Conf. on Electronics, Telecommunications and Computers, CETC 2013
Lisboa, Portugal

DISCRETE CAMERA AUTOCALIBRATION CONSISTENT WITH THE FRAME OF THE ROBOTIC PAN-TILT BASIS

R. Galego[1], R. Ferreira[1], A. Bernardino[1], E. Grossmann[2], J. Gaspar[1]
[1]*ISR, Instituto Superior Tecnico, Lisboa, Portugal*
[2]*Intel Corp., Menlo Park, USA*
{*rgalego,ricardo,alex,jag*}*@isr.ist.utl.pt, etienne@egdn.net*

Keywords: Sensor topology, Multidimensional scaling (MDS), Isomap.

Abstract: In this paper we approach the problem of calibrating topologically a discrete camera mounted on a robotic pan-tilt basis. In particular this work is focused in choosing the coordinate system of the camera consistently with the basis. The topological calibrating methodology of the sensor is based on multidimensional scaling (MDS). The choice of the coordinate system is based in performing known pan and tilt movements with the robotic basis. These movements allow estimating the arbitrary unitary transformation introduced by MDS like algorithms, and therefore compensate image rotation and vertical or horizontal mirroring. Given a consistent camera coordinate system and the odometry of the robotic basis during the capture of sequence of images, allows building image mosaics with discrete cameras.

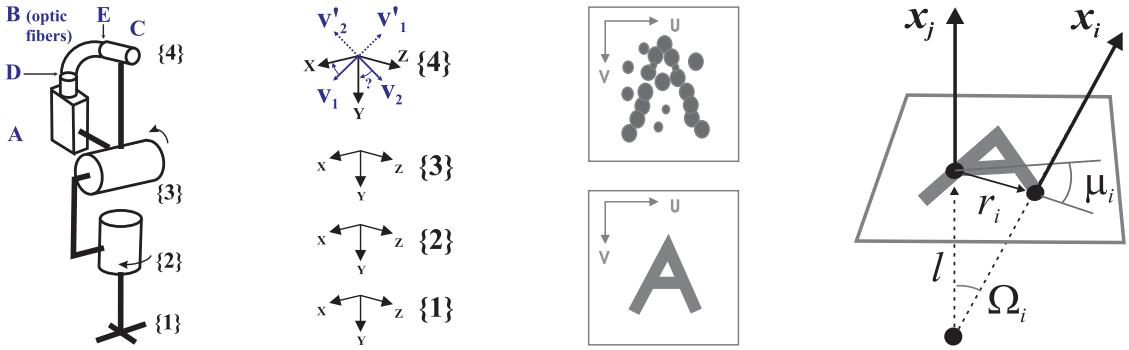
1 Introduction

Traditional imaging sensors are formed by pixels precisely placed in a rectangular grid, and thus look like calibrated sensors for many practical purposes such as localizing image edges or corners. In contrast, the most common imaging sensors found in nature are the compound eyes, collections of individual photo cells which clearly do not form rectangular grids, but are still interesting to mimic as they are very effective in solving various tasks at hand.

In the cases where the sensor topology is not a rectangular grid, one can not use traditional calibration methodologies [1, 12, 10, 7]. Despite not forming in general regular grids, recent works show that compound imaging systems, i.e. discrete cameras, can be autocalibrated. In the seminal work [8] Pierce and Kuipers have shown that is possible to reconstruct the topology of a group of sensors just by knowing their output. Grossmann *et al.* [5] shown that the function relating signal-correlation and distance-angles of the photosensors of a camera can help calibrating another camera. Recently Galego *et al.* [6] proposed a methodology to autocalibrate a central imaging sensor which is effective even without a priori calibration information.

In this work we want to do auto-calibration of central sensors with a number of pixels orders of magnitude larger than [5]. In addition, we assume that the sensors are mounted on robotic pan-tilt basis, and therefore want the camera to have a reference frame consistent with (linked to) the reference frame of the basis. We approach the computational complexity with Multi Dimensional Scaling (MDS) like algorithms [2, 11, 9]. Classical MDS [2] allows finding a representation of a data set on a given dimensionality from the knowledge of all interpoint Euclidean distances. We approach the problem of matching reference frames by performing pan-tilt movements with the robotic basis and observing the flow in the topologically calibrated camera. The flow vectors allow computing a unitary transformation which matches the camera and the basis frames.

