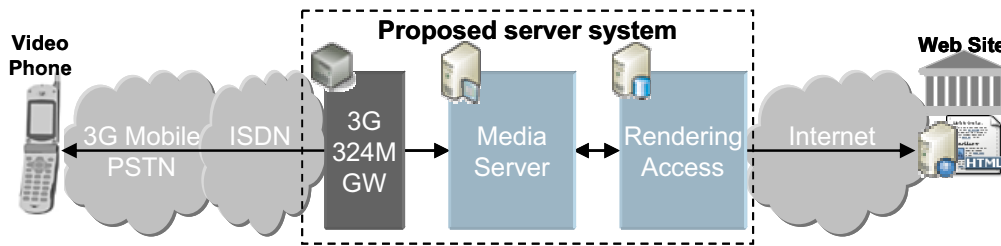


Fig.2 Proposed server rendering system for videophone

to an arbitrary unique number which maps to the URL of corresponding web pages.

(1-3) The terminal identification system

The media server carries out the session management by using mobile phone numbers (the incoming call numbers) and simultaneously processing into multiple terminals.

(1-4) The method of converting web pages to video data

The rendering server has all the functions required for web browsing. In order to avoid the complicated implementation of the web browser, the browser function on the server accesses the web site and converts the display screen to image data. In addition, this function will overlay several static images and convert them into video data. Since the rendering processing requires a heavy load (refer to the section 4.1), in this method we distributed the load between the rendering server and the media server.

3.2 The Method of Viewing Operation

Most web browsers for mobile phones work with key operations in which the page viewing and the link selection are combined. However, in the proposed method of creating videos from static images would slow down the process of displaying selected links. To solve this issue, page viewing and link selection are divided into separate modes to enable the user to view and operate each mode by changing it with the DTMF signal (0 key).

(2-1) The method of web page viewing operation

In the viewing mode, simple key operations enable the enlarging and minimizing the display area and scrolling in both horizontally and vertically.

(2-2) The method of link selection

In the selection mode, all hyperlinks on the screen are surrounded by a rectangle frame and marked with a two digit number, starting with 01. By pressing the target number, the selection of the link is done (Fig.3). When a link is selected, the web page corresponding to the link is displayed. This enables the user to select a link with one video process.

The web browser of a PC contains the link coordinate information (X-axis, Y-axis, the height, the width) and the URL of the link inside of it, so that links can be selected

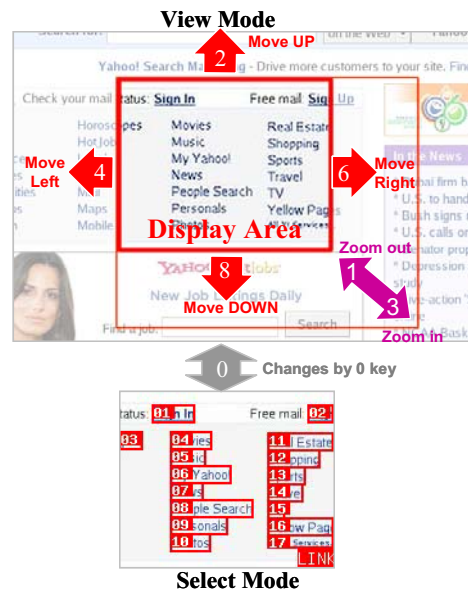


Fig.3 Proposed page browsing method

correctly with a pointing device like a mouse (Fig.4). In our proposed method, these data is extracted and kept as the link information.

3.3 The Method of Making Video Menus

In the interactive mode on the videophone, it is necessary to prepare a video menu and VoiceXML[24] for call-controlling in advance. Because of this requirement, it is difficult to dynamically create a menu of video data. As a solution, we propose the method of creating a video data for videophones by making the HTML.

Compared with the existing web browsing method described in the section 3.2, the following issues need to be solved when creating a video menu.

- Since there is no link display, key entry is the only way of selection.
- Call controlling corresponding to the key entry is necessary by VoiceXML.
- Users will expect the background music.

