(54) Title: GESTURE/SPEECH INTEGRATED RECOGNITION SYSTEM AND METHOD

(57) Abstract: Provided is a gesture/speech integrated recognition system and method including: a speech feature extraction unit extracting speech feature information by detecting a start point and an end point of an order in an input speech; a gesture feature extraction unit extracting gesture feature information by detecting an order section from a gesture of taken images by using information on the detected start point and the end point; and an integrated recognition unit outputting the extracted speech feature information and the gesture feature information as integrated recognition data by using a learning parameter set in advance. Accordingly, it is possible to increase a performance of order recognition by integrating the speech and the gesture in a noise environment.
RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.


(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

Published: — with international search report
Description

GESTURE/SPEECH INTEGRATED RECOGNITION SYSTEM
AND METHOD

Technical Field

The present invention relates to a integrated recognition technology, and more particularly, to a gesture/speech integrated recognition system and method capable of recognizing an order of a user by extracting feature information on a gesture by using an end point detection (EPD) value of a speech and integrating feature information on the speech with the feature information on the gesture, thereby recognizing the order of the user at a high recognition rate in a noise environment.

This work was supported by the IT R&D program of MIC/IITA[2006-S-03 1-01, Five Sences Information Processing Technology Development for Network Based Reality Service]

Background Art

Recently, as a multimedia technology and an interface technology have been developed, researches on multimodal recognition for easily implementing a simple interface between the human and machines by using facial expressions or directions, lip movements, eye-gaze tracking, hand movements, speech, and the like have been actively carried out.

Particularly, among man-machine interface technologies, a speech recognition technology and a gesture recognition technology are used as the most convenient interface technology. The speech recognition technology and the gesture recognition technology have a high recognition rate in a limited environment. However, performances of the technologies are not good in a noise environment. This is because environmental noise affects the performance of the speech recognition, and a change in lighting and a type of gesture affect the performance of a camera-based gesture recognition technology. Therefore, the speech recognition technology needs the development of a technology of recognizing speech by using an algorithm robust against noise, and the gesture recognition technology needs the development of a technology of extracting a particular section of a gesture including recognition information. In addition, when general gestures are used, a particular section of the gesture cannot be easily identified, so that recognition is difficult.

In addition, when a speech and a gesture are integrated to be recognized, there is a problem in that since a speech frame rate is about 10ms/frame and an image frame rate is about 66.7ms/frame, the rates for processing the speech frame and the image frame are different. In addition, a section of a gesture requires more time as compared with a
section of a speech, so that lengths of the speech section and the gesture section are
different, and this causes a problem of synchronizing the speech and gesture.

Disclosure of Invention

Technical Problem

An aspect of the present invention provides a means for extracting feature information by detecting an order section from a speech of a user by using an algorithm robust against environmental noise, and detecting a feature section of a gesture by using information on an order start point of the speech, thereby easily recognizing the order in the gesture that is not easily identified.

An aspect of the present invention also provides a means for synchronizing a gesture and a speech by applying optimal frames set in advance to an order section of the gesture detected by a speech end point detection (EPD) value, thereby solving a problem of a synchronization difference between the speech and the gesture while integrated recognition is performed.

Technical Solution

According to an aspect of the present invention, there is provided a gesture/speech integrated recognition system including: a speech feature extraction unit extracting speech feature information by detecting a start point and an end point of an order in an input speech; a gesture feature extraction unit extracting gesture feature information by detecting an order section from a gesture of taken images by using information on the detected start point and the end point; and an integrated recognition unit outputting the extracted speech feature information and the gesture feature information as integrated recognition data by using a learning parameter set in advance.

In the above aspect of the present invention, the system may further include a synchronization module including: a gesture start point detection module detecting a start point of the gesture from the taken images by using the detected start point; and an optimal frame applying module calculating and extracting optimal image frames by applying the number of optimal frames set in advance from the start point of the gesture. Here, the gesture start point detection module may detect the start point of the gesture by checking a start point that is an end point detection (EPD) plug of the detected speech from the taken images.

