



Discrete camera calibration from pixel streams

Etienne Grossmann^{a,*}, José António Gaspar^b, Francesco Orabona^c

^a Tyzx, Inc., 660 Partridge Ave, Menlo Park, CA 94025, USA

^b ISR, Instituto Superior Técnico, Lisbon, Portugal

^c Idiap Research Institute, Martigny, Switzerland

ARTICLE INFO

Article history:

Received 11 March 2008

Accepted 6 March 2009

Available online 20 March 2009

Keywords:

Non-parametric camera calibration

Statistical signal processing

Information theory

Distance geometry

Sphere embedding

ABSTRACT

We consider the problem of estimating the relative orientation of a number of individual photocells – or pixels – that hold fixed relative positions. The photocells measure the intensity of light traveling on a pencil of lines. We assume that the light-field thus sampled is changing, e.g. as the result of motion of the sensors and use the obtained measurements to estimate the orientations of the photocells.

Our approach is based on correlation and information-theory dissimilarity measures. Experiments with real-world data show that the dissimilarity measures are strongly related to the angular separation between the photocells, and the relation can be modeled quantitatively. In particular we show that this model allows to estimate the angular separation from the dissimilarity. Although the resulting estimators are not very accurate, they maintain their performance throughout different visual environments, suggesting that the model encodes a very general property of our visual world. Finally, leveraging this method to estimate angles from signal pairs, we show how distance geometry techniques allow to recover the complete sensor geometry.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Early experiments in psychological research, such as the famous experiments with the prism and the inverted glasses [16,32,31], revealed that much of the geometry of the human's vision is known beforehand and is saved in the brain as imaging experiences. In particular, those psychological experiments showed that a person wearing distorting glasses for a few days, after a very confusing and disturbing period, could learn the necessary image correction to restart interacting effectively with the environment.

This amazing learning capability of the human's vision system clearly contrasts with current calibration methodologies of artificial vision systems, which are still strongly grounded to the a priori knowledge of the projection models. Calibration is typically bootstrapped by using some basic properties of the imaging system. In particular, in all the cases that we are aware of, the local topology is known: pixel (i,j) is a neighbor of $(i+1,j)$; the triplet $(i,j), (i+1,j), (i+2,j)$ is approximately aligned. It is possible to detect and localize edges, extrema, corners and features in uncalibrated images, just as well as on calibrated images. In short, it is undeniable that locally, and for many practical purposes, an uncalibrated image looks just like a calibrated image.

This paper departs from traditional computer vision on this point: we do not assume that pixels (i,j) and $(i+1,j)$ are neigh-

bors. In fact, we are given pixels indexed by a single index i . Since such a camera is peculiar by the absence of a priori known topological information, it is only natural to call it a *discontinuous* or *discrete camera*.

The problem that we address is to determine the geometry of a discrete camera that produced a stream of values $x(i,t)$, where the integer index i denotes the photocell or pixel, and t , the time. This problem may seem too general to pertain to computer vision. Indeed, digital cameras are unlikely to ever produce discontinuous images.¹ Yet we must also consider the computation that occur in biological systems, where it is unclear how much of the geometry of the sensor is known beforehand, and how much is determined by processing visual stimuli. We must also consider robots equipped with arrays of photocells, and new visual sensors [28] that may have variable geometry.

Our goal here is thus to determine the 3D direction pointed by each pixel, under the assumption that pixels sample light that travels along rays that intersect in a unique *center of projection* (see Fig. 1). In the impossibility of performing local image processing, traditional calibration techniques [34,12] are out of the question.

Less traditional non-parametric methods that assume a smooth image mapping and smooth motion [20,10] can obviously not be applied either. By using controlled-light stimuli or known calibration, matches could be obtained, allowing to use match-based

* Corresponding author.

E-mail address: etienne@isr.ist.utl.pt (E. Grossmann).

¹ Except in rare cases such as a camera with a stereo adapter or split mirror.

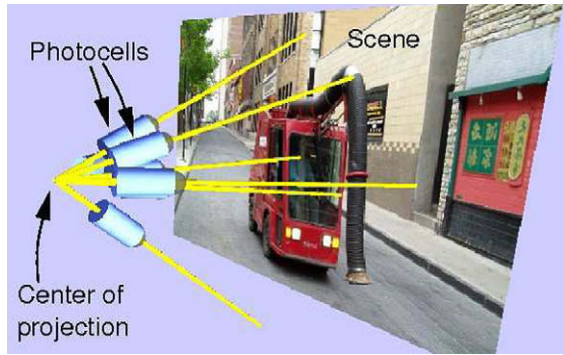


Fig. 1. A discrete camera consists of a number of photocells (pixels) that measure the light traveling along pencil of lines. More generally, one could consider that each photocell is characterized by a point spread function defined on the 3D sphere.

non-parametric techniques [25]. In the technical report [6], a non-central projection sensor samples the light-field in a discontinuous way: the response at each pixel is the convolution of the light-field with a possibly multimodal function (this point is not mentioned explicitly); this “point-spread function” is only sufficiently estimated to reconstruct images on a fixed plane, and calibration is done with controlled stimuli. In this study however, we wish to exclude known calibration objects and other controlled stimuli.

