(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 May 2013 (30.05.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/076583 A2

(51) International Patent Classification:
*H04N 13/02* (2006.01)

(21) International Application Number:
PCT/IB2012/002901

(22) International Filing Date:
21 November 2012 (21.11.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/563,649 25 November 2011 (25.11.2011) US

(71) Applicants: **UNIVERSITE DE STRASBOURG** [FR/FR]; (Etablissement Public National a Caractère Scientifique, Culturel et Professionnel), 4 rue Blaise Pascal, F-67000 Strasbourg (FR). **CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE** [FR/FR]; (Etablissement Public National à Caractère Scientifique et Technologique), 3, rue Michel Ange, F-75016 Paris (FR).

(72) Inventors: **DOIGNON, Christophe**; 8 rue de la Forêt, F-67140 Barr (FR). **MAURICE, Xavier**; 38 rue des Sports, Bât. Escaut, F-31200 Toulouse (FR).

(74) Agent: **CABINET NUSS**; 10 rue Jacques Kablé, F-67080 Strasbourg Cedex (FR).

(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, IS, JP, KE, KG, KM, 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, 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.

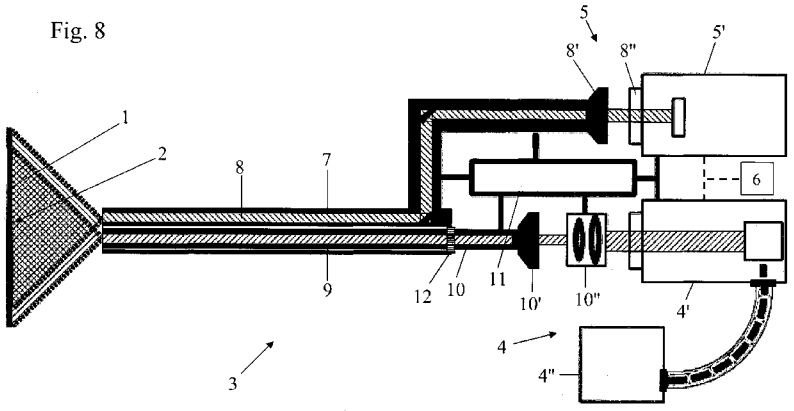
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, 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, ML, MR, NE, SN, TD, TG).

**Published**:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) Title: ACTIVE VISION METHOD FOR STEREO IMAGING SYSTEM AND CORRESPONDING SYSTEM

Fig. 8




(57) Abstract: The present invention concerns an active vision method comprising mainly the steps of projecting a structured light-based pattern with embedded features on a scene and of acquiring or capturing images of said scene, in particular for a single shot 3-D reconstruction, preferably in real-time or near real-time, by using a stereo or multiview system comprising at least one light projecting means and at least one image capturing or acquiring means, said system showing epipolar geometry features, method characterised in that it consists in producing a single- shot structured light pattern (1) which is coded using a 2-D neighborhood and 1-D coding scheme with uniqueness constraint, and in that the or at least some lines of the coded pattern (1), i.e. the resulting matrix of generated codes, are, at least substantially, aligned with the nearly parallel epipolar lines of the stereo or multiview system (3).

WO 2013/076583 A2

# Active vision method for stereo imaging system and corresponding system

The present invention is related to the field of active vision, using a structured light-based stereo or multiview imaging system and aiming for a 3D-reconstruction, in particular in connection with real-time imaging.

The present invention concerns more particularly an active vision method with a highly efficient pattern coding process, and possibly system distortion compensating means, as well as an endoscopic stereo system implementing said method.

Documents illustrating the state of the art in relation to the present invention, or illustrating more precisely some theoretical or practical, and at least partially known, aspects of the invention disclosed herein are mentioned at the end of the present specification and are quoted by numerical references between square brackets.

3-D is gaining more and more place in the industry, and the ability to capture the depth of a priori unknown surfaces is of high importance. To this end, the advantages of active vision (at least one projector) over passive vision are numerous. The projected features are more homogeneously distributed, their geometry and spectrum are known, which makes the matching easier, faster and more robust, and poorly textured surfaces can be reconstructed where passive vision would fail.

But when the scene is highly textured, or when specular surfaces or shadows are present, active vision will meet some difficulties. Indeed, all these spectral variations, together with geometric discontinuities, which are an advantage in passive vision, will be considered as perturbations in active vision, as they will modify the geometry and/or the spectral reflectance of the projected pattern features. Then the correspondence problem will be more difficult to solve.

To tackle these difficulties, it is possible to use an invisible projection. One well-known recent example is the Kinect® device, from the Microsoft® company, which projects an infra-red (IR) active pattern with diffractive optics (see reference [1]), to reconstruct very heterogeneous scenes with one pattern. Nevertheless this solution severely limits the number of different colors that can be used and in some scenes, like intra-

abdominal surfaces, near-red spectrum diffuses too much to be used. To provide systems as robust as possible, numerous coding schemes have recently been developed which ensure the uniqueness of each projected feature of structured light.

Many recent studies (see references [4], [5], (6], [7], [8], [9]) sum-up and compare the different coding strategies. Among them, each feature can be directly, spatially or temporally coded and they can be spectrally and/or geometrically differentiated. Projected features are most often grey-level (see reference [10]) or colored stripes (see references [11], [12], [13]), or punctual spots (see references [14], [15]), but different monochromatic shapes are also used (see references [16], [17]) for more robustness towards spectral perturbations.

It is also possible to choose not to code the features and to use geometrical constraints only (epipolar geometry, local contiguity) like in reference [18], or to use non formal coding like Perlin Noise, in reference [19], where features matching is done with a block matching in two cameras views, coupled with a graph-cut. In reference [20], the pattern stripes are globally coded from the two central stripes which are more spaced than the other, but only few geometrical discontinuities are permitted.

In a direct codification scheme (or continuous coding in reference [8]), each feature embeds itself its own unique information. And to code a high number of features, continuously varying waves are used to ensure the uniqueness of codes. Such approaches are often sensitive to noise due to the very small contrast between successive features and therefore such approaches are suitable for scene having a very constant and homogeneous spectral reflectance.

On the opposite, discrete methods use highly contrasted features and are more suitable to tackle noise and spectral perturbations. Among them, time-multiplexing (see references [10], [21], [22]) consists in the projection of successive patterns to reconstruct scenes in a coarse to fine way. The coding often uses (binary) Gray codes combined with a phase shifting to increase the resolution. As the contrast between projected black and white stripes is optimized, the segmentation is generally very accurate, but many patterns are necessary, which restricts this approach to static scenes.

Finally, in a spatial coding scheme, each feature is coded thanks to its spatial 1-D or 2-D neighborhood. Then all the information is coded in one single pattern, and one-shot, real-time reconstructions, suitable for moving scenes, are possible.

It can be admitted that there is no absolute pattern design which will perform better than any other for all applications. Direct codifications can yield very dense reconstructions while time-multiplexing is adapted for static scenes where a high accuracy is needed and moving objects are better handled with spatial coding. So a choice has to be made depending on the considered application.

In a 1-D and 2-D spatial encoding scheme, k symbols are used to numerically represent the k projected features. Then a neighborhood of size u x v is considered to uniquely encode each of the features in a m x n matrix. The two main strategies using spatial encoding are either 1-D or 2-D based.

In 1-D based coding (u = 1), a sequence of k symbols is generated in which each subsequence of length u, is unique. Then associated features are most often colored stripes. In 2-D based coding, a matrix of k symbols is generated in which each sub-matrix of length u x v is unique. Then associated features are punctual and can be colored, grey-level spots or different shapes can be used.

More than ensuring the uniqueness of codes within the sequences/matrices, recent works strive to impose a high redundancy thanks to the use of the Hamming distance (i.e. the number of positions is which the symbols of two codes differ) well-known in information theory. Indeed, with a high redundancy, systems can cope with misclassified features (when relating segmented features to a symbol) that can occur due to the geometric and spectral perturbations in the scene. The minimal Hamming distance ($H_{Min}$) (i.e. the smallest Hamming distance between all couple of codes of the m x n matrix) is often used as the redundancy criterion as it ensures that if the code has $H_{Min} \geq 2M+1$, then the decoding can be robust up to M erroneous classifications (that is when relating segmented visual features to one of the symbols).

Apart from redundancy, a high matrix/sequence resolution is often suited for a finer reconstruction and today projection/acquisition devices can handle high resolution patterns. A small number of symbols (k) will increase the geometrical/spectral contrast between each symbol and

will, as a consequence, lead to less classification errors. Yet, with less symbols there will be less redundancy in the code, so a balance has to be found. The size of subsequences/sub-matrices should also remain small to be able to decode features by the examination of a restricted area, which allows more geometrical curvature discontinuities.

Some spatial codes generation techniques use direct methods with De-Bruijn sequences in 1-D (in e.g. references [2], [28]) or Galois Fields (in e.g. references [3], [29]) and their 2-D extensions (in e.g. references [30], [31]), to generate Perfect Maps (PM) (figure 2), in which all the submatrices are present exactly once. But as all submatrices are present, then $H_{Min}$ cannot be higher than 1, hence no redundancy at all. As there is no direct method that allows to constrain the $H_{Min}$, so-called "brute-force" algorithms have been developed and used in 1-D coding (in reference [13]), and 2-D coding (in references [14], [15], (17]). Brute-force algorithms proceed by trials and error, checking the $H_{Min}$ and/or any other constraints (central symmetry in reference [17], different adjacent symbols in reference [13], rotational invariance in references [15] and [24]) each time a new symbol (S) is added to the matrix, which is initially empty. As not all submatrices are used, this coding is called Perfect SubMaps (PSM).

When considering "brute-force" coding algorithms, two main issues must be addressed, namely: 1) what is the search behaviour and 2) how is the unicity or uniqueness test done.

When the uniqueness test, which checks the $H_{Min}$ constraint, fails, the search behaviour defines how the algorithm behaves.

While a random approach was used in reference [14] and a full exhaustive search was proposed in reference [15], a better search behaviour was obtained in reference [32]. The latter method used a trade-off between exploration and exploitation of the search space thanks to a locally exhaustive search in a “correction window”.

Furthermore, in reference [32], a way the uniqueness test could be realised is addressed. In this document, it is done between groups of features having nearby epipolar lines thanks to the use of ambiguous areas. This introduced the SubPerfect SubMaps (SPSM), which are PSM where the test of uniqueness does not hold in the whole matrix. Also, in reference [23], it is proposed to implement the test of uniqueness in the codes index space instead of the classical matrix space (by comparison of already selected codes). The results were interesting for PSM only, as those only

have a large amount of codes to be checked in the matrix space, contrary to SPSM.

In view of the previously presented state of the art, the first aim of the present invention is to propose, within the aforementioned active vision context, a coding scheme, preferably using a discrete spatial coding method, which shows higher performances in comparison to existing coding schemes, in particular concerning the size of the used patterns, the Hamming distance, the number of coding symbols, the generation speed and/or the decoding speed.

Therefore, the present invention concerns an active vision method comprising mainly the steps of projecting a structured light-based pattern with embedded features on a scene and of acquiring or capturing images of said scene, in particular for a single shot 3-D reconstruction, preferably in real-time or near real-time, by using a stereo or multiview system comprising at least one light projecting means and at least one image capturing or acquiring means, said system showing epipolar geometry features,

method characterised in that it consists in producing a single-shot structured light pattern which is coded using a 2-D neighborhood and 1-D uniqueness constraint coding scheme, and in that the or at least some lines of the coded pattern, i.e. the resulting matrix of generated codes, are, at least substantially (i.e. up to estimation errors in the epipolar geometry identification), aligned with the nearly parallel epipolar lines of the stereo or multiview system.

By providing a coding scheme with a 2-D neighborhood and a 1-D uniqueness property, the invention actually provides a totally novel hybrid 1.5-D coding scheme.

Thus, the basic principle underlying the invention consists in taking into account the stereo system geometry already at the pattern design step and to exploit advantageously specific properties of said geometry, in particular related to the epipolar lines homography (independence from the scene).

By aligning, fully or at least substantially, the projected coded pattern with the epipolar geometry of the system, the coding constraint can be lessened (increasing of its size, increasing of its redundancy, decreasing the number of coding symbols, hence decreasing the probability of potential misclassifications and, by consequence, decreasing the probability of

decoding errors). Indeed, the uniqueness (or unicity) of the codes or codewords needs only to be verified between codes associated to the same line.

Furthermore, the skilled person is aware that, during the acquisition, when using 2-D neighborhoods (3 x 3 submatrix, in most cases), a non-trivial task is to search for the contributing segmented features belonging to each neighborhood, so as to compute the corresponding code. To do this, most often a proximity criterion is considered: given a visual feature (the central one in the 3 x 3 neighborhood), the closest ones in the acquired image on the top, bottom, left, right sides and corners are approached as the top, bottom, left, right and corners neighbors, respectively, to retrieve the original code. But when surfaces orientations are off-plane, this criterion is less efficient.