In addition, the speech feature extraction unit may include: an EPD module detecting a start point and an end point in the input speech; and a hearing model-based speech feature extraction module extracting speech feature information included in the detected order by using a hearing model-based algorithm. In addition, the speech feature extraction unit may remove noise from the extracted speech feature information.
In addition, the gesture feature extraction unit may include: a hand tracking module tracking movements of a hand in images taken by a camera and transmitting the movements to the synchronization module; and a gesture feature extraction module extracting gesture feature information by using the optimal image frames extracted by the synchronization module.

In addition, the integrated recognition unit may include: an integrated learning DB (database) control module generating the learning parameter on the basis of an integrated learning model and an integrated learning DB set in advance; an integrated feature control module controlling the extracted speech feature information and the gesture feature information by using the generated learning parameter; and an integrated recognition module generating a result controlled by the integrated feature control module as a recognition result. Here, the integrated feature control module may control feature vectors of the extracted speech feature information and the gesture feature information through the extension and the reduction of the node numbers of input vectors.

According to another aspect of the present invention, there is provided a gesture/speech integrated recognition method including: a first step of extracting speech feature information by detecting a start point that is an EPD value and an end point of an order in an input speech; a second step of extracting gesture feature information by detecting an order section from a gesture in images input through a camera by using the detected start point of the order; and a third step of outputting the extracted speech feature information and gesture feature information as integrated recognition data by using a learning parameter set in advance.

In the above aspect of the present invention, in the first step, the speech feature information may be extracted from the order section obtained by the start point and the end point of the order on the basis of a hearing model.

In addition, the second step may include: the second step comprises: a B step of detecting an order section from the gesture of the movements of the hand by using the transmitted EPD value; a C step of determining optimal frames from the order section from the gesture by applying optimal frames set in advance; and a D step of extracting gesture feature information from the determined optimal frames.

**Advantageous Effects**

The gesture/speech integrated recognition system and method according to the present invention increases a recognition rate of a gesture that is not easily identified by detecting an order section of the gesture by using an end point detection (EPD) value that is a start point of an order section of a speech. In addition, optimal frames are applied to the order section of the gesture to synchronize the speech and the
gesture, so that it is possible to implement integrated recognition between the speech and the gesture.

**Brief Description of the Drawings**

[17] FIG. 1 is a view illustrating a concept of a gesture/speech integrated recognition system according to an embodiment of the present invention.

[18] FIG. 2 is a view illustrating a configuration of a gesture/speech integrated recognition system according to an embodiment of the present invention.

[19] FIG. 3 is a flowchart illustrating a gesture/speech integrated recognition method according to an embodiment of the present invention.

**Best Mode for Carrying Out the Invention**

[20] Exemplary embodiments of the present invention will now be described in detail with reference to the accompanying drawings.

[21] In the description, the detailed descriptions of well-known functions and structures may be omitted so as not to hinder the understanding of the present invention.

[22] FIG. 1 is a view illustrating a concept of a gesture/speech integrated recognition system according to an embodiment of the present invention.

[23] *Referring to FIG. 1, in a gesture/speech integrated recognition technology, orders by a speech or a gesture of a person are integrated to be recognized, and a device of representing the five senses by using a control instruction generated by a result of the recognition is controlled.

[24] Specifically, a person 100 orders through a speech 110 and a gesture 120. Here, as an example of the order of the person, in order for the person to order goods through a cyberspace, the person may indicate a corn bread with a finger while saying "select a corn bread" as an order for selecting the corn bread from among the displayed goods.

[25] After the person 100 orders by the speech 110 and the gesture 120, feature information on the speech order of the person is recognized through speech recognition 111, and feature information on the gesture of the person is recognized through gesture recognition 121. The recognized feature information on the gesture and the speech is recognized through integrated recognition 130 as a single user order in order to increase a recognition rate of the speech affected by environmental noise and the gesture that cannot be easily identified.