Given the problem, a mixture statistical signal processing and geometry seems the most natural approach and this is indeed how we proceed.

1.1. Related work

Our approach to calibration is closely related to the work of Pierce and Kuipers [23], who measure the dissimilarity, or distance, between sensor elements that are not necessarily light sensors. The elements are then embedded in a metric space using metric scaling [17], which also determines the dimension of the space. A relaxation method then improves this embedding, so that the Euclidean distance between sensor elements better matches the dissimilarity between the sensor inputs. Getting very close to the problem addressed in the present paper, the authors use this method to reconstitute the geometry of a rectangular array of visual sensors that scans a fronto-parallel image.

Going further, Olsson et al. [21] use the information metric of [4] as a more appropriate method to measure the distance between visual or other sensor elements. They also show how visual sensors – the pixels of the camera of a mobile robot – can be mapped to a plane, either using the method of [23], or their own, that embeds sensor elements specifically in a square grid.

The works of Olsson et al. and of Pierce and Kuipers are very interesting to computer vision researchers, but they cannot calibrate an arbitrary discrete camera, since the geometry of the embedding space is either abstract or fixed to a grid. In either case, it lacks an explicit connection to the geometry of the sensor. Filling this gap is one of the motivations for the present work.

Since the present work exploits statistical properties of the light-field of the world surrounding a light sensor, this article is also related to research on the statistical properties of real-world images [13]. In this field a theoretical framework which is gaining more and more support considers the properties of the visual system to be reflections of the statistical structure of natural images, because of evolutionary adaptation processes. Our approach is similar in spirit, trying to replicate the plasticity of the visual system of biological systems. That research has put in evidence fundamental properties, in terms of local, global and spectral statistics, of real-world images, that has been also exploited for computer vision

tasks, such as classification [33], image restoration [7] and 3D inference [24].

Although these results are of great interest, they are not directly applicable in our case, mainly because we lack images.

Moreover, these statistics are about images formed on a planar image plane, which is a hindrance in our case: first, we do not want to exclude the case of visual sensor elements that are separated by more than 180 degrees, such as the increasingly popular omnidirectional cameras. Also, the local statistical properties of perspective images depend of the orientation of the image plane with respect to the scene, except in special constrained cases such as the fronto-parallel “leaf world” of Wu et al. [35]. Defining images on the unit sphere thus appears as a natural way to render image statistics independent of the sensor orientation, at least with proper assumptions on the surrounding world and/or the motion of the sensor.

1.2. Proposed approach

In the present work, we assume that the statistical properties of the data streams produced by pairs of sensor elements depends only on the angular separation between the elements. This assumption, when one considers the images as observations of a random field defined on the sphere, is equivalent [26] to assuming that the random field is homogeneous² – we could say “isotropic” to use the common image processing vocabulary.

This assumption of homogeneity clearly does not hold in an anisotropic world, unless the orientation of the sensor is uniformly distributed amongst all unitary transformations of the sphere. We thus adopt this last assumption, in order to ensure the homogeneity of the observations. In terms of computer vision and robotics, our assumption amounts to saying that the sensor is randomly oriented, so that each photocell is just as likely to sample the light-field in any direction.

Algorithm 1. Procedure for estimating the geometry of a discrete camera

Input: Signals produced by N rigidly connected light sensors at T time instants: $x(i, t)$, $1 \leq i \leq N$, $1 \leq t \leq T$. The light sensors are rigidly connected and point towards unknown 3D directions $X_1, \dots, X_N, X_i \in \mathbb{R}^3, X^\top X = 1$.

Output: Estimates of the directions X_1, \dots, X_N .