In this content, it must be emphasized that the projection of an epipolar-aligned grid pattern is very suitable to that purpose: even if the visual features embed a direction information, the proximity criterion is inefficient for handling many deformations while looking for nearest neighbors instead of taking account for a local affine displacement of the neighbors. Thanks to the epipolar-aligned arrangement of all the visual features, these deformations can be modelled with only two parameters leading to a more efficient search of neighbors, thus a more reliable code retrieval.

The method according to the invention may also incorporate or show one or several of the following additional features or alternative embodiments:

- the constraint of uniqueness is verified or computed for a limited number of neighborhood lines, preferably only one (i.e. computed for one line only but ensured for all lines in the pattern);

- the step of producing the matrix of codes consists in first coding a number v of first lines $L_i$ of the matrix, preferably using a sizeable correcting window (for example like in reference [32]) and optimal codes resulting from a fully exhaustive search, and then producing or filling the remaining lines $L_i$ of the matrix, step by step, by copying the codes of the preceding and previously filled lines $L_{i-v}$, in such a way that the constraint of uniqueness holds also for the copied lines (for example $v = 3$, as in figure 1, or $v = 5$);

- the projected pattern is an adaptative pattern, adapting itself, in particular in terms of resolution and/or spectral properties, to local features and/or specificities of the scene to be imaged, by updating the numerical coding of the pattern to said scene specificities and/or features;

- a given periodicity is provided along the lines of codes of the matrix;

- the projected pattern, for example a regular grid of visual symbols or a similar two-dimensional regular pattern, is subjected to a global rotation such that the average orientation of the lines of the grid or coded matrix, or of the features of the pattern, corresponds to the average orientation of the epipolar lines in the plane of the projecting means;

- a central symmetry constraint is included during coding of the pattern, the central symmetry of the features of the pattern being preserved in case of a global rotation of the projected pattern;

- in view of the 3-D reconstruction of each image of the scene, a decoding is performed on the basis of a belief propagation scheme from feature correspondences with a high belief, using an adjacency graph to define the frontiers of propagation and a selection function based on a lower Hamming distance test;

- additional image data is provided through at least one passive vision component, in particular image data related to the texture of the scene, and mixing in said passive vision image data with the collected active vision image data;

- the stereo or multiview pattern projecting and image capturing system is a video system, comprising a video projecting means and a video acquiring means, the acquired image data being preferably exploited to perform visual control, robot guidance or servoing;

- the inventive method also comprises a preliminary process for estimating and/or correcting the geometric distortions of the stereo or multiview system, in particular video system, preferably of the endoscopic type, by performing a separate estimation, and a possible separate correction or compensation, of the distortions produced or induced respectively by the projecting means and by the acquiring or capturing means;

- the estimation and the possible rectification of the distortions of the projecting and/or capturing means are performed by using the plumb-line principle;

- the method consists, after having estimated and possibly corrected the geometrical distortions linked to the image acquiring or capturing means (known as such in the state of the art), in performing an a priori software based estimation and correction of the geometrical distortions of the projecting means, preferably by realising multiple projections and captures of successive distorted patterns so as to gradually improve the straightness of the projected pattern and aiming towards a fully straight projected pattern;

- the estimation and rectification of the distortions of the projecting means consists, using an adequately programmed software, in estimating and correcting the geometrical distortions related to the pattern projecting means by: a) projecting a regular pattern, for example a regular grid with horizontal and vertical lines, on a planar or substantially planar item or object (or a planar surface), b) acquiring the image with the capturing means, c) evaluating distortion parameters by using an adequate distortion model, d) repeating steps a) to c) by acquiring at each cycle the image of the pattern as distorted by the updated parameters of the used distortion model, until distortion parameter values are reached which lead to a minimised distortion of the projected pattern, or at least to a distortion within a tolerance level;

- the estimation and rectification of the distortion parameters of the projecting means are performed by means of an image based optimisation function incorporating a cost reduction function and by using an adapted distortion model, such as the Brown distortion model;

- when the stereo or multiview system is an endoscopic system, an estimation of the endoscopic circle (i.e. the endoscopic channel borders appearing inside the visual field of the endoscopic image capturing means) is performed by using an optimisation function with a cost reduction function and taking into account the contrast between the inside and the outside of the considered circle.

The present invention also concerns, according to an other aspect, a stereo endoscopic system for projecting and capturing a video flow through a single physical passage, for example a single trocar in a mini-invasive context, said system comprising at least means for projecting a structured light-based pattern on a scene and means for capturing or acquiring in real time images of said scene,

system characterised in that it also comprises means, in particular computer and software means, for implementing the method as described before.

Preferably, the system comprises one projecting means and one image acquiring means.

The present invention will be better understood thanks to the following description and drawings of one embodiment of said invention given as non limitative example thereof, wherein:

- Figure 1 illustrates schematically the 1.5-D coding scheme of the invention showing the coding of the HA first lines of symbols, the copying of their content in the following HM-HA lines of the matrix and the mapping of the code lines of the matrix to epipolar lines;

- Figures 2A and 2B illustrate schematically a possible design of a 50 x 50 pattern built upon the epipolar geometry, with figure 2A showing the relocalisation of the pattern (red) from a grid (green) onto the closest epipolar lines (black) in the image of the projector ($e_c$: the epipole/k: number of symbols / $H_{Min}$: minimum Hamming distance) and figure 2B representing a final pattern in the image of the projector after features alignment (translation, rotation) along chosen epipolar lines (resolution : $50^2$, $k = 8$, $H_{Min} = 6$).

- Figure 3 shows another example of the alignment of a pattern of cuneiform symbols (figure 3A) with the epipolar lines of the system (figures 3B and 3C);

- Figure 4 illustrates schematically a pinhole model of a stereo system (with one projector and one camera) showing the epipolar geometry and the geometrical transformation of the pattern before projection and after/during acquisition (Epipoles: $e_p$, $e_c$ / Optics: $O_p$, $O_c$ / Optical centers: $C_p$, $C_c$);

- Figures 5A to 5C illustrate an example of estimation and correction of the optical distortions of the capturing means (camera);

- Figures 6A to 6C illustrate an example of estimation and a priori correction of the optical distortions of the projecting means (projector);

- Figures 7A and 7B illustrate the projection of epipolar lines on a scene with geometric discontinuities (figure 7A) and the valid epipolar lines homography (figure 7B - camera distortions not corrected, hence continuous but curved epipolar lines);

- Figure 8 is a schematical and functional representation of an embodiment of a stereo endoscopic system according to the invention;

- Figure 9A shows the projection of a 100 x 150 pattern on a hand (with $H_{Min}$ = 4 and C = 4 symbols - associated to cuneiforms) and figure 9B illustrates four successive images of the reconstruction of a real-time closing hand;

- Figures 10A to 10D illustrate, through various images, the robustness of the method with relation to geometrical difficulties (figures 10A and 10B - occlusions by thin objects) and to spectral difficulties (figures 10C and 10D);

- Figure 11A illustrates the projection of a 70 x 70 pattern on a head sculpture showing two faces with different expressions ($H_{Min}$ = 2 and k = 2 symbols) and figure 11B illustrates the reconstructed image of the faces allowing to distinguish the two facial expressions;

- Figure 12 illustrates, by way of four images (a) to (d), in vivo, mini-invasive reconstructions of the stomach (bottom left), the liver (right) and the gallbladder (center-blue);

- Figure 13 illustrates in vivo, mini-invasive, organs reconstructions and rendering (the inner surface (b,c,f), the stomach (a,d), the liver (c,d), the intestine (a,b,e), the gallbladder (d,g,h)) with $M^2$: the resolution of the SPSM coded aligned pattern, K: the number of symbols, $H_{Min}$: the minimal Hamming distance, PP: the number of projected features, PC: the number of features captured by the camera, PD: the number of decoded features and MC: the estimated number of misclassifications;

- Figure 14 illustrates perfect maps variants with (a): A 2-ary perfect map (5,5; 2,2); (b): A 3-ary perfect submap (5,5; 3,3) with $H_{Min}$ = 5; (c): A 3-ary subperfect submap (5,5; 3,3) with $H_{Min}$ = 9, uniqueness holds between neighborhoods on the same line only;

- Figure 15 is a comparative table showing the coding results for PSM in reference [25] (column C), in reference [14] (column M) and in reference [23] (column X);

- Figure 16 is a table showing the coding results with the method according to the invention (N.B.: in figures 13 and 14: k is the number of symbols, $H_{Min}$ is the minimal Hamming distance and u x v = 3 x 3).

As mentioned before, and as illustrated in particular in figures 1 to 4, the present invention concerns primarily an active vision method

comprising mainly the steps of projecting a structured light-based pattern 1 with embedded features or symbols on a scene 2 and of acquiring or capturing images of said scene 2, in particular for a single shot 3-D reconstruction, preferably in real-time or near real-time, by using a stereo or multiview system 3 comprising at least one light projecting means 4 and at least one image capturing or acquiring means 5, said system 3 showing epipolar geometry features.

According to the invention, said method consists in producing a single-shot structured light pattern 1 which is coded using a 2-D neighborhood and 1-D uniqueness constraint coding scheme, and in that the or at least some lines of the coded pattern 1, i.e. the resulting matrix of generated codes, are, at least substantially, aligned with the nearly parallel epipolar lines of the stereo or multiview system 3.

Thus, the proposed inventive approach relies on the 1-D homography which relates epipolar lines in the IPP (Image Plane of the Projecting means) to those in the IPC (Image Plane of the Capturing means), whatever the geometry of the scene. If a set of epipolar lines $P_p$ intersect at the epipole $e_p$, in the IPP, then corresponding epipolar lines in the IPC, $P_c$ will intersect at $e_c$ and verify:

$$P_p \equiv F[e_c] \times P_c$$

with F, the Fundamental matrix.

The invention proposes to exploit this geometrical constraint for the pattern coding, the pattern alignment through the choice and positioning of features and, finally, for the 3-D reconstruction.

Indeed, if the pattern is geometrically aligned on the epipolar geometry, that is, rows of features are mapped to epipolar lines (figure 3), then the uniqueness of features has to hold only for features aligned on the same epipolar lines. Moreover, for stripes projection, or features having cross shapes, the orthogonality between epipolar lines and stripes can be maximized and so the features segmentation accuracy.

The herein mentioned epipolar-aligned patterns can be seen as the equivalent of rectification in the camera plane (figure 3A), but instead of laying epipolar lines horizontally as with a rectification, initially horizontal alignment of features (figure 3A) are given, the position and orientation of epipolar lines (figure 3B).

As the epipolar geometry relies on the pinhole projection model, geometrical optical distortions have to be corrected both on the

camera and projector image planes. Therefore, the inventors have developed, in connection with the invention, an appropriate methodology to estimate those distortion parameters for the projector and lead to patterns where the geometrical distortions of the projector are compensated a priori (figure 3B).

This point is of practical importance in endoscopic applications as the optics of the endoscopes yield significant distortions.

The used coding defines both the pattern size, the number of symbols, the size of neighborhoods and the redundancy of the projected pattern ($H_{Min}$).

As indicated and according to the invention, a new 1.5-D coding scheme is provided. The main idea is to use a 2-D neighborhood to have longer codes than in 1-D, but to hold the uniqueness in 1-D only, that is along matrix lines that will be aligned with epipolar lines. This results in a far less constrained coding problem and thus produces very interesting results in terms of redundancy and/or matrix size.

If the epipole is sufficiently far from the visible part of the camera image plane, then epipolar lines will appear nearly parallel and lines of the matrix of features can be directly mapped to regularly spaced epipolar lines without an important loss of homogeneity of their spatial distribution. Then the uniqueness has to hold between the neighborhoods that are on the same SPSM line only. Indeed, due to the epipolar lines homography, those neighborhoods only can lead to matching ambiguity. Thanks to a Hamming distance property and a line copy mechanism, only the first line of neighbors have to be coded.

Indeed, as according to the invention a recopy mechanism is used to fill the entire matrix, only the HA first lines of symbols, for a neighboring size of HA x WA, need to be coded. As the Hamming distance is computed position-wise and the uniqueness is 1-D, along each line, the first line, recopied after to the $Ha^{th}$ line will yield a new line of HA x WA neighborhoods with same Hamming distance properties as the HA first.

Thus, as described and illustrated by way of example in figure 1, and in case the uniqueness has to hold only among neighborhoods on the same row (the epipolar geometry is accurately estimated), the inventive method proposes advantageously to proceed as follows: Once the first row of HA x WA neighborhoods is coded, that is, HA (3 in this example) rows of the coding matrix are filled with numerical symbols, the first row of

symbols is entirely re-copied (from symbol S1,1 to S1,Wm) to the HA + 1th row (from symbol S4,1 to S4,Wm), the second row is re-copied to the HA + 2th row, the third to the HA + 3th row etc... with HA (= 3) rows offset each time until the whole matrix is filled. This way, the copied rows of symbols will generate new (3 x 3) neighbors which respect minimal Hamming distance property (among all the neighborhoods on their respective row), because the Hamming distance is computed position-wise along each row and the symbols position is respected along each row during the copy mechanism.