[26] The present invention provides a technology of integrated recognition for a speech and a gesture of a person. The recognized order is transmitted by a controller to a speaker 170, a display apparatus 171, a diffuser 172, a touching device 173, and a tasting device 174 which are output devices for the five senses, and the controller controls each device. In addition, the recognition result is transmitted to a network, so
that data of the five senses as the result is transmitted to control each of the output
devices. The present invention relates to integrated recognition, and a configuration
after the recognition operation can be modified in various manners, so that a detailed
description thereof is omitted.

[28] FIG. 2 is a view illustrating a configuration of the gesture/speech integrated rec-
ognition system according to an embodiment of the present invention.

[29] Referring to FIG. 2, the gesture/speech integrated recognition system includes a
speech feature extraction unit 210 for extracting speech feature information by
detecting a start point and an end point of an order from a speech input through a mi-
crophone 211, a gesture feature extraction unit 220 for extracting gesture feature in-
formation by detecting an order section from a gesture of images taken by a camera by
using information on the start point and the end point detected by the speech feature
extraction unit 210, a synchronization module 230 for detecting a start point of the
gesture from the taken images by using the start point detected by the speech feature
extraction unit 210, and calculating optimal image frames by applying the number
of optimal frames set in advance from the detected start point of the gesture, and an in-
tegrated recognition unit 240 outputting the extracted speech feature information and
the gesture feature information as integrated recognition data by using a predetermined
learning parameter set in advance. Now, each component is described in detail.

[30] The speech feature extraction unit 210 includes the microphone 211 receiving a
speech of a user, an end point detection (EPD) module 212 detecting a start point and
an end point of an order section from the speech of the user, and a hearing model-
based speech feature extraction module 213 extracting speech feature information from
the order section of the speech detected by the EPD module 212 on the basis of the
hearing model. In addition, a channel noise reduction module (not shown) for
removing noise from the extracted speech feature information may further be included.

[31] The EPD module 212 detects the start and the end of the order by analyzing the
speech input through the wired or wireless microphone.

[32] Specifically, the EPD module 212 acquires a speech signal, calculates an energy
value needed to detect the end point of the speech signal, and detects the start and the
end of the order by identifying a section to be calculated as the order from the input
speech signal.

[33] The EPD module 212 firstly acquires the speech signal through the microphone and
converts the acquired speech into a type for frame calculation. In the meanwhile, when
the speech is input wirelessly, a problem of data loss or signal deterioration due to
signal interference may occur, so that an operation to deal with the problem during the
signal acquisition is needed.

[34] The energy value needed to detect the end point of the speech signal by the EPD
module 212 is calculated as follows. A frame size used to analyze the speech signal is 160 samples, and frame energy is calculated by the following equation.

\[ \text{FrameEnergy} = \log_{10} \]

[35] Here, \( S(n) \) denotes a vocal cord signal sample, and \( N \) denotes the number of samples per frame.

[36] The calculated frame energy is used as a parameter for detecting the end point.

[37] The calculated energy is used as a parameter for detecting the end point.

[38] The EPD module 212 determines a section to be calculated as the order after calculating the frame energy value. For example, the operation of calculating the start point and the end point of the speech signal is determined by four energy thresholds using the frame energy and 10 conditions. Here, the four energy thresholds and the 10 conditions can be set in various manners and may have values obtained in experiments in order to obtain the order section. The four thresholds are used to detect the start and end of every frame by an EPD algorithm.

[39] The EPD module 212 transmits information on the detected start point (hereinafter, referred to as an "EPD value" of the order to a gesture start point detection module 231 of the synchronization module 230.

[40] In addition, the EPD module 212 transmits information on the order section in the input speech to the hearing model-based speech feature extraction module 213 to extract speech feature information.

[41] The hearing model-based speech feature extraction module 213 that receives the information on the order section in the speech extracts the feature information on the basis of the hearing model from the order section detected by the EPD module 212. Examples of an algorithm used to extract the speech feature information on the basis of the hearing model include an ensemble interval histogram (EIH) and a zero-crossings with peak amplitudes (ZCPA).