(3-1) The method of making the video menu

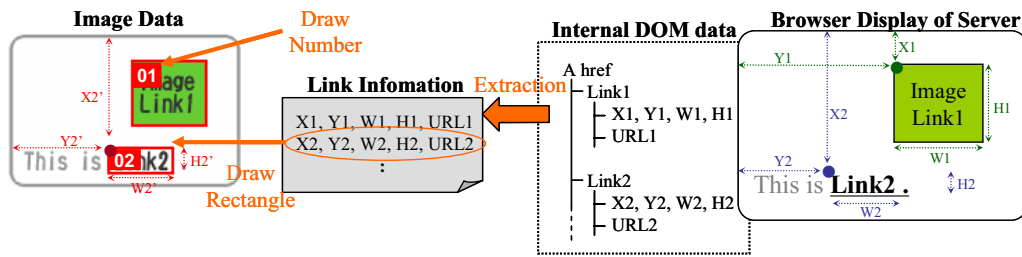


Fig.4 Extraction of link information

In this proposal, we define HTML tags as shown in Fig.5 to allow the coexistence of a video menu and a browser mode.

- A) Web browser can switch between the menu and the browser with META tags. If the META tag NAME='ocmpmenu' is on, it acts as a menu. Moreover, video data extended or reduced by the content value is created according to the number of the pixels displayed (Fig.6).
- B) Music data (WAV) defined by 'embed' in the video data is included in the created video data.
- C) After receiving a key entry defined by 'accesskey' (DTMF signal), the call control will be carried out by defining VoiceXML as specified by 'href' tag.

```

<META NAME=" ocmpmenu" CONTENT=" 70%" > ... A)
<embed src=" bgsound.wav" > ... B)
.
<a href=" key1.vxml" accesskey=" 1" > ... C)
<a href=" key2.vxml" accesskey=" 2" >
.
    
```

Fig.5 Extended HTML sample for video menu

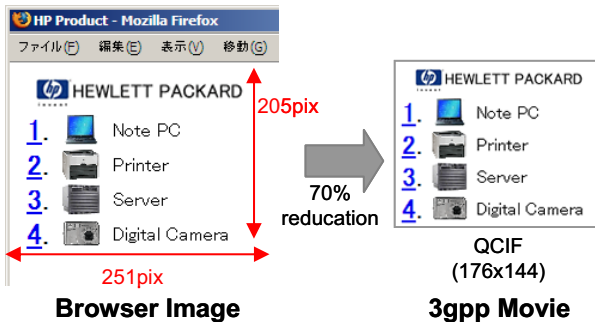


Fig.6 Convert HTML to Video menu

This enhancement enables the dynamic creation of HTML and VoiceXML data through the web application server. Furthermore, it enables to dynamically create a customized menu like a portal site.

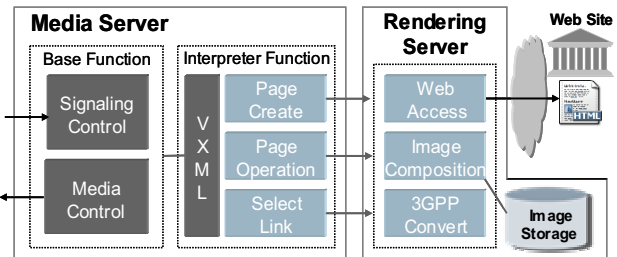


Fig.7 Server function

3.4 Processing Mechanism

Fig.7 shows the server-function configuration of the server conversion-type web browser. Fig.8 shows the processing sequence for displaying the initial page.

The media server has a base function, consists of the Signaling Control which carries out the call controlling, and the Media Control which streams video. In addition, the media server has an interpreter function and uses the logic to process actions, such as the web page creation (Page Create), the movement and zooming of the display area (Page Operation), and the display and the selection of links (Select Link).

The rendering server includes the functions below to access web sites and create video data. Fig.9 shows the basic configuration and the processing procedures.

