

Topological Auto-Calibration of Central Imaging Sensors

R. Galego¹, R. Ferreira¹, A. Bernardino¹, E. Grossmann², and J. Gaspar¹

¹ Institute for Systems and Robotics, IST/UTL, Portugal
{rgalego,ricardo,alex,jag}@isr.ist.utl.pt,
² Intel Corp., Menlo Park, USA
etienne@egdn.net,

Abstract. This paper departs from traditional calibration in the sense that the pixels forming the camera have a completely unknown topology. Previous works have shown that the statistical properties of natural scenes, and a uniform motion of a camera both in translation and rotation, allow determining the topology of a central camera [10, 6]. Here we show that there is a quasi-linear relationship between time-correlation and angular inter-pixel distance, considering small angles and a simple scenario encompassing one bright light on a dark background. The topology reconstruction algorithm is therefore based on correlating time series (pixel streams) acquired by the pixels of the moving camera. Correlations are converted to inter-pixel distances using a fixed linear transformation. Distances are finally embedded on a plane using a manifold learning methodology, namely Isomap. Experiments on real datasets have been conducted and have shown that the theoretical derivations are accurate for the considered scenario.

1 Introduction

Traditional imaging sensors are formed by pixels placed on a rectangular grid, and thus look like calibrated sensors for many practical purposes such as localizing local extrema, 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 very effective for solving various tasks at hand and thus have inspired the design of many artificial systems. Völkel et al. studied several types of eyes and discussed the miniaturization of imaging systems [15]. Neumann et al. [9] proposed a compound eye vision sensor for 3D ego motion computation. Recently, Micro-Electro-Mechanical Systems fabrication technologies were applied to build artificial compound eyes on planar surfaces [5].

In other words, novel fabrication technologies allow creating sensors with pixel arrangements (topologies) tuned for the tasks at hand. In the cases where the sensor topology is not a rectangular grid using traditional calibration methodologies [1, 16, 13] will not be possible. Hence, the question arising here is: how to calibrate sensors with unknown topologies? In the case that the sensors are

mounted on mobile robots the question can be restated as: can we calibrate an unknown topology of a moving sensor just with the data acquired by the sensor?

Pierce and Kuipers have shown that it is possible to reconstruct the topology of a group of sensors just by knowing their output [12]. They use natural properties of an agent’s world in order to infer the structure of its sensors. Olsson et al. improved the methodologies introduced by Pierce and Kuipers by adding information distances and, in particular, Hamming metrics [10, 11]. They compute the position of several sensors of a Sony Aibo robot, which has, among other sensors, one camera sub-sampled to 8×8 pixels. Hyvarinen et al. showed that imaging natural scenes allow defining a *neuronal topography* using Independent Subspace Analysis [7]. In [6], Grossmann et al. proposed a method for calibrating a central imaging sensor based on a number of photocells. They need to know (estimate) a priori a function curve relating correlation (or information-distance) and distance-angles. Their algorithm has been tested on a small set of pixels (photocells), about one hundred, as otherwise the computation time and memory would be too large. Recently, Censi and Scaramuzza [2] showed that it is possible to dispense with the prior estimation of the correlation-to-angle curve. Their method estimates this curve and calibrates the sensor simultaneously.

In this work we want to do auto-calibration of central sensors with a number of pixels orders of magnitude larger than [11, 6] and 50% larger than [2]. We approach the computational complexity with Multi Dimensional Scaling (MDS) like algorithms. A relatively old but very effective in the presence of noise free data is the Classical MDS [3], based on Euclidean distances. Its goal is to find a representation of a data set on a given dimensionality from the knowledge of all interpoint distances. Several new algorithms evolved from MDS, such as Isomap [14], where geodesic distances induced by a neighborhood graph are used instead of Euclidean distances.

The structure of the paper is the following: in Sec. 2 we describe Isomap applied to topological calibration; in Sec. 3 we study a simple black and white scenario and show there is a linear relationship between the correlation of the time series acquired by pairs of pixels and the inter-pixel angle; in Sec. 4 we show some experimental results, and finally in Sec. 5 we draw some conclusions.

2 Auto-Calibration Methodology

A discrete camera is defined as a collection of pixels, corresponding to a central (perspective) projection model, organized in a general topology. In this work we assume point pixels, meaning that the integration area of each pixel collapses to a point. Without loss of generality we denote pixel orientations as a vector on the unit sphere $\mathbf{x}_i \in \mathbb{S}^2$. The collection of pixel orientations defines the camera topology, which is unknown and is to be found.