[42] A channel noise reduction module (not shown) removes noise from the speech feature information extracted by the hearing model-based speech feature extraction module 213, and the speech feature information from which the noise is removed is transmitted to the integrated recognition unit 240.[245->240]

[43] The gesture feature extraction unit 220 includes a face and hand detection module 222 for detecting a face and a hand from the images taken by a camera 221, and a gesture feature extraction module 224 for tracking and transmitting movements of the detected hand to the synchronization module 230 and extracting feature information on the gesture by using optimal frames calculated by the synchronization module 230.

[44] The face and hand detection module 222 detects the face and the hand that gives a gesture, and a hand tracking module 223 continuously tracks the hand movements in the images. In the description, the hand tracking module 223 tracks the hand. However, it will be understood by those of ordinary skill in the art that the hand tracking module
223 may track various body portions having movements recognized as a gesture.

The hand tracking module 223 continuously stores the hand movements as time elapses, and a part recognized as a gesture order in the hand movement is detected by the synchronization module 230 by using the EPD value transmitted from the speech feature extraction unit 210. Now, the synchronization module 230 which detects a section recognized as the gesture order in the hand movements by using the EPD value and applies the optimal frames for synchronization between speeches and gestures will be described.

The synchronization module 230 includes the gesture start point detection module 231 for detecting the start point of the gesture by using the EPD value and the images showing the hand movements, and an optimal frame applying module 232 for calculating optimal image frames needed for integrated recognition by using a start frame of the gesture calculated by using the detected start point of the gesture.

When an EPD value of a speech is detected by the EPD detection module 212 while a speech signal and an image signal are input in real-time, the gesture start point detection module 231 of the synchronization module 230 checks a speech EPD plug in the image signal. In this method, the gesture start point detection module 231 calculates the start frame of the gesture. In addition, the optimal frame applying module 232 calculates optimal image frames needed for integrated recognition by using the calculated start frame of the gesture and transmits the optimal image frames to the gesture feature extraction module 224. In order to calculate the optimal image frames needed for integrated recognition applied by the optimal frame applying module 232, the number of frames which are determined to have a high recognition rate of gestures is set in advance, and when the start frame of the gesture is calculated by the gesture start point detection module 231, the optimal image frames are determined.

The integrated recognition unit 240 includes an integrated model generation module 242 for generating an integrated model for effectively integrating the speech feature information and the gesture feature information on the basis of a learning model, an integrated learning database (DB) 244 implemented to be proper for integrated recognition algorithm development based on a statistical model, an integrated learning DB control module 243 for controlling learning by the integrated model generation module 242 and the integrated learning DB 244 and a learning parameter, an integrated feature control module 241 for controlling the learning parameter and feature vectors of the input speech feature information and the gesture feature information, and an integrated recognition model 245 for providing various functions by generating recognition results.

The integrated model generation module 242 generates a high-performance in-
integrated model in order to effectively integrate the speech feature information with the
gesture feature information. In order to determine the high-performance integrated
model, various learning algorithms such as hidden Markov model (HMM), neural
network (NN), dynamic time warping (DTW) are implemented and experiments are
performed. Particularly, according to the embodiment of the present invention, a
method of optimizing NN parameters by determining an integrated model on the basis
of the NN to obtain high-performance integrated recognition may be used. However,
the generation of the high-performance integrated model has a problem of syn-
chronizing two modalities that have different frame numbers in the learning model.

The problem of the synchronization in the learning model is referred to as a learning
model optimization problem. According to the embodiment of the present invention,
an integration layer is provided, and a method of connecting the speech and the gesture
is optimized in the layer. An overlapping length between the speech and the gesture is
calculated with respect to a time axis for the optimization, and the synchronization is
performed on the basis of the overlapping length. The overlapping length is used to
search for a connection method having a highest recognition rate through a recognition
rate experiment.