- (a) Web Access: Initiates web browsing, acquires web pages, captures the browser image of the rendering web pages, and creates an image data. Furthermore, it extracts the linked information contained in the web browser (the coordinates of hyperlinks and URL links).
- (b) Image Composition: Calculates the next display area from the currently displayed area of the whole web page, and a partly cropped image data is created. Moreover, when displaying links, it also draws rectangle frames and assigns numbers to indicate the presence of links using the coordinate data of the link information.
- (c) 3GPP Convert: Converts the cropped image data above to MPEG video data in the 3GPP file format. With this mechanism, to create the video data for 2 seconds with 7.5fps, the video data in the mpeg4 format is created by repeating fifteen identical jpeg

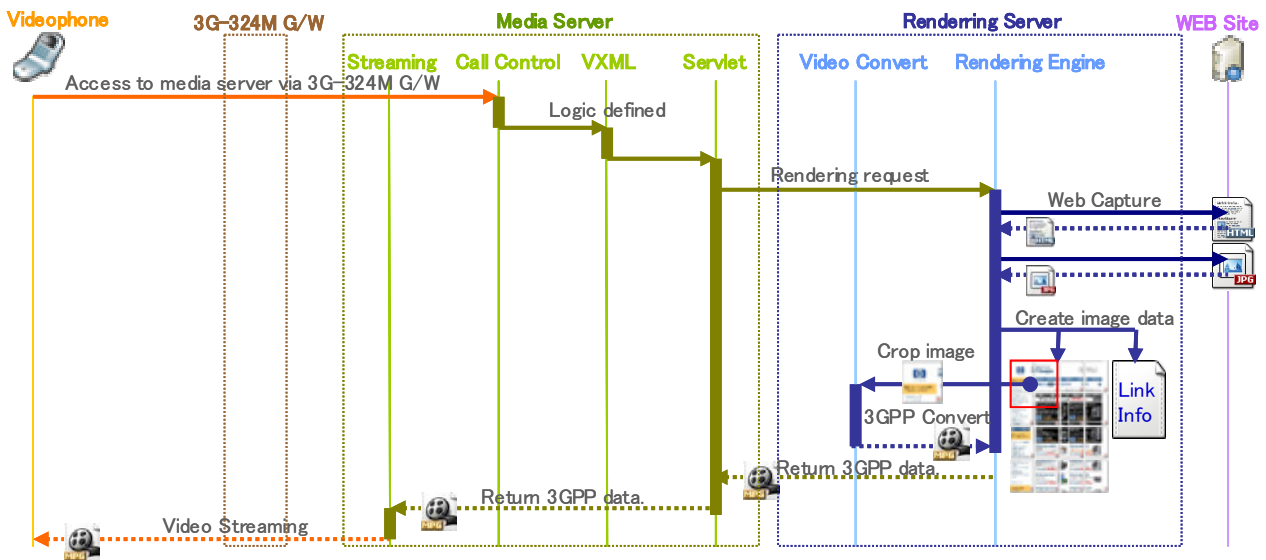


Fig.8 Processing mechanism

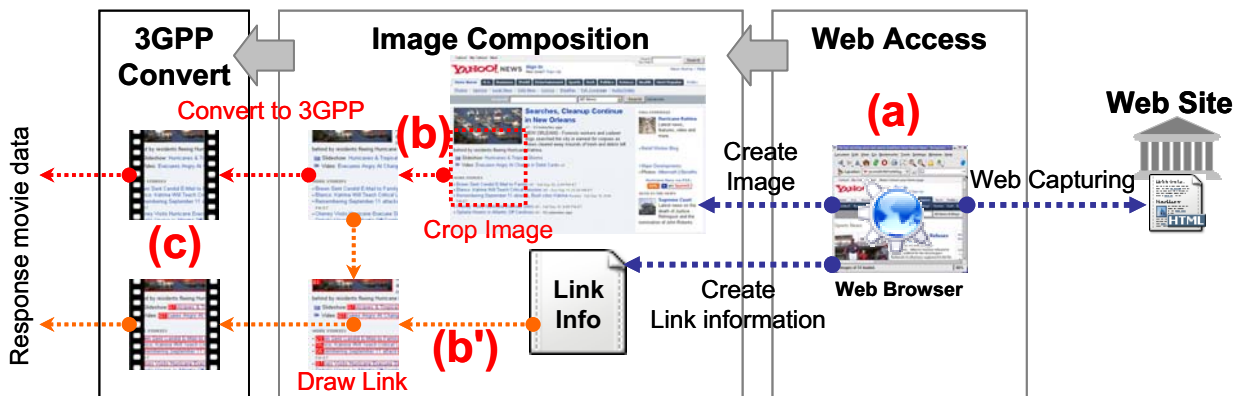


Fig.9 Rendering server mechanism

data. The media server then streams the video image to the client by repeatedly playing the video data.

When the media server receives the DTMF signal, the logic defined by VoiceXML is initialized, and the rendering server processing is executed. The video data is received and streamed to the videophone as a result of this process. The logic of the media server and how this relates to each function of the rendering server will be described in the next section.

- Page Create: Executed when the web page is first displayed. Here the Web Access, Image Composition, and 3GPP Convert of the rendering server are executed. Fig.9 (a)-(b)-(c).
- Page Operation: Executed when the vertical or horizontal movement in the display area and zoom in or zoom out operation are performed. Here Image

Composition and 3GPP Convert of the rendering server are executed. Fig.9 (b)-(c).

- Select Link: Executed when a link is selected. Here the link drawing process in Image Composition and 3GPP Convert of the rendering server are executed. Fig.9 (b')-(c).

4 IMPLEMENTATION AND EVALUATION

The actual implementation of the above system was done to confirm its operation. The media server required OpenCall Media Platform3.1(OCMP3.1)[25] on the Linux(RedHat Enterprise Linux AS3.0). After extending the “weview” system as a way of creating images of the web pages and extracting the link information, the rendering server required the followings: PHP and C++ on



