



- (51) International Patent Classification:
G06K 9/46 (2006.01) *G06T 7/00* (2006.01)
- (21) International Application Number:
PCT/EP2015/065031
- (22) International Filing Date:
1 July 2015 (01.07.2015)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
14306127.3 10 July 2014 (10.07.2014) EP
- (71) Applicant: THOMSON LICENSING [FR/FR]; 1-5 rue
Jeanne d'Arc, 92130 Issy-les-Moulineaux (FR).
- (72) Inventors: SORGL, Lorenzo; Gehaegestrasse 20K, 30655
Hannover (DE). DUMUSC, Gwenaël; Baumaz 9, CH-
1422 Grandson (CH).
- (74) Agent: SCHMIDT-UHLIG, Thomas; Deutsche Thomson
OHG, Patent Operations, Karl-Wiechert-Allee 74, 30625
Hannover (DE).
- (81) Designated States (*unless otherwise indicated, for every
kind of national protection available*): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG,
MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM,
PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC,
SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) Designated States (*unless otherwise indicated, for every
kind of regional protection available*): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,
TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,

[Continued on next page]

(54) Title: METHOD AND DEVICE FOR IMAGE EXTRACTION FROM A VIDEO

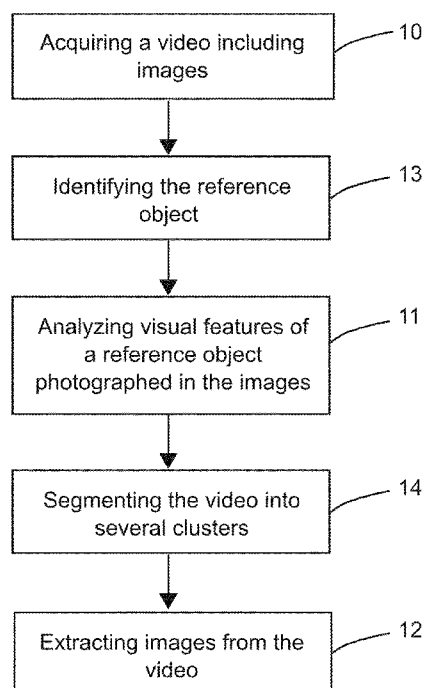


Fig. 2

(57) Abstract: A method for extracting images from a video and a device (20) configured to perform the method are described. An acquiring unit (21) of the device acquires (10) a video including a plurality of images in which a reference object is captured. An analysing unit (22) analyses (11) features of the reference object in the images of the video, segments (14) the video into clusters, and determine one key image for each cluster based on the analysis of the features. An extracting unit (23) then extracts (12) the key images from the video.



TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

METHOD AND DEVICE FOR IMAGE EXTRACTION FROM A VIDEO

TECHNICAL FIELD

5 A method and a device for image extraction from a video are proposed. The extracted images provide an image dataset required for camera calibration, which is a field that collects algorithms and technologies aimed at the characterization of a mathematical projection model for the
10 image formation process of a camera.

BACKGROUND

A camera calibration process in general consists of
15 radiometrical and geometrical stages. The radiometrical calibration is to acquire information on how a camera distorts the luminous properties of a scene, e.g. color and luminance. It plays a fundamental and important role for applications such as astronomical imaging and color
20 processing, but is generally bypassed in most of typical 3D vision applications. The geometrical calibration leads to the estimation of a suitable model for the image formation geometry, namely the camera projection and the optical distortion, and is crucial for most 3D vision applications.

25

Calibration techniques can be generally classified into self-calibration and object-based calibration methods. Self-calibration attempts to infer the camera model from the transformation of the image appearance under the action
30 of a rigid motion, while object-based calibration relies on a certain a-priori known calibration object [I, II].

The most common object-based techniques assume the availability of images of a specific object, which has a

- 2 -

known shape and can be easily detected. The calibration object is captured by a camera from different points of view. This prerequisite eases the collection of an error-free set of correspondences between 3D points and 2D image
5 projections for a subsequent camera calibration procedure [III].

The pioneer technique in this field was proposed by Tsai [IV] and has been followed by a large number of other
10 algorithms, which differentiate from each other for attributes such as the geometry of the calibration object, the features represented on the object surface, the number of required images, the constraints on camera motion and the estimated camera projection model [V]. Among the huge
15 proliferation of calibration algorithms, Zhang's approach [VI] deserves a remark as it has become the basis of many open-source as well as commercial calibration tools. Meanwhile, with a large number of standard solutions, several tools addressing specific problems like fisheye
20 lenses and omnidirectional imaging systems and camera clusters calibration are also available.

Despite the large number of solutions, not much attention has been given to the preliminary processing task of
25 gathering an optimal calibration image dataset from a video, which is actually crucial for and can influence the result and accuracy of the expected camera calibration. In most of the techniques, the task of collecting the image dataset is committed to end users. However, an end user is
30 normally without specific knowledge about how to acquire an optimal image dataset for camera calibration. In addition, many low cost cameras are equipped merely with a minimal interface, sometimes even without a live-view feature, which makes it impossible for a user to control the images

- 3 -

actually captured. The manual capture and decision for the image dataset is a tedious and time consuming task especially when a minimal interface is used.

- 5 To acquire an optimal calibration image dataset, one essential step is to identify the reference object captured in the given video or images. It is sometimes burdensome and tedious for an operator to perform a manual identification. For example, when a checkerboard is used as
10 the reference object, which is common because of its easily detectable a-priori known geometry, the number of pixel points to be identified is tremendous and the identification result might be poor and thus unreliable.
- 15 An alternative to the manual identification is a calibration tool equipped with a fully automatic function for checkerboard corner detection or a light-assisted one, which is nowadays commonly available on the market, e.g. Camera Calibration Toolbox for Matlab, OcamCalib,
20 EasyCamCalib, Omnidirectional Calibration Toolbox, etc. Nevertheless, most of the tools are limited to certain conditions and are not feasible in some cases, e.g. when the calibration object is only partially visible in the images and when highly distorted images are used. Since
25 wide angle lenses introduce a severe nonlinear distortion into the image formation process, all those techniques that assume a roughly regular geometry of the checkerboard pattern would be inevitably infeasible for such cases, especially when a linear homography mapping between the
30 object space and the image space is used for an approximate location of the checker corners. The high curvature of the checkerboard structure on the image indeed makes the corner prediction unreliable.

- 4 -

Another problem is the partial visibility of the calibration object. A strong calibration of a camera system requires a well-spread distribution of the 2D/3D correspondences, namely the source 3D points should span a
5 volume with sufficient extension in depth with respect to the camera location, and the corresponding 2D image projections should cover with approximately equal density the whole retinal plane. This is an essential condition to achieve an accurate estimation of the calibration
10 parameters, which requires a detection system able to cope with partially occluded patterns. In these cases, most of the calibration tools do not support a human interaction for a removal of false detections or for the integration of supplementary data for missed detections.

15

SUMMARY

Therefore, it is an objective to propose an improved solution for extracting a calibration image dataset from a
20 video. More specifically, the proposed method and device is for automatic image extraction from a video.

According to the one embodiment, a method for extracting images from a video is introduced. The video includes a
25 plurality of images in which a reference object is captured. The method comprises: acquiring the video; analyzing features of the reference object in the images of the video; segmenting the video into clusters, each cluster including at least one image; determining one key image for
30 each cluster based on the analysis of the features; and extracting the key images from the video.

In one embodiment, segmenting the video into clusters includes using the techniques of K-means clustering.

- 5 -

In one embodiment, the reference object includes a reference checker of a planar checkerboard, and the features of the reference object include one or more of
5 velocity, orientation and area of the reference checker, and a representing nodal point closest to the geometric center of the reference checker.

In one embodiment, in particular, analyzing the features
10 includes labeling the images of the video in a tridimensional space defined according to the orientation, area and closest representing nodal point of the reference checker in the images. Segmenting the video into clusters includes using the techniques of K-means clustering on the
15 labeled images in the tridimensional space. Determining one key image for each cluster includes selecting the image in which the reference checker has a lowest velocity for each cluster.

20 In one embodiment, the method further comprises identifying the reference object in the video. Preferably, identifying the reference object includes using Gauss-Laguerre Circular Harmonic Filters.

25 Accordingly, a device configured to extract images from a video is introduced. The device comprises: an acquiring unit configured to acquire the video; an analyzing unit configured to analyze features of the reference object in the images of the video, to segment the video into clusters
30 and to determine one key image for each cluster based on analysis of the features; and an extracting unit configured to extract the key images from the video.

- 6 -

In one embodiment, the device further comprises a user input unit configured to receive a user input for identifying the reference object.

5 Also, a computer readable storage medium has stored therein instructions enabling extracting images from a video, which when executed by a computer, cause the computer to: acquire a video including a plurality of images in which a reference object is captured; analyze features of the
10 reference object in the images of the video; segment the video into clusters, each cluster including at least one image; determine one key image for each cluster based on the analysis of the features; and extracting the key images from the video.

15

The embodiments of the proposed method provide an improved solution to automatically acquire a calibration image dataset. Since the method can be performed automatically in a single device, it is time saving and can provide a higher
20 suitability of the extracted image dataset. The images are extracted according to some optimality criteria.

BRIEF DESCRIPTION OF THE DRAWINGS

25 For a better understanding the proposed solution shall now be explained in more detail in the following description with reference to the figures. It is understood that the solutions are not limited to this disclosed exemplary embodiments and that specified features can also
30 expediently be combined and/or modified without departing from the scope of the present solutions as defined in the appended claims. In the figures:

- 7 -

- Fig. 1 is a flow chart illustrating one preferred embodiment of the proposed image extraction method.
- 5 Fig. 2 is a flow chart illustrating other preferred embodiments of the image extraction method.
- Fig. 3 is a flow chart illustrating an exemplary tracking of a reference object according to one
10 exemplary embodiment of the method.
- Fig. 4 shows the tracking results obtained from the exemplary embodiment of the method shown in Fig. 3.
15
- Fig. 5 shows one exemplary measurement for a grid of nodal points implemented in one exemplary embodiment of the method.
- 20 Figs. 6-7 show two implementation examples of the proposed image extraction method.
- Fig. 8 is a flow chart illustrating the checkerboard detection of one exemplary embodiment of the
25 image extraction method.
- Figs. 9-11 show one implementation example of one embodiment of the proposed method for automatic identification of a calibration object.
30
- Fig. 12 shows an exemplary result obtained from the implementation example shown in Figs. 9-11.