As a result, and in order to compute the previous tasks, a possible algorithmic scheme is advantageously as follows:

1) Code the first line of (for example: 3 x 3) neighborhoods of the SPSM (the first v lines of symbols are filled). To do so, the locally exhaustive approach presented in reference [32] may be used.

2) Fill as many other lines $L_i$ as needed, step by step, by copying the symbols of the line $L_{i-v}$, where $L_{i-v}$ has been previously filled. In this manner, the same minimum Hamming distance property must hold line-wise between new 2-D neighborhoods created when copying new lines of symbols.

3) Map successive lines of the SPSM to successive epipolar lines in the projector image plane. Epipolar lines should be preferably chosen so as to be in the camera field of view and such that their distribution is as homogeneous as possible.

This approach highly reduces the coding complexity as:

- the uniqueness constraint has to hold in 1-D only
- only the first line of codes has to be coded, the other are obtained by a copy mechanism.

As the number of rows can be virtually infinite (as first coded lines can be infinitely recopied), those results are far better, in terms of resolution and redundancy, other parameters being equals, with what can be obtained with PSM as in references [25], [14] and [23]. Even results generated in real-time with a simple C implementation (no SIMD or GPU) outperform PSM results. This allows to update the pattern numerical coding to scene specificities (features density, symbols number) in an adaptive pattern framework. This was not the case in previous works dealing with adaptive patterns (cf. references [36], [18], [34]).

Additionally, two alternative embodiments (in comparison to the [first row coding + recopy] solution) must also be considered in relation to the invention, even if less efficient than the preferred solution:

- instead of using a recopy mechanism, all other (than the HA first) rows of the matrix might be coded independently from the HA first ones;

- if the epipolar geometry is not accurately estimated and a margin has to be considered, the uniqueness might hold among neighbourhoods belonging to multiple adjacent rows.

Thanks to the 2-D to 1-D complexity reduction, usable optimal codes can be generated. Those are obtained by a fully exhaustive search (without any random component, like in [14] or [32], thus ensuring that all configurations have been tested and no bigger size can be obtained for the k, $H_{Min}$ and u x v = (3 x 3) input parameters.

Example of optimal code: k = 2; $H_{Min}$ = 3; 39 columns:

0 0 0 1 0 0 0 1 0 1 0 0 1 1 0 0 0 1 1 1 1 0 1 1 0 1 1 1 0 1 0 1 1 0 0 0 1 0 0
0 0 0 1 1 1 1 1 0 0 1 1 0 0 0 1 0 0 0 0 1 0 1 0 0 1 1 0 1 0 1 0 1 1 0 1 1 0 0
0 0 1 1 0 0 1 0 1 0 1 0 0 1 1 1 1 1 1 0 1 0 0 1 1 0 0 0 0 0 1 1 0 0 1 1 0 1 0 1

Example of optimal code: k = 3; $H_{Min}$ = 6; 26 columns:

0 1 0 1 0 0 0 2 1 0 0 2 2 0 1 1 1 1 2 2 0 2 1 2 0 1
0 2 2 1 1 2 0 0 0 0 1 1 2 1 0 1 2 0 2 0 2 2 2 1 0 2
0 0 1 1 2 1 0 1 2 0 2 0 2 2 2 1 0 2 2 1 1 2 0 0 0 0

With 26 and 39 columns and a virtually infinite number of rows, those resolutions are sufficient to be used in practice and offer a very high redundancy for very few symbols. This was not the case in reference [15] where obtained PSM optimal codes where too small for a practical use.

Previously, as in reference [32], epipolar lines proximity was exploited to restrict the area in which the uniqueness test has to hold, but yet all the features had to be coded independently from the others while, according to the invention, preferably only the first line of codes has to be coded, hence an important gain in coding speed.

Further coding improvement could be obtained by combining 1.5-D SPSM coding and introducing a periodicity along lines of codes, as this is done with 1-D coding to benefit from limited disparities of the features.

The present invention also deals with the problems caused by the geometrical distortions generated within the stereo or multiview system 3, in particular when the latter is a video system of the endoscopic type.

Indeed, as the processing pipeline highly depends on the epipolar geometry, and this is based on the pinhole model, geometric distortions have to be corrected for both the projector and the camera. In the literature, we did not find any method to estimate and correct the geometric distortions for the projector without making a global optimization, like in [37] and [2]. In a global optimization, the projector is viewed as an inverted camera. In most of the times, the projectors distortions are corrected by the manufacturer, but in some cases, the manufactured miniaturized projection system yields significant pincushion distortion due to the endoscopic optics. This is also the case for the Kinect® device mentioned before.

To overcome these problems, the present invention proposes the following distortion estimation and correction method.

As the Fundamental matrix, which fully characterizes the epipolar geometry, can be computed with points correspondences only (see reference [38]), an image-based approach is chosen to estimate the distortions for both the camera and the projector. The Brown distortion model as described in reference [39] and the "Plumb Line" principle were used: straight lines in 3-D space have to remain straight when projected on the image plane of the camera and of the projector. Nevertheless, other distortion models can be used, as it is known to the man skilled in the art.

Using the "plumb line" approach for the camera has been already studied in the literature (e.g. references [40], [41], [42]). The distortion model parameters (for example $C_X$, $C_Y$, $K_1$, $K_2$, $P_1$, $P_2$ in the Brown model) are estimated by an optimization process which minimizes the straightness error of segmented lines in the grabbed images. Nonetheless this method cannot be directly applied to the projector which does not make any acquisition. Indeed, one can consider that once the distortions parameters related to the camera optics have been found and corrected (e.g. using a regular grid in the 3-D space), then, if a regular grid is projected on a plane, one can consider that the curved lines of the captured grid are now due to the projector optical distortions. But the distortions parameters found using the acquired images of distorted grid would not give any clue on how to modify the projected grid. Indeed, the whole stereo geometry (intrinsic and extrinsic parameters) would have to be known.

So the following solution is proposed in connection with the invention. Instead of projecting a regular grid and trying to find the

projector distortion parameters once and for all with a single acquisition, the projected grid will be continuously distorted in the projector plane though a sequence of projections until the distortion parameters -that minimize or nullify the lines straightness errors in the camera image- are found. To find those parameters, the cost function of the optimization function previously used for the camera will be modified, for each call of the cost function, as follows:

1) The cost function is called from the optimization process with a new set of distortion parameters to evaluate for the projector and projected on a planar item or object, preferably facing the stereo system.

2) The projected grid pattern, PG, is distorted in the projector plane according to the distortion parameters to try.

3) This modified grid is captured, the camera trigger must be synchronized with PG for the acquisition.

4) The previously estimated camera-related distortions are corrected on the acquired image.

5) The segmentation and straightness evaluation steps are performed on the image corrected from camera distortions. Hence the computed straightness score is related to the projector distortions only.

6) The cost function returns this straightness score to the optimization process.

Once the projector optimization process is completed (after a target straightness score has been reached to a user-defined threshold), the displayed grid pattern is distorted in the projector plane (figure 6B) with the set of distortion parameters that yielded the less distorted grid in the camera image plane. With this distorted projection, the projector optical distortions are compensated a priori, and, thus, lines are straight in the 3-D scene (figure 6C), which was not the case when a regular, undistorted grid pattern was projected (figure 6A). With this a priori compensation in place, the pinhole model does apply with minimal errors, in the stereo system 3, and the Fundamental matrix can be computed more accurately from point correspondences.

One important feature of this method is that, as the camera and the projector distortion parameters are evaluated independently, no coupling is expected between the distortions of the two devices. This is not the case with a general optimization method where all parameters are mixed in the same process.

The present invention also concerns, as shown schematically in figures 4 and 8, a stereo endoscopic system 3 for projecting and capturing a video flow through a single physical passage, for example a single trocar in a mini-invasive context, said system 3 comprising at least means 4 for projecting a structured light-based pattern 1 on a scene 2 and means 5 for capturing or acquiring in real time images of said scene 2.

According to the invention, said system 3 is characterised in that it also comprises means 6, in particular computer and software means, for implementing the method as described before.

According to an advantageous embodiment of the inventive system, this latter comprises a first endoscopic device 7 with an optical passage or channel 8 and at least one other or second channel or passage 9, for example an operator channel, and a second endoscopic device 10 which is fitted or accommodated within said other or second channel 9 of the first endoscopic device 7, one 7 or 10 of said endoscopic devices being functionally connected to the projecting means 4, in particular projecting means for time-varying patterns like video projector, and the other 10 or 7 of said endoscopic devices being functionally connected to the image acquiring means 5.

Thus, only one physical passage is needed when using the system 3.

In order to structure the composite system as a single rigid construction, made of commercially available components, the invention can provide that the pattern projecting means 4 comprise a miniaturised video projector 4' incorporating or connected to a light sources 4", in that the image capturing means 5 comprise a video camera 5' and in that both endoscopic devices 7 and 10 and the projecting and capturing means 4 and 5 are physically, and preferably rigidly, linked together by a contention system 11, possibly incorporating or supported by a bearing arm or stand.

Preferably, in order to optimise the optical performances of the system 3, the optical channel 8 of the first endoscopic device 7 and the second optical endoscopic device 10 comprise optical connectors 8' and 10' at their ends which are connected to the projecting or capturing means 4, 5, preferably through interposed interface optics 8", 10".

According to an other advantageous feature of the invention, the pattern projecting means 4 incorporate means able to perform a global rigid rotation of the projected pattern 1 or of its coded matrix in order to

achieve a global alignment with the average epipolar lines direction of the stereo system 3, the central symmetry of the features of the pattern 1 being maintained. Thus making this epipolar-aligned approach better suited to the use of diffractive optics elements (DOE), where a central symmetry allows a better illumination efficiency. In such a case, the uniqueness of codes would possibly hold between more than one row of neighborhoods in the pattern as epipolar alignment would be only approximated for some pattern rows.

As shown on figure 8, and in relation to a first embodiment of the invention, the first endoscopic device 7 is an angled rigid endoscope with a straight operator channel 9 and in that the second endoscopic device 10 is a straight endoscopic device with at least an optical channel, the insertion port of the second endoscopic device comprising a sealing means 12.

According to a second alternative embodiment not shown on the figure, at least the second endoscopic device, preferably both endoscopic devices, is(are) flexible endoscope(s).

The inventors have, as an example, implemented an endoscopic integrated composite system 3 wherein the first device 7 is a Karl Storz angled rigid endoscope of 10 mm diameter with an operating channel of 6 mm diameter, the second device 10 is a Karl Storz straight rigid endoscope of 5 mm diameter, the video camera 5' is a Karl Storz image 1 HD camera head and the projector is a DLP projector (for example: HP Notebook Companion®). Furthermore, a set of additional interface optical lenses 10" can be provided between the optical connector 8' and the projector exit in order to adjust the field of view of the projector 4' and its focus for a working distance adapted for mini-invasive scene imaging (for example 2 - 20 cm for intra-abdominal scenes).

The light source 4", for example a Storz Xenon device, can be connected to the projector 4' by a light guide, as shown on figure 8.

A computer or CU based system 6 which is able to perform the inventive method can be provided as a separate device or integrated with the composite system 3. Said system 6 may also control the means 4 and 5, and possibly process and/or exploit the acquired images.

In order to evaluate the robustness of the method and of the system, the inventors have used the two stereo configurations presented previously with scenes containing geometrical and spectral perturbations.

Reconstructions with a classical stereo system have been realised with an epipolar-aligned monochromatic 100 x 150 pattern using $H_{Min}$ = 4 and 4 cuneiform shapes. Reconstructions depicted in figure 10 show the ability of the system to reconstruct scenes with significant geometric discontinuities (figures 10A-B) and spectral disturbances (figures 10C-D). They were executed at 27 reconstructions per second with a GPU (NVidia GeForce 470 GTX) implementation. In figure 9, a closing hand is captured in real-time, thanks to the pattern high resolution with respect to those usually generated with this approach, the efficiency of the proposed coding scheme allows sufficiently high resolution pattern to capture the fingers during the whole sequence. That would not be the case with typical patterns resolutions using 2-D coding.

Reconstructions with the endoscopic stereo system have been realised with an epipolar-aligned monochromatic 30 x 30 to 70 x 70 pattern using $H_{Min}$ = 2 and 2 shapes (2 symbols) were the energy gradient is maximized in the vertical direction (that is perpendicular to the epipolar lines). Due to additional processing steps to deal with the significant noise due to the very small size of the endoscopic optics, they were executed at 18 reconstructions per second with a GPU implementation. Figure 3C depicts such a 30 x 30 pattern. Reconstructions depicted in figure 11 show the ability of the system to capture sufficiently fine surface details to recognize the face expressions on an ex-vivo scene with a 70 x 70 pattern.

Mini-invasive, in vivo, reconstructions of intra-abdominalscenes of a pig have also been realised. The coding conditions were the same as with ex vivo scenes.

