[Continued on next page]

(54) **Title:** METHOD AND APPARATUS FOR VALIDITY DETERMINATION OF A DATA DIVIDING OPERATION

Fig. 3
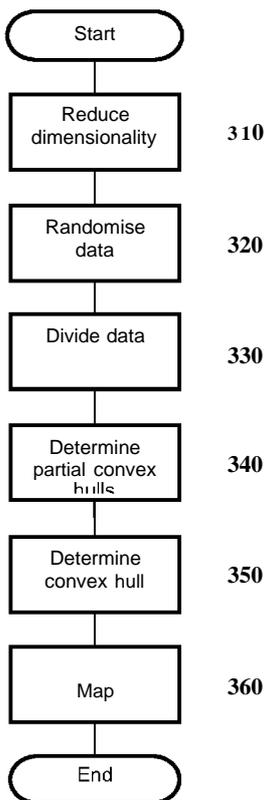
(57) **Abstract:** Embodiments of the present invention provide a method of determining whether data exceeds a validity range of an operation, comprising dividing operation data indicative of a range of validity of the operation into a plurality of chunks, determining a partial convex hull for each of the plurality of chunks, determining a convex hull for the operation data based on the plurality of partial convex hulls, and determining whether the data exceeds the validity range of the operation data based on the convex hull.

w o 2016/015752 A i |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

METHOD AND APPARATUS FOR VALIDITY DETERMINATION OF A DATA
DIVIDING OPERATION

Background

5   In the field of imaging, such as image correction, printing etc., it is often required to
apply solutions such as correction techniques, colour models, colour mapping
algorithms etc., to input data. However such solutions have a range of validity for
which they are known to work. The range of validity often, although not exclusively,
derives from measured data from which the solution is derived. For example a
10   correction method may be derived from a range of reflectances, a colour mapping
algorithm may be based upon a range of ink-vectors and a colour profile may be based
upon a range of measurements. The data on which a solution is based will hereinafter
be referred to as operation data.

15   If a solution is applied to image data outside of the operation data on which the
solution is based, then artefacts may be present in output data or failure of the solution
may arise.

It is an object of embodiments of the invention to at least mitigate one or more of the
20   problems of the prior art.

Brief Description of the Drawings

Embodiments of the invention will now be described by way of example only, with
25   reference to the accompanying figures, in which:

Figure 1 shows an illustration of computing time for computing a convex hull for
various data;

30   Figure 2 shows a further illustration of computing time for computing a convex hull
for various data;

Figure 3 shows a method according to an embodiment of the invention;

Figure 4 shows a system according to an embodiment of the invention;

Figure 5 illustrates coverage of reflectance data having differing numbers of dimensions; and

5

Figure 6 illustrates a performance of a method according to an embodiment of the invention.

Detailed Description of Embodiments of the Invention

10

Embodiments of the present invention provide methods and apparatus for validity checking. Validity checking comprises determining whether input data, on which it is intended to perform an operation, falls outside of a domain or range of operation data on which the operation is based or is known to be valid. For example, to determine

15 whether the input data falls outside of data on which an algorithm is based. The operation may be a correction method, a colour model, performance of a reference system etc. If the input data falls outside of the domain or range of the operation data then a warning may be generated, the operation stopped or an alternative, sometimes less accurate, operation may be performed which may be more robust to

20 extrapolation.

A convex hull or convex envelope may be used to determine a boundary or extent of a set of data, that is the data's extreme vertices, such as operation data. For data having a large number of dimensions, such as $N>5$ where N is the number of dimensions of

25 the data set, computation of a convex hull can be problematic. Most convex hull computation methods are intended for use with data having two or three dimensions. However reflectance data is typically 16 or 31, or higher, dimensional; ink vectors may have between 4 and 12 dimensions; and 3D object descriptions for print may also have additional dimensions such as material properties and colorimetry resulting in a

30 total of 4 (3D coordinates + material), 7 (+ XYZ colorimetry) or more dimensions. It will be appreciated that these numbers of dimensions are merely illustrative. Similarly computing a convex hull for a large number of data points, or a large number of data points also having a large number of dimensions, can be equally problematic, as will be appreciated.

Figure 1 illustrates computation times for a convex hull using a double-description method *{cdd)* by Komei Fukuda for data sets comprising various numbers of samples and dimensions. As can be appreciated, increasing numbers of samples and

5   dimensions of the data significantly increases the computation time. For a relatively small set of data (in this case 3000 samples) this method is useful up to 16 dimensions and perhaps beyond. However for rank-deficient data which is not fully dimensional or is clustered, or for large numbers of samples, the method slows down considerably. Therefore the speed of convex hull determination using *cdd* depends upon one or

10   more of: the true dimensionality of the data (uniform random points compute faster than, for example, reflectance data); the nominal dimensionality of the data; and the number of sample points.