- 8 -

Fig. 13 is a schematic diagram illustrating a device configured to perform the image extraction method.

5 DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Fig. 1 schematically illustrates a preferred embodiment of the proposed method for extracting images from a video. The method comprises: acquiring 10 a video including a plurality of images in which a reference object is captured; analyzing 11 visual features of the reference object in the images of the video; and extracting 12 images from the video based on the analysis of the features. The method is preferably performed in a single device such as a personal computer, a server, a mobile device, a camera, or the like. In other words, the method is advantageously and automatically conducted by the device without external inputs.

20 In one preferred embodiment, acquiring 10 the video includes filming the reference object, particularly by a same device which performs the analyzing and extracting steps. In this case, the device is equipped with the functions such as filming and editing a video, and thus can acquire a video or an image sequence by its own. For example, a multi-functional mobile phone or a mobile device can be used for the implementation of this preferred embodiment. Of course, the video can also be acquired from an external device such as a supplementary camera, a database, or any other suitable apparatus. If the captured video is used for a subsequent camera calibration procedure, the same is captured by the camera which is to be calibrated.

- 9 -

As shown in Fig. 2, in another preferred embodiment, the method further comprises identifying 13 the reference object in the video. The identification 13 of the reference object can be automatically performed by the device or
5 alternatively by a user via a user input.

In yet another preferred embodiment, the method optionally comprises segmenting 14 the video into several clusters each including at least one image. In this embodiment,
10 extracting 12 images from the video includes determining one key image for each of the clusters and then extracting the key images.

In the following one exemplary embodiment of the proposed
15 method is described in detail. In this embodiment, a video is acquired by capturing a calibration object from various distances and orientations using a camera. A checkerboard is used as the calibration object for this embodiment, while a central checker is defined as the reference object
20 which preferably remains visible across the images of the whole video. Of course, any one of the checkers of the checkerboard can be defined and chosen as the reference object.

25 In this embodiment, the central checker is identified 13 by a user through a user input by localization of the four corners thereof in a clockwise or a counterclockwise order in the first image of the video. Alternatively, the reference object can be automatically defined by an
30 operation device depending on different situations and demands. More details about automatic identification are given below in an implementation example.

- 10 -

The analysis 11 of the features of the reference object, i.e., the central checker in this embodiment, advantageously comprises temporally tracking the reference object across the images of the video and extracting and
 5 analyzing the features of the reference object. In particular, the positions of the four checker corners of the reference checker are tracked and extracted, which are subsequently used to extract four features related to the motion state and visual appearance of the reference object.

10

Fig. 3 shows a flowchart illustrating the procedure of an exemplary tracking of the reference checker. First, the locations of the corners of the reference checker are predicted 15. The predicted position, the estimated
 15 position and the velocity of the i -th corner in the j -th image of the video are denoted as $(\hat{p}_{i,j}, \tilde{p}_{i,j}, v_{i,j})$. In the first image of the video, the predicted positions are identified via a user interaction or by a device automatically, and the velocity is initialized to zero. Then the tracking of
 20 the reference checker proceeds across the successive images using a prediction-measurement scheme. A constant velocity dynamical model is applied to compute the predicted corner locations in the images,

25

$$\hat{p}_{i,j} = \tilde{p}_{i,j-1} + v_{i,j-1} ,$$

where the velocity is computed as

$$v_{i,j-1} = \tilde{p}_{i,j-1} - \tilde{p}_{i,j-2} .$$

30

The measurement step is implemented by searching 17 for the local maxima of a corneriness measure within a small search region, denoted as W_i , centered on the corner predicted

- 11 -

location $\hat{p}_{i,j}$. To compute the corneriness measure, many different solutions are available in the literature. Among them, it is opted in this embodiment for the magnitude of the complex response of the image to the Gauss-Laguerre
 5 Circular Harmonic Filter $(2,0)$, denoted as $I_0^2(p)$, since it has been shown to be an effective detector of a checker pattern.

In order to achieve a subpixel accuracy of the corner
 10 detection, the center of mass of all pixels with response magnitude higher than a given threshold is computed,

$$\tilde{p}_{i,j} = \frac{\sum_{p \in \Omega} |I_0^2(p)| p}{\sum_{p \in \Omega} |I_0^2(p)|}$$

where Ω denotes the set of pixels fulfilling the
 15 condition. In this implementation,

$$th_{\Omega} = 0.9 \times \max_{p \in W_i} \{|I_0^2(p)|\}$$

is used as threshold for the definition of the set Ω .

20 Fig. 4 shows an exemplary result of the tracking of the reference checker, wherein Fig. 4(a) shows the shape of the corner detection filter, Fig. 4(b) shows the magnitude of the response and Fig. 4(c) shows the extracted corner position, which is accurately located at the corner of the
 25 checkers.

Subsequently, the temporal tracking of the reference checker is used to compute four quality measures of the reference checker in the images of the video and to
 30 populate the calibration image dataset from the video.

- 12 -

The four quality measures are defined according to the features that are preferably retrieved in the final image dataset, namely low speed, slanted and front-to-parallel orientation of the reference object, exhaustive coverage of the image retinal plane with the reference object, and the distance of the reference object from the camera.

The overall measure of the velocity of the reference checker is defined as the sum magnitudes of the velocities of the checker corners:

$$v_j = \sum_{k=1, \dots, 4} \|v_{k,j}\|$$

It is assumed that the formula provides an indirect indication of image blur likelihood and, therefore, is used to identify the sharpest images of the video, where the corner localization can reach the highest accuracy.

The second measure deals with the orientation of the reference checker. It is observed that in an ideal case where a camera is not affected by optical distortion, a front-to-parallel configuration maintains the rectangular shape of the checker in the captured images. Therefore, in this embodiment, the projective distortion of the reference checker is defined and taken as an indicator of the relative slant between the camera and the reference object. The distortion measure, denoted as θ_j , is derived from two simple geometrical features, namely the ratio between the checker diagonals and the ratio between the longest and shortest checker side:

$$\theta_j = \min(\theta_{s,j}, \theta_{d,j}), \text{ where } \begin{cases} \theta_{s,j} = \frac{\min_i \{s_{i,j}\}}{\max_i \{s_{i,j}\}} \\ \theta_{d,j} = \frac{d_j}{D_j} \end{cases}$$

- 13 -

where d_j and D_j are the short and long diagonals, and $s_{i,j}$ is the length of the i -th checker side.

5 An exhaustive coverage is indeed helpful in order to guarantee a reliable estimation of the camera projection model also in the image periphery. For this purpose, a regular grid of nodal points are denoted as c_k and defined as representatives of the target image areas, and each
 10 image is labeled according to the nodal point closest to the reference checker:

$$k_j = \underset{k=1,\dots,K}{\operatorname{argmin}} \{ \|\bar{p}_j - c_k\| \}$$

where $\bar{p}_j$ is the geometric center of the reference checker. Fig. 5 shows the grid of nodal points used in this
 15 embodiment and also the resulting space partitioning as a Voronoi diagram.

The fourth measure deals with the distance of the reference object from the camera. The area of the central reference
 20 checker, denoted as σ_j , is used as a simple distance indicator. It is reasonable to assume that the bigger the checker is, the closer the reference object is. The area is simply computed using a diagonal to split the checker in two triangles and summing up the corresponding areas.

25

After the above computation of the four quality measures, the same are analyzed in order to decide which image of the video should be selected and extracted from the given
 video.

30

A tridimensional labeling space $L = L_\theta \times L_\sigma \times L_k$ is defined and implemented for the whole given video, where $L_\theta = \{\theta_s, \theta_p\}$

- 14 -

is used to indicate the slanted or front-to-parallel orientation of the reference object, $L_\sigma = \{\sigma_N, \sigma_F\}$ is the distance of the reference object from the camera, and $L_k = \{1, \dots, K\}$ is the corresponding image region.

5

The L_k labeling is a relatively simple task as it corresponds to the closest nodal point label k_j . The labeling L_θ and L_σ serve as two numerical thresholds that are required to be compared with the distance and
 10 distortion measures θ_j and σ_j . K-means clustering algorithm is used in this exemplary embodiment, which is aimed to partition the 2D observation vectors (θ_j and σ_j) in $k = 4$ clusters. The cluster centers, needed as algorithm initialization, are provided by the four possible
 15 combinations of the extreme values assumed by θ_j and σ_j , i.e., four extremities spanned by the scaled measures of θ_j and σ_j .

Once the whole acquired video is labeled, the image of the
 20 video with lowest velocity is selected and extracted from each cluster of the video. The extracted images then provide the image dataset for subsequent camera calibration process.

25 Figs. 6 and 7 show two implementation examples for the above method, where a Panasonic stereo camcorder with normal 50mm-like field of view (Fig. 6) and a GoPro HD Hero2 camera with a 170° wide-angle lens (Fig. 7) are respectively used. The black squares shown in the curve
 30 diagrams represent the extracted images which are displayed at the bottom of the respect figures.

- 15 -

In an ideal case, the extracted images from the video should contain a representative for each label combination, in order to meet the requirements regarding the volumetric extension of the virtual calibration structure and the coverage of image retinal plane. This leads to an image dataset with a maximum cardinality of $4 \times K$. In practice, it sometimes occurs that not all labels are represented in the final dataset. However, based on the above exemplary embodiment of the proposed method, the extracted image dataset is the one which best meets the suitability requirements, given the available video.