The structure of the paper is the following: in Sec.2 we describe camera model; in Sec.3 we describe MDS/Isomap applied to topological calibration; in Sec.4 is proposed a rotation and/or mirroring correction methodology for the coordinate frame of the imaging sensor; in Sec.5 summarizes the complete calibration methodology; in Sec.6 we show some experimental results, and finally in Sec.7 we draw some conclusions and propose some future work.



(a) Camera and basis (b) Coordinate frames (c) Images at D and E (d) Projection model

Figure 1: Model of a discrete camera mounted on a pan-tilt basis. The optic-fiber bundle, points E to D in (a), twists the input image (d). Vectors v_1 and v_2 allow computing a unity transform to obtain $\{4\}$ (b). Projection model and raxels notation (d).

2 Discrete Camera Model

Discrete cameras, as conventional (standard) cameras, are described geometrically by the pin-hole projection model. Differently from standard cameras, discrete cameras are simply composed of collections of pixels organized as pencils of lines with unknown topologies.

Figure 1 shows a model of a discrete camera. The discrete camera is composed by one conventional CCD camera (see Fig. 1(a) label *A*), one cable of optic fibers mounted in front of the camera (*B*), and one extra lens mounted in front of the fibers (*C*). The fibers of the optic fiber cable are randomly mixed inside the cable, meaning that a fiber has a position (u_e, v_e) at one end (*E*) and at the other end (*D*) has a different position (u_d, v_d), i.e. $(u_e, v_e) \neq (u_d, v_d)$.

Grossberg and Nayar[4] have introduced the concept of raxel to allow representing more general cameras. A raxel is in simple terms an abstraction of the position of a light (punctual) sensor. Instead of representing the real position of the light (punctual) sensor, a raxel is just representing the direction of the chief ray associated to the sensor. A raxel is characterized as a 3*D* position and a direction vector.

In this work we consider central discrete cameras. These cameras are represented as collections of raxels. Since the cameras are central, the 3*D* position associated to each raxel can be the same for all raxels. Raxels can therefore be represented as vectors on the unit sphere $\mathbf{x_i} \in \mathbb{S}^2$. Each raxel can also be characterized by two values, (Ω_i, μ_i). Ω_i is the angle between the principal axis and the incoming ray (see Fig. 1(d)). The azimuthal angle μ_i characterizes both the angular location of an imaged point and the plane containing a raxel.

In uncalibrated discrete cameras is not possible to define corner points or image lines since the topology is unknown. The only information available for camera calibration are a set of pixel streams, $\{f_i\}$, where each pixel-stream f_i corresponds to a time series of brightness values captured by i^{th} photocell (pixel). Galego *et al.* [6] have shown that the normalized cross correlation of two pixels streams, $C(f_i, f_j)$, is affine to the angular distance of pixels, $d(\mathbf{x}_i, \mathbf{x}_j)$, when observing a circular object. Having all-to-all raxel angular distances, one can use MDS to estimate the full topology. In calibration methodology can therefore be summarized as (i) receive a set of pixels streams, (ii) compute the cross correlation between all pixels streams, (iii) convert all correlation values into angular distances, and finally (iv) estimate the topology using the distances previously computed. See Fig. 2.

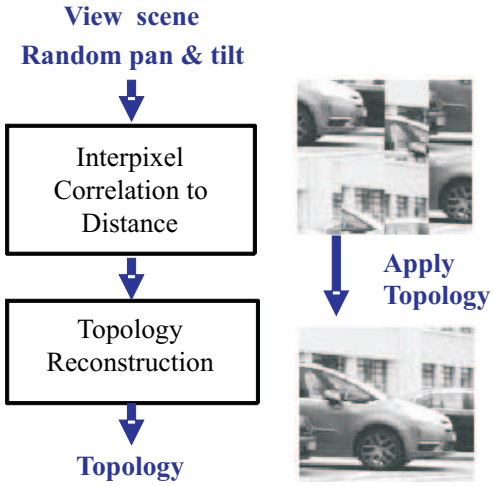


Figure 2: Overview of the calibration process

3 Topology Reconstruction

The topology reconstruction block of Fig. 2 transforms distances into a topology. This is done using Multidimensional Scaling (MDS) or a MDS derived algorithm.