The classical Multiple Dimensional Scaling (MDS) algorithm [3] provides a simple way of embedding a set of points in an Euclidean space given the interpoint distances. Noting that the squared distance function between two points

is a linear transformation of the inner product of the points

$$d^2(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i - \mathbf{x}_j, \mathbf{x}_i - \mathbf{x}_j \rangle = \langle \mathbf{x}_i, \mathbf{x}_i \rangle - 2\langle \mathbf{x}_i, \mathbf{x}_j \rangle + \langle \mathbf{x}_j, \mathbf{x}_j \rangle \quad (1)$$

allows one to write all distances as inner products [4]. More precisely, collecting all squared distances in a matrix, $D^2 = [d^2(\mathbf{x}_i, \mathbf{x}_j)]$ a 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 $X = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_n]$, then $G = X^T X$. Therefore X can 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 [14]. 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.

In our particular case, this algorithm is used to provide a pixel embedding given the inter-pixel distances estimated from the pixel stream correlations as detailed in the next section.

3 Distance Between Pixels

Prior works have experimentally shown a relationship between the normalized cross correlation of pixel streams and the angular inter-pixel distance. This section further explores this coupling by providing analytical results for a particular observed static scenario and camera motion.

Each camera pixel observes the environment at every time instance producing a scalar value corresponding to the observed brightness. A pixel-stream f_i is the time-series obtained from the observations over time by the i th camera pixel. When two of these streams f_i and f_j are considered, one can define the normalized cross correlation as

$$C(f_i, f_j) = \frac{\mathbb{E}[f_i f_j] - \mathbb{E}[f_i] \mathbb{E}[f_j]}{\sqrt{(\mathbb{E}[f_i^2] - \mathbb{E}^2[f_i]) (\mathbb{E}[f_j^2] - \mathbb{E}^2[f_j])}} \quad (2)$$

Grossmann et al. [6] proposed that the inter-pixel distance between two pixels \mathbf{x}_i and \mathbf{x}_j is related to the normalized cross correlation of their pixel streams $C(f_i, f_j)$. Although the existence of this relation is proposed, no specific transformation is provided and in [6] the authors propose to learn it as a look up table from observations of the environment.

Black and White Scene - As referred, to better understand the relationship between correlation and angular distance a particular simple environment is proposed: (i) the camera is at the center of a unit sphere, which is black everywhere except for a hard edged white circular hub-cap with angular radius ρ , (ii) the camera motion covers uniformly the rotation configuration space $\mathbb{S}\mathbb{O}(3)$ and acquires an infinite time series with each of the pixels.

Equation (2) states that the theoretical correlation value can be computed from $\mathbb{E}[f_i]$, $\mathbb{E}[f_j]$, $\mathbb{E}[f_i^2]$, $\mathbb{E}[f_j^2]$ and $\mathbb{E}[f_i f_j]$. Since the camera motion is uniform, there is no difference in the expected values of the isolated pixel streams, i.e. $\mathbb{E}[f_i] = \mathbb{E}[f_j]$ and $\mathbb{E}[f_i^2] = \mathbb{E}[f_j^2]$. Without loss of generality it is assumed that each pixel reports a value of 1 when it is pointed towards the white circular cap and a value of -1 when pointed towards the black area (note that the definition of correlation is invariant to the choice of these values). This particular choice leads to $\mathbb{E}[f_i^2] = \mathbb{E}[f_j^2] = 1$, and only $\mathbb{E}[f_i]$ and $\mathbb{E}[f_i f_j]$ are left to be computed.

Expected value of a pixel stream, $\mathbb{E}[f_i]$ - Since the camera pose distribution is uniform, the probability p of a point pixel observing the white hub-cap, corresponds to the surface area that the hub-cap occupies [8], $A_d = 2\pi(1 - \cos(\rho))$, divided by the total surface area of the observed unit sphere, $A_s = 4\pi$, i.e. $p = (1 - \cos(\rho))/2$. The expected value for the pixel stream is thus

$$\mathbb{E}[f] = 1 \cdot p + (-1)(1 - p) = -\cos(\rho). \quad (3)$$

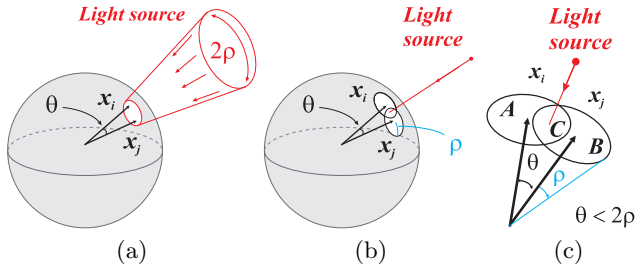


Fig. 1. Illumination and the information acquired by pixels x_i and x_j . A light source is seen as an illuminating hubcap (a). Instead of infinitesimal pixels, one can consider enlarged pixels and an infinitesimal illumination (b). The intersection of the pixel areas $C = A \cap B \neq \emptyset$ if $\theta < 2\rho$ (c).