Algorithm:

1. Estimate the distance d_{ij} between the temporal signals $x(i, \cdot)$ and $x(j, \cdot)$, for each $1 \leq i, j \leq N$.
 2. Estimate the angular separations $\theta_{ij} = \cos(X_i^\top X_j)$.
 3. Embed the angular separation in the 3D sphere: find X_1, \dots, X_N s.t. $X_i^\top X_i = 1$, $X_i^\top X_j = \cos(\theta_{ij})$, $1 \leq i, j \leq N$.
-

The great practical utility of this assumption is that, as a consequence, the statistical properties of a pair of data streams generated by two photocells depend only on the angle separating them. In this situation, one can envision estimating the angular separation from a measure of dissimilarity (e.g. correlation or information distance) between the streams, and this is precisely what we do (Section 3). In order to estimate angles from dissimilarity measures, we empirically observe the relation between these two quantities. We may then (Section 4.1) embed these angular estimates using techniques from distance geometry [5]. The whole process is outlined in Fig. 2 and Algorithm 1. Finally Section 5 pre-

² The definition of a homogeneous random field defined on the sphere is that the covariance between two random variables sampled at two locations of the sphere depends only on the angular separation between the two points and that the expectation does not depend on the sampling location.

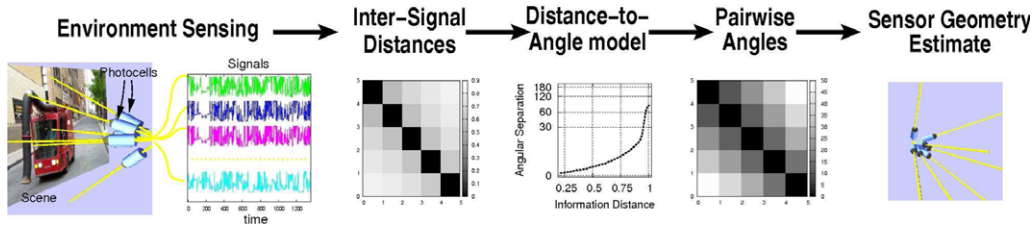


Fig. 2. The process of estimating the geometry of an unknown discrete camera.

sents some conclusions and a list of questions to be addressed by future research.

This work is the continuation of [11] and also includes some of the improvements brought by our previous work [9]. In the latter, it is shown that a metric between signals based on the correlation can just as well be used, instead of a metric based on information distance. The embedding method used in [9] is described more thoroughly in the present paper, allowing others to reproduce our results more easily.

2. Discrete camera model

Before entering into the details of our methodology for estimating the sensor geometry, we define the discrete camera and explain how to simulate it using an omnidirectional image sensor.

We define a *discrete camera* as a set of N photocells indexed by $i \in \{1, \dots, N\}$, pointing in directions $X_i \in \mathbb{R}^3$. T, on the number of pixels N , acquiring along time t , brightness measurements $x(i, t)$ in the range $\{0, \dots, 255\}$. The directions of the light rays, contrarily to conventional cameras, are not necessarily organized in a regular grid. Many examples of cameras can be found under these definitions. One example is the linear camera, where all the X_i are co-planar. Another example is the conventional perspective camera which comprises a rectangular grid of photocells that are enumerated in our model by a single index i ,

$$\left\{ X_i \mid X_i \sim K^{-1} \begin{bmatrix} i \% W \\ \lfloor i / W \rfloor \\ 1 \end{bmatrix}, 0 \leq i < HW \right\}$$

where W, H are the image width and height, K is the intrinsic parameters matrix, $\%$ represents the integer modulo operation and $\lfloor \cdot \rfloor$ is the lower-rounding operation. Cameras equipped with fisheye lenses, or having log-polar sensors, can also be modeled again by setting X_i to represent the directions of the light-rays associated to the image pixels. In the same vein, omnidirectional cameras having a single projection center, as the ones represented by the unified projection model [8], also fit in the proposed model. In this paper

we use a calibrated omnidirectional camera to simulate various discrete cameras.

2.1. Simulated image sensor

We simulate a discrete camera with known Euclidean geometry by sampling a calibrated panoramic image with unique projection center at fixed locations. Since the camera is calibrated, it is straightforward to locate the position (u, v) in the panoramic image corresponding to the 3D direction X_i of a photocell that is part of the simulated discrete camera. In the present work, we use bilinear interpolation to measure the graylevel value at non-integer coordinates (u, v) .

Images are acquired by a VStone catadioptric camera consisting of a perspective camera fitted to a hyperbolic mirror, shown in Fig. 3, left. This system is modeled as single projection center camera [8] with a $360^\circ \times 210^\circ$ field of view and a $\sim 45^\circ$ blind spot at the south pole (Fig. 3, right). The mirror occupies a 453×453 pixel region of the image. The angular separation between neighboring pixels in the panoramic image is usually slightly smaller than 0.5° . Also, some mild vignetting occurs, that could be corrected. Apart for these minor inconveniences, simulating a discrete camera by an omnidirectional camera presents many advantages: no other specialized hardware is needed and each omnidirectional image can be used to simulate many discrete camera “images”, as in Fig. 4(a). With respect to perspective cameras, the available field of view allows to study very-wide-angle discrete cameras.