The classical Multidimensional Scaling algorithm [2] provides a simple way of embedding a set of points in Euclidean space given their inter-distances. This is done by collecting all squared distances in a matrix $D^2 = [d^2(\mathbf{x}_i, \mathbf{x}_j)]$. D can be transformed into a matrix of inner products by inverting the previous linear relation and forcing the resulting embedding to have zero mean [3]. More precisely, the matrix of inner products G is obtained from D^2 through the transformation $G = -JD^2J/2$, where $J = (I - 1/n)$, I is the $n \times n$ identity matrix and n is the number of the elements of the data. Next, one observes that if the desired point embedding is collected in a matrix $X = [\mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_n]$, then

$$G = \begin{bmatrix} \langle \mathbf{x}_1, \mathbf{x}_1 \rangle & \dots & \langle \mathbf{x}_1, \mathbf{x}_n \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{x}_n, \mathbf{x}_1 \rangle & \dots & \langle \mathbf{x}_n, \mathbf{x}_n \rangle \end{bmatrix} = X^T X. \qquad (1)$$

X can thus be obtained (reconstructed) up to a unitary transformation using an SVD decomposition, as $G = X^T X = U\Sigma U^T = U\sqrt{\Sigma}\sqrt{\Sigma}U^T$, and thus $X^T = U\sqrt{\Sigma}$.

The classical MDS works well when the distances are Euclidean and when the structures are linear, however, when the manifolds are nonlinear, the classical MDS fails to detect the true dimensionality of the data set. Isomap is built on classical MDS but instead of using Euclidean distances it uses an approximation of geodesic distances [11]. These geodesic distance approximations are defined as a series of hops between neighboring points in the Euclidean space using a shortest path graph algorithm such as Dijkstra's.

4 Rotation and Mirror Correction

MDS, and derived methods as Isomap, provide a reconstruction of the vectors collected in X up to a unitary transformation. Since the camera is mounted on a mobile robot, we propose to fix the unitary transformation in accordance with the motion degrees of freedom of the robot.

Having reconstructed the topology of the imaging sensor allows doing 2D interpolation and therefore computing (approximated) directional derivatives and finding feature points using standard image processing techniques. Then, considering for example that a camera has experienced from t_1 to t_2 a leftwards pan motion and from t_3 to t_4 an upwards tilt motion, where t_i denote timestamps, allows computing two median optical-flows (or disparities), v_1 and v_2. The two flow vectors allow therefore setting the coordinates of a pixel location to be first horizontal, growing right, and the second to be vertical, growing down:

$$X_f = TX = [\hat{v}_1\ \hat{v}_2]^{-1}X \qquad (2)$$

where $\hat{v}_1$ and $\hat{v}_2$ denote normalization to unit length of v_1 and v_2. Note that noise prevents perfect orthogonality, i.e. $v_1^T v_2 \neq 0$, in which case we rotate both vectors in opposing directions to meet orthogonality. Having $v_1^T v_2 = 0$ with nonzero v_1 and v_2, implies $|\det(T)| = 1$, where $\det(T) = -1$ indicates a mirroring effect found in the reconstructed topology.