Fig.10 Screenshots of proposed browser

the Linux(Fedora Core3), and khtml[26] of KDE3.3 as web browser. In addition, the video conversion involved ffmpeg[27] and mp4creator[28], and PDF conversion involved xPDF[29]. The mobile phones used for this experimental development was a FOMA[30] 3G mobile phone from NTT DoCoMo.

The following are the steps to use the browser via a video phone. Fig.10 shows the display screens on the mobile phone used for this testing.

- (1) When a user dials a particular phone number, the web page that corresponds to the number is displayed on the screen (Fig.10 First view). The cropped portion of the actual web page on the upper left-hand side is displayed.
- (2) By pressing the 2, 8, 4, or 6 key on the numeric keypad, the user can move up, down, left, and right. (Fig.10 Move down). By pressing the 1 or the 3 key, the display area can be zoomed in or zoomed out (Fig.10 Zoom out). These simple key operations allow the user to browse a web page with a videophone.
- (3) The user can select links by pressing the 0 key which activates the link view mode (Fig.10 Link view). Pressing the 0 again returns to the browsing mode.
- (4) By entering the assigned number of a link displayed in the link view mode, the corresponding page for the selected link is displayed. Once the new page is displayed, the user can repeat the operation from the step (2).

The key operations described above enable users to browse web pages for PC's on a small screen of a mobile phone and select links by using its limited number of keys (numeric keys only).

4.1 The Execution Performance and the Processing Capability

We measured the execution speed by using HP DL 380 (Xeon3.0GHz DUAL, Memory 4GB) on the media server and the rendering server. With this evaluation we obtained the acceptable action speed: It takes 10 seconds for processing the initial page access (Fig.10 First view), 3 to 4 seconds for moving (Fig.10 Move Down), enlarging, minimizing (Zoom out), and switching to the link mode



Fig.11 Screenshots of various videophone

(Fig.10 Link view). There is an approximately 1 second delay on the videophone circuit switch, which causes an approximately 2 second delay from receiving the input to showing the result. The capability of the media server is 50 channels, and the rendering server can process approximately 15 requests simultaneously within 10 seconds (1.5cases/second).