2.2. Data acquisition

For the purpose of studying the effect of angular separation on joint photocell signal statistics, we use the 31-pixel planar camera (or “probe”) shown in Fig. 4(b). This probe design allows to study the effect of angular separations ranging from 0.5 to 180 degrees and each sample provides $465 = 31(31 - 1)/2$ pixel pairs. In the “tighter” part of the discrete camera layout, there exists a slight linear dependence between the values of consecutive pixels due to aliasing.

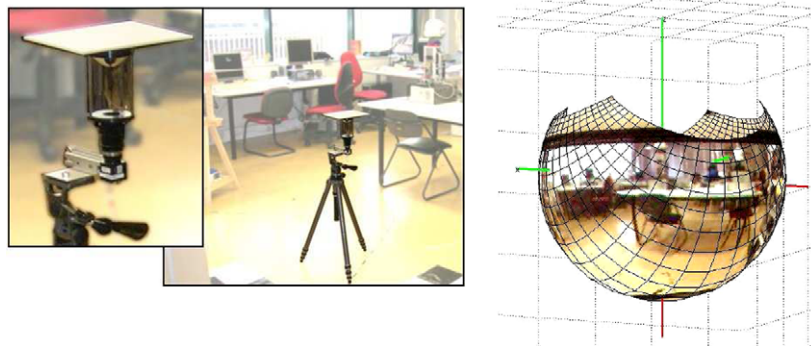


Fig. 3. Left: The camera used to sample omnidirectional images (image mirrored). Right: A calibrated omnidirectional image mapped to a sphere.

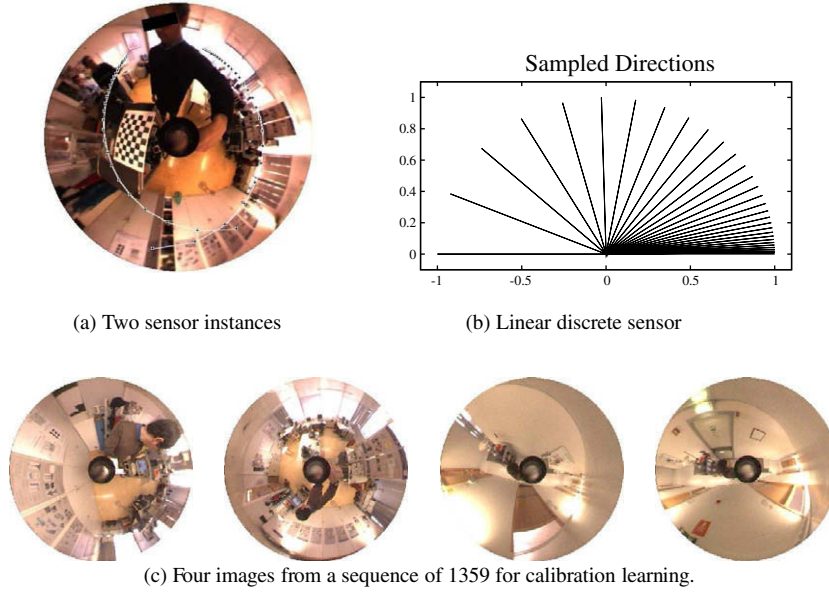


Fig. 4. (a) Two instances of the linear discrete camera, inserted in an omnidirectional image. Pixels locations are indicated by small crosses connected by white lines. (b) Geometry of a discrete camera consisting of a planar array of thirty one (31) pixels, spanning 180° in the plane. The first two pixels are separated by 0.5° , the separation between consecutive photocells increases geometrically (ratio ≈ 1.14), so that the 31st photocell is antipodal with respect to the first. (c) Four images of a set to be used for calibration learning.

The camera is hand-held and undergoes “random” general rotation and translation, according to the author’s whim, while remaining near the middle of the room, at 1.0–1.8 m from the ground. We acquired three sequences consecutively, in very similar conditions and joined them in a single sequence totaling 1359 images, i.e. approximately 5 min of video at 4.5 frames per second.