Numerous organs surfaces could be reconstructed despite the numerous spectral perturbations (specularities, inter-organs light reflections, mirror effect due to wet organs, the blood system, and very noisy images). To tackle those perturbations, the decoding was realised by belief propagation from features correspondences with a high belief (which have a low Hamming distance between codes in the projector and the camera images) to the locally adjacent ones. To define the frontiers of propagation, an adjacency graph is built in which frontiers are defined by features neighborhoods which maximize the elongation S and skew angle $\alpha$ local variations. If two different correspondences are propagated to a feature, the one which yields the lower Hamming distance with the proposed SPSM code is selected.

As it is obvious for figure 14, reconstruction statistics show that despite the high rate of estimated misclassifications (when assigning a segmented feature to symbol), between 1/4 and 1/3 of features, the belief propagation decoding approach enables the decoding and reconstruction of nearly all the segmented features.

Herein, there are presented, in relation to the invention, two new uses of the epipolar geometry before and after the classical matching step where it is usually used.

First, it is proposed to use it at the pattern design step for a coded structured light-based approach for a new numerical 1.5-D Sub Perfect SubMap coding scheme that yields very interesting coding results. This is done thanks to the geometrical alignment of the pattern to the epipolar geometry in order to exploit the epipolar lines homography which is a projective invariant.

Second, thanks to an epipolar-aligned grid projection, the inventors could figure out that the proximity criterion is not sufficient to find features vertical neighbors in a significant number of scenes configurations, more particularly containing off-plane orientations.

The herein presented mono-trocar stereo system 3, implementing the invention method, is able to reconstruct scenes robustly to geometrical and spectral disturbances and intra-abdominal organs surfaces could be reconstructed in real-time.

The correction of the associated camera and projector lens distortions has been described through a new method based on a new grid pattern projection/acquisition inside each call of the cost function of the optimization process.

A practical implementation of the invention is reported in X. Maurice, C. Albitar, C. Doignon and M. de Mathelin, "A structured light-based laparoscope with real-time organs' surface reconstruction for minimally invasive surgery", Proceedings of the IEEE International Conference on Engineering in Medicine and Biology, San Diego, August 27 - September 3, 2012.

The following bibliographic references have been previously quoted in this specification, to illustrate the state of the art, possibly in relation to some specific aspects of the invention:

[1] A. Shpunt and B. Pesach, "Optical pattern projection," in *US Patent 20100284082*, 2010.

[2] J. Salvi, J. Batlle, and E. Mouaddib, "A robust-coded pattern projection for dynamic 3d scene measurement," *Pattern Recognition Letters*, vol. 19, pp. 1055–1065, 1998.

[3] Z. Song and C.-K. Chung, "Determining both surface position and orientation in structured-light-based sensing," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 10, pp. 1770-1780, October 2010.

[4] J. Batlle, E. Mouaddib, and J. Salvi, "Recent progress in coded structured light as a technique to solve the correspondence problem: a survey," *Pattern Recognition*, vol. 31, no. 7, pp. 963-982, 1998.

[5] J. Salvi, J. Pagès, and J. Batlle, "Pattern codifications strategies in structured light systems," *Pattern Recognition*, vol. 37, pp. 827-849, 2004.

[6] M. Ribo and M. Brandner, "State of the art on vision-based structured light systems for 3d measurements," *International Workshop on Robotic and Sensor Environments*, pp. 2-6, 2005.

[7] B. Pan, Q. Guan, X. Wang, and S. Chen, "Strategies of real-time 3d reconstruction by structured light," in *Pattern Recognition (CCPR), 2010 Chinese Conference on*, October 2010, pp. 1-5.

[8] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado, "A state of the art in structured light patterns for surface profilometry," *Pattern Recogn.*, vol. 43, pp. 2666-2680, August 2010.

[9] J. Geng, "Dlp-based structured light 3d imaging technologies and applications," *Proc. SPIE 7932, 79320B*, January 2011.

[10] J. Posdamer and M. Altschuler, "Surface measurement by space-encoded projected beam systems," *Computer Graphics and Image Processing*, vol. 18, no. 1, pp. 1-17, 1982.

[11] K. L. Boyer and A. C. Kak, "Color-encoded structured light for rapid active ranging," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-9, no. 1, pp. 14-28, January 1987.

[12] J. Pages, J. Salvi, and J. Forest, "A new optimised de bruijn coding strategy for structured light patterns," in *Pattern Recognition, 2004. ICPR 2004.Proceedings of the 17th International Conference on*, vol. 4, August 2004, pp. 284-287 Vol.4.

[13] S. Yamazaki, A. Nukada, and M. Mochimaru, "Hamming color code for dense and robust one-shot 3d scanning," *In Proc. British Machine Vision Conference*, August 2011.

[14] A. Morano, C. Ozturk, R. Conn, S. Dubin, S. Zietz, and J. Nissanov, "Structured light using pseudo-random codes," *IEEE Int'l Trans. on*

*Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 322-327, 1998.

[15] K. Claes and H. Bruyninckx, "Robot positioning using structured light patterns suitable for self calibration and 3d tracking," in *Int'l Conf. on Advanced Robotics*, Jeju, Korea, 2007.

[16] S. Yee and P. Griffin, "Three-dimensional imaging system," *Optical Engineering*, vol. 33, no. 6, pp. 2070-2075, 1994.

[17] C. Albitar, P. Graebling, and C. Doignon, "Robust structured light coding for 3d reconstruction," in *Int'l Conf. on Computer Vision*, Rio de J., Brazil, 2007.

[18] X. Li and L. Yaping, "Optimization design for structured light system," in *Intelligent Computing and Intelligent Systems (ICIS), 2010 IEEE International Conference on*, vol. 3, oct. 2010, pp. 602-605.

[19] J. Harvent, F.Bugarin, J.J.Orteui, M.Devy, P. Barbeau, and G. Marin, "Inspection of aeronautics parts for shape defect detection using a multi-camera system," in *International Symposium on Optical Metrology in Industrial, Medical and Daily Life Applications*, Orlando, June 2008.

[20] P. Graebling, A. Lallement, D. Y. Zhou, and E. Hirsch, "Optical high-precision three-dimensional vision-based quality control of manufactured parts by use of synthetic images and knowledge for image-data evaluation and interpretation," *Applied Optics*, vol. 41, no. 14, pp. 2627-2643, 2002.

[21] K. Keller and J. Ackerman, "Real-time structured light depth extraction," *proceedings of SPIE on Three Dimensional Image Capture and Applications III*, vol. 3958, pp. 11-18, 2000.

[22] J. Guhring, "Dense 3-d surface acquisition by structured light using off-the-shelf components," in *Proc. Videometrics and Optical Methods for 3D Shape Measurement*, 2001, pp. 220-231.

[23] X. Maurice, P. Graebling, and C. Doignon, "A pattern framework driven by the hamming distance for structured light-based reconstruction with a single image." *CVPR 2011, IEEE International conference on Computer Vision and Pattern Recognition*, June 2011.

[24] A. Adán, A. Vázquez, C. Cerrada, and S.Salamanca, "Moving surface extraction based on unordered hexagonal perfect submaps projection: Applications to 3d feature tracking," *Image Vision Computing*, vol. 27, no. 8, pp. 1083-1096, 2009.

[25] K. Claes, "Structured light adapted to control a robot arm," in *PhD thesis, Kath. Univ. Leuven*, 2008.

[26] P. M. Griffin, L. S. Narasimhan, and S. R. Yee, "Generation of uniquely encoded light patterns for range data acquisition," *Pattern Rec.*, vol. 25, no. 6, pp. 609-616, 1992.

[27] H. Morita, K. Yajima, and S. Sakata, "Reconstruction of surfaces of 3-d objects by m-array pattern projection method," *ICCV*, vol. 88, pp. 468-473, 1988.

[28] R. Sagawa, Y. Ota, Y. Yagi, R. Furukawa, and N. Asada, "Dense 3d reconstruction method using a single pattern for fast moving object," in *Int'l Conf. Computer Vision*, Kyoto, Japan, 2009.

[29] J. Xu, N. Xi, C. Zhang, and Q. Shi, "Real-time 3d shape measurement system based on single structure light pattern." in *ICRA'10*, May 2010, pp. 121-126.

[30] T. Etzion, "Constructions for perfect maps and pseudorandom arrays," *IEEE Int'l Trans. on Information Theory*, vol. 34, no. 5/1, pp. 1308-1316, 1988.

[31] F. J. MacWilliams and N. J. A. Sloane, "Pseudo-random sequences and arrays," *Proceedings of the IEEE*, vol. 64, no. 12, pp. 1715-1729, 1976.

[32] X. Maurice, P. Graebling, and C. Doignon, "Real-time structured light coding for adaptive patterns," *Journal of Real-Time Image Processing*, pp. 1-10, 2011.

[33] D. Caspi, N. Kiryati, and J. Shamir, "Range imaging with adaptive color structured light," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 5, pp. 470-480, May 1998.

[34] Q. Li, M. Biswas, M. Pickering, and M. Frater, "Two-shot sparse depth estimation using adaptive structured light," *Electronics Letters*, vol. 47, no. 13, pp. 745-746, 23 2011.

[35] A. Adan, F. Molina, and L. Morena, "Disordered patterns projection for 3d motion recovering," in *Proceedings of the 3D Data Processing, Visualization, and Transmission, 2nd International Symposium*, ser. 3DPVT '04.1em plus 0.5em minus 0.4emWashington, DC, USA: IEEE Computer Society, 2004, pp. 262-269.

[36] T. Koninckx and L. V. Gool, "Real-time range acquisition by adaptive structured light," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 432-445, March 2006.

[37] R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Journal on Robotics and Automation*, vol. 3, no. 4, pp. 323-344, August 1987.

[38] X. Armangué and J. Salvi, "Overall view regarding fundamental matrix estimation," *Image and Vision Computing*, vol. 21, pp. 205-220, 2003.

[39] D. Brown, "Decentering distortion of lenses," *Photogrammetric Engineering.*, vol. 32, no. 3, pp. 444-462, 1966.

[40] F. Devernay and O. Faugeras, "Straight lines have to be straight: automatic calibration and removal of distortion from scenes of structured environments," *Mach. Vision Appl.*, vol. 13, pp. 14-24, August 2001. [Online]. Available: http://dl.acm.org/citation.cfm?id=513534.513536 =0pt

[41] T. Thormaehlen, H. Broszio, and I. Wassermann, "Robust line-based calibration of lens distortion from a single view," in *Proceedings of Mirage 2003*, 2003, pp. 105-112.

[42] G.-Y. Song and J.-W. Lee, "Correction of radial distortion based on line-fitting," International Journal of Control, Automation, and System, vol. 8, no. 3, pp. 615-621.

The present invention is of course not limited to the preferred embodiments described and represented herein, changes can be made or equivalents used without departing from the scope of the invention.

# CLAIMS

1. Active vision method comprising mainly the steps of projecting a structured light-based pattern with embedded features on a scene and of acquiring or capturing images of said scene, in particular for a single shot 3-D reconstruction, preferably in real-time or near real-time, by using a stereo or multiview system comprising at least one light projecting means and at least one image capturing or acquiring means, said system showing epipolar geometry features,

method characterised in that it consists in producing a single-shot structured light pattern (1) which is coded using a 2-D neighborhood and 1-D uniqueness constraint coding scheme, and in that the or at least some lines of the coded pattern (1), i.e. of the resulting matrix of generated codes, are, at least substantially, aligned with the nearly parallel epipolar lines of the stereo or multiview system (3).

2. Method according to claim 1, characterised in that it consists in computing the constraint of uniqueness for a limited number of neighborhood lines, preferably only one.

3. Method according to claim 1 or 2, characterised in that the step of producing the matrix of codes consists in first coding a number v of first lines $L_i$ of the matrix, preferably using a sizeable correcting window and optimal codes resulting from a fully exhaustive search, and then producing or filling the remaining lines $L_i$ of the matrix, step by step, by copying the codes of the preceding and previously filled lines $L_{i-v}$, in such a way that the constraint of uniqueness holds also for the copied lines.

4. Method according to claim 3, characterised in that $v = 3$ or 5.

5. Method according to anyone of claims 1 to 4, characterised in that the projected pattern (1) is an adaptative pattern, adapting itself, in particular in terms of resolution and/or spectral properties, to local features and/or specificities of the scene (2) to be imaged, by updating the numerical coding of the pattern to said scene specificities and/or features.

6. Method according to anyone of claims 1 to 5, characterised in that it consists in providing a periodicity along the lines of codes of the matrix.

7. Method according to anyone of claims 1 to 6, characterised in that the projected pattern (1), for example a regular grid of visual

symbols or a similar two-dimensional regular pattern, is subjected to a global rotation such that the average orientation of the lines of the grid or coded matrix, or of the features of the pattern, corresponds to the average orientation of the epipolar lines in the plane of the projecting means (4).

8. Method according to anyone of claims 1 to 7, characterised in that a central symmetry constraint is included during coding of the pattern (1), the central symmetry of the features of the pattern being preserved in case of a global rotation of the projected pattern.

9. Method according to anyone of claims 1 to 8, characterised in that, in view of the 3-D reconstruction of each image of the scene (2), a decoding is performed on the basis of a belief propagation scheme from feature correspondences with a high belief, using an adjacency graph to define the frontiers of propagation and a selection function based on a lower Hamming distance test.