The integrated learning DB 244 implements an integrated recognition DB to be
proper for the integrated recognition algorithm development based on the statistical
model.

For example, data of ten words in various age groups is synchronized and collected
by a stereo camera and a wireless microphone. Table 1 shows an order set defined to
integrate gestures and speeches. The defined order set is obtained on the basis of
natural gestures that people can understand without learning.

Table 1
Here, the speech is recorded at a sampling rate of 16kHz/16bits per sample by using a waveform in a pulse code modulation (PCM) method using the channel number of 1. In addition, 24bits BITMAP images having a size of 320x240 are recorded with a blue screen background in a light condition having four fluorescent boxes at 15 frames per second by using a STH-DCSG-C stereo camera. Since the stereo camera does not have a speech interface, a speech collection module and an image collection module are independently provided, and a synchronization program of the image and speech to perform a method of controlling an image collecting process through inter-process communications (IPC) in a speech recording program controls is written to collect data. The image collection module is configured by using an open source computer vision library (OpenCV) and a small vision system (SVS).

Stereo camera images have to be adjusted to a real recording environment through an additional calibration process, and in order to acquire optima images, associated gain, exposure, brightness, red, and blue parameter values are modified to control color, exposure, and WB values. Calibration information and parameter information are stored in an additional ini file so that the image storage model calls and use the information.

The integrated learning DB control module 243 generates the learning parameter on the basis of the integrated learning DB 244 that is related to the integrated model generation module 242 to be generated and stored in advance.

The integrated feature control module 241 controls the learning parameter generated by the integrated learning DB control module 243 and the feature vectors of the feature information on the speech and the gesture extracted by the speech feature extraction
unit 210 and the gesture feature extraction unit 220. The control operation is associated with the extension and the reduction of the node number of input vectors. The integrated feature control module 241 has the integration layer, and the integration layer is developed to effectively integrate lengths of the speech and the gesture that have different values and propose a single recognition rate.

The integrated recognition module 245 generates a recognition result by using a control result obtained by the integrated feature control module 241. In addition, the integrated recognition module 245 provides various functions for interactions with a network or a integration representing unit.

FIG. 3 is a flowchart illustrating a gesture/speech integrated recognition method according to an embodiment of the present invention.

Referring to FIG. 3, the gesture/speech integrated recognition method includes three threads to be operated. The three threads include a speech feature extraction thread 10 for extracting speech features, a gesture feature extraction thread 20 for extracting gesture features, and an integrated recognition thread 30 for performing integrated recognition on the speech and the gesture. The three threads 10, 20, and 30 are generated at a time point when the learning parameter is loaded and cooperatively operated by using a thread plug. Now, the gesture/speech integrated recognition method in which the three threads 10, 20, 30 are cooperatively operated is described.

When the user orders by a speech and a gesture, the speech feature extraction thread 10 continuously receives the speech through a wired or wireless microphone (operation S31). Next, the gesture feature extraction thread 20 continuously receives images including the gesture through a camera (operation S320). Speech frames of the speech continuously input through the microphone are calculated (operation S312), and the EPD module 212 detects a start point and an end point (speech EPD value) of an order included in the speech (operation S313). When the speech EPD value is detected, the speech EPD value is transmitted to a synchronization operation 40 of the gesture feature extraction thread 20. In addition, when an order section of the speech is determined by using the start point and the end point of the order included in the speech, the hearing model-based speech feature extraction module 213 extracts a speech feature from the order section on the basis of the hearing model (operation S314) and transmits the extracted speech feature to the integrated recognition thread 30.

The gesture feature extraction thread 20 detects a hand and a face from the images continuously input through the camera (operation S321). When the hand and face are detected, a gesture of the user is tracked (operation S322). Since the gesture of the user continuously changes, a gesture having a predetermined length is stored in a buffer (operation S323).