To simulate the discrete camera, we randomly choose an orientation (i.e. half a great circle) such that all pixels of the discrete camera fall in the field of view of the panoramic camera. Fig. 4(b) shows two such choices of orientations. For each choice of orientation, we produce a sequence of 31 samples $x(i, t)$, $1 \leq i \leq 31$, $1 \leq t \leq 1359$, where each $x(i, t) \in \{0, \dots, 255\}$. Choosing 100 different orientations, we obtain 100 discrete sensors and 100 arrays of data. Appending these arrays we obtain 31 signals $x(i, t)$ of length to 135,900.

Considering any pair of pixel (indices) $1 \leq i, j \leq 31$, the angular separation θ_{ij} is the main object to estimate by our method, and it is known in our test setup. We define in the next section two possible metrics $d(i, j)$ between signals $x(i, t)$ and $x(j, t)$, $1 \leq t \leq 135,900$ and, as a result, get a dataset, \mathcal{D} of angle–distance pairs:

$$\mathcal{D} = \{(\theta_{ij}, d(i, j)) \mid 1 \leq i, j \leq 31\}. \quad (1)$$

We will then use this dataset to build a functional relation that allows to estimate the angular separation θ_{ij} from the information distance $d(i, j)$.

3. Distances between pairs of signals versus angular separations

In this section, we define the measures of distance between signals, namely correlation and information distance, and use them to build a functional relation between the angle separations of photocells and the distances of pixel streams they acquire.

3.1. Correlation distance

We call correlation distance between signals $x(t)$ and $y(t)$, $1 \leq t \leq T$, the quantity

$$d_c(x, y) = \frac{1}{2}(1 - C(x, y)),$$

where $C(x, y)$ is the correlation between the signals. It is easy to verify that $d_c(\cdot, \cdot)$ is a distance.

For the task considered in this paper, it is natural to prefer the correlation distance over the covariance or the Euclidean distance $\|x - y\|$, because both vary with signal amplitude (and offset, for the latter), whereas $d_c(\cdot, \cdot)$ is offset- and scale-invariant.

3.2. Information distance

Given two random variables x and y (in our case, the values produced by individual pixels of a discrete camera) taking values in a discrete set $\{1, \dots, Q\}$, the *information distance* between x and y is [4]:

$$d(x, y) = H(x|y) + H(y|x) = 2H(x, y) - H(x) - H(y), \quad (2)$$

where $H(x, y)$ is the Shannon entropy of the paired random variable (x, y) , and $H(x)$ and $H(y)$ are the entropies of x and y , respectively. It is easy to show that Eq. (2) defines a distance over random variables. This distance is bounded by $H(x, y) \leq \log_2 Q$, and can be conveniently replaced by the *normalized information distance*:

$$\bar{d}(x, y) = d(x, y)/H(x, y). \quad (3)$$

The advantage of (3) over (2) is that its properties are more invariant to Q : Eq. (2) is bounded by $\log_2 Q$ and is not necessarily equal to its upper bound when x and y are independent. In contrast, Eq. (3) is bounded by 1 independently of Q and is exactly 1 if and only if x and y are independent [4].

3.3. Estimating the information distance

Caution should be taken when estimating the information distance (3) from finite samples $x(t), y(t)$, $1 \leq t \leq T$: it is relatively common knowledge that replacing unknown probabilities $p_x(q)$ by sample frequencies $\hat{p}_x(q) = |\{t|x(t) = q\}|/T$ in the expressions of the entropy results in the biased *plug-in estimator* $\hat{H}(x)$, with expectancy

³ $|\cdot|$ denotes the set cardinal.

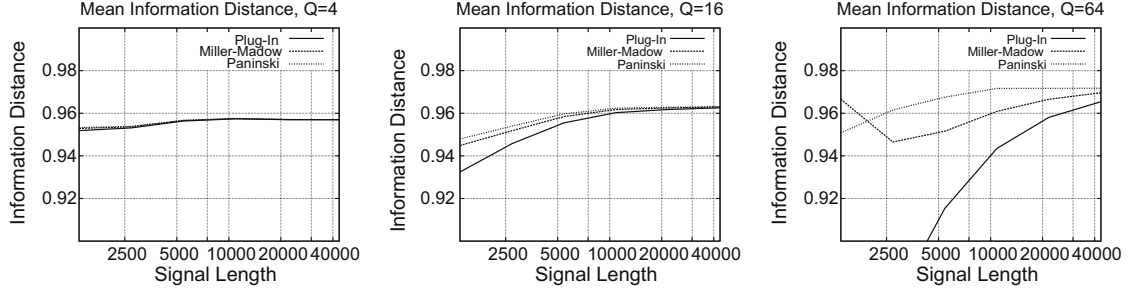


Fig. 5. Each of