10. Method according to anyone of claims 1 to 9, characterised in that it consists in providing additional image data through at least one passive vision component, in particular image data related to the texture of the scene (3), and mixing in said passive vision image data with the collected active vision image data.

11. Method according to anyone of the previous claims, characterised in that the stereo or multiview pattern projecting and image capturing system (3) is a video system, comprising a video projecting means (4) and a video acquiring means (5), the acquired image data being preferably exploited to perform visual control, robot guidance or serving.

12. Method according to anyone of claims 1 to 11, characterised in that it also comprises a preliminary process for estimating and/or correcting the geometric distortions of the stereo or multiview system (3), in particular video system, preferably of the endoscopic type, by performing a separate estimation, and a possible separate correction or compensation, of the distortions produced or induced respectively by the projecting means (4) and by the acquiring or capturing means (5).

13. Method according to claim 12, characterised in that the estimation and the possible rectification of the distortions of the projecting and/or capturing means (4, 5) are performed by using the plumb-line principle.

14. Method according to anyone of claims 12 and 13, characterised in that it consists, after having estimated and possibly

corrected the geometrical distortions linked to the image acquiring or capturing means (5), in performing an a priori software based estimation and correction of the geometrical distortions of the projecting means (4), preferably by realising multiple projections and captures of successive distorted patterns (1) so as to gradually improve the straightness of the projected pattern and aiming towards a fully straight projected pattern.

15. Method according to anyone of claims 12 to 14, characterised in that the estimation and rectification of the distortions of the projecting means (4) consists, using an adequately programmed software, in estimating and correcting the geometrical distortions related to the pattern projecting means (4) by: a) projecting a regular pattern (1), for example a regular grid with horizontal and vertical lines, on a planar item or object, b) acquiring the image with the capturing means (5), c) evaluating distortion parameters by using an adequate distortion model, d) repeating steps a) to c) by acquiring at each cycle the image of the pattern (1) as distorted by the updated parameters of the used distortion model, until distortion parameter values are reached which lead to a minimised distortion of the projected pattern (1), or at least to a distortion within a tolerance level.

16. Method according to anyone of claims 12 to 15, characterised in that the estimation and rectification of the distortion parameters of the projecting means (4) are performed by means of an image based optimisation function incorporating a cost reduction function and by using an adapted distortion model, such as the Brown distortion model.

17. Method according to anyone of claims 12 to 16, characterised in that, when the stereo or multiview system (3) is an endoscopic system, an estimation of the endoscopic circle is performed by using an optimisation function with a cost reduction function and taking into account the contrast between the inside and the outside of the considered circle.

18. Stereo endoscopic system for projecting and capturing a video flow through a single physical passage, for example a single trocar in a mini-invasive context, said system comprising at least means for projecting a structured light-based pattern on a scene and means for capturing or acquiring in real time images of said scene,

system (3) characterised in that it also comprises means (6), in particular computer and software means, for implementing the method according to anyone of claims 1 to 17.

19. Stereo endoscopic system according to claim 18, characterised in that it comprises a first endoscopic device (7) with an optical passage or channel (8) and at least one other or second channel or passage (9), for example an operator channel, and a second endoscopic device (10) which is fitted or accommodated within said other or second channel (9) of the first endoscopic device (7), one (7 or 10) of said endoscopic devices being functionally connected to the projecting means (4), in particular projecting means for time-varying patterns like video projector, and the other (10 or 7) of said endoscopic devices being functionally connected to the image acquiring means (5).

20. Stereo endoscopic system according to claim 18 or 19, characterised in that the pattern projecting means (4) comprise a miniaturised video projector (4') incorporating or connected to a light sources (4"), in that the image capturing means (5) comprise a video camera (5') and in that both endoscopic devices (7 and 10) and the projecting and capturing means (4 and 5) are physically, and preferably rigidly, linked together by a contention system (11), possibly incorporating or supported by a bearing arm or stand.

21. Stereo endoscopic system according to anyone of claims 18 to 20, characterised in that the optical channel (8) of the first endoscopic device (7) and the second optical endoscopic device (10) comprise optical connectors (8' an 10') at their ends connected to the projecting or capturing means (4, 5), preferably through interposed interface optics (8", 10").

22. Stereo endoscopic system according to anyone of claims 18 to 21, characterised in that the first endoscopic device (7) is an angled rigid endoscope with a straight operator channel (9) and in that the second endoscopic device (10) is a straight endoscopic device with at least an optical channel, the insertion port of the second endoscopic device comprising a sealing means (12).

23. Stereo endoscopic system according to anyone of claims 18 to 21, characterised in that at least the second endoscopic device, preferably both endoscopic devices, is(are) flexible endoscope(s).

24. Stereo endoscopic system according to anyone of claims 18 to 23, characterised in that the pattern projecting means (4) incorporate means able to perform a global rigid rotation of the projected pattern (1) or of its coded matrix in order to achieve a global alignment with the epipolar

lines of the stereo system (3), the central symmetry of the features of the pattern (1) being maintained.

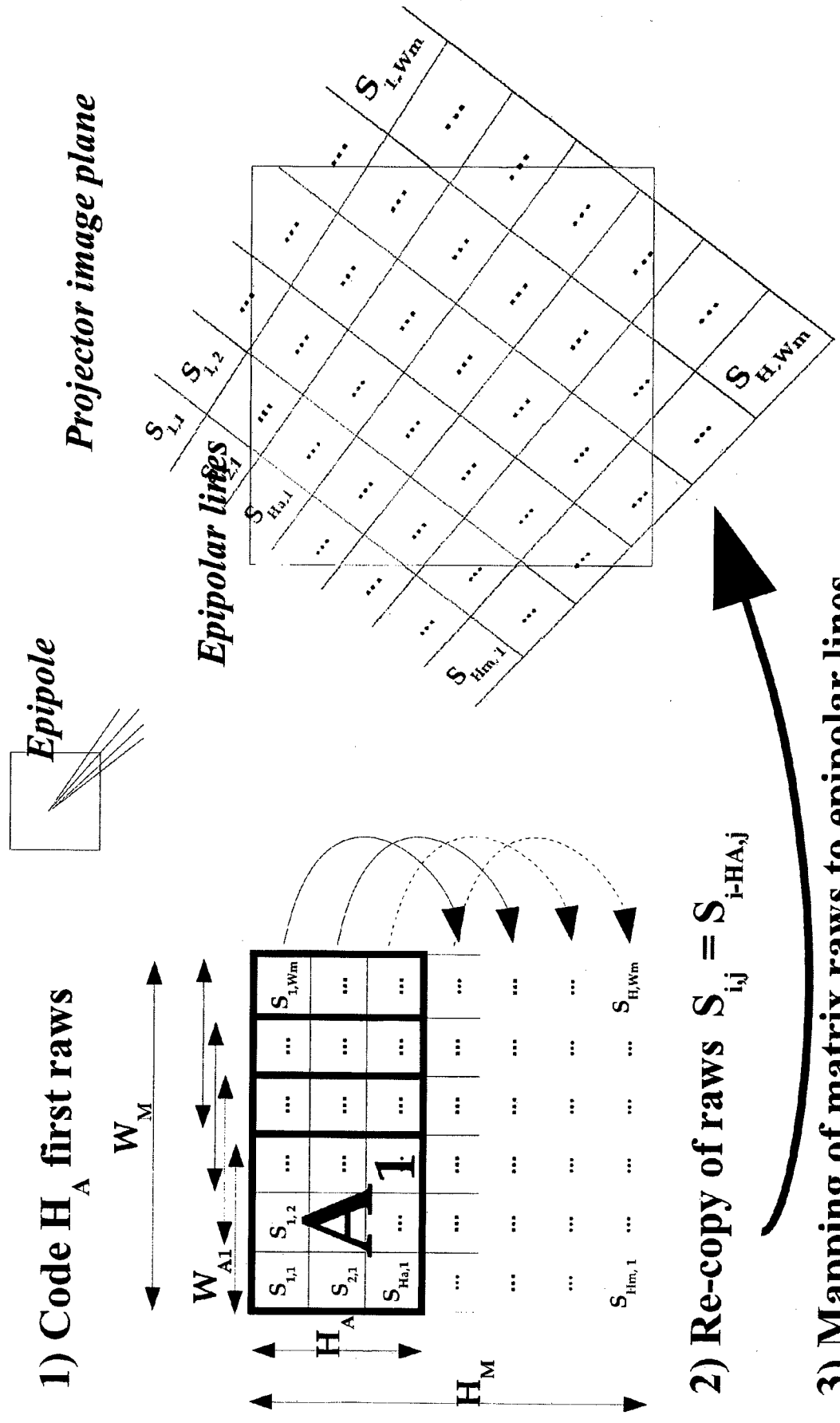

1) Code H_A first raws
W_M
W_A1
H_A
H_M
A1
S_1,1
S_1,2
S_2,1
S_Ha,1
S_1,Wm
S_Hm,1
S_H,Wm
2) Re-copy of raws S_i,j = S_i-HA,j
3) Mapping of matrix raws to epipolar lines
Epipole
Epipolar lines
Projector image plane