Expected value of the product of two pixel-streams, $\mathbb{E}[f_i f_j]$ - To compute $\mathbb{E}[f_i f_j]$ first notice that the pointwise result of $f_i f_j$ is either 1, when both pixels observe the same color, or -1 when one pixel observes the interior of the circular hubcap and the other does not.

Figure 1 shows two pixels x_i and x_j separated by an angular distance θ . A pixel observes white whenever the center of the white hubcap falls within a

distance ρ of the pixel (within circular region A in the figure or within circular region B, depending on the considered pixel). Thus the probability q of exactly one pixel observing the white circle is given by:

$$q = (\text{Area}(A) + \text{Area}(B) - 2 \text{Area}(C)) / \text{Area}(\text{Sphere}) \quad (4)$$

where the intersection area C is removed since in this area both pixels measure white. Here, $\text{Area}(A)$ and $\text{Area}(B)$ are equal to the previously computed A_d . The intersection area of two hub-caps, $\text{Area}(C)$, has not been found in the literature and was obtained as a surface integral (due to space constraints the details are not given here). The result is zero when $\theta \geq 2\rho$ and, if $\theta < 2\rho$, then

$$\text{Area}(C) = 2\pi (1 - \cos(\rho)) - 4 \operatorname{asin}\left(\frac{\sin(\theta/2)}{\sin(\rho)}\right) + 4 \cos(\rho) \operatorname{asin}\left(\frac{\tan(\theta/2)}{\tan(\rho)}\right) \quad (5)$$

Combining (5) with (4) and substituting into $\mathbb{E}[f_i f_j] = 1(1 - q) + (-1)q$ yields

$$\mathbb{E}[f_i f_j] = \begin{cases} 1 - \frac{4}{\pi} \operatorname{asin}\left(\frac{\sin(\theta/2)}{\sin(\rho)}\right) + \frac{4}{\pi} \cos(\rho) \operatorname{asin}\left(\frac{\tan(\theta/2)}{\tan(\rho)}\right) & \text{if } \theta < 2\rho \\ 2 \cos(\rho) - 1 & \text{if } \theta \geq 2\rho \end{cases} \quad (6)$$

Computing the correlation - Now that all the values are available, $C(f_i, f_j)$ can be computed as in expression (2):

$$C(f_i, f_j) = \begin{cases} \frac{-\frac{4}{\pi} \operatorname{asin}\left(\frac{\sin(\theta/2)}{\sin(\rho)}\right) - \frac{4}{\pi} \cos(\rho) \operatorname{asin}\left(-\frac{\tan(\theta/2)}{\tan(\rho)}\right) + \sin^2(\rho)}{\sin^2(\rho)} & \text{if } \theta < 2\rho \\ 1 - \frac{2(\cos(\rho) - 1)}{\sin^2(\rho)} & \text{if } \theta \geq 2\rho \end{cases} \quad (7)$$

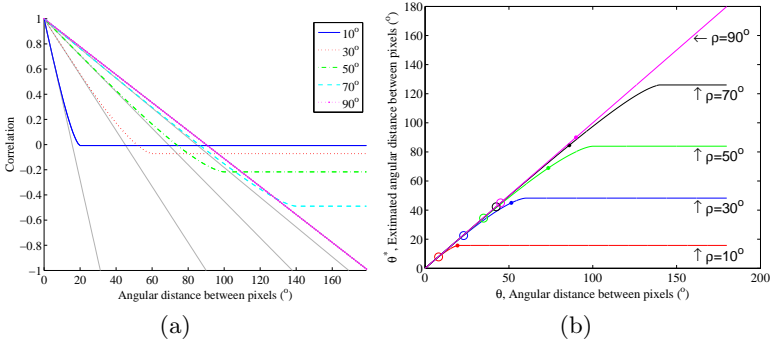


Fig. 2. Correlation v.s. angular distance between pixels for several observed circle sizes. Theoretical values computed from equation (7) with the tangents at the origin to each curve overlaid in light gray (a). Approximation error of using the second order Taylor approximation, where for each curve empty circles correspond to correlation equal to 0.5 and full dots correspond to zero correlation (b).