The proposed image extraction method provides an improved solution to automatically acquire a calibration image dataset by a simple and normal device or computer. The initial video input of the method can be easily acquired from a supplementary camera or by a device itself directly without special requirements or limitations, which is also feasible for end consumers. For example, the video can be captured by simply waving a reference object in different orientations and distances in front of a filming camera, where the camera can stand at one fixed position without moving.

Given the video, the image extraction method simplifies the procedure to acquire an image dataset from the video and optimize the dataset suitability for a subsequent camera calibration process. Comparing to typical solutions which acquire an image dataset from a video by manual review and selection of the images from an operator, the proposed image extraction method is time saving and can be widely used in either normal or critical situations such as an outdoor environment where a qualitative visual analysis of a calibration dataset might not be possible.

Automatic identification of calibration object

As described above, in the proposed image extraction
5 method, the preliminary identification of the reference
object can be advantageously and automatically performed by
a device, a computer, or the like. In the following
implementation example of the automatic identification, a
checkerboard is used as the calibration object.

10

Fig. 8 shows an exemplary procedure of the checkerboard
detection in this example. First, the identification of
seed corners, e.g. three corners of a central checker, is
performed, either by a user via a user input or
15 automatically by a device. This information is used to
initialize the reference corner location and the analysis
of local velocity in the direction of the grid axis, which
represent the seed for the following image exploration. An
automatic labeling of the whole set of checkerboard corners
20 is subsequently performed by means of a simple corner
extractor integrated within a recursive framework.
Essentially, the checkerboard is progressively discovered
by walking along the grid edges from corner to corner in
the direction of the four connected neighbors. The
25 exploration directions and the labels of the searched
corners are completely defined by the information available
for the initiating corner. The local response to a pair of
circular harmonic filters, which are specifically tailored
to edges and corners, is used to update the discovered
30 velocities and to localize the corners with sub-pixel
precision. By means of this framework, the geometrical
information are propagated from one seed corner, and the
set of detected corners is progressively updated using only

- 17 -

the local image information, which make the technique insensible to image distortion or partial visibility.

The grid dataset is denoted as $G = \{n_i = [X_i, Y_i, x_i, y_i, v_{iX}, v_{iY}]\}_{i=1, \dots, N}$,
 5 where N is the number of corners, (X_i, Y_i) and (x_i, y_i) are the coordinates of the i -th corner in object and image spaces and (v_{iX}, v_{iY}) are the local velocity in X and Y directions, namely the directions in image space towards the next corners $(X_i + 1, Y_i)$ and $(X_i, Y_i + 1)$.

10

For the identification of seed corners, three corners $(\hat{x}_j, \hat{y}_j)_{j=1,2,3}$ of the reference checker are selected and used to instantiate the seed corners and the local velocities:

$$15 \quad \begin{cases} G = \{n_1\} \\ n_1 = [0, 0, x_1, y_1, v_{1X}, v_{1Y}] \\ (x_1, y_1) = \text{ref}(\hat{x}_1, \hat{y}_1) \\ v_{1X} = \frac{[\hat{x}_2, \hat{y}_2]^T - [\hat{x}_1, \hat{y}_1]^T}{\|[\hat{x}_2, \hat{y}_2]^T - [\hat{x}_1, \hat{y}_1]^T\|} \\ v_{1Y} = \frac{[\hat{x}_3, \hat{y}_3]^T - [\hat{x}_1, \hat{y}_1]^T}{\|[\hat{x}_3, \hat{y}_3]^T - [\hat{x}_1, \hat{y}_1]^T\|} \end{cases} \quad (1)$$

where function $\text{ref}(\cdot)$ identifies the sub-pixel refinement of the corner localization.

20 Starting from this seed the checkerboard is discovered by recursively calling a corner-to-corner detection method which alternates edge walking and velocity update steps until the next corner is detected or the image border is reached. This recursion is invoked from the calling node in
 25 the direction of the corners belonging to the four-connected neighborhood. For example, from a node n_i , the search for the corners $(X_i, Y_i + 1)$ and $(X_i, Y_i - 1)$ is invoked using the velocities v_{iY} and $-v_{iY}$, which is analogously done for the other two connected nodes in the X direction.

The edge walking step is performed in a filtering based context using the Gauss-Laguerre Circular Harmonic Filters (GL-CHF). The set of GL polynomials $\{L_k^n(\rho, \theta); n, k \in N\}$ provides

5 a complex basis suitable for the orthogonal expansion of square integrable functions represented in polar coordinates, under the weighting of a Gaussian window

$V(\rho) = e^{-\pi\rho^2}$ [VII]:

$$f(\rho, \theta)V\left(\frac{\rho}{\sigma}\right) = \sum_{n,k} \hat{f}_k^n(\sigma) \frac{1}{\sigma} L_k^n\left(\frac{\rho}{\sigma}, \theta\right)$$

10 In literature the set of coefficients $\hat{f}_k^n(\sigma)$ is referred as the Gauss Laguerre Transform (GLT) of $f(\rho, \theta)$ at scale σ and is obtained by projection of $f(\rho, \theta)$ onto the basis provided by the Gaussian-weighted GL polynomials. Equivalently, in the case of digital images, the GLT can be obtained by

15 filtering through a bank of GL-CHFs.

The GLT provides a representation of an image as a stack of sub-bands, and the first of which are selectively tuned on different low level features, such as corners, borders or

20 lines. In particular, the sub-bands $\hat{f}_0^1$ and $\hat{f}_0^2$ may be accounted as edgeness and cornerness maps and are used for the edge walking and the corner localization in this implementation example.

25 The high selectivity of the GL-CHFs to specific image features suggests the possibility to exploit the local response $\hat{f}_0^1(x)$ to follow an edge between two consecutive corners. Indeed, the magnitude of the local response supports the edge localization and provides an estimate for

30 its local direction [VIII]. Fig. 9(a) shows a sample template of a checker corner of a checkerboard, and Fig.

- 19 -

9(b) shows the magnitude and phase of the edgeness map extracted from the sample template of Fig. 9(a) using GL-CHFs techniques. Since the edges are not straight on the image plane due to distortion introduced by the optical system, the edge walking is performed by iteratively taking short steps and updating the the edge direction according to the local response $\hat{f}_0^1(x)$.

Given the current edge position $p = (x, y)$, a step is taken in the direction of the local velocity v_p . The predicted edge location $p' = p + \delta v_p$ is used to identify a segment denoted as L and orthogonal to the edge itself. Along this segment, the pixel with the highest edgeness magnitude is selected and the local velocity is update according to the local phase of $\hat{f}_0^1(x)$:

$$\begin{cases} p^+ = \arg \max_{p \in L} \|\hat{f}_0^1(p)\| \\ v_{p^+} \leftarrow \frac{\pi}{2} + \angle(\hat{f}_0^1(p^+)) \end{cases}$$

Fig. 10 illustrates an exemplary graphic for the above step. The point A represents the location p' of the edge predicted using the current velocity estimate. From this point a search 1D space (the solid line) segment is inspected in order to refine the edge location by searching for the maximum of the local edgeness map (point B). The local velocity is updated using the phase of the local edgeness map. The edge is travelled until the update of the direction of the local velocity exceeds a given threshold. This happens when a corner is crossed, as the inversion of the checker colors determines a rotation of approximate π radians in the phase of the filter response.

- 20 -

Once the edge walking termination condition is met, the corner location is locally refined using the cornerness map $\hat{f}_0^2(x)$. The corner candidates are extracted by searching for the local maxima of the cornerness map within a small neighborhood of the current edge location. In highly distorted images and in particular in the peripheral areas, it may happen that a set of multiple candidates is detected, from which only the most likely ones are retained. For this purpose, the distribution of the edgeness map around each corner candidate is inspected and the coherence with the grid topology is measured.

The set of corner candidates is denoted as $C = \{c_j\}_{j=1, \dots, N_C}$, with Γ_i denoted as the closed square contour centered on each corner candidate and $D_j = \{m_{j,k} \in \Gamma_j\}_{k=1, \dots, K_j}$ is the set of local maxima of the edgeness map detected on the same contour. If the candidate is correctly centered on the checkerboard corner, the set of velocities directly sampled from the edgeness on the location given by D_j should be aligned with the directions analytically defined by the coordinate of the points in a same set D_j , which is schematically shown in Fig. 11(a).

Following the above assumption, a quality measure q is defined for each corner candidate:

$$\begin{cases} q_j = \min_{k=1, \dots, K_j} \{1 - |\langle v_{j,k}, \hat{v}_{j,k} \rangle|\} \\ \hat{v}_{j,k} = \frac{m_{j,k} - c_j}{\|m_{j,k} - c_j\|} \\ v_{j,k} = \frac{\hat{f}_0^1(m_{j,k})}{|\hat{f}_0^1(m_{j,k})|} \end{cases}$$

The candidate with the highest score is selected. To achieve a sub-pixel accuracy of the pixel position, an

- 21 -

elliptical paraboloid is fitted using a 3×3 window centered on the selected candidate and the local minima are analytically computed. This simple refinement step is introduced as described above in (1) with the notation
5 *ref(.)*.

Fig. 12 shows an exemplary result of this implementation example for automatic identification of a checkerboard. Fig. 12(a) shows the identification of the seed corners and
10 Fig. 12(b) shows the final result of the checkerboard identification. The images are captured using a GO-PRO Hero camera which is equipped with a very wide angle lens able to operate in two different configurations, 120° and 170° field of view. The latter, i.e., the narrowest
15 configuration, is used in this implementation example.