Figure 2 illustrates computation times for a convex hull using a process provided in

15   MATLAB called *convhulln* (which in turn is based on an open source *qhull* library) for data sets comprising various numbers of samples and dimensions. It will be noted that, unlike Figure 1, the time axis is logarithmic. Whilst this technique is faster at six dimensions it can be appreciated that it rapidly slows above eight dimensions. Therefore embodiments of the invention aim to improve useability with data having

20   larger numbers of dimensions, particularly $N>6$, and/or numbers of data points.

Figure 3 illustrates a method 300 according to an embodiment of the invention. The method 300 is a method of determining a convex hull for operation data to enable validity checking of input data.

25

The method shown in Figure 3 will be described with reference to a system 400 illustrated in Figure 4. The system 400 comprises a processing unit 410 for executing instructions forming computer software for implementing a method according to an embodiment of the invention. The processing unit 410 may comprise a plurality of

30   processors and/or processing cores each capable of independently executing instructions. The processing unit 410 may comprise one or more CPUs, GPUs, FPGAs or the like. The software may be stored in a memory 420 of the system which may comprise a read-only portion storing the computer-executable instructions. The memory 420 may also comprise a rewritable portion for storing data during operation

of the system 400. The memory 420 may store input data for which it is required to determine the validity of an operation according to an embodiment of the invention. The memory 420 may alternatively or additionally store the data on which the operation is based, such as the operation data, for which a convex hull will be

5   determined. The data on which the operation is based will hereinafter be referred to as operation data. It will be realised that the memory 420 may only store a representative portion of the data on which the operation is based, such that the validity of the operation may be checked using the representative operation data. The system 400 may further comprise, in some embodiments, a I/O unit 430 and/or a 440

10   printing unit 440.   The I/O unit 440 may comprise one or more of a device for communicating data with other apparatus such as via a network which may be a wired or wireless connection; a device for receiving a user input such as via one or more keys or buttons; and an output device for outputting information to a user such as in the form of a display device. In some embodiments the device for receiving the user

15   input and the output device may be combined, for example in the form of a touch-sensitive display device.   The printing unit 440 is for printing an image corresponding to image data which may be stored in the memory 420. The image data may be based upon the input data for which it is required to validity check the operation.

20

Returning to Figure 3, the method 300 comprises a step 310 of reducing a dimensionality of the operation data.   Step 310 may comprise sub-steps of determining a lower dimensionality of the operation data and mapping the operation data to a model on the basis of the determined lower number of dimensions.

25

The operation data may have a number of dimensions defined by N x N_{smp}. The output of step 310 is data having dimensions N x N_D where $N_D < N_{smp}$. The operation data may be stored in the memory 420 of the system 400 and the data output by step 310 may also be stored in the memory 420.

30

Step 310 may comprise performing linear or non-linear dimensionality reduction on the operation data. The linear dimensionality reduction may comprise performing Principal Component Analysis (PCA) on the operation data to determine a "true" dimensionality of the operation data. In other words, in the first sub-step of step 310

it is determined whether a number of dimensions of the operation data can be reduced. The PCA determines a value for the number of reduced dimensions $N_¾$ such that above a predetermined variance is represented in the data where the predetermined variance may be above 90%, 95%, substantially 1*00%* or exactly 1*00%*. If N_D is less

5   than N_{smp} then the second sub-step of step 310 is performed, or else if $N_¾$ is equal to *Nsmp* then the method moves to step 320.

The second sub-step of step 310 comprises performing a linear mapping of the N x *Nsmp* operation data in N_{smp} space to the lower-dimensional $N_¾$ space. This reduces

10   the operation data from N x N_{smp} to N x N_D where N_D is less than N_{smp} thereby reducing the number of dimensions of the operation data.

The second sub-step may be referred to as linear subspace projection. In this sub-step a plurality of data points in A^-dimensions are mapped to closest points in a $A_¾$-

15   dimensional subspace defined by a set of $N_{s_{o'P}}$-vectors in this space (a N_D x N_{smp} matrix B where the N_D rows of the matrix are the vectors that define the axes of the N^-dimensional subspace, hence N_D of the N_{smp}-vectors). Where these arise from PCA the vectors are also orthonormal. If we denote this set of vectors B then the projection matrix is defined as P = B*B^T*B (where B^T is the transpose). Since B is

20   orthonormal B^T*B is the identity matrix and hence P = B and if S is the data set of N points in N_{smp} dimensions represented in a N x N_{smp} matrix then its representation in the reduced dimensional domain $N_¾$ is Q = S^T*B such that Q is N x *No-*

Figure 5 illustrates coverage of spectral reflectance data having differing numbers of

25   dimensions. The coverage in % is determined using principal component analysis (PCA) and also Multi-Illuminant Parameter Delta E (MIPE) colour difference (DE00). The reflectance data comprises 54,000 samples. It can be appreciated from Figure 5 that lower dimensional data can represent data having higher dimensions well until a lower cut-off point.