4.2 Evaluation of the Typical Handsets

In this method, since videophones are connected through the circuit switching, most handsets or mobile operators can use these videophone functions. Therefore, this method can be used with various users from all over the world. An action testing was done on Vodafone's 3G handset 702NK¹ and 703SH¹, and on DoCoMo's FOMA handset P2101V¹, m1000, F2102V and SH902i. Fig.11 shows the display

¹ Since there is no function to send DTMF signaling on videophone, the test was done by voice input instead.



Fig.12 Screenshots of multilingual WEB pages and a PDF file

screen using this function. With this method, we confirmed that it is possible to view the display screen using various handsets with videophone capability, including some old models.

4.3 The Display of Multi languages and Documents

In our method, the data is converted to images and video on the server. This makes possible to expand contents on the display screen by adding the functions on the server. For example, multi-language web pages can be displayed by adding fonts on the server.

Additional enhancement was implemented for a PDF contents to convert it to image, after the PDF downloading was done on the server, and to the video displayed on the videophone. Fig.12 shows display examples of multi-language and PDF contents.

5 CONSIDERATION

Table.1 shows the function comparison of the browser by each method. For the client system such as NetFront, the program size needs to be large enough to almost do the same functions as a PC browser. For the server/client system such as “jig”, it is difficult to implement the JavaScript functions because of the limited program size.

For the server conversion-type system, the applicable functions are limited since the client’s system depends on the existing mobile browsers. In addition, in these methods the use of the applicable fonts (size and type) on CSS is limited, since only those fonts embedded on the handsets can be used. On the other hand, the proposed method uses PC browsing on the server and it is possible to display the same image of the PC browser.

With an intensive evaluation, the following issues of the proposed method were identified.

- A) The user may not be able to view some web pages, because there is no method to input URL.
- B) Some handsets do not support DTMF.
- C) FORM input and JavaScript cannot be processed.
- D) The browser operation is slow due to the delay of circuit switch.
- E) It is difficult to display the whole page, because the display screen QCIF(176x144) has limited space.

As a solution for the issue A), we’ve implemented the function to display a specific web page by sending e-mail. If an e-mail is sent to the server with a phone number as the title and a URL as the body text the server places a call to the phone number, and the web page of the URL is displayed when the call is answered.

B) to E) are the remaining issues for the future enhancement.

Table.1 Comparison of basic browsing method and function

Processing method	Client	Server-Client	Server conversion for mobile browser	Server conversion for videophone
Reference application	NetFront	jig	jig browser web	Proposed method
Client implementation	Native App	Java App	Mobile Browser	Videophone
Transmission method	Packet based	Packet based	Packet based	Circuit based
Browser Function	HTML	○	○	○
	CSS	△	△	×
	Frame	○	○	×
	JavaScript	○	△	×
	Form	○	○	○

○: Supported, △: Partially Supported, ×: Not Supported

6 SUMMARY

The server-side web browser with the use of a videophone described in this document performed well. Since the input with voice is always possible on the videophone, a new operating method by voice recognition of keywords, which does not require any keystrokes, is under investigation. The implementation of another new function with image recognition[31] is also in progress. This function displays a web page corresponding to the image taken by the camera on the videophone. It makes possible to display the related web page by simply taking a picture of a logo or landmark (Fig.13). Furthermore, a future enhancement of the browsing method on a videophone is being considered not only by converting HTML and PDF, but also by converting various types of document, sound, and video data. Our focus now is to enhance the server function to improve the usability.