When the speech EPD value is detected and transmitted while the gesture is stored in
the buffer, a speech EPD plug in the gesture images stored in the buffer is checked (operation S324). The start point and the end point of the gesture including the feature information on the images are searched for by the speech EPD plug (operation S325), and the searched gesture feature is stored (operation S326). The stored gesture feature is not in synchronization with the speech, so that optimal frames are calculated from a start frame of the gesture by applying optimal frames set in advance. In addition, the calculated optimal frames are used to extract gesture feature information by the gesture feature extraction module 224, and the extracted gesture feature information is transmitted to the integrated recognition thread 30.

When the speech feature extraction thread 10 and the gesture feature extraction thread 20 successfully extract feature information on the speech and the gesture, the speech/gesture feature extraction threads 10 and 20 are in a sleep state while the integrated recognition thread checks a recognition result (operations S328 and S315).

The integrated recognition thread 30 generates a high-performance integrated model in advance by using the integrated model generation module 245 before receiving the speech feature information and the gesture feature information and controls the generated integrated model and the integrated learning DB 244 so that the integrated learning DB control module 243 generates and loads the learning parameter (operation S331). When the learning parameter is loaded, the integrated recognition thread 30 is maintained in the sleep state until the speech/gesture feature information is received (operation S332).

When the integrated recognition thread 30 that is in the sleep state receives a signal associated with the feature information after the extraction of the feature information on the speech and the gesture is completed (operation S333), the integrated recognition thread 30 loads each feature in a memory (operation S334). When the feature information on the speech and the gesture is loaded, the recognition result is calculated by using the optimized integrated learning model and the learning parameter that are set in advance (operation S335).

When the recognition result is calculated by the integrated recognition unit 240, the speech feature extraction thread 10 and the gesture feature extraction thread 20 that are in the sleep state perform operations of extracting feature information from a speech and an image that are input.

While the present invention has been shown and described in connection with the exemplary embodiments, it will be apparent to those skilled in the art that modifications and variations can be made without departing from the spirit and scope of the invention as defined by the appended claims.
Claims

[1] A gesture/speech integrated recognition system comprising:
   a speech feature extraction unit extracting speech feature information by
detecting a start point and an end point of an order in an input speech;
a gesture feature extraction unit extracting gesture feature information by
detecting an order section from a gesture of taken images by using information
on the detected start point and the end point; and
an integrated recognition unit outputting the extracted speech feature information
and the gesture feature information as integrated recognition data by using a
learning parameter set in advance.

[2] The system of claim 1, further comprising a synchronization module comprising:
a gesture start point detection module detecting a start point of the gesture from
the taken images by using the detected start point; and
an optimal frame applying module calculating and extracting optimal image frames by applying the number of optimal frames set in advance from the start point of the gesture.

[3] The system of claim 2, wherein the gesture start point detection module detects
the start point of the gesture by checking a start point that is an EPD (end point
detection) plug of the detected speech from the taken images.

[4] The system of claim 1, wherein the speech feature extraction unit comprises:
an EPD module detecting a start point and an end point in the input speech; and
a hearing model-based speech feature extraction module extracting speech feature information included in the detected order by using a hearing model-based algorithm.

[5] The system of claim 4, wherein the speech feature extraction unit removes noise
from the extracted speech feature information.

[6] The system of claim 3, wherein the gesture feature extraction unit comprises:
*a hand tracking module tracking movements of a hand in images taken by a
camera and transmitting the movements to the synchronization module; and
a gesture feature extraction module extracting gesture feature information by
using the optimal image frames extracted by the synchronization module.

[7] The system of claim 1, wherein the integrated recognition unit comprises:
an integrated learning DB (database) control module generating the learning
parameter on the basis of an integrated learning model and an integrated learning
DB set in advance;
an integrated feature control module controlling the extracted speech feature information and the gesture feature information by using the generated learning
parameter; and
an integrated recognition module generating a result controlled by the integrated
feature control module as a recognition result.

[8] The system of claim 7, wherein the integrated learning model is generated on the
basis of a neural network learning algorithm.