30

It will be noted that operating at a lower dimensionality does not imply lower precision. Extreme vertices of lower dimension data are likely to be a subset of extreme vertices at higher dimensions as well, although a part of a convex hull volume may be sacrificed by the reduction in dimensions.

In optional step 320 the operation data is re-ordered to reduce any structure of the data. It will be realised that step 320 may be performed prior to, or after as shown in Figure 3, step 310. In step 320 the data set is re-ordered to increase a randomisation of the data. If the data set has an intrinsic order this may be detrimental to later steps of the method.

In optional step 330 it is determined whether the data has greater than a predetermined number of samples. If the data has greater than a predetermined number K of samples where K may be 3000, 5000 or another value of samples may be chosen, the data is divided into chunks of data. Each chunk of data may be equally sized and may comprise a predetermined number of samples. The predetermined number of samples is a number of samples which may be processed by each processor or core of the processing unit 410. The number of samples may also be determined depending on the available number of processors or CPU or GPU cores that can compute in parallel. Each chunk comprises K samples although it will be realised that other numbers may be chosen.

In step 340 a partial convex hull is determined for each of the chunks determined in step 330. The determination of each partial convex hull is allocated to a respective core or processor, hereinafter generally referred to as a processing unit, of the processing unit 410. The determination of each partial convex hull may therefore be performed at least partly in parallel by the respective processing units in order to reduce a processing time of step 340.

Each respective processing unit may determine the partial convex hull (a convex hull for each respective chunk of data) using a process such as *cdd, qhull* or *convhulln* as referred to above. It will also be realised that other processes may be used. Furthermore it will be realised that each processing unit may determine a partial convex hull for a plurality of chunks i.e. the number of chunks may be more than the number of processing units and that each processing unit may not be allocated the same number of chunks. Step 340 thus determines extreme vertices of a plurality of partial convex hulls. For every i-th data chunk $K_i \, x \, N_{smp}$ the convex hull computation returns a new set of extreme vertices $P_i \, x \, N_{smp}$ where $P_i <= K_i$ (it is equal if and only

if all points of the chunk are already on the convex hull). The extreme vertices may then be joined into a single L x N_D data set whereby L is the sum of the partial convex hull computations on the K_i x $N_{3/4}$ data sets such that $L = P_i + P_2 + P_3$ ... and such that $L <= N$ (the sum of extreme vertices is less than or equal to all initial data points, equality being the case if and only if all initial points were on the convex hull to begin with).

If the number of samples L is larger than a predetermined threshold, steps 330 and 340 may be repeated. The repetition further divides the data points into chunks and computes partial convex hulls for those chunks. Furthermore if L is similar to N after the division into chunks and subsequent partial convex hull computation, the original data set was already largely on the convex hull and further repetition is not useful. Therefore the method 300 may comprise a decision step in which it is determined whether to repeat steps 330-340 one or more times.

In step 350 a total convex hull of the vertices of the partial convex hulls arising from step 340 is determined. That is, in step 350 a convex hull of joint convex hulls from step 340 is determined. The result is a set of M x $N_{3/4}$ points in a linear model basis whereby M is the number of vertices strictly on the convex hull of the joint data set that resulted from the partial convex hull computations on the K_i x N_D chunks, and $M <= L$ (the overall convex hull has at most as many vertices as the sum of the partial convex hulls).

In step 360 the points having N_D dimensions are projected or mapped back to N_{smp} dimensions to return data in the form M x N_{smp}.

In a further step, not shown in Figure 3, which may be performed after step 360, input data can be compared against the convex hull in M x N_{smp} in order to determine whether the input data falls outside of the convex hull for the operation data. This may be achieved in a number of ways, for example re-expressing the convex hull as a set of half-spaces (linear inequalities) which allows for fast inclusion checking by means of evaluating a new data point against all linear inequalities, if the data point satisfies all inequalities (that define the convex hull), this means that it is inside the

convex hull, otherwise it is outside. Other methods, such as a determinant-based method etc. are known to the skilled person.