The plot of $C(f_i, f_j)$, shown in figure 2, reveals an almost perfect affine function for small angular values.

Property 1 *The correlation based distance $d(\mathbf{x}_i, \mathbf{x}_j) = 1 - C(f_i, f_j)$ is (i) a linear function of the inter-pixel angle $\theta = \widehat{\mathbf{x}_i \mathbf{x}_j}$ when the observer has uniform random motion in all rotation d.o.f. and the spherical scene has one hemisphere light and the other dark, i.e. $\rho = \pi/2$. Otherwise, (ii) the distance $d(\mathbf{x}_i, \mathbf{x}_j)$ has a linear second order Taylor series expansion around $\theta = 0$ when $\rho \neq \pi/2$ (i.e. zero second order term).*

Proof. Replacing $\rho = \pi/2$ in eq.(7) one obtains $d(\mathbf{x}_i, \mathbf{x}_j) = 1 - C(f_i, f_j) = 1 - (1 - 2\theta/\pi) = 2\theta/\pi$ as stated in (i).

Demonstration of (ii) is based in the second order Taylor expansion of the correlation function (eq:(7) case $\theta < 2\rho$), $d(\mathbf{x}_i, \mathbf{x}_j) = 1 - C(f_i, f_j) \approx 2\theta/(\pi \sin(\rho)) + 0\theta^2$ which asserts that the second order Taylor expansion has a null second order term, having therefore a linear relationship with the angular distance θ . Figure 2 (b) illustrates this property is valid for values of $C > 0.5$ (open circle mark) and even when $C > 0$ (dot mark) the estimated distance is close to the real distance. Figure 2 (a) shows a function of correlation in terms of distance where the Taylor expansion for each curve is shown in gray. \square

For the particularly simple scenery described in this section it is clear that for correlation values above 0.5 the look up table needed in [6] can be replaced by a single slope parameter dependent on the size of the observed hubcap. We propose to further drop this requirement for small sensors by embedding the locations in a plane up to a scale factor. The proposed method consists of directly converting the obtained correlation values to distance values and then applying the Isomap algorithm [14] to obtain the reconstruction. Hence we will consider that the distance between two pixels whose correlation is above 0.5 is given by $d(\mathbf{x}_i, \mathbf{x}_j) = 1 - C(f_i, f_j)$.

4 Results

In order to test the proposed topology estimation methodology two experiments have been conducted using a Point-Grey Flea camera and a cable of optical fibers, Fig. 3 (a). The 2475 fibers observed by the camera are the longest of the whole cable (the remaining being too small to be used), Fig. 3 (a) red dashed line. The ground truth of the topology was estimated using 2 periodic signals one vertical and the other horizontal. One can use the phase of each fiber stream to build a good estimation of the topology.

The calibration experiments were performed observing the circular projection of a laser beam targeting the fiber bundle. A set of 7978 images were obtained with the laser illuminating different sets of fibers in each case.

In order to check the dependence of the obtained topology on the number of neighbors, Isomap was run for different values of this parameter. Figure 3 (b) shows the resulting Procrustes error between ground truth and each estimated topology. We conclude that better results are obtained when less than 10% of neighbors are used. This is due to the fact that when more neighbors are used, we

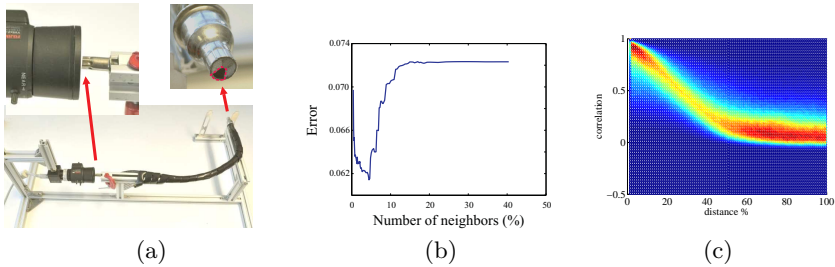


Fig. 3. (a) Experimental setup. Part of the fiber bundle used in the experiments is marked by the red dashed line; (b) Error vs the number of neighbors; (c) Histogram of the correlation vs distance.

leave the proposed correlation linear work zone ($C > 0.5$), as shown in Fig. 3 (c), and theoretically predicted in Fig.2 (a).

Figure 4 validates empirically the topology obtained for the fibers bundle using 20 closest neighbors. Figure 4(a) shows the image observed when a letter 'P' is shown to the fiber bundle. Without the topology correction the letter is unreadable. Figure 4(b) shows the image obtained with the ground truth topology. As expected, the letter 'P' is now readable. Figure 4(c) shows the results of the topology estimation using the proposed methodology. The letter 'P' is reconstructed with almost no difference relative to the ground truth.