Fig. 1

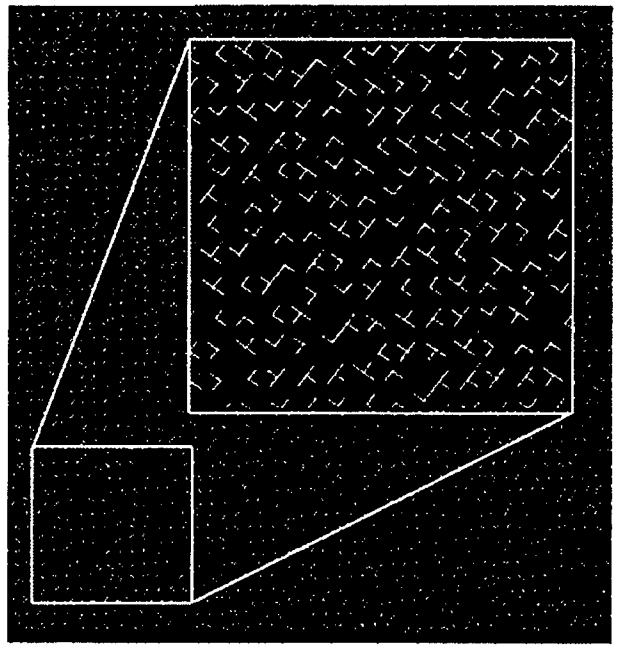


Fig. 2B

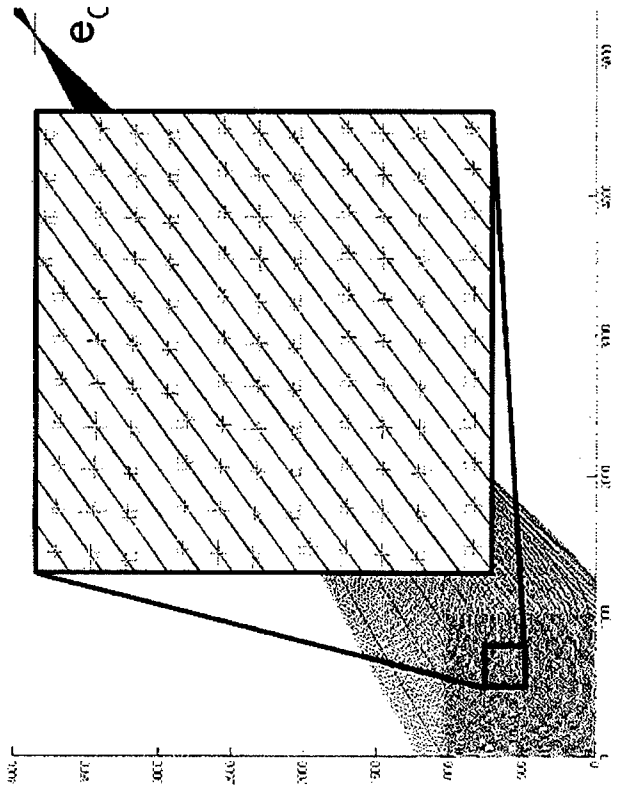



Fig. 2A

Fig. 2

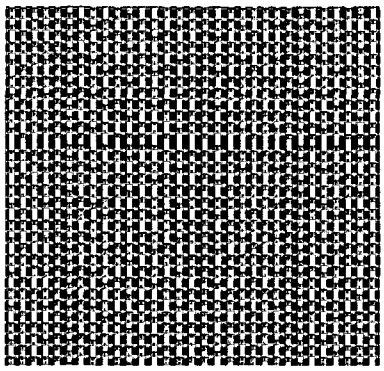


Fig. 3A

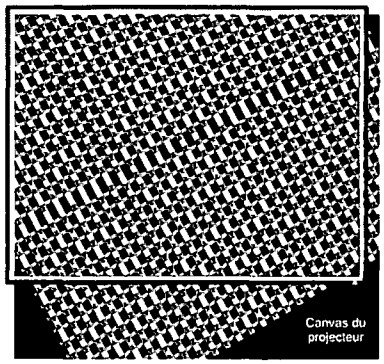



Fig. 3B

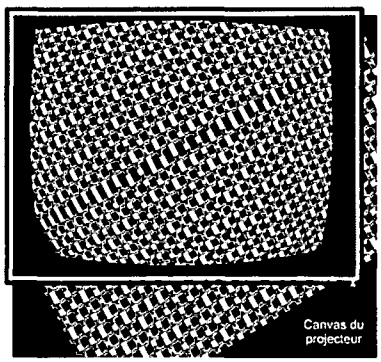



Fig. 3C

Fig. 3

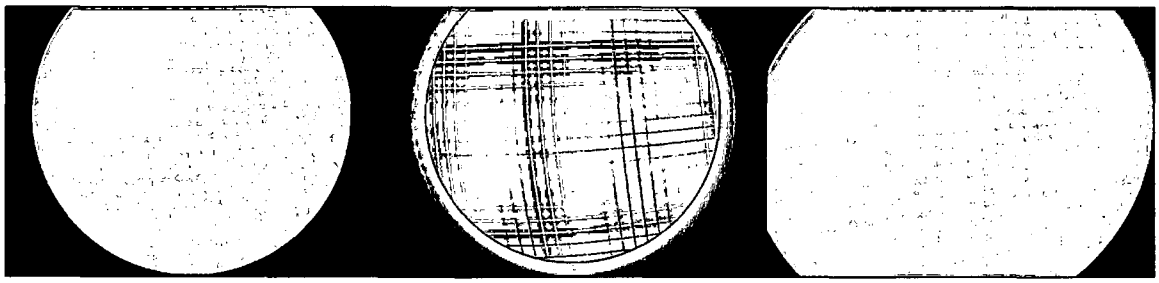


Fig. 5A Fig. 5B Fig. 5C

Fig. 5

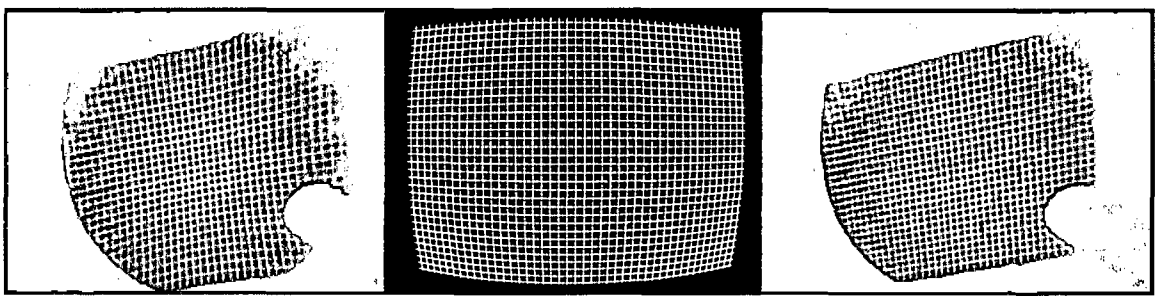


Fig. 6A Fig. 6B Fig. 6C

Fig. 6

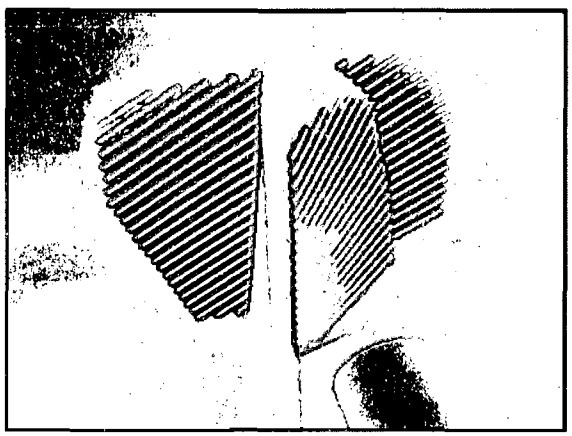


Fig. 7A

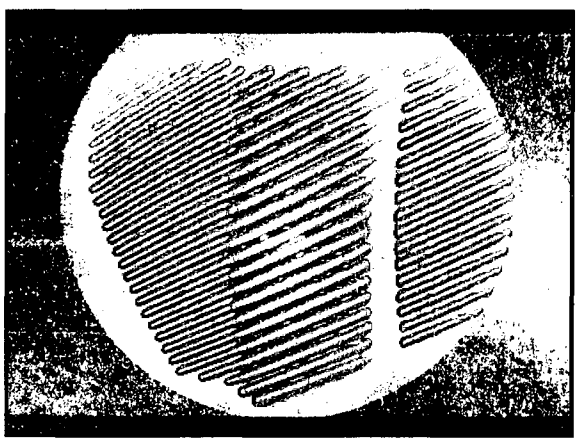


Fig. 7B

Fig. 7

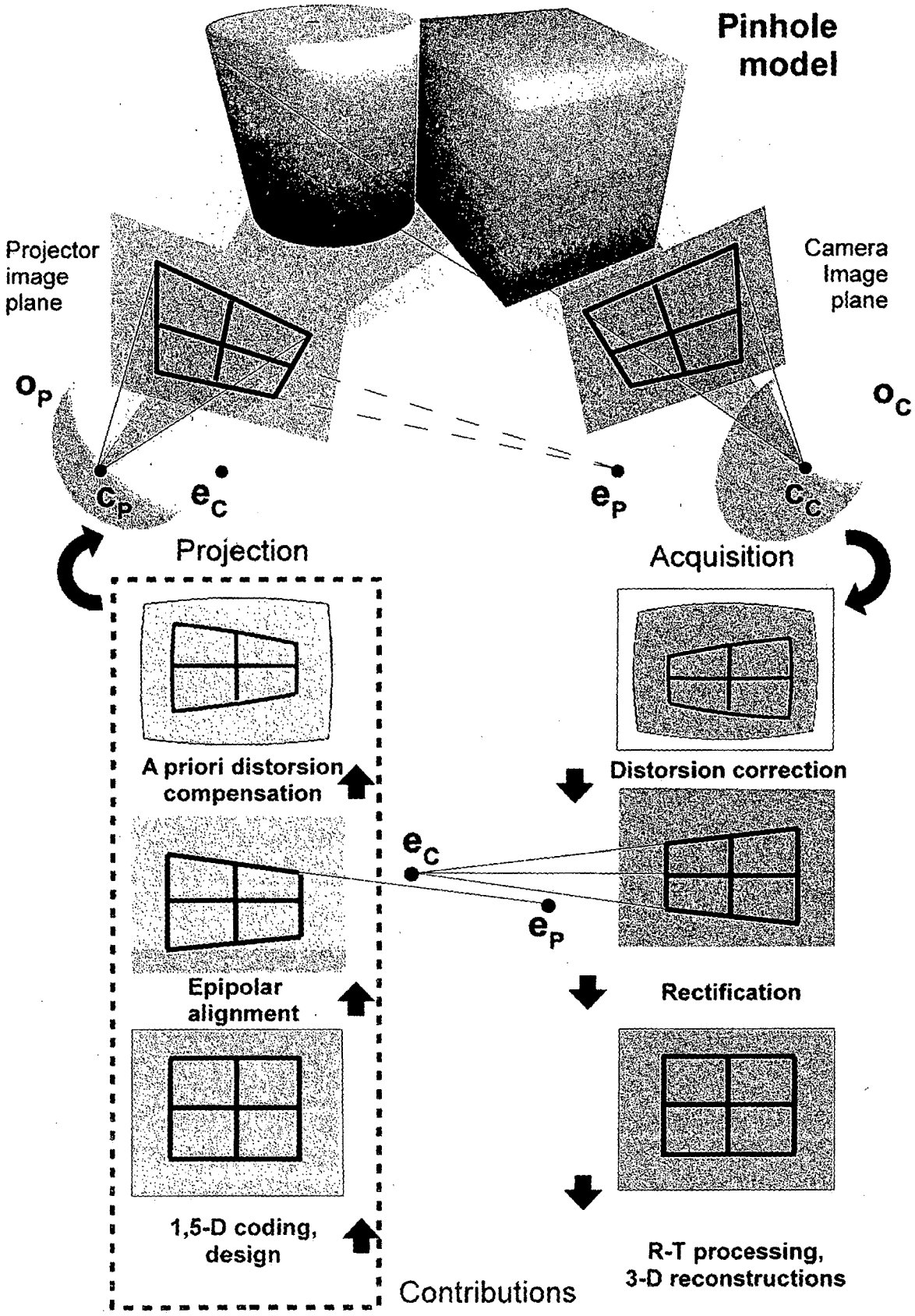

Pinhole model
Projector image plane
Camera Image plane
$O_P$
$O_C$
$C_P$
$e_C$
$e_P$
$C_C$
Projection
Acquisition
A priori distorsion compensation
Distorsion correction
$e_C$
$e_P$
Epipolar alignment
Rectification
1,5-D coding, design
R-T processing, 3-D reconstructions
Contributions