If the input data falls outside, at least partly, the convex hull then a warning may be output that the a validity of an operation to be performed on the input data is not guaranteed. In another embodiment the operation may be stopped if the input data falls outside of the convex hull, or only partial results may be computed for a portion of the input data that satisfies the convex hull (i.e. falls within the convex hull). Alternatively a different method or operation may be performed on the input data which falls outside of the convex hull. In this case a warning may be provided, for example a warning about a heterogeneity of methods used.

Figure 6 illustrates a comparison of *cdd* against a method of determining a convex hull according to an embodiment the present invention for data having a variety of dimensions. As can be appreciated from Figure 5 for reflectance data the method according to an embodiment of the invention significantly outperforms the *cdd* method.

Whilst the above embodiments have been described with reference to reflectance data, it will be realised that embodiments of the invention are not necessarily restricted to use with such data. Embodiments of the invention may be used with any of the following:

Reflectance data which is typically high dimensional (common formats are 16, 31 or 101 sampling dimensions), typically well representable by PCA.

Colorimetric data (CIE XYZs, LABs, etc.) which while not high dimensional may comprise large numbers of samples.

Ink-vectors, the dimensionality of which depends directly on a number of inks, e.g. 4 to 12, although more may be envisaged.

HANS NPacs which describes print as 'Neugebauer Primary area coverages' which are effectively vectors not in ink-space but in Neugebauer Primary' space which is the

8

domain of all ink-overprints. For example for a 4-ink binary (some ink, no ink) printing system there exists $2^4 = 16$ dimensions, wherein if there are two drop sizes (no ink, 1 drop, 2 drops) there exists $3^4 = 81$ dimensions. As can be appreciated there may be a large number of dimensions. In a further example a colour look-up table indexed in RGB may comprise (e.g. $17^3 = 4913$ samples) in an 81 dimensional domain.

3D printing data which may extend beyond 3 dimensions *(x, y, z* coordinates) particularly when involving further properties at every spatial location, such as material or colour. Therefore the data comprises 4 (3 coordinates + 1 material), 7 (3 coordinates, 1 material, 3 color) or more dimensions and also involves very large sets of data points (10K+).

As can be appreciated embodiments of the present invention provide methods and apparatus for determining a validity of an operation. The methods and apparatus comprise determining a convex hull based upon a plurality of partial hulls.

It will be appreciated that embodiments of the present invention can be realised in the form of hardware, software or a combination of hardware and software. Any such software may be stored in the form of volatile or non-volatile storage such as, for example, a storage device like a ROM, whether erasable or rewritable or not, or in the form of memory such as, for example, RAM, memory chips, device or integrated circuits or on an optically or magnetically readable medium such as, for example, a CD, DVD, magnetic disk or magnetic tape. It will be appreciated that the storage devices and storage media are embodiments of machine-readable storage that are suitable for storing a program or programs that, when executed, implement embodiments of the present invention. Accordingly, embodiments provide a program comprising code for implementing a system or method as claimed in any preceding claim and a machine readable storage storing such a program. Still further, embodiments of the present invention may be conveyed electronically via any medium such as a communication signal carried over a wired or wireless connection and embodiments suitably encompass the same.

All of the features disclosed in this specification (including any accompanying claims, abstract and drawings), and/or all of the steps of any method or process so disclosed, may be combined in any combination, except combinations where at least some of such features and/or steps are mutually exclusive.

5

Each feature disclosed in this specification (including any accompanying claims, abstract and drawings), may be replaced by alternative features serving the same, equivalent or similar purpose, unless expressly stated otherwise. Thus, unless expressly stated otherwise, each feature disclosed is one example only of a generic

10   series of equivalent or similar features.

The invention is not restricted to the details of any foregoing embodiments. The invention extends to any novel one, or any novel combination, of the features disclosed in this specification (including any accompanying claims, abstract and

15   drawings), or to any novel one, or any novel combination, of the steps of any method or process so disclosed. The claims should not be construed to cover merely the foregoing embodiments, but also any embodiments which fall within the scope of the claims.

CLAIMS

1.   A method of determining whether data exceeds a validity range of an operation, comprising

        dividing operation data indicative of a range of validity of the operation into a plurality of chunks;

        determining a partial convex hull for each of the plurality of chunks;

        determining a convex hull for the operation data based on the plurality of partial convex hulls; and

        determining whether the data exceeds the validity range of the operation data based on the convex hull.

2.   The method of claim 1, comprising determining a reduced number of dimensions of the operation data.

3.   The method of claim 2, wherein the determining comprises performing a principal component analysis (PCA) on the operation data.