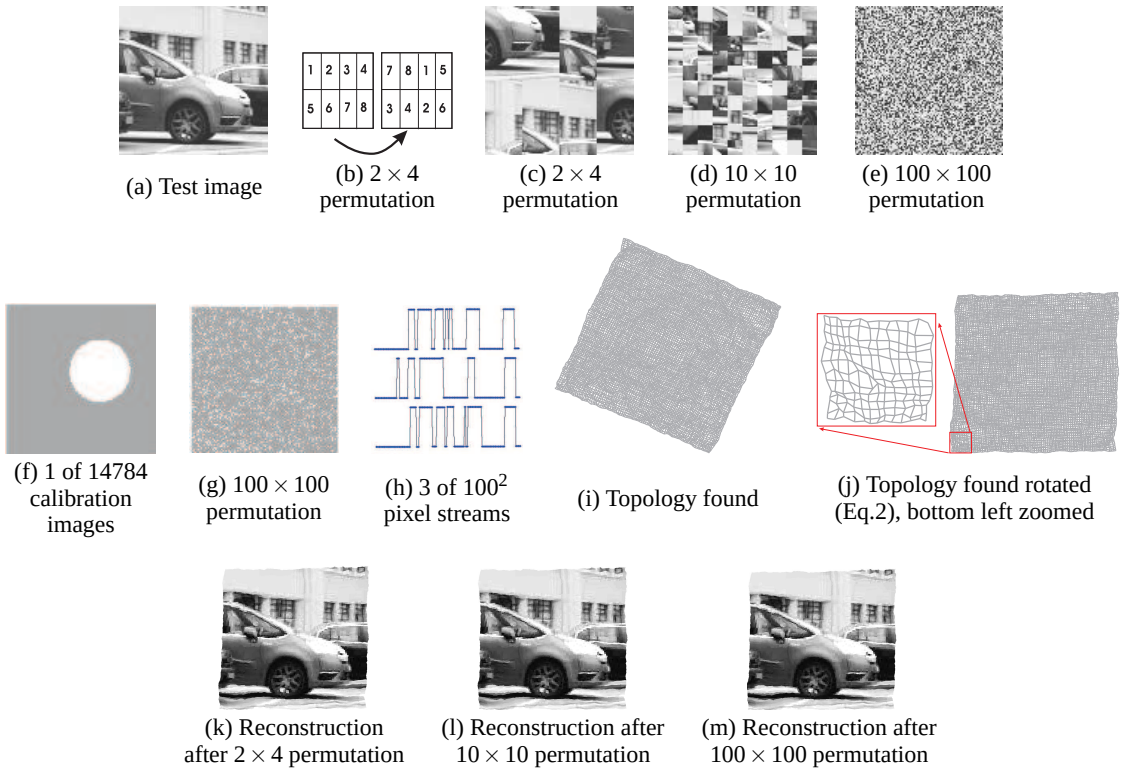
5 Complete Calibration Methodology

Summarizing the previous sections, estimating and embedding the topology of a central imaging sensor involves acquiring a set of images. Since the correlation of pixel-streams, $C(f_i, f_j)$ is invariant to shuffling of the time-series (as long as both time series are affected by the same shuffling), these images can be acquired either as a continuous sequence (i.e. a video) or as discrete individual images. In the end one wants to obtain the embedded pixel locations, $X_f = [\mathbf{x}_1\ \mathbf{x}_2\ \cdots \mathbf{x}_N]$. The required steps are the following:

1. Binarize data using a fixed threshold. Each value within the pixel stream is set to 1 or 0.
2. Compute the normalized-correlations between each pixel-stream and all the others.
3. Convert the inter-pixel correlations into distances, using the linear transformation $d(\mathbf{x}_i, \mathbf{x}_j) = 1 - C(f_i, f_j)$.
4. Use Isomap to compute the topology of the sensor.
5. Choose a coordinate system for the camera based on the supporting robot motion (Eq.2).

6 Results

In order to test the proposed topology estimation methodology four experiments have been conducted. In the first three experiments we use a standard camera so that the results can be easily compared with the ground truth. In the fourth experiment we use a prototype discrete camera based in a (twisted) cable of optic fibers.



(a) Test image (b) 2×4 permutation (c) 2×4 permutation (d) 10×10 permutation (e) 100×100 permutation

(f) 1 of 14784 calibration images (g) 100×100 permutation (h) 3 of 100^2 pixel streams (i) Topology found (j) Topology found rotated (Eq.2), bottom left zoomed

(k) Reconstruction after 2×4 permutation (l) Reconstruction after 10×10 permutation (m) Reconstruction after 100×100 permutation

Figure 3: Estimating the topology of a 100×100 sensor. Test image before permutation (a). Permutation example of 2×4 blocks (b). Permutations of pixel-streams in 2×4 blocks, 10×10 blocks, or 100×100 (all) pixels, illustrated on the test image, (c,d,e). Typical image used during calibration (f). Typical image used during calibration when pixels exist pixel-streams permutation of 100×100 (g). Example of tree pixels streams given to the algorithm (h). Output topology after the MDS algorithm (i). Output topology after rotation and mirror correction (j). Estimated topology applied to reconstruct the test image after the three permutations (k,l,m).

6.1 Standard Camera

In this first experiment, and the next two, we use a Nikon D5000 camera in video mode, selecting just a central region of 100x100 pixels. Hence, in this case we have the ground truth information that the sensor is composed by square pixels forming a regular square grid. The main purpose of the current experiment is to show that the proposed topology reconstruction algorithm correctly handles an arbitrarily sorted list of pixel-streams.

In the first experiment the camera was pointed to a bright circle on a dark background, following the suggestion of [6]. The data acquisition was performed at 24fps for about ten minutes (14784 frames), while panning and tilting the camera. In this case no roll motion was performed. An example of the images acquired and used to obtain a calibration can be seen in Fig. 3(f). Figure 3(h) shows three pixels streams, one can note that the top and the bottom streams are highly correlated thus geometrically close to each other. Figures 3(i,j) show the estimated topology, approximately forming a regular square grid, close to the ground truth. A test image, acquired in daylight (Fig. 3(a)), was then used to illustrate more clearly that the sequencing of the pixel-streams (see pixel permutations in Figs. 3(b,c,d,e)) does not influence the perceptual quality of the estimated topology and image reconstruction (Figs. 3(k,l,m)). A calibration image changed by a 100×100 permutation is shown in Figs. 3(g). Note that the permuted calibration images (g) have the same information as (f). Figure 3(j) shows that our method found the correct rotation of the topology.