Fig. 4

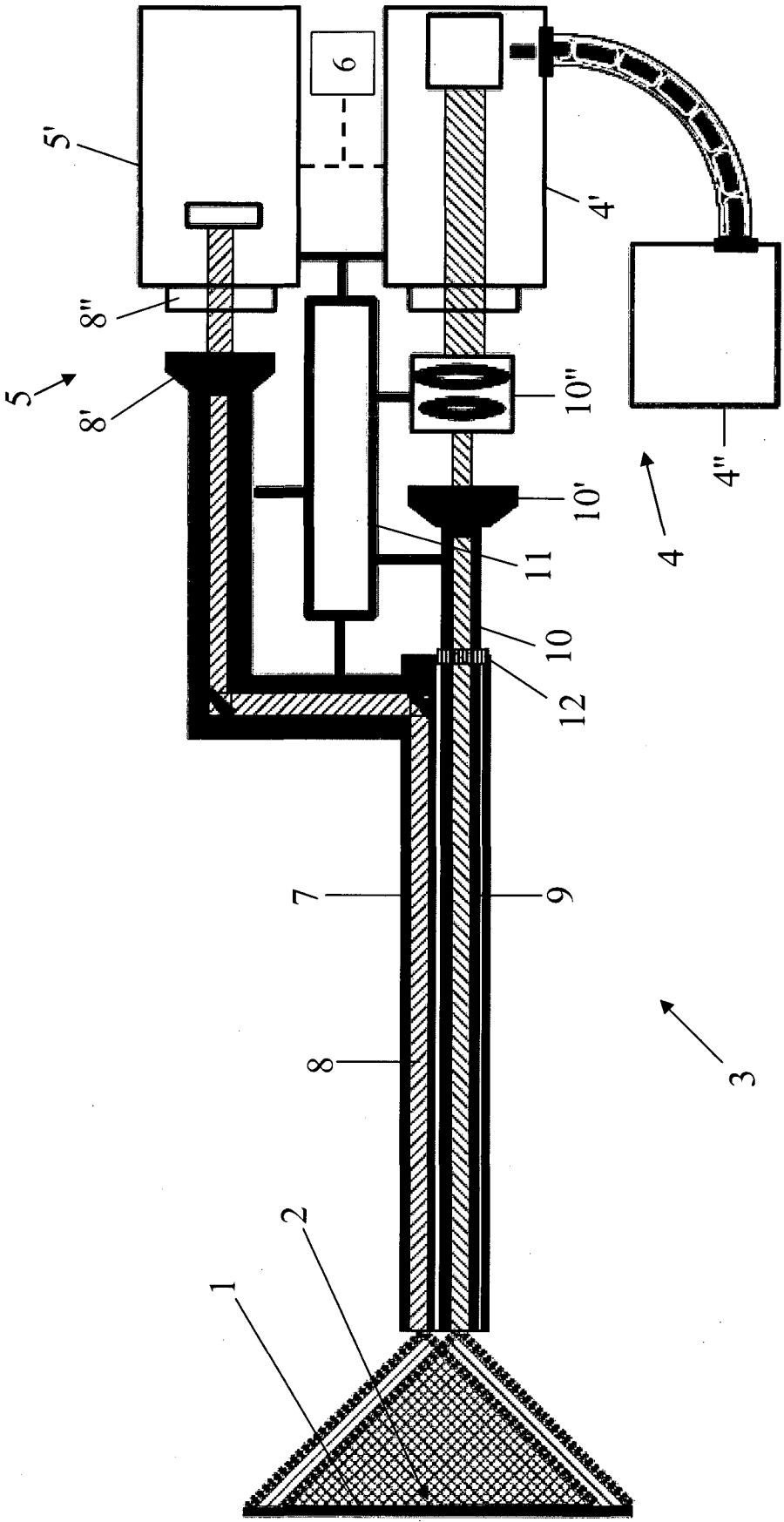



Fig. 8

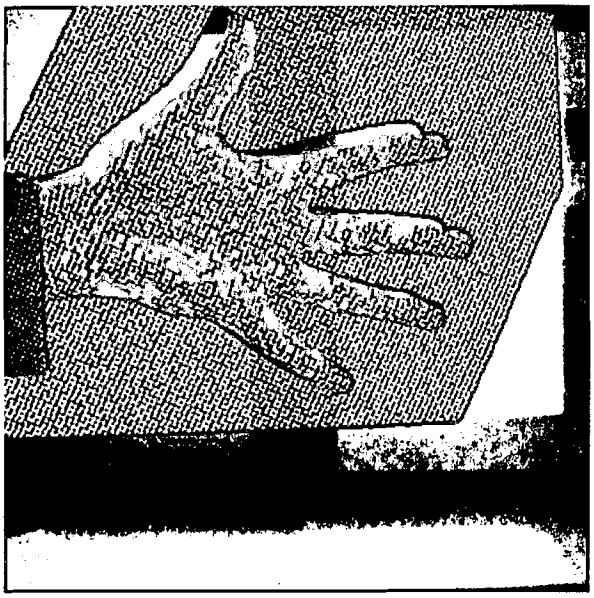


Fig. 9A

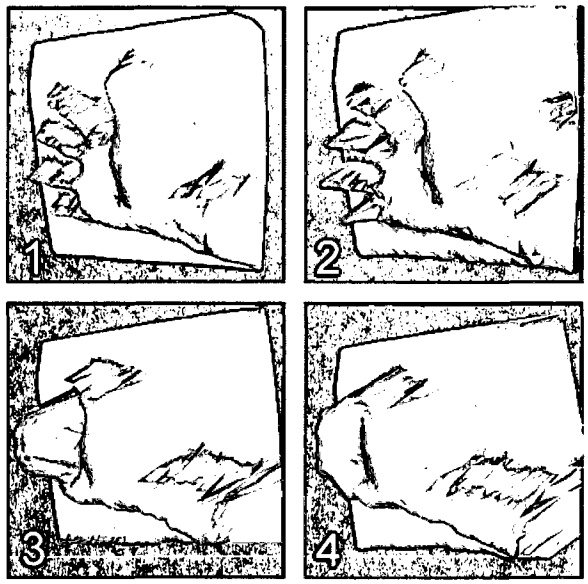



Fig. 9B

Fig. 9

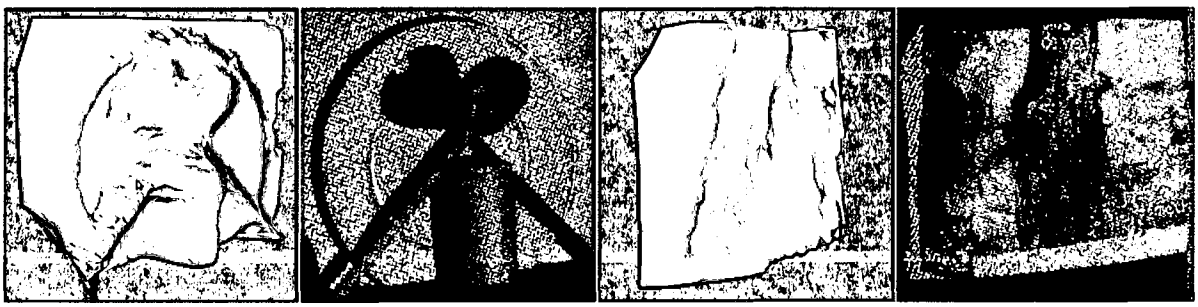


Fig. 10A Fig. 10B Fig. 10C Fig. 10D

Fig. 10

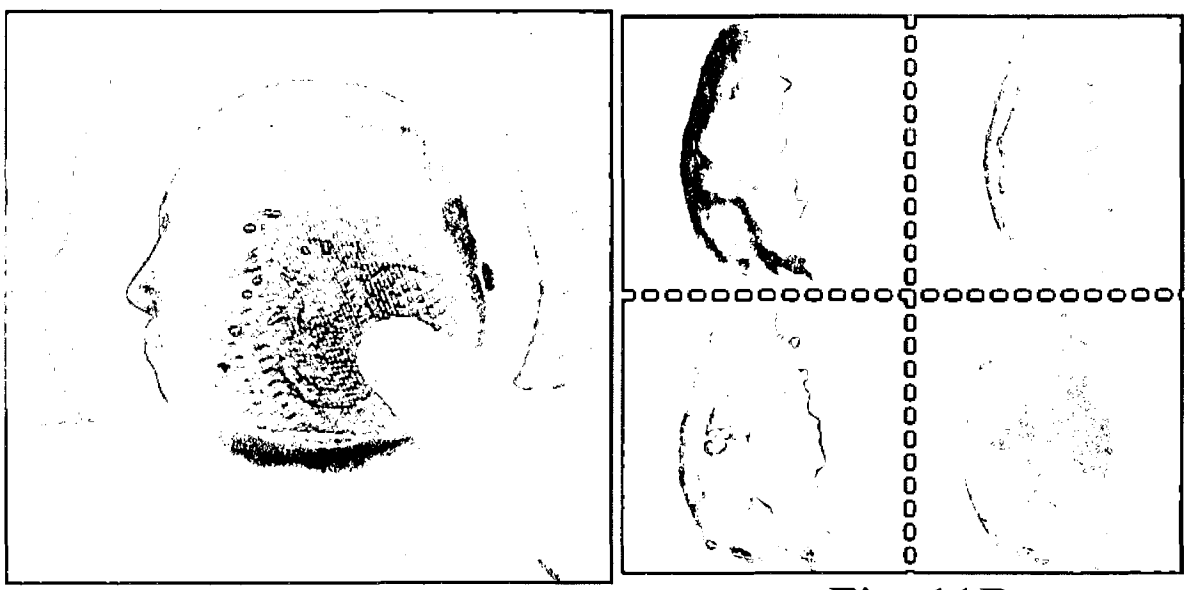


Fig. 11A Fig. 11B

Fig. 11

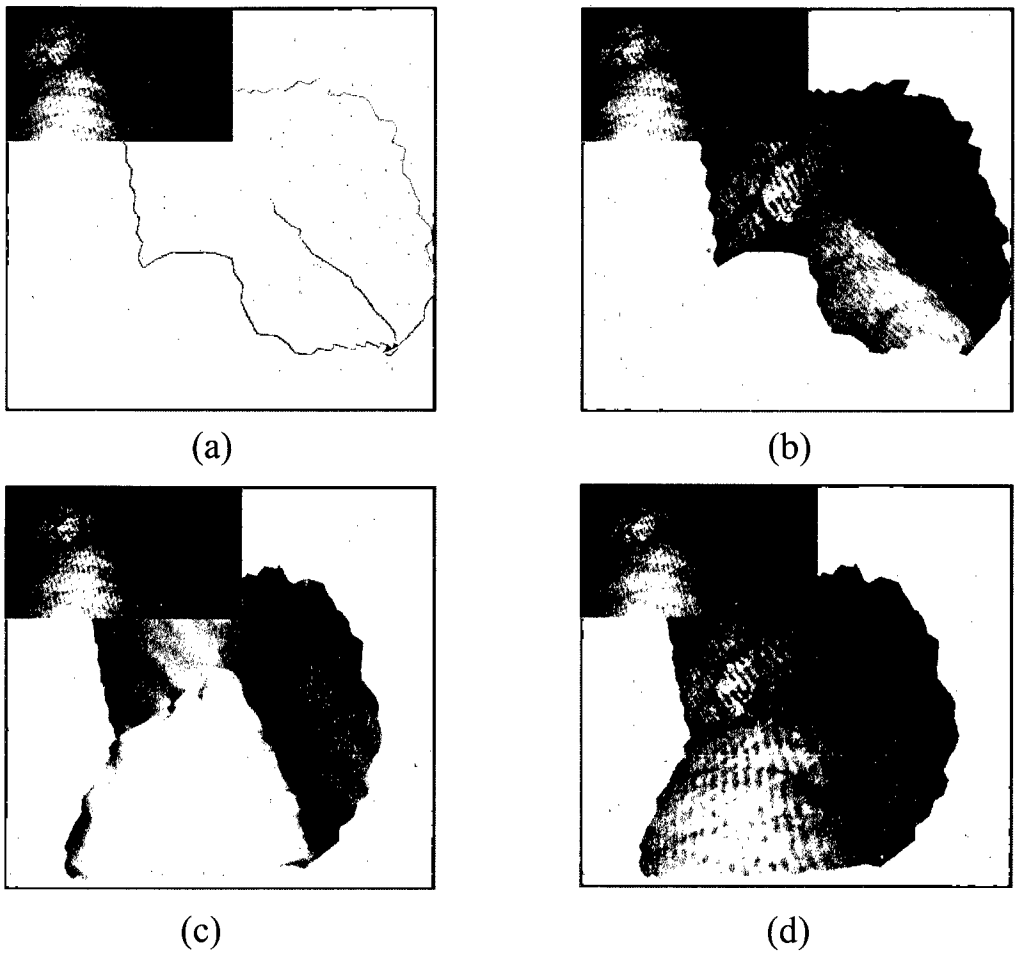



Fig. 12

| | | | | |
|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 1 |
| 1 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 1 | 0 |
| 1 | 1 | 0 | 1 | 1 |
| 1 | 0 | 1 | 1 | 1 |

(a)

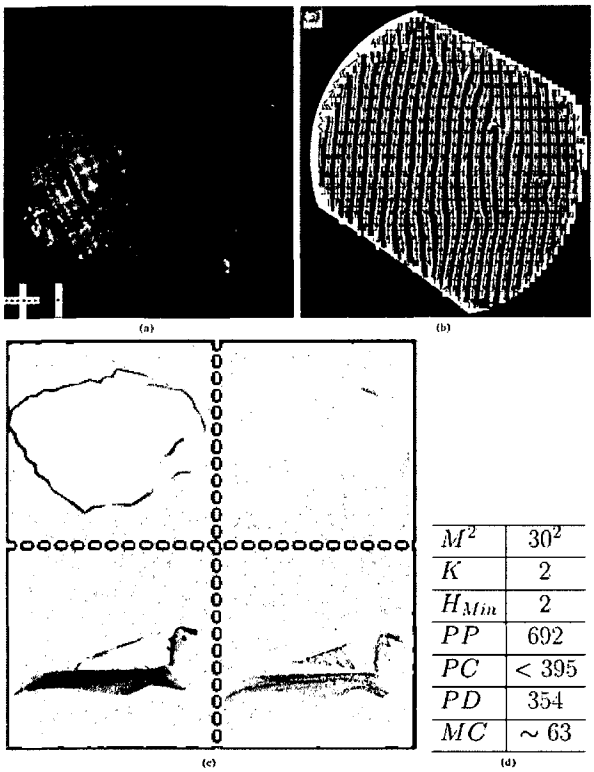
| | | | | |
|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 |
| 1 | 2 | 2 | 1 | 0 |
| 1 | 2 | 0 | 2 | 1 |
| 0 | 1 | 2 | 2 | 1 |
| 1 | 1 | 1 | 1 | 1 |

(b)

| | | | | |
|---|---|---|---|---|
| 1 | 2 | 0 | 1 | 2 |
| 1 | 2 | 0 | 1 | 2 |
| 1 | 2 | 0 | 1 | 2 |
| 1 | 2 | 0 | 1 | 2 |
| 1 | 2 | 0 | 1 | 2 |

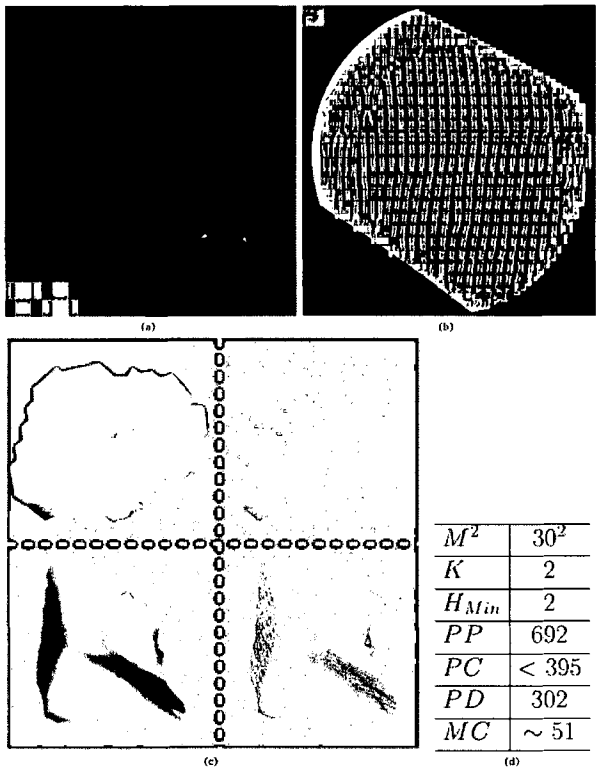
(c)