It can be seen that the checker corners are accurately identified, even including those located close to the image border. In addition, even though the captured checkerboard
20 is severely distorted due to the wide angle lens and is only partially visible, the detection and identification of the checkerboard are still reliable. The missed detection rate of the checker corners is highly reduced, contributing to a superior collection of 2D/3D correspondences in
25 subsequent camera calibration procedures. In other words, the automatic identification technique can automate the creation of the 2D/3D correspondences dataset and is suitable for exploitation with a wide range of lenses, from the regular field of view up to very wide-angle fisheye
30 lenses and full panoramic catadioptric systems.

Fig. 13 schematically shows a device 20 configured to perform the embodiments of the proposed image extraction

- 22 -

method. The device 20 comprises an acquiring unit 21 configured to acquire a video; an analyzing unit 22 configured to analyze features of a reference object captured in the images of the video; and an extracting unit
5 23 configured to extract images from the video based on the analysis of the features.

Preferably, the acquiring unit 21 is configured to film a reference object. More preferably, the analyzing unit 22 is
10 configured to identify the reference object and segment the video into several clusters each including at least one image, and the extracting unit 23 is configured to determine one key image for each of the clusters and extract the key images. Optionally, the device 20 further
15 comprises a user input unit 24 configured to receive a user input for identifying the reference object captured in the images of the given video. The analyzing unit 22 and the extracting unit 23 can be optionally combined into a single device or function separately.

20

In one embodiment, the reference object includes a reference checker of a planar checkerboard. The analyzed unit 22 is configured to analyze 11 the features by labeling the images of the video in a tridimensional space
25 defined according to the orientation, area and closest representing nodal point of the reference checker in the images, and to segment 14 the video into clusters by using the techniques of K-means clustering on the labeled images in the tridimensional space. The extracting unit 23 is
30 configured to determine one image for each cluster by selecting the image in which the reference checker has a lowest velocity for each cluster.

REFERENCES

- [I] D. Liebowitz, "Camera Calibration and Reconstruction of Geometry from Images," D.Phil. thesis, University of Oxford, 2001
- [II] E. Hemayed, "A survey of camera self-calibration," Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance, pp.351-357, 2003
- [III] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision (2 ed.)*. Cambridge University Press, New York, NY, USA, 2003
- [IV] R.Y. Tsai, "An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 364-374, 1986
- [V] J. Salvi, X. Armangué and J. Batlle, "A comparative review of camera calibration methods with accuracy evaluation," Pattern Recognition, Vol. 35, Issue 7, pp. 1617-1635, 2002
- [VI] Z. Zhang, "A flexible new technique for camera calibration." IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11):1330-1334, 2000
- [VII] G. Jacovitti, A. Neri, "Multiresolution circular harmonic decomposition", IEEE Transactions on Signal Processing, 48(11):3242-3247, 2000
- [VIII] L. Sorgi, "Edgelet tracking using Gauss-Laguerre Circular Harmonic filters". ICIP 2011: 2897-2900, 2009

CLAIMS

1. A method for extracting images from a video including a plurality of images in which a reference object is captured, the method comprising:
 - acquiring (10) the video;
 - analyzing (11) features of the reference object in the images of the video;
 - segmenting (14) the video into clusters based on the analysis of the features, each cluster including at least one image;
 - determining one image for each cluster based on the analysis of the features; and
 - extracting (12) the determined images from the video.
2. The method of claim 1, wherein segmenting (14) the video into clusters includes using the techniques of K-means clustering.
3. The method of claim 1 or 2, wherein the reference object includes a reference checker of a planar checkerboard, and the features of the reference object include one or more of velocity, orientation and area of the reference checker, and a representing nodal point closest to the geometric center of the reference checker.
4. The method of claim 3, wherein
 - analyzing (11) the features includes labeling the images of the video in a tridimensional space defined according to the orientation, area and closest representing nodal point of the reference checker in the images,

- 25 -

- segmenting (14) the video into clusters includes using the techniques of K-means clustering on the labeled images in the tridimensional space, and
 - determining one image for each cluster includes selecting the image in which the reference checker has a lowest velocity for each cluster.
- 5
5. The method of one of the preceding claims, wherein analyzing (11) features of the reference object includes tracking the reference object across the images of the video and extracting features of the tracked reference object.
- 10
6. The method of one of the preceding claims, wherein acquiring (10) the video includes filming the reference object.
- 15
7. The method of one of the preceding claims, further comprising identifying (13) the reference object in the video.
- 20
8. The method of claim 7, wherein identifying (13) the reference object includes using Gauss-Laguerre Circular Harmonic Filters.
- 25
9. A device (20) configured to extract images from a video including a plurality of images in which a reference object is captured, the device comprising:
 - an acquiring unit (21) configured to acquire (10) the video;
 - an analyzing unit (22) configured to analyze (11) features of the reference object in the images of the video, to segment (14) the video into clusters based on
- 30

- 26 -

the analysis of the features, each of the clusters including at least one image; and

5 - an extracting unit (23) configured to determine one image for each cluster based on the analysis of the features and to extract (12) the determined images from the video.

10. The device (20) of claim 9, wherein the reference object includes a reference checker of a planar checkerboard, and the analyzing unit (22) is configured to

10 - analyze (11) the features by labeling the images of the video in a tridimensional space defined according to the orientation, area and closest representing nodal point of the reference checker in the images, and

15 - segment (14) the video into clusters by using the techniques of K-means clustering on the labeled images in the tridimensional space,

and the extracting unit (23) is configured to

20 - determine one image for each cluster by selecting the image in which the reference checker has a lowest velocity for each cluster.

11. The device (20) of claim 9 or 10, wherein the analyzing unit (22) is configured to track the reference object across the images of the video and to extract features of the tracked reference object.

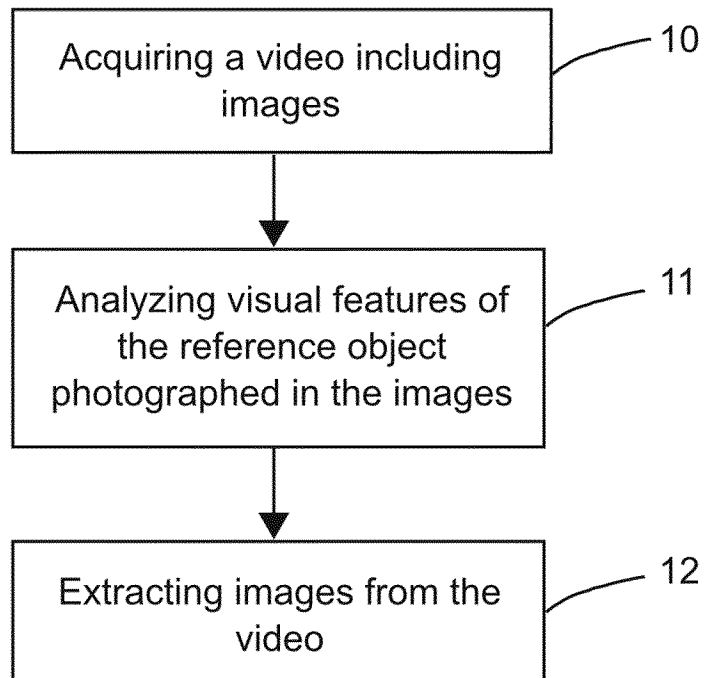
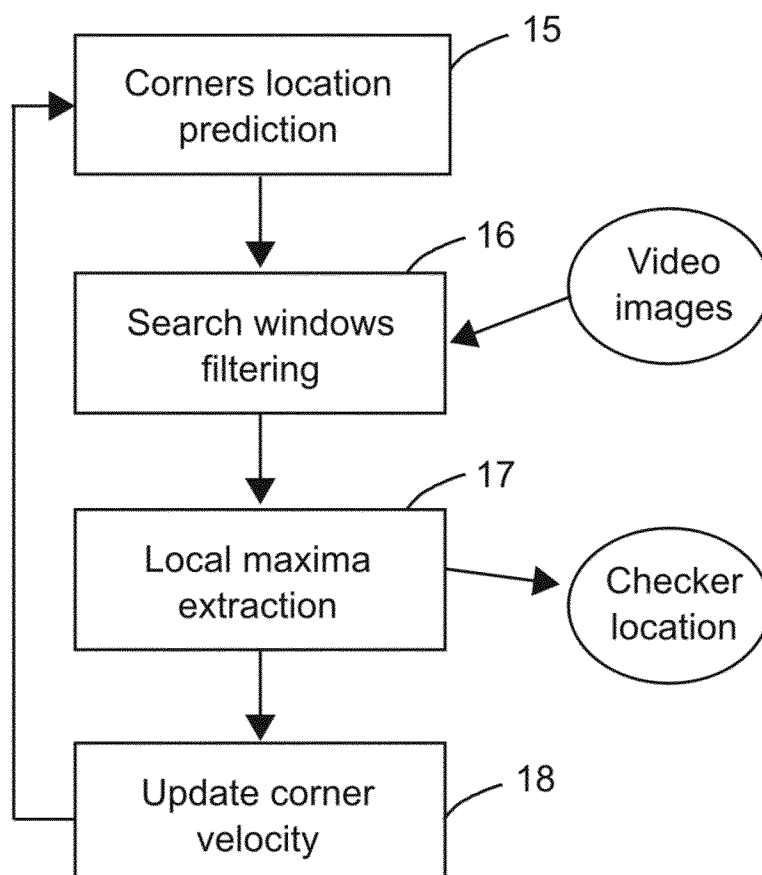
25