4.   The method of claim 2, comprising mapping the operation data to a model based on the reduced number of dimensions.

5.   The method of claim 4, wherein the model is a linear model.

6.   The method of claim 1, comprising randomising the operation data.

7.   The method of claim 1, wherein at least some of the partial convex hulls are determined in parallel.

8.   The method of claim 1, wherein the determining whether the data exceeds the validity range of the operation data comprises mapping vertices of the convex

hull to an original dimension of the operation data and comparing the data against the vertices in the original dimension.

9.   The apparatus of claim 1, comprising outputting an indication in dependence on the determination of whether the data exceeds the validity range of the operation data.

10.   An apparatus comprising:

        a memory; and

        a processing unit, wherein the processing unit is arranged to execute instructions to perform a method, comprising:

        dividing operation data stored in the memory indicative of a range of validity of an operation into a plurality of chunks;

        determining a partial convex hull for each of the plurality of chunks;

        determining a convex hull for the operation data based on the plurality of partial convex hulls; and

        determining whether the data exceeds the validity range of the operation data based on the convex hull.

11.   The apparatus of claim 10, wherein the processing unit is arranged to determine a reduced number of dimensions of the operation data.

12.   The apparatus of claim 11, wherein the processing unit is arranged to map the operation data to a model based on the reduced number of dimensions.

13.   The apparatus of claim 10, wherein the processing unit is arranged to randomise the operation data.

14.   The apparatus of claim 10, comprising a plurality of processing units, wherein each processing unit is arranged to determine a partial convex at least partly in parallel.

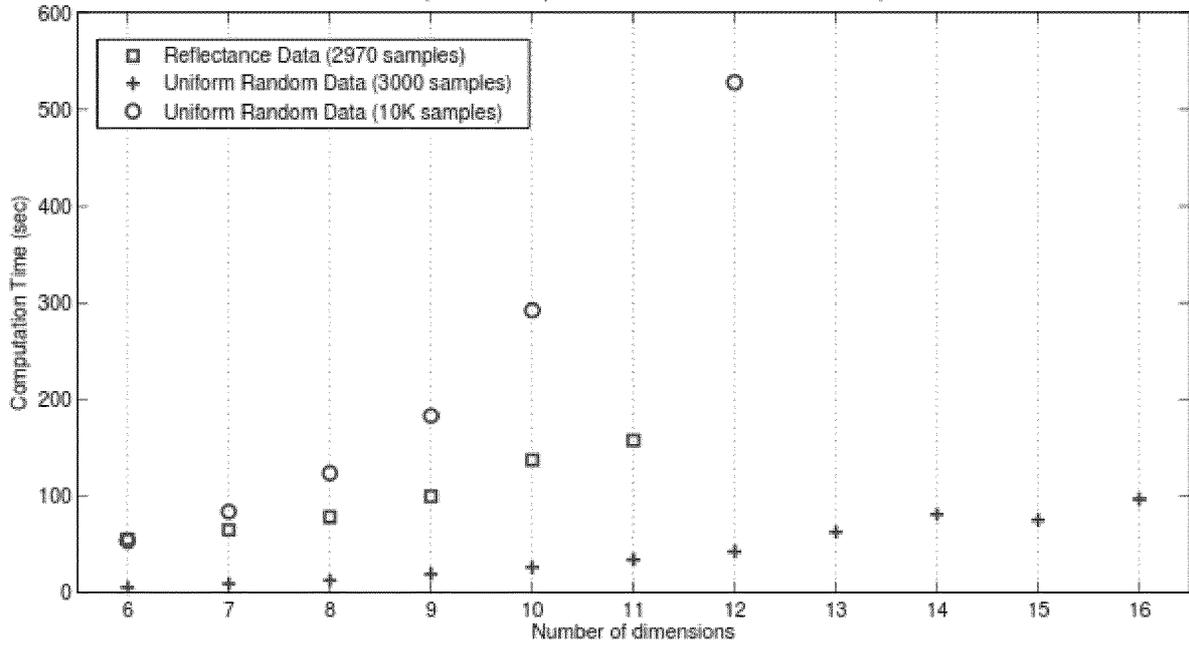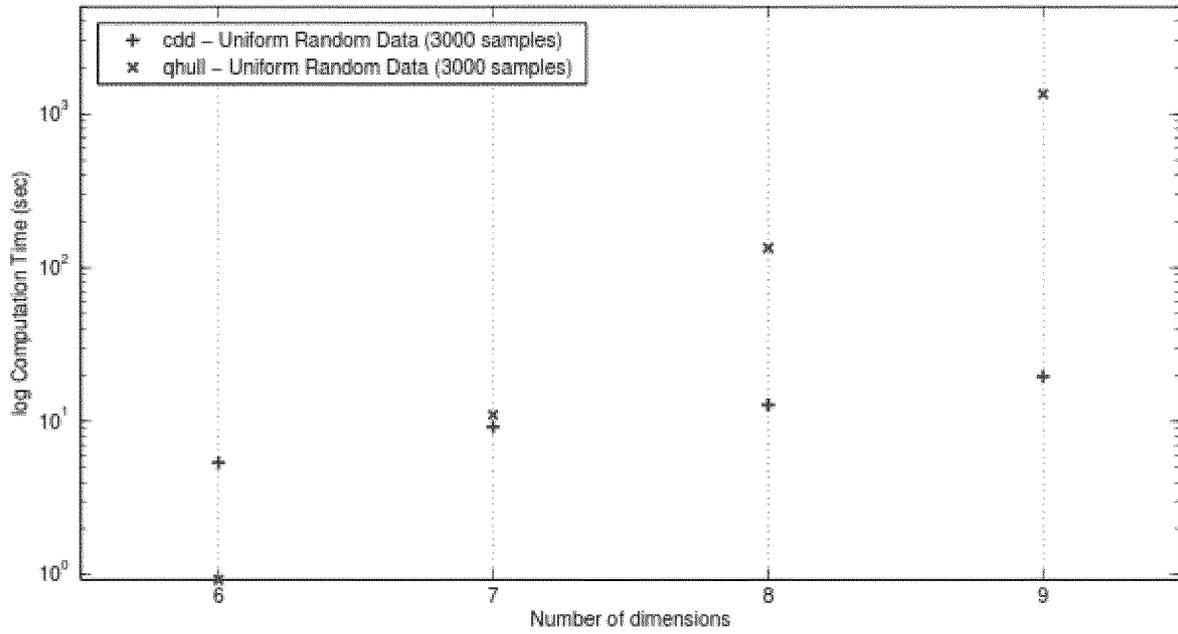5   15.   The apparatus of claim 10, wherein the apparatus is an image processing device.

