

(51) International Patent Classification:
G06K 9/34 (2006.01)(21) International Application Number:
PCT/US2014/035589(22) International Filing Date:
27 April 2014 (27.04.2014)

(25) Filing Language: English

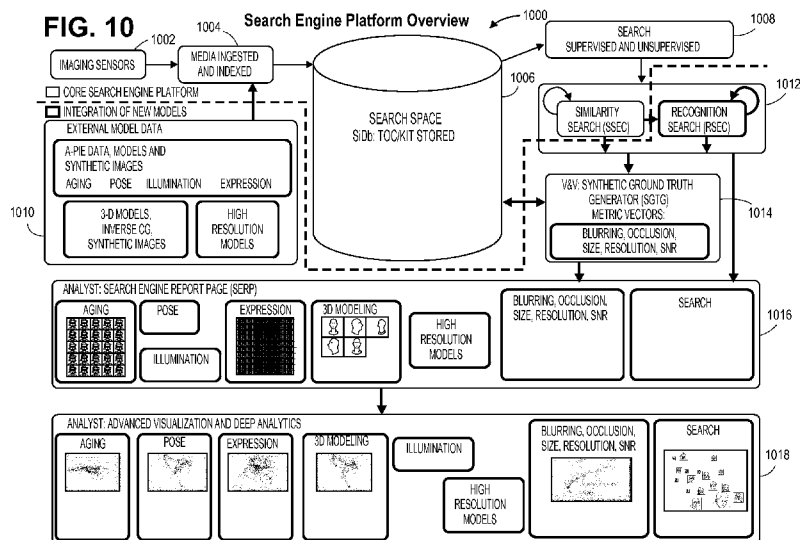
(26) Publication Language: English

(30) Priority Data:
61/816,719 27 April 2013 (27.04.2013) US
14/262,756 27 April 2014 (27.04.2014) US(71) Applicant: DATAFISSION CORPORATION [US/US];
2526 Qume Drive, Suite 22, San Jose, CA 95131 (US).(72) Inventors: TREASE, Harold; 2526 Qume Drive, Suite
22, San Jose, CA 95131 (US). TREASE, Lynn; 2526
Qume Drive, Suite 22, San Jose, CA 95131 (US). HER-
RERA, Shawn; 2526 Qume Drive, Suite 22, San Jose, CA
95131 (US).(74) Agent: KALER, Stuart, P.; 2040 16th Avenue, San Fran-
cisco, CA 94116 (US).(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,
SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, 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, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ,
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))
— before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments (Rule 48.2(h))(88) Date of publication of the international search report:
22 January 2015

(54) Title: CONTENT BASED SEARCH ENGINE FOR PROCESSING UNSTRUCTURED DIGITAL



(57) Abstract: Systems and methods for receiving and indexing native digital data and generating signature vectors for subsequent storage and searching for such native digital data in a database of digital data are disclosed. Native digital data may be transformed into associated transform data sets. Such transformation may comprise entropy-like transforms and/or spatial frequency transforms. The native and associated transform data sets may then be partitioned in to spectral components and those spectral components may have statistical moments applied to them to create a signature vector. Other systems and methods for processing non-image digital data are disclosed. Non-image digital data may be transformed into an amplitude vs time data set and a spectrogram may then be applied to such data sets. Such transformed data sets may then be processed as described.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2014/035589

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G06K 9/34 (2014.01)

CPC - G06K 9/34 (2014.09)

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 17/30; G06K 9/00, 34, 46 (2014.01)

USPC - 382/100, 173; 386/201

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
CPC - G06F 17/30; G06K 9/00, 34, 46 (2014.09) (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Google, Orbit, Google Patents

Search terms used: Signature, vector, database, audio, video, image, search, match, compare, input, discontinuity

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X		1,2,7,9
----		-----
Y	US 2002/0126872 A1 (BRUNK et al) 12 September 2002 (12.09.2002) entire document	3-6,8,10-16
Y	US 6600874 B1 (FUJITA et al) 29 July 2003 (29.07.2003) entire document	3,13
Y	US 2009/0265024 A1 (DITTMAR et al) 22 October 2009 (22.10.2009) entire document	4-6
Y	US 2009/0018897 A1 (BREITER) 15 January 2009 (15.01.2009) entire document	8
Y	US 5046112 A (CHANG et al) 03 September 1991 (03.09.1991) entire document	10-12
Y	US 2010/0329547 A1 (CAVET) 30 December 2010 (30.12.2010) entire document	12
Y	BAL et al. Interactive degraded document enhancement and ground truth generation. 2008. [retrieved on 2014-10-29]. Retrieved from the Internet:<URL: http://ir.cs.georgetown.edu/publications/downloads/DRR08-INTERACTIVE.pdf >. entire document.	14-16
Y	WHITING et al. Creating Realistic, Scenario-Based Synthetic Data for Test and Evaluation of Information Analytics Software. 2008. [retrieved on 2014-10-29]. Retrieved from the Internet:<URL: https://www.purdue.edu/discoverypark/vaccine/assets/pdfs/publications/pdf/Creating%20Realistic,%20Scenario-Based.pdf >. entire document.	15,16

☐

Further documents are listed in the continuation of Box C.

☐

* 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 another 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

28 October 2014

Date of mailing of the international search report

26 NOV 2014

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2014/035589

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

See Extra Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-16

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☒ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2014/035589

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claims 1-16, drawn to a system for searching digital data.

Group II, claims 17-22, drawn to a method for generating signature vectors.

The inventions listed as Groups I and II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical feature of the Group I invention: an indexing module, a signature generation module capable of generating transform data sets, the signature vector comprising a spectral decomposition and statistical decomposition, a TOC database, a searching module as claimed therein is not present in the invention of Group II. The special technical feature of the Group II invention: applying an entropy transform, applying a spatial frequency transform, partitioning the digital data set and applying a set of statistical moments as claimed therein is not present in the invention of Group I.

Groups I II lack unity of invention because even though the inventions of these groups require the technical feature of generating a signature vector from a native digital data set comprising receiving a native digital data set and applying a transform on the data set to generate a signature vector, this technical feature is not a special technical feature as it does not make a contribution over the prior art. Specifically, US 2002/0126872 A1 (BRUNK et al) 12 September 2002 (12.09.2002) teaches generating a signature vector from a native digital data set (deriving a content signature from a signal provided, abstract, see Fig. 1, para 22) comprising receiving a native digital data set (para 20-21, Fig. 1) and applying a transform on the data set to generate a signature vector (transforming signal, para 22, Fig. 1).

Since none of the special technical features of the Group I or II inventions are found in more than one of the inventions, unity of invention is lacking.