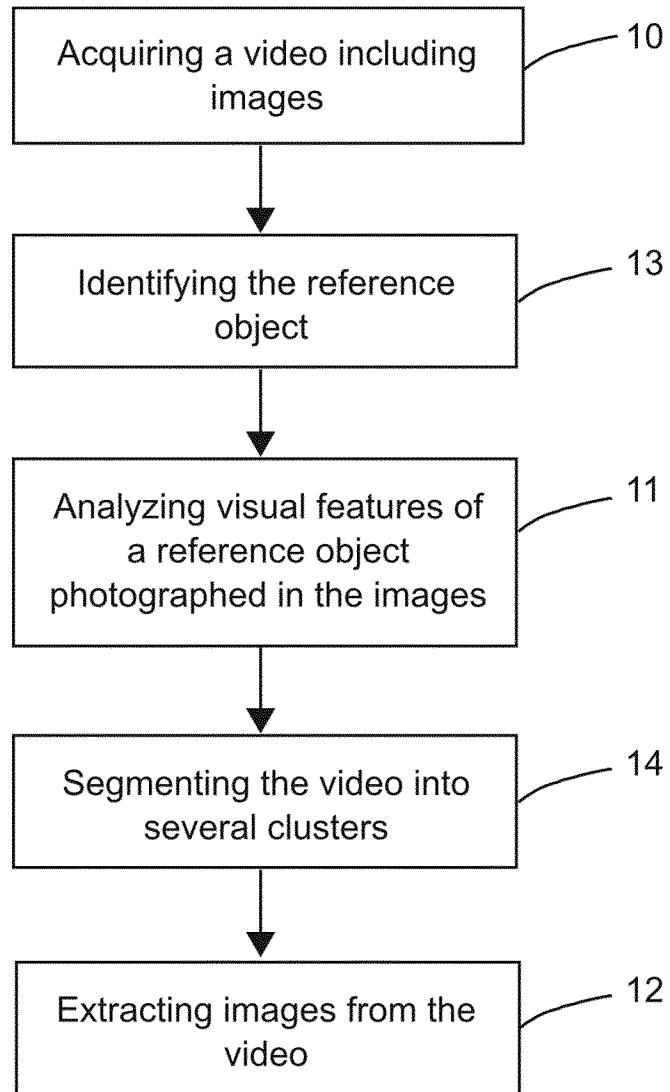
12. The device (20) of one of claims 9 to 11, wherein the acquiring unit (21) is configured to film the reference object.

30

- 27 -

13. The device (20) of one of claims 9 to 12, wherein the analyzing unit (22) is configured to identify (13) the reference object in the video.
- 5 14. The device (20) of one of claims 9 to 13, further comprising a user input unit (24) configured to receive a user input for identifying the reference object.
- 10 15. A computer readable storage medium having stored therein instructions enabling extracting images from a video, which when executed by a computer, cause the computer to:
- acquire a video including a plurality of images in which a reference object is captured;
 - 15 - analyze features of the reference object in the images of the video;
 - segment the video into clusters based on the analysis of the features, each cluster including at least one image;
 - 20 - determine one image for each cluster based on the analysis of the features; and
 - extracting the determined images from the video.