Fig. 1

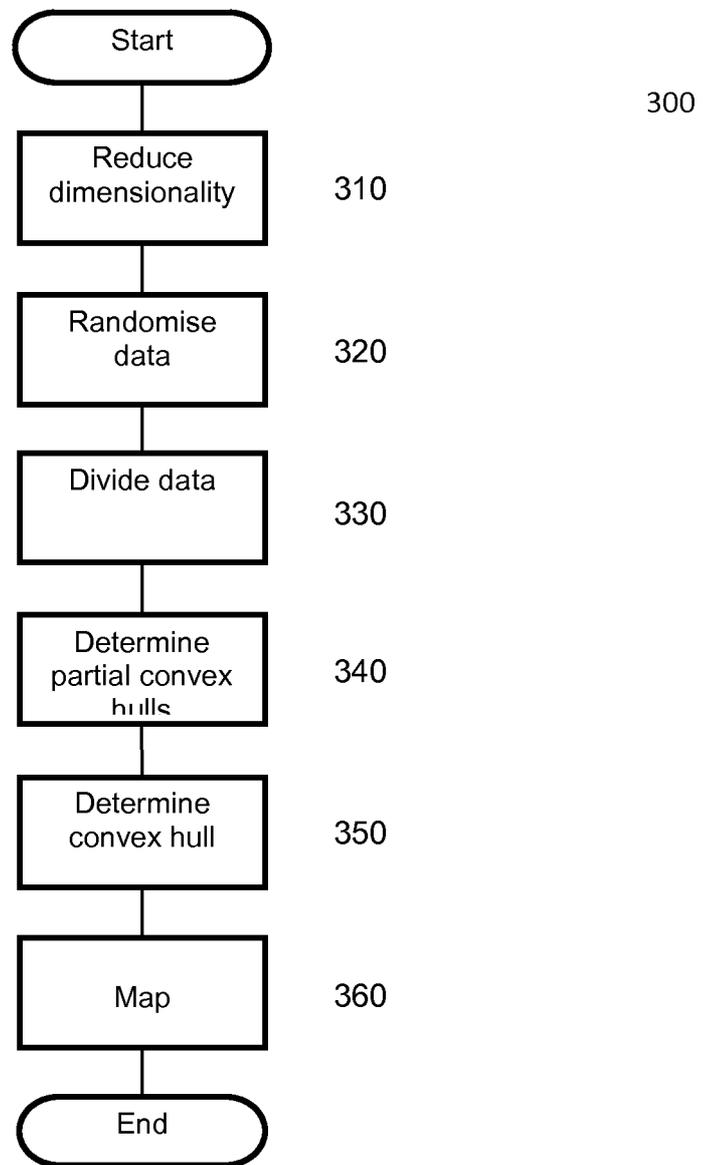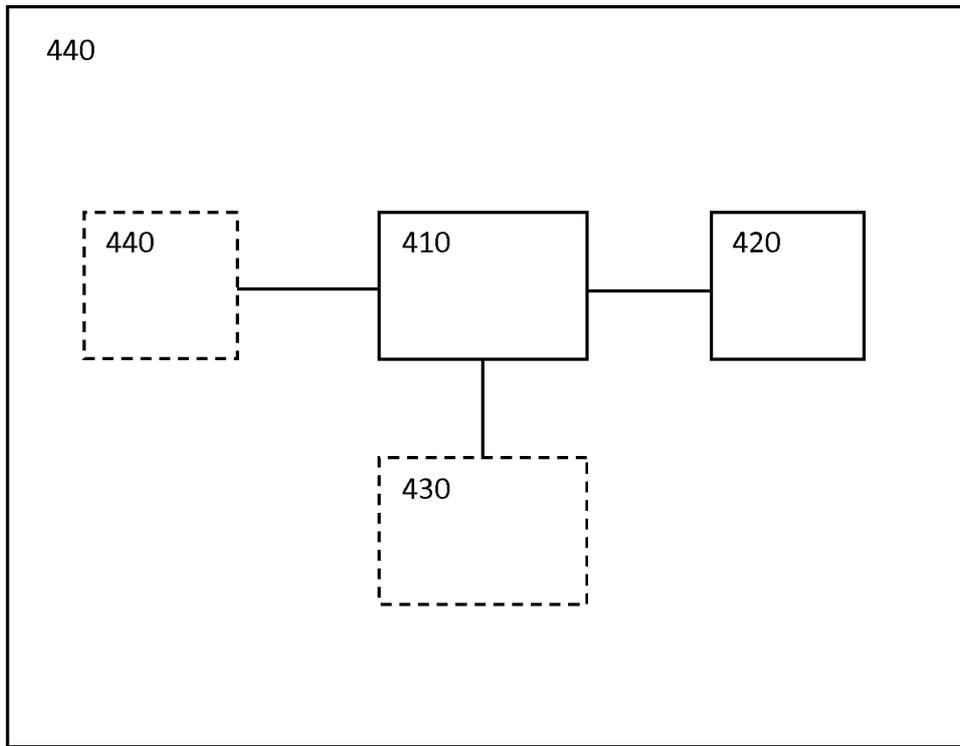Fig. 2

```
                  ┌──────────────┐
                  │    Start     │
                  └──────────────┘
                         │                          300
                  ┌──────────────┐
                  │    Reduce    │
                  │dimensionality│        310
                  └──────────────┘
                         │
                  ┌──────────────┐
                  │  Randomise   │
                  │    data      │        320
                  └──────────────┘
                         │
                  ┌──────────────┐
                  │ Divide data  │
                  │              │        330
                  └──────────────┘
                         │
                  ┌──────────────┐
                  │  Determine   │
                  │partial convex│        340
                  │    hulls     │
                  └──────────────┘
                         │
                  ┌──────────────┐
                  │  Determine   │
                  │ convex hull  │        350
                  └──────────────┘
                         │
                  ┌──────────────┐
                  │     Map      │        360
                  └──────────────┘
                         │
                  ┌──────────────┐
                  │     End      │
                  └──────────────┘
```