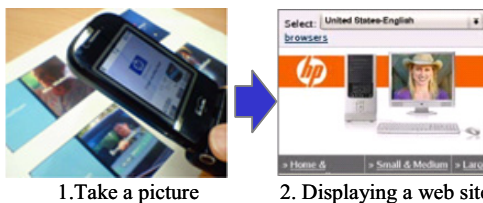


Fig.13 Choosing a web site using the video feature

REFERENCES

- [1]W3C, Compact HTML for Small Information Appliances, <http://www.w3.org/TR/1998/NOTE-compactHTML-19980209/>
- [2]W3C, XHTML 1.0 The Extensible HyperText Markup Language (Second Edition). A Reformulation of HTML 4 in XML 1.0. W3C Recommendation 26 January 2000, revised 1 August 2002, <http://www.w3.org/TR/xhtml1/>
- [3]Wireless Application Protocol Forum, Ltd., Wireless Markup Language Specification, <http://www.openmobilealliance.org/tech/affiliates/wap/wap-191-wml-20000219-a.pdf>
- [4]Nikkei Business Publications, Inc., From "Full browser cellular phone" questionnaire survey , <http://itpro.nikkeibp.co.jp/free/TIS/keitai/20050608/162330/>
- [5] Third Generation Partnership Project, <http://www.3gpp.org/>
- [6]Opera software, Opera for Mobile, <http://www.opera.com/products/mobile/>
- [7]ACCESS CO. LTD, NetFront, <http://www.access.co.jp/english/products/nf.html>
- [8]jig.jp, jig browser, <http://br.jig.jp/pc>
- [9]Ubiquitous Entertainment Inc., SiteSneaker, <http://www.uei.co.jp/sitesneaker/>
- [10]IBIS INC., ibisBrowser, <http://www.i-products.net/browser>
- [11]Programmers' Factory, Inc., Scope, <http://www.programmer.co.jp/scope.shtml>
- [12] Makoto OKITA, Norio YAMAGUCHI, Takayuki SHIGEMATSU, Osamu TAKAHASHI, Eiichi MIYAMOTO, Development of Server Client WEB Browser for Cellular Phone, IPSJ General Conference, pp.397-.398, March. 2005.
- [13] Makoto OKITA, Norio YAMAGUCHI, Takayuki SHIGEMATSU, Osamu TAKAHASHI, Eiichi MIYAMOTO, Implementation and Evaluation of WEB Browser for Cellular Phone using Server Client System, Special Issue on Multimedia, Distributed, Cooperative and Mobile (DICOMO2005), pp.733-736, July.2005.
- [14]jig.jp, jig browser web, <http://bw.jig.jp/pc/?p=1>
- [15] Tatsuo TAKAHASHI, Osamu TAKAHASHI, Tadanori MIZUNO: "A Study of a Thin Client System for Mobile Computing", Information Processing Society of Japan thesis magazine, Vol45, No.5, pp.1417-.1431(May. 2004).
- [16]PartySpirits, "iVNC", <http://ivnc.ptysp.jp/>
- [17]J2ME VNC, "J2ME VNC Client", <http://j2-mevnc.sourceforge.net/>
- [18]TOSHIBA, "Ubiquitous viewer", http://www.toshiba.co.jp/about/press/2005_01/pr_j1801.htm
- [19] Japanese Patent Office, P2001-168923A
- [20]ITU-T, Multifrequency push-button signal reception, <http://www.itu.int/rec/T-REC-Q.24/en>
- [21]IETF, Session Initiation Protocol, <http://www.ietf.org/html.charters/sip-charter.html>
- [22]IETF, RTP Profile for Audio and Video Conferences with Minimal Control, <http://www.ietf.org/rfc/rfc3551.txt>
- [23]3GPP, 3GPP TR 26.911, <http://www.3gpp.org/ftp/Specs/html-info/26911.htm>
- [24]W3C, Voice Extensible Markup Language (VoiceXML) Version 2.0, <http://www.w3.org/TR/2001/WD-VoiceXML20-20011023/>
- [25]HEWLETT PACKARD, HP OpenCall, <http://devresource.hp.com/drc/topics/OpenCall.jsp>
- [26]KDE, KHTML - KDE's HTML library, <http://developer.kde.org/documentation/library/kdeqt/kde3arch/khtml>
- [27]FFMPEG Project, ffmpeg, <http://ffmpeg.sourceforge.net/index.php>
- [28]MPEG4IP, mp4creator, <http://mpeg4ip.sourceforge.net/docs>
- [29]xPDF Project, xpdf, <http://www.foolabs.com/xpdf>
- [30]NTT DoCoMo, FOMA Products, <http://www.nttdocomo.co.jp/english/product/foma/>
- [31] Clementec Co.,Ltd., M4 Service, http://clementec.com/service_m4.htm