**Fig. 1****Fig. 3**

**Fig. 2**

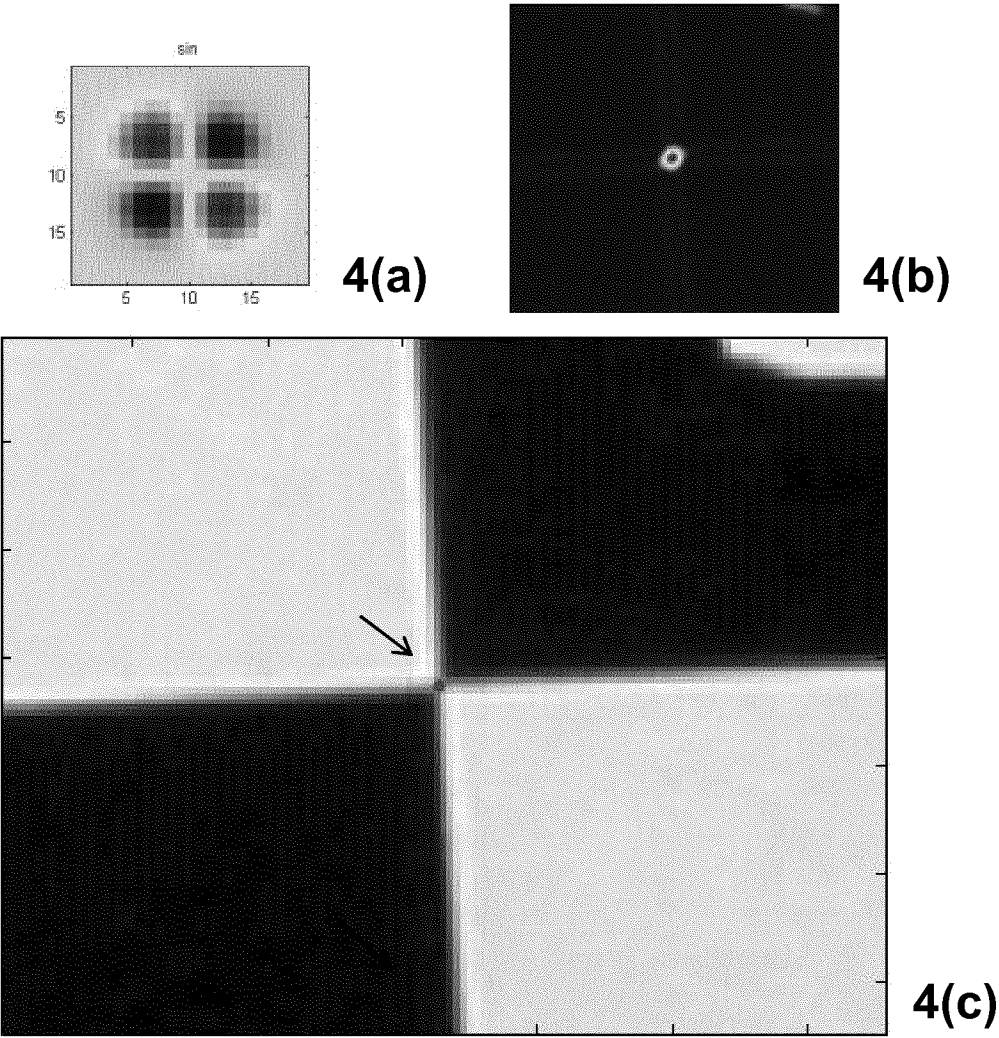


Fig. 4

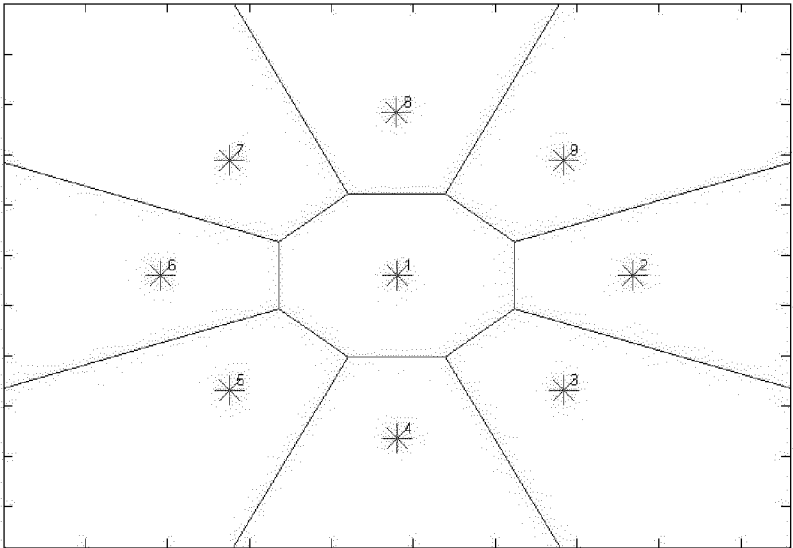
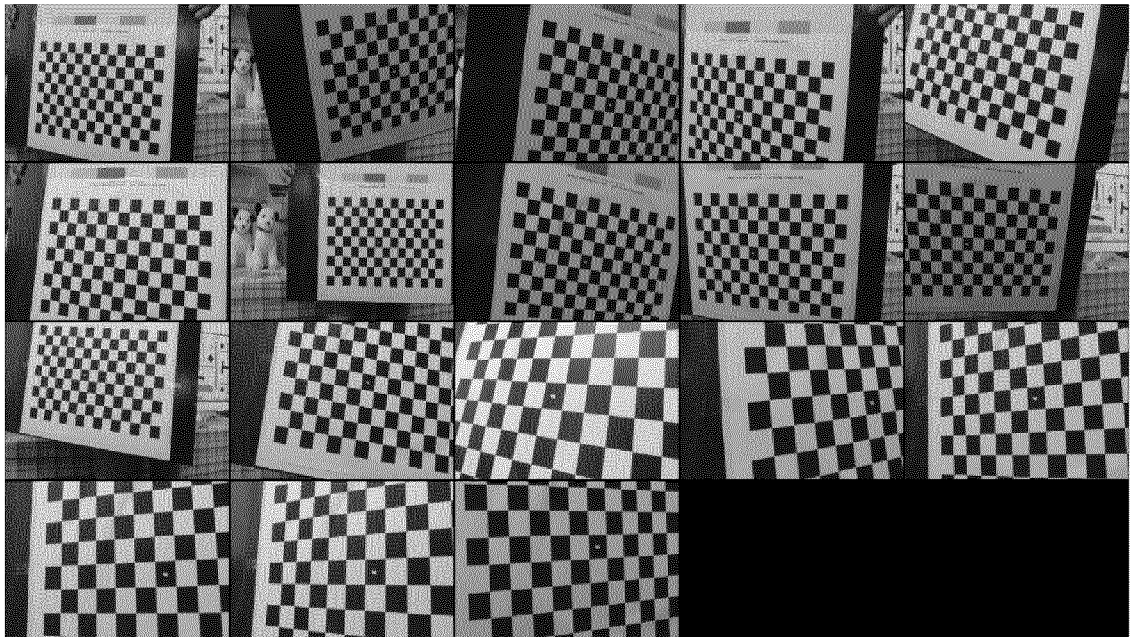
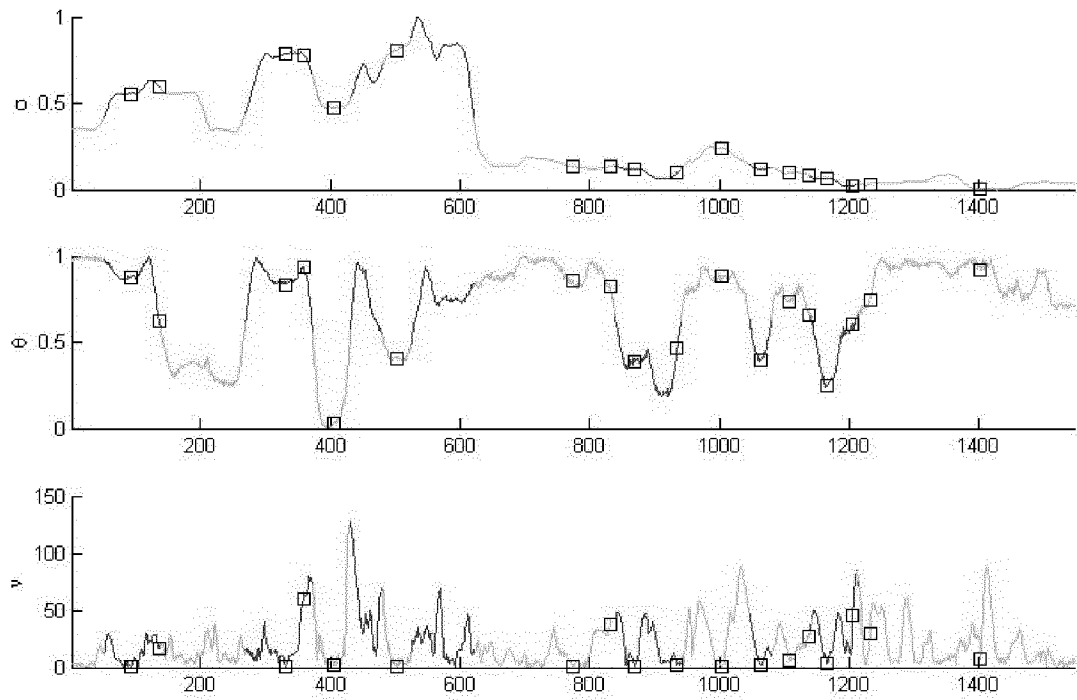
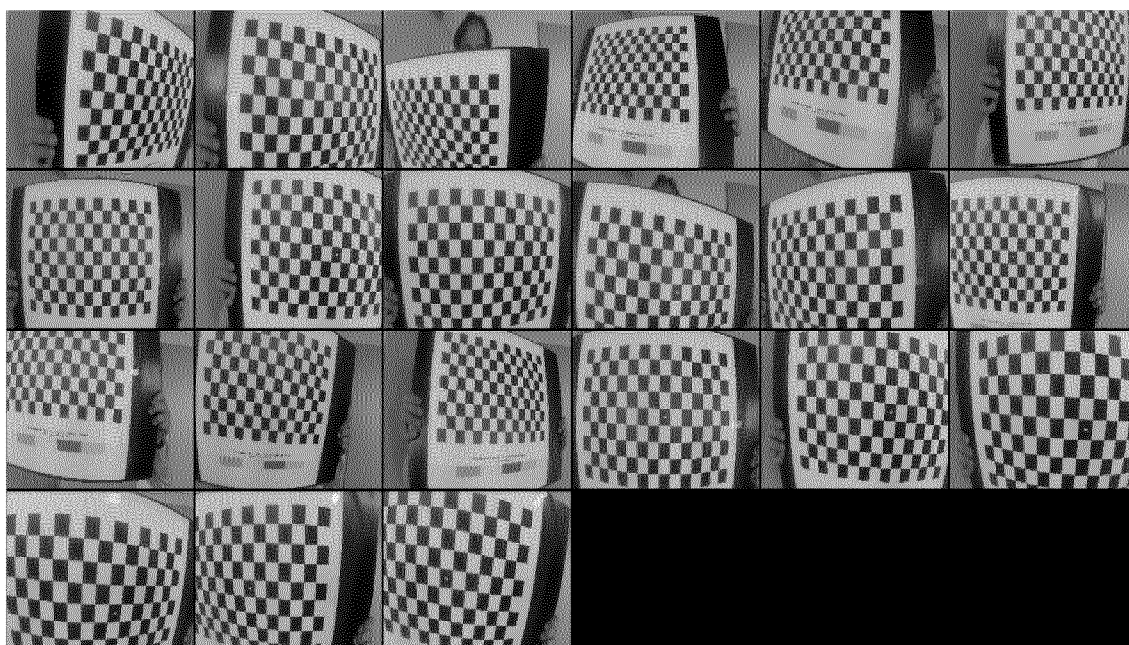
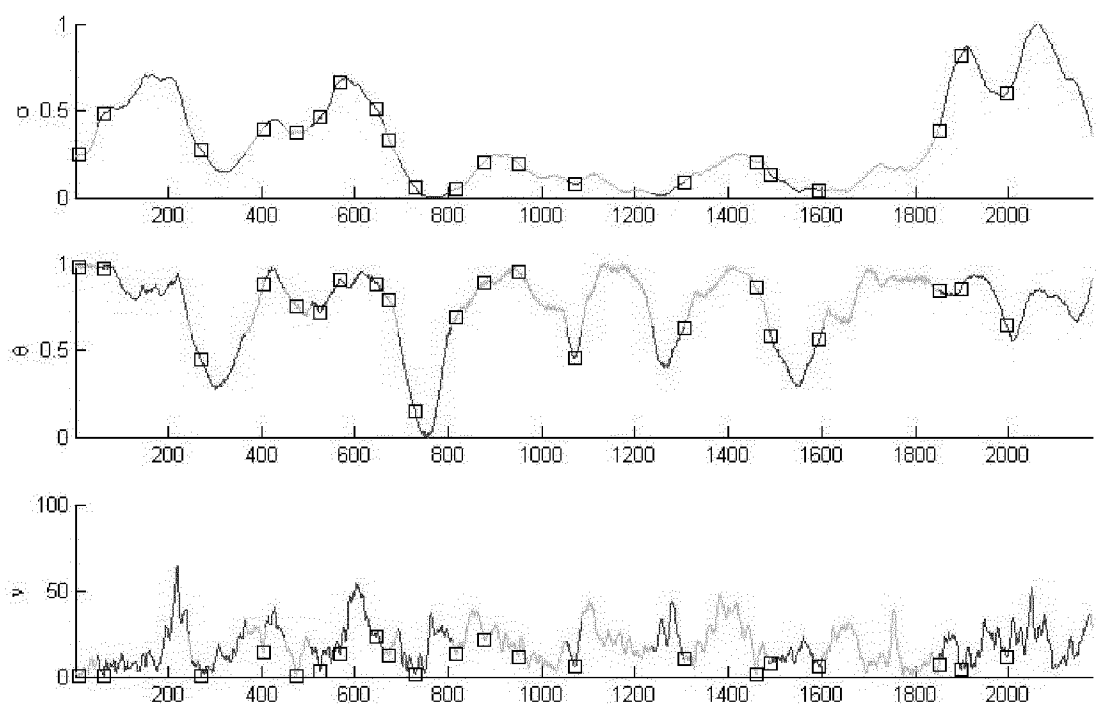
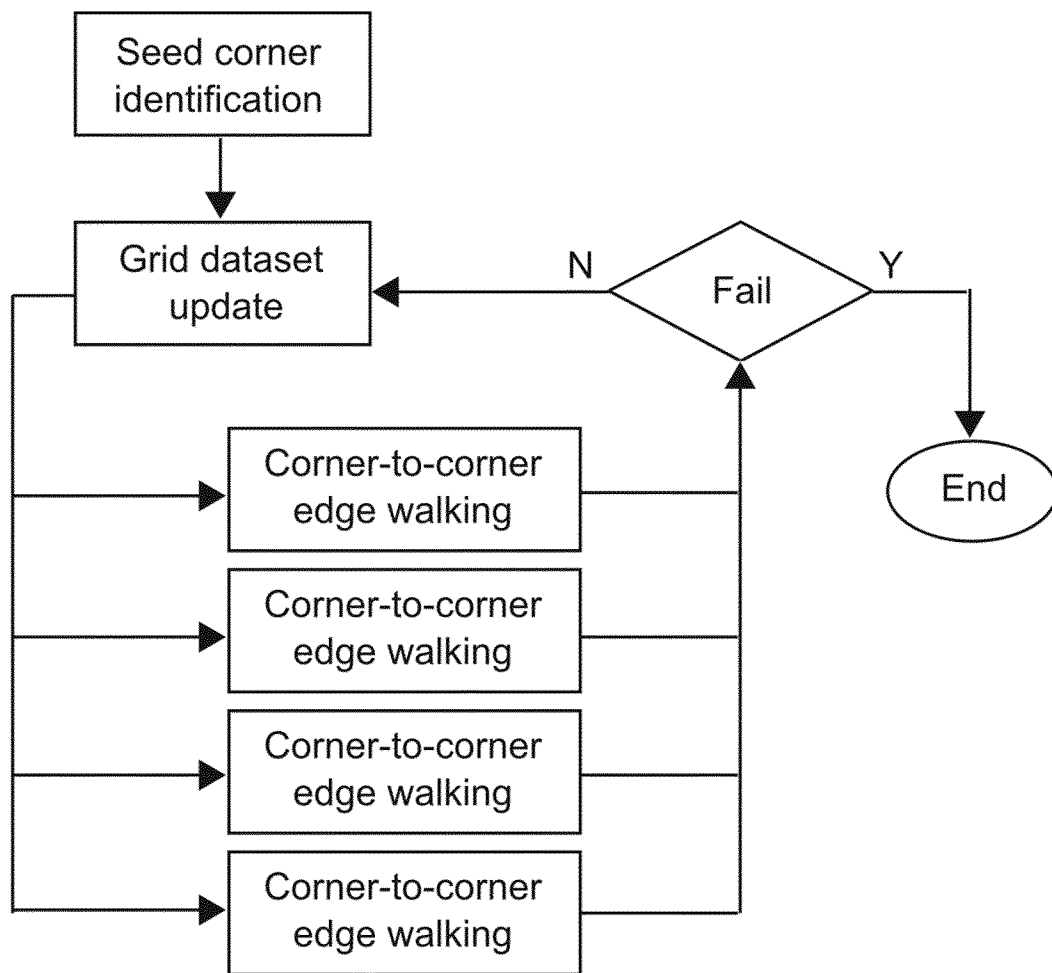
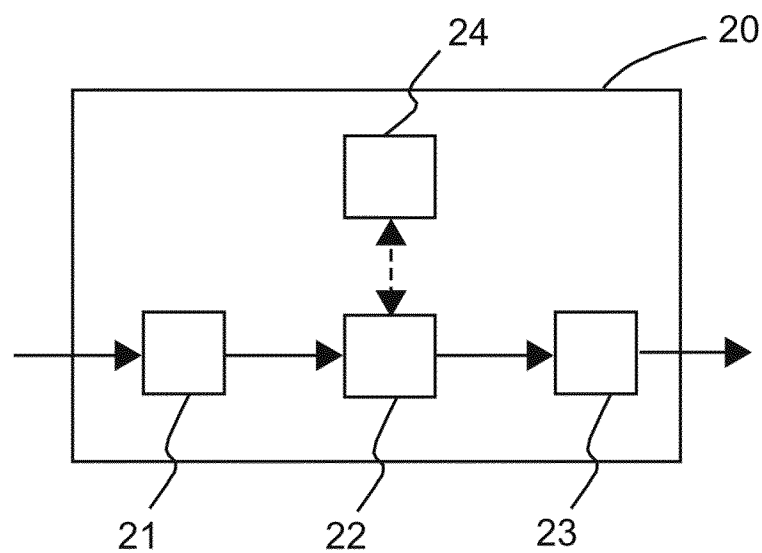
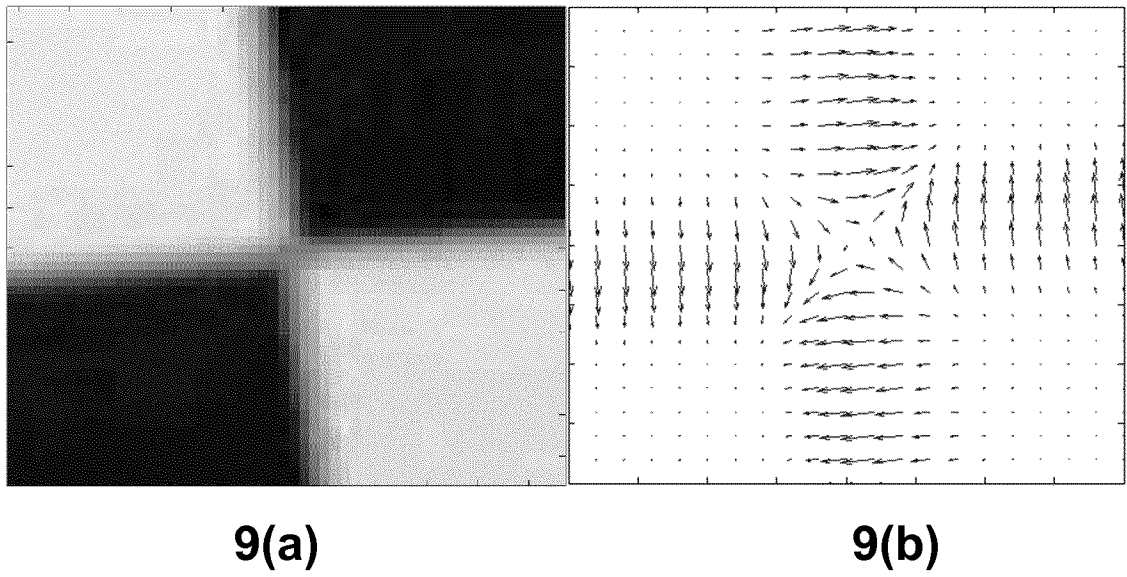
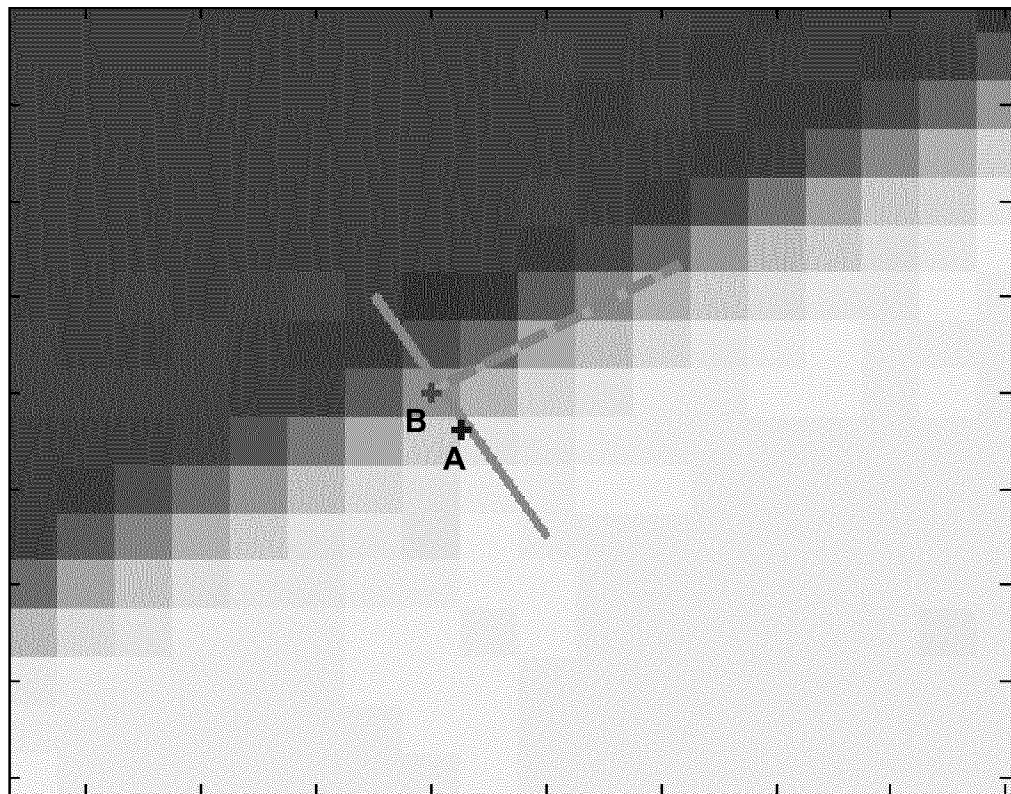