Fig. 3

Fig. 4
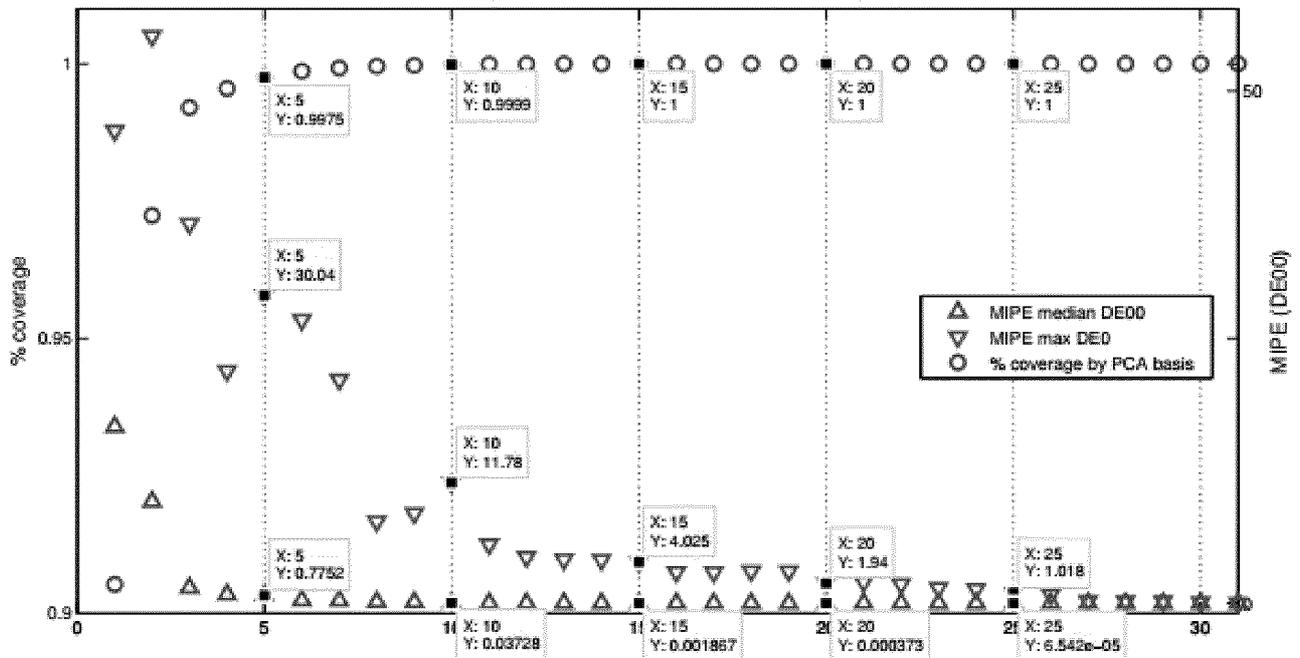
Fig. 5

Fig. 6

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

INV. G06K9/46   G06T13/00
ADD. H04N1/46

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)

G06K   H04N   G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

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

EPO-Internal , WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ARNE M BAKKE ET AL: "Multispectral gamut mapping and visualization: a first attempt" , PROCEEDINGS OF SPIE, vol. 5667, 17 January 2005 (2005-01-17) , XP055180567 , ISSN: 0277-786X, DOI: 10.1117/12.586225 the whole document ----- -/- ·	1-15

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* 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" documentwhich may throw doubts on priority claim(s) orwhich 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	Date of mailing of the international search report
1 April 2015	10/04/2015

Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Bakker, Jeroen

C(Continuation).   DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Rong Liu ET AL: "Convex Hull Covering of Polygonal Scenes for Accurate Collision Detection in Games", Proceedings of graphics interface 2008, 28 May 2008 (2008-05-28) , pages 203-210, XP055180823 , Toronto, Ont. , Canada ISBN: 978-1-56-881423-0 Retrieved from the Internet: URL:https ://www. es.sfu. ca/~haoz/pubs/l iu_zhang _gi08.pdf [retrieved on 2015-04-01] the whole document	1, 10

X	Wo 2012/122309 A2 (UNIV CONNECTICUT [US] ; TEHRANI POOR MOHAMMAD [US] ; WANG XIA0XIA0 [US] ;) 13 September 2012 (2012-09-13) paragraph [0112] - paragraph [0113]	1, 10

X	US 2011/033108 AI (TIN SIU-KEI [US]) 10 February 2011 (2011-02-10) paragraph [0062] - paragraph [0063] paragraph [0122] - paragraph [0128] ; figures 19,20	1, 10

A	MOROVIC P ET AL: "Revisiting spectral printing: a data driven approach" , TWENTIETH COLOR AND IMAGING CONFERENCE. COLOR SCIENCE AND ENGINEERING SYSTEMS, TECHNOLOGIES, AND APPLICATIONS; 12-16 NOV. 2012 ; LOS ANGELES, CA, USA, IS&T: SOCIETY FOR IMAGING SCIENCE & TECHNOLOGY, SPRINGFIELD, VA, USA, 1 January 2012 (2012-01-01) , pages 335-340, XP008172408, ISBN: 978-0-89208-303-9 the whole document	1-15

A	KOMEI FUKUDA ET AL: "Double description method revisited" , 3 July 1995 (1995-07-03) , COMBINATORICS AND COMPUTER SCIENCE, SPRINGER BERLIN HEIDELBERG, BERLIN , HEIDELBERG, PAGE(S) 91 - III , XP047002163 , ISBN: 978-3-540-61576-7 cited in the application the whole document	1-15

1

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 2012122309 A2	13-09-2012	US	2013019324 A1	17-01-2013
		WO	2012122309 A2	13-09-2012
US 2011033108 A1	10-02-2011	NONE		