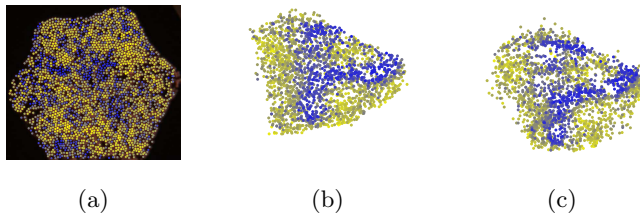


Fig. 4. Topology estimation. (a) Image acquired by the camera when the letter 'P' is shown in front of the fiber bundle; (b) Ground truth; (c) Estimated topology.

5 Conclusions and Future Work

In this paper we consider cameras as simple collections of fibers, arranged with an unknown topology. We show that the topology can be estimated with a fairly simple experimental setup. On the theoretical side we demonstrated that, in a simple scenario where the camera is in the center of a spherical-surface, half-white and half-black, the correlation of pixel-streams is linearly related with the inter-pixel angle. In the cases where the white hub-cap is smaller than a hemisphere, we show that the relationship between correlation and inter-pixel angle is

approximately linear for high correlation values. Combining the inter-pixel angle estimation, directly from correlation, with manifold learning for embedding all the inter-pixel angles, has proved to be an effective methodology for topology estimation. Promising results have been obtained with the proposed topology estimation methodology applied to a prototype camera. Future work will focus on the demonstration of linearity for larger classes of scenarios.

Acknowledgments

This work has been partially supported by the FCT project PEst-OE / EEI / LA0009 / 2011, by the FCT project PTDC / EEACRO / 105413 / 2008 DCCAL, and by the project High Definition Analytics (HDA), QREN - I&D em Co-Promoção 13750. We thank João Mendanha Dias, IST Lisbon, the help to build the microscopic input, and Ricardo Nunes, IST/ISR Lisbon, the help to build the setup based on a cable of optic fibers.

References

1. Agapito, L., Hartley, R., Hayman, E.: Linear calibration of a rotating and zooming camera. In: CVPR. pp. 15–21 (June 1999)
2. Censi, A., Scaramuzza, D.: Calibration by correlation using metric embedding from non-metric similarities. IEEE T-PAMI (to appear 2013)
3. Cox, T., Cox, M.: Multidimensional scaling. Chapman & Hall/CRC (2001)
4. Dattorro, J.: Convex Optimization and Euclidean Distance Geometry. Meboo Publishing USA (2010)
5. Di, S., Lin, H., Du, R.: An artificial compound eyes imaging system based on mems technology. In: Robotics and Biomimetics, IEEE Int. Conf. on (Dec 2009)
6. Grossmann, E., Gaspar, J., Orabona, F.: Discrete camera calibration from pixel streams. Computer Vision and Image Understanding 114(2), 198–209 (Feb 2010)
7. Hyvriinen, A., Hurri, J., Hoyer, P.O.: Natural Image Statistics: A Probabilistic Approach to Early Computational Vision. Springer, 1st edn. (2009)
8. Li, S.: Concise formulas for the area and volume of a hyperspherical cap. Asian Journal of Mathematics and Statistics 4, 66–70 (2011)
9. Neumann, J., Fermuller, C., Aloimonos, Y., Brajovic, V.: Compound eye sensor for 3d ego motion estimation. In: IROS. IEEE (April 2004)
10. Olsson, L., Nehaniv, C.L., Polani, D.: Sensory channel grouping and structure from uninterpreted sensor data. In: NASA/NoD Conf. on Evolvable Hardware (2004)
11. Olsson, L., Nehaniv, C.L., Polani, D.: Measuring informational distances between sensors and sensor integration. Tech. rep., In: Artificial Life X (2006)
12. Pierce, D., Kuipers, B.J.: Map learning with uninterpreted sensors and effectors. Artif. Intell. 92(1-2), 169–227 (May 1997)
13. Sinha, S.N., Pollefeys, M.: Towards calibrating a pan-tilt-zoom camera network. In: CVIU. pp. 91–110 (2006)
14. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science 290 (2000)
15. Völkel, R., Eisner, M., Weible, K.J.: Miniaturized imaging systems. Microelectron. Eng. 67-68(1), 461–472 (Jun 2003)
16. Zhang, Z.: Flexible camera calibration by viewing a plane from unknown orientations. In: in ICCV. pp. 666–673 (1999)