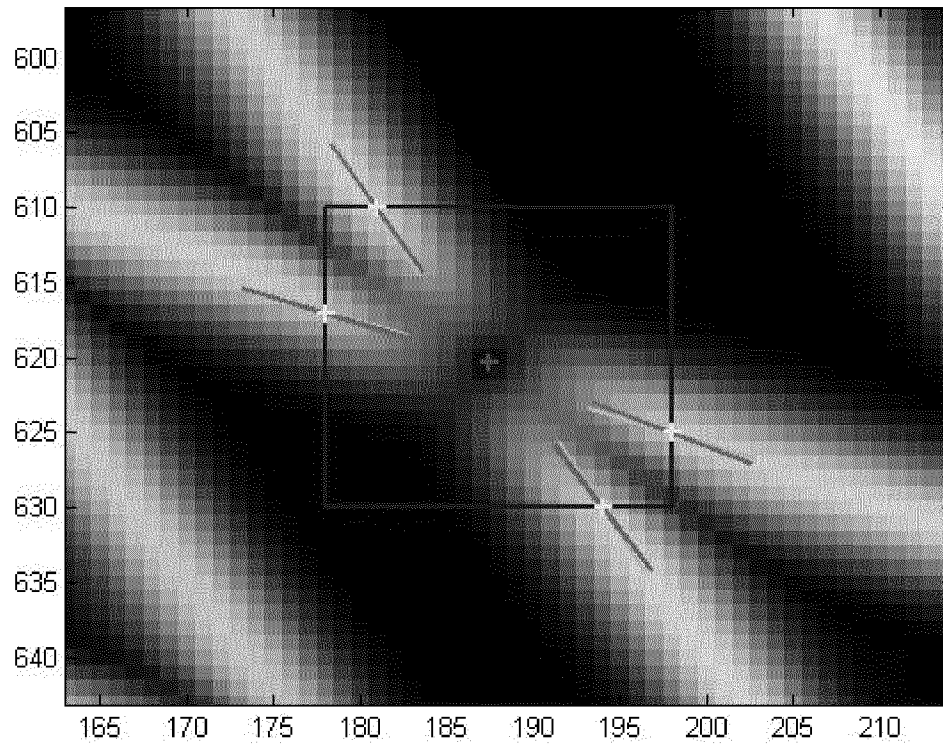
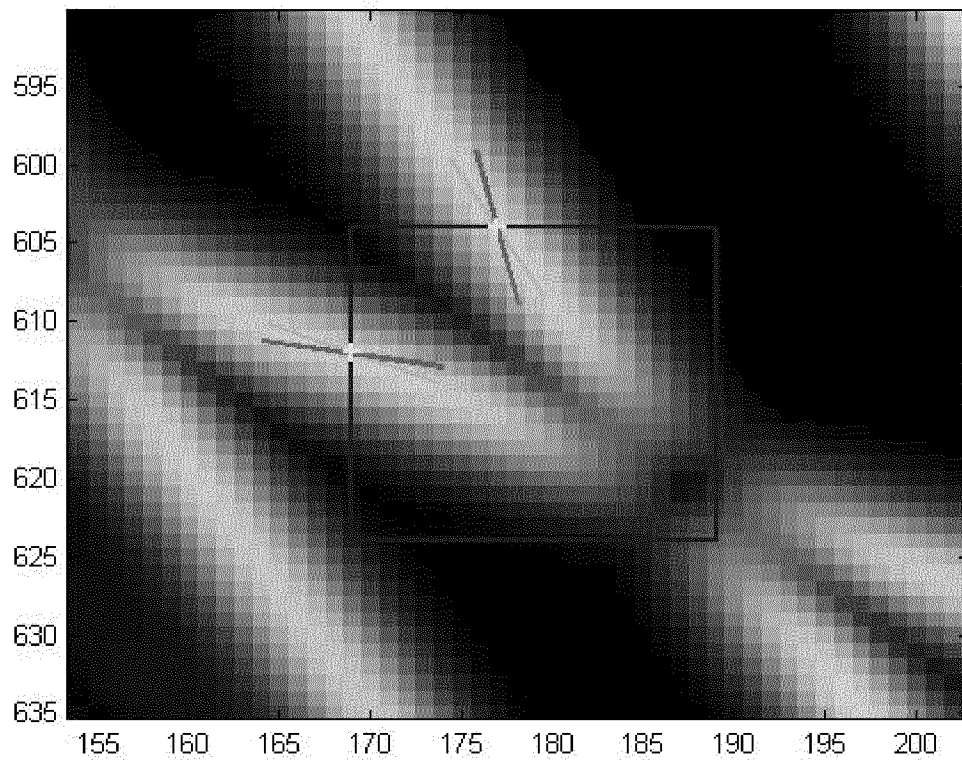
Fig. 5