Fig. 14



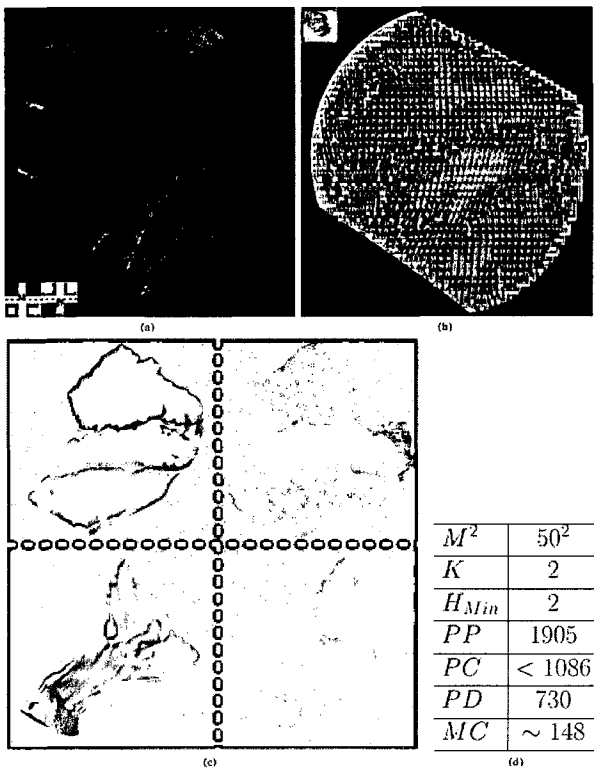
| | |
|---|---|
| $M^2$ | $30^2$ |
| $K$ | 2 |
| $H_{Min}$ | 2 |
| $PP$ | 692 |
| $PC$ | $< 395$ |
| $PD$ | 354 |
| $MC$ | $\sim 63$ |

(a)



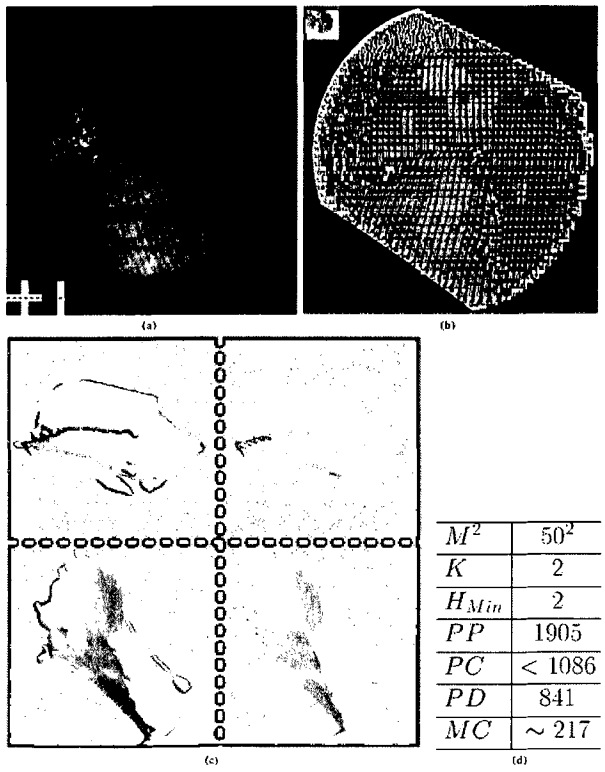
| | |
|---|---|
| $M^2$ | $30^2$ |
| $K$ | 2 |
| $H_{Min}$ | 2 |
| $PP$ | 692 |
| $PC$ | $< 395$ |
| $PD$ | 302 |
| $MC$ | $\sim 51$ |

(b)



| | |
|---|---|
| $M^2$ | $50^2$ |
| $K$ | 2 |
| $H_{Min}$ | 2 |
| $PP$ | 1905 |
| $PC$ | $< 1086$ |
| $PD$ | 730 |
| $MC$ | $\sim 148$ |

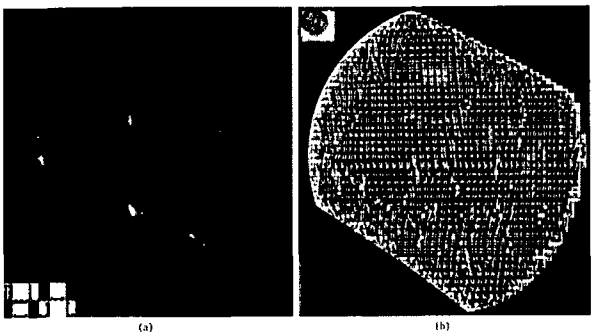
(c)

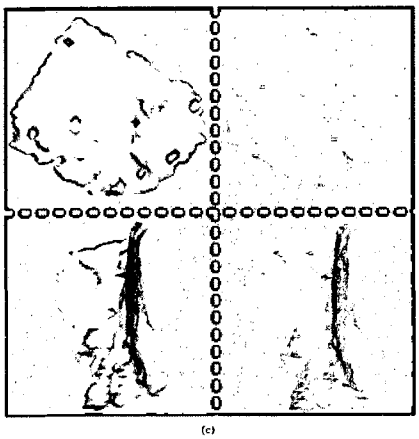


| | |
|---|---|
| $M^2$ | $50^2$ |
| $K$ | 2 |
| $H_{Min}$ | 2 |
| $PP$ | 1905 |
| $PC$ | $< 1086$ |
| $PD$ | 841 |
| $MC$ | $\sim 217$ |

(d)

Fig. 13

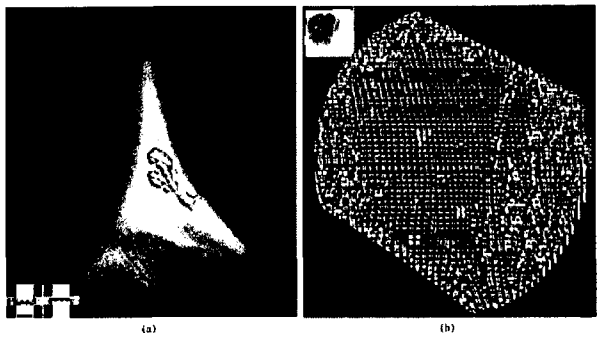


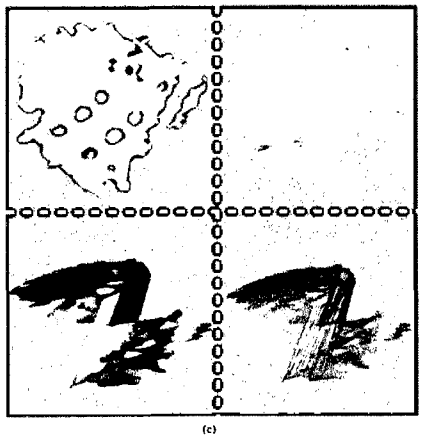




| | |
|---|---|
| $M^2$ | $50^2$ |
| $K$ | 2 |
| $H_{Min}$ | 2 |
| $PP$ | 1905 |
| $PC$ | $< 1086$ |
| $PD$ | 1022 |
| $MC$ | $\sim 275$ |

(d)

(e)

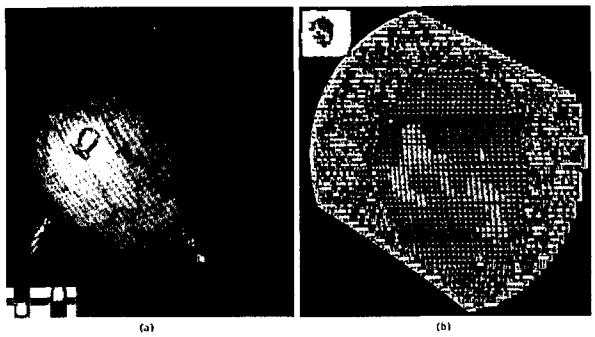


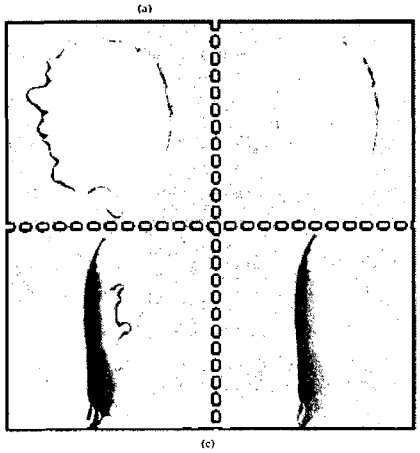




| | |
|---|---|
| $M^2$ | $70^2$ |
| $K$ | 2 |
| $H_{Min}$ | 2 |
| $PP$ | 3721 |
| $PC$ | $< 2121$ |
| $PD$ | 1132 |
| $MC$ | $\sim 474$ |

(d)

(f)

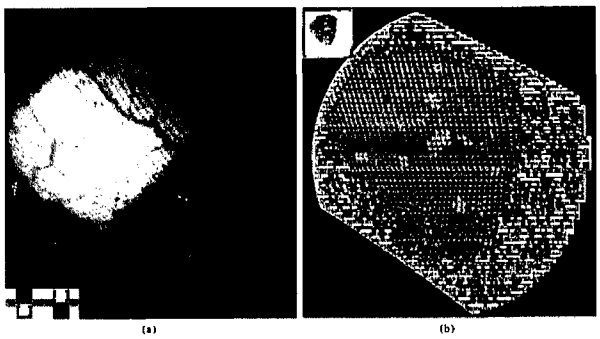


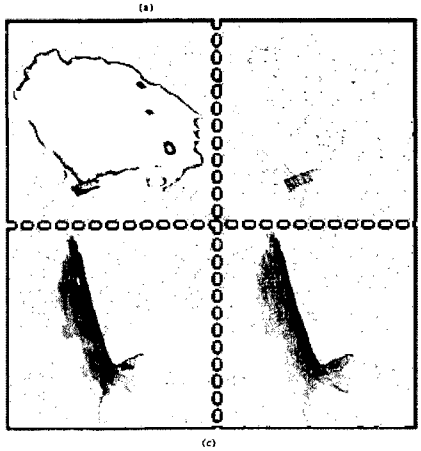




| | |
|---|---|
| $M^2$ | $70^2$ |
| $K$ | 2 |
| $H_{Min}$ | 2 |
| $PP$ | 3721 |
| $PC$ | $< 2121$ |
| $PD$ | 884 |
| $MC$ | $\sim 180$ |

(d)

(g)




| | |
|---|---|
| $M^2$ | $70^2$ |
| $K$ | 2 |
| $H_{Min}$ | 2 |
| $PP$ | 3721 |
| $PC$ | $< 2121$ |
| $PD$ | 1024 |
| $MC$ | $\sim 301$ |

(d)

(h)

Fig. 13

| k/H Min | 1 | | | 2 | | 3 | | | 4 | | 5 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Notes | | | | C. : N.A. | | | | | C. : N.A. | | M. : N.A. | |
| Algo. | X. | C. | M. | X. | M. | X. | C. | M. | X. | M. | X. | C. |
| 2 | 22² | 19x25 | N.D. | 11² | N.D. | 7² | 6x8 | N.D. | 5² | N.D. | N.D. | 4x5 |
| 3 | 110² | 106x114 | >= 45² | 36²- | < 45² | 16² | 14x18 | < 45² | 9² | < 45² | 7² | 6x8 |
| 4 | 390² | 316x421 | >= 45² | 104² | < 45² | 37² | 33x44 | < 45² | 17² | < 45² | 10² | 9x11 |
| 5 | 1100² | 429x572 | >= 45² | 262² | >= 45² | 73² | 68x90 | < 45² | 27² | < 45² | 14² | 12x16 |
| 6 | 2464² | 496x661 | >= 45² | 548² | >= 45² | 133² | 124x165 | < 45² | 43² | < 45² | 20² | 18x23 |
| 7 | 5200² | 676x901 | >= 45² | 1023² | >= 45² | 228² | 204x272 | < 45² | 64² | < 45² | 27² | 24x32 |
| 8 | 9500² | 696x925 | >= 45² | 1460² | >= 45² | 344² | 292x389 | >= 45² | 92² | < 45² | 35² | 32x42 |

Biggest pattern, size increase : < 100% or N.A. 100% - 500% > 500%

Fig. 15

| $k/H_{Min}$ | 1 | | 2 | | 3 | | 4 | | 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Algo** | *<40ms* | *SPSM aligné Max* | *<40ms* | *SPSM aligné Max* | *<40ms* | *SPSM aligné Max* | *<40ms* | *SPSM aligné Max* | *<40ms* | *SPSM aligné Max* |
| 2 | 503 | **513** | 103 | 113 | 30 | 33 | 17 | 18 | 7 | 7 |
| 3 | 2079 | >5000 | 1583 | 2588 | 327 | 352 | 83 | 90 | 28 | 32 |
| 4 | 1910 | >5000 | 1825 | >5000 | 1285 | 2473 | 376 | 399 | 90 | 96 |
| 5 | 1781 | >5000 | 1751 | >5000 | 1565 | >5000 | 883 | 1429 | 240 | 260 |
| 6 | 1702 | >5000 | 1694 | >5000 | 1606 | >5000 | 1120 | 4233 | 486 | 597 |
| 7 | 1658 | >5000 | 1648 | >5000 | 1602 | >5000 | 1246 | >5000 | 659 | 1246 |
| 8 | 1629 | >5000 | 1615 | >5000 | 1596 | >5000 | 1322 | >5000 | 744 | 2425 |

Fig. 16