Despite having obtained the results in Figs. 3(k,l,m) after different calibrations, and thus subject to different random selections of landmark pixels, the differences of the estimated topologies are small. Using a 2D Procrustes, to register the three reconstructed topologies with a square 1-pixel-steps grid, resulted in inter-pixel (four nearest neighbors) distance-error distribution with a standard deviation

of 0.566, 0.563 and 0.563 pixels, respectively.

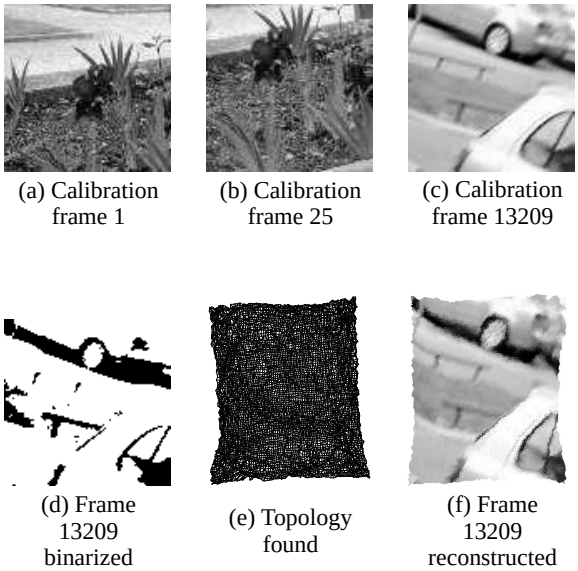
6.2 Mirror Effect Correction

(a) (b) (c)

Figure 4: Mirror effect example. Test image (a). Topology corrected with detecting the mirror effect (b). Image with mirror effect corrected (c).

In the second experiment we gave to our algorithm a mirrored topology, Figure 4 (b). If we had no movement information it would be impossible to know which of the topologies shown was the correct one, however with motion information we can determine that the correct topology is in fact an mirror of Fig. 4 (b). The correct topology can be seen in Fig. 4(c).

6.3 Unconstrained Calibration Scenario



(a) Calibration frame 1 (b) Calibration frame 25 (c) Calibration frame 13209

(d) Frame 13209 binarized (e) Topology found (f) Frame 13209 reconstructed

Figure 5: Topology estimation applied to 17448 images random images of the University campus

In the third experiment, the main purpose is to explore more general calibration scenarios, while keeping the topology estimation and the reference frame accurate. Hence we considered a garden and car park scenario, Fig. 5(a,b,c), where the vegetation provides many edge directions. For this data set we filmed about twelve minutes, at 24fps, having acquired 17448 frames. In this case we do pan and tilt motions, as well as roll and translation.

Figure 5(d) shows the a binary image used in the calibration algorithm. Figure 5(e) shows the topology reconstruction considering binary level pixel-streams.

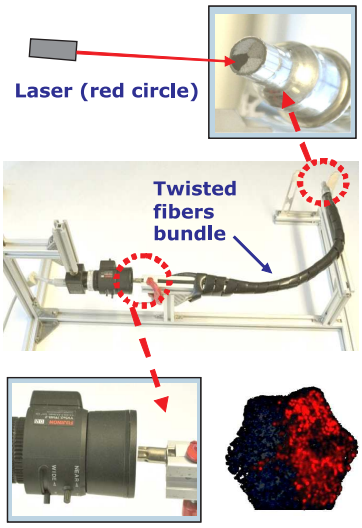
Note that Fig. 5(e) is the direct result of the Isomap, while Fig. 5(f) is the result after using Eq.2, and thus show a detected and corrected mirror effect. Fig. 5(a,b) are also example of the motion used, in this case vertical movement, to determine the rotation. Note that since this calibration was not done using the ideal scenario, according to [6], the topology found was indeed not has good as in the previews experiments, however our rotation and mirror algorithm work perfectly.