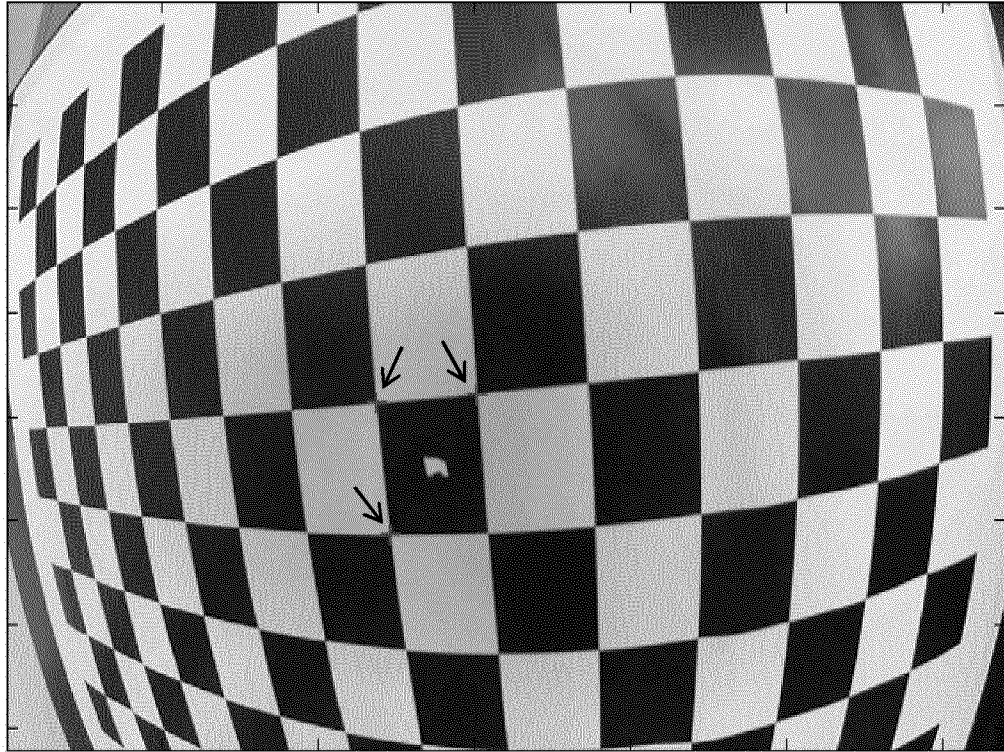
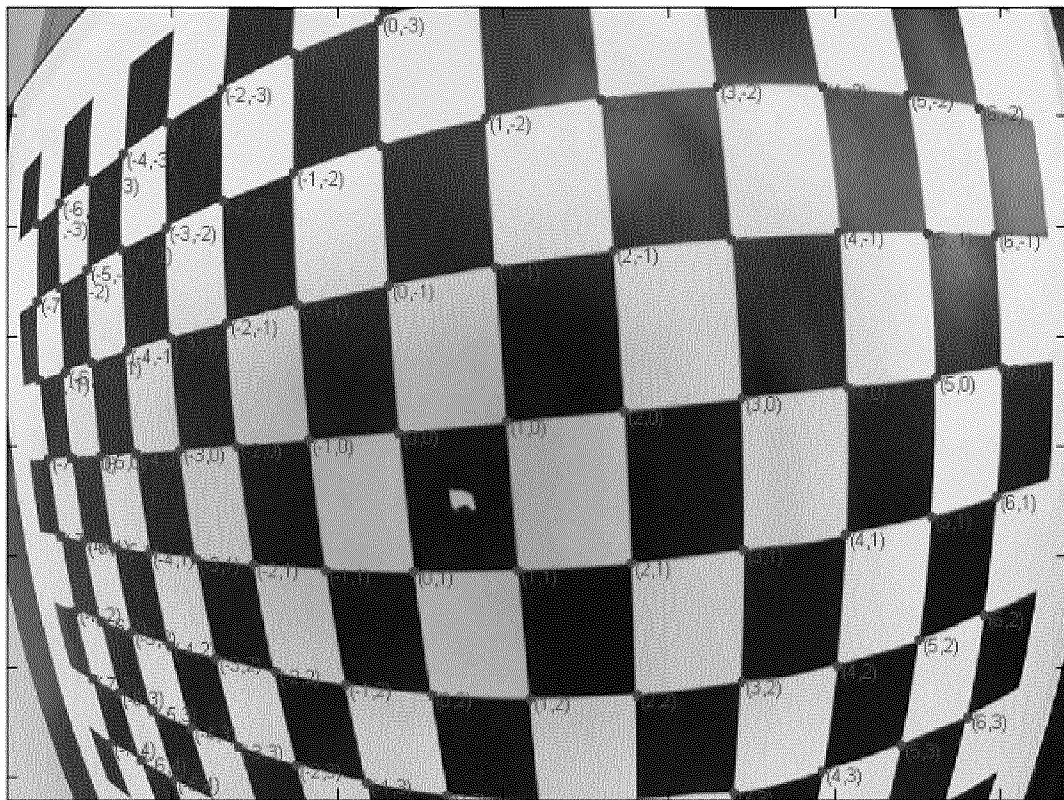
**Fig. 6**

**Fig. 7**

**Fig. 8****Fig. 13**

**Fig. 9****Fig. 10**

**11(a)****11(b)****Fig. 11**

**12(a)****12(b)****Fig. 12**

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2015/065031

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06K9/46 G06T7/00
ADD.

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 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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JIANHUI CHEN ET AL: "Calibration for high-definition camera rigs with marker chessboard", COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS (CVPRW), 2012 IEEE COMPUTER SOCIETY CONFERENCE ON, IEEE, 16 June 2012 (2012-06-16), pages 29-36, XP032201688, DOI: 10.1109/CVPRW.2012.6238905 ISBN: 978-1-4673-1611-8	1-3,5-7, 9,11-15
Y	Section 4.2, Figure 3 ----- -/--	4,8,10



Further documents are listed in the continuation of Box C.



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" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

9 September 2015

Date of mailing of the international search report

17/09/2015

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

Boltz, Sylvain

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2015/065031

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	DANIEL HERRERA C ET AL: "Joint Depth and Color Camera Calibration with Distortion Correction", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE COMPUTER SOCIETY, USA, vol. 34, no. 10, 1 October 2012 (2012-10-01), pages 2058-2064, XP011490734, ISSN: 0162-8828, DOI: 10.1109/TPAMI.2012.125 Section 4, second paragraph, Figure 1.; page 2062	4,10
Y	----- LORENZO SORGI: "Edgelet tracking using Gauss-Laguerre Circular Harmonic filters", IMAGE PROCESSING (ICIP), 2011 18TH IEEE INTERNATIONAL CONFERENCE ON, IEEE, 11 September 2011 (2011-09-11), pages 2897-2900, XP032080270, DOI: 10.1109/ICIP.2011.6116264 ISBN: 978-1-4577-1304-0 Section 1 Section 2 Section 4	8
Y	----- YOONSU PARK ET AL: "Calibration between Color Camera and 3D LIDAR Instruments with a Polygonal Planar Board", SENSORS, vol. 14, no. 3, 17 March 2014 (2014-03-17) , pages 5333-5353, XP055209758, ISSN: 1424-8220, DOI: 10.3390/s140305333 page 5336, paragraph 1; figure 17	4,10
Y	----- Davide Scaramuzza: "OCamCalib: Omnidirectional Camera Calibration Toolbox for Matlab - Davide Scaramuzza", 4 July 2013 (2013-07-04), XP055209763, Retrieved from the Internet: URL:https://web.archive.org/web/20130704095948/https://sites.google.com/site/scarabotix/ocamcalib-toolbox [retrieved on 2015-08-26] page 4 ----- -/--	4,10

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2015/065031

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
T	RUAN MINGHAO ET AL: "Calibration of 3D Sensors Using a Spherical Target", 2014 2ND INTERNATIONAL CONFERENCE ON 3D VISION, IEEE, vol. 1, 8 December 2014 (2014-12-08), pages 187-193, XP032733114, DOI: 10.1109/3DV.2014.100 [retrieved on 2015-02-06] page 187, left-hand column, paragraph first -----	