[9] The system of claim 7, wherein the integrated learning DB is implemented to be
proper to an integrated recognition algorithm based on a statistical model by in-
tegrating feature information on speeches and gestures in various ages groups by
using a stereo camera and a wireless microphone.

[10] The system of claim 7, wherein the integrated recognition module includes an in-
tegration layer for integrating the extracted speech feature information with the
gesture feature information.

[11] The system of claim 7, wherein the integrated feature control module controls
feature vectors of the extracted speech feature information and the gesture
feature information through the extension and the reduction of the node numbers
of input vectors.

[12] A gesture/speech integrated recognition method comprising:
a first step of extracting speech feature information by detecting a start point that
is an EPD value and an end point of an order in an input speech;
a second step of extracting gesture feature information by detecting an order
section from a gesture in images input through a camera by using the detected
start point of the order; and
a third step of outputting the extracted speech feature information and gesture
feature information as integrated recognition data by using a learning parameter
set in advance.

[13] The method of claim 12, wherein in the first step, the speech feature information
is extracted from the order section obtained by the start point and the end point of
the order on the basis of a hearing model.

[14] The method of claim 12, wherein the second step comprises:
an A step of tracking a gesture of movements of a hand from the image input
through the camera;
a B step of detecting an order section from the gesture of the movements of the
hand by using the transmitted EPD value;
a C step of determining optimal frames from the order section from the gesture
by applying optimal frames set in advance; and
a D step of extracting gesture feature information from the determined optimal
frames.

[15] The method of claim 12, wherein the first step further comprises removing noise
from the extracted speech feature information.
INTERNATIONAL SEARCH REPORT
PCT/KR2007/006189

A. CLASSIFICATION OF SUBJECT MATTER

G06F 3/00(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 8 G06F, G09G

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean Utility models applications for Utility Models since 1975, IPC as above
Japanese Utility Models and application for Utility Models since 1975, IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKIPASS, IEEEExpl, Google, gesture, speech, recognition, integration, user interface,

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category*</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>EP0594129 A2 (HITACHI, LTD ) 19 October 1993 See Abstract, Claims 1&amp;1 and Figures 1,9,20</td>
<td>1-15</td>
</tr>
<tr>
<td>X</td>
<td>KRI 0-2001-0075838 A (KOREA ELECTRONICS &amp; TELECOMMUNICATIONS RESEARCH INSTITUTE) 11 September 2001 See Abstract and Figures 1-4</td>
<td>1-15</td>
</tr>
</tbody>
</table>

* Special categories of cited documents
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance, the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance, the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search
29 FEBRUARY 2008 (2902 2008)

Date of mailing of the international search report
29 FEBRUARY 2008 (29.02.2008)

Name and mailing address of the ISA/KR
Korean Intellectual Property Office
Government Complex-Daejeon, 139 Seomsa-ro, Seogu, Daejeon 302-701, Republic of Korea
Facsimile No 82-42-472-7140

Authorized officer
SEONG, KYOUNG A
Telephone No 82-42-481-8171

Form PCT/ISA/210 (second sheet) (April 2007)
<table>
<thead>
<tr>
<th>Patent document cited in search report</th>
<th>Publication date</th>
<th>Patent family member(s)</th>
<th>Publication date</th>
</tr>
</thead>
<tbody>
<tr>
<td>US20030001908A 1</td>
<td>02.01.2003</td>
<td>CN1265625C</td>
<td>19.07.2006</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CN1520685A</td>
<td>11.08.2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP1405509A1</td>
<td>07.04.2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP16531183A</td>
<td>07.10.2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KR10200400 15001</td>
<td>18.02.2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO2003003728A1</td>
<td>09.01.2003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP594129A3</td>
<td>05.10.1994</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP06131437A</td>
<td>13.05.1994</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US5600765A</td>
<td>04.02.1997</td>
</tr>
<tr>
<td>JP07306772A</td>
<td>21.11.1995</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>KR1020010075838A</td>
<td>11.08.2001</td>
<td>None</td>
<td></td>
</tr>
</tbody>
</table>