6.4 Fiber Bundle Based Camera

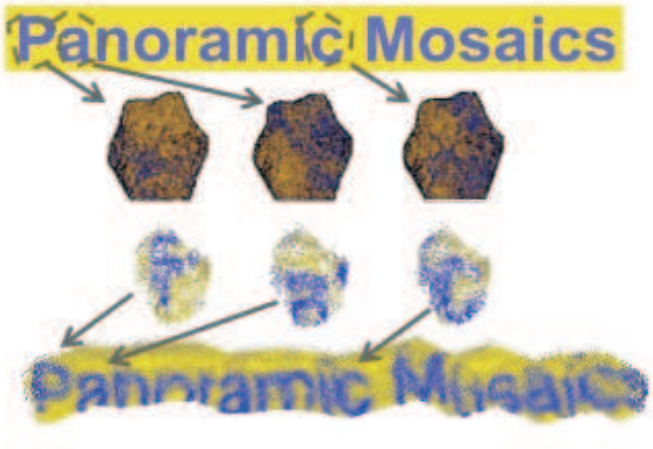
The final experiment was conducted with a fiber bundle of 2475 fibers. At one end of the fiber bundle we show several images and record the data at the other end of the fiber Fig.6 (a). As one can see some of the typical images of the fiber bundle in Fig.6 (b) (second row), the twisting of the fibers makes the images unreadable. When using the complete calibration methodology, with rotation and mirror correction, one can see easily what the camera is imaging, in this case several letters, Fig.6 (b) (third row). In order to assess qualitatively the accuracy of the topology and the rotation found, we show a mosaic in figure 6 (b) (bottom). The mosaic was made from a set of 258 images, given the motion of the camera. In the original images show to the camera one figure 6 (b) (top), one can see that is written ”Panoramic Mosaics”. The final result, figure 6 (b) (bottom), shows that our calibration methodology works with irregular topologies by building a mosaic. One can see the overall mosaic is clear, meaning that we had a good rotation correction.

7 Conclusions

In this paper we consider cameras as collection of raxels, each raxel represents a photosensor that acquires a pixel-stream. We presented a topological calibration of the raxels consistent with the global coordinates of a mobile platform. Knowing that a calibration algorithm that uses MDS or its MDS based



(a) Discrete camera.



(b) Mosaic.

Figure 6: Discrete camera based in an optic fibers bundle (2k fibers). Calibration data, image of a laser pointer, i.e. a red circular footprint distorted due to the twisting of the fibers (a). Topology correction and mosaicing of 258 images given the camera odometry (b).

as an inherent problem, unknown rotation and possible mirror effect. We proposed to solve this problem, assuming that the sensor is on top of a mobile robot. Knowing the motion of the robot one can correct a rotation and mirror ambiguity in the topology found. The correction is done by analyzing the optical flow of the sensor. We show that this method works in conventional rectangular shaped sensors. It is also shown that our method can work in irregular topologies, as it was shown in the fiber bundle case.

Acknowledgments

This work was supported by the FCT project PEst-OE/EEI/LA0009/2013.

REFERENCES

[1] L. Agapito, R. Hartley, and E. Hayman. Linear calibration of a rotating and zooming camera. In *CVPR*, pages 15–21, June 1999.

[2] T.F. Cox and M.A.A. Cox. *Multidimensional scaling*. Chapman & Hall/CRC, 2001.

[3] Jon Dattorro. *Convex Optimization and Euclidean Distance Geometry*. Meboo Publishing USA, 2010.

[4] Michael D. Grossberg and Shree K. Nayar. The raxel imaging model and ray-based calibration, 2005.

[5] Etienne Grossmann, Jose Gaspar, and Francesco Orabona. Discrete camera calibration from pixel streams. *Computer Vision and Image Understanding*, 114(2):198–209, Feb. 2010.

[6] R. Galego R. Ferreira A. Bernardino E. Grossmann and J. Gaspar. Topological auto-calibration of central imaging sensors. In *IbPRIA*, Jun. 2013.

[7] Juho Kannala and S.S. Brandt. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(8):1335–1340, 2006.

[8] David Pierce and Benjamin J. Kuipers. Map learning with uninterpreted sensors and effectors. *Artif. Intell.*, 92(1-2):169–227, May 1997.

[9] Vin De Silva and Joshua B. Tenenbaum. Global versus local methods in nonlinear dimensionality reduction. In *Advances in Neural Information Processing Systems 15*, pages 705–712. MIT Press, 2003.

[10] S. N. Sinha and M. Pollefeys. Towards calibrating a pan-tilt-zoom camera network. In *CVIU*, pages 91–110, 2006.

[11] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290, 2000.

[12] Zhengyou Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *in ICCV*, pages 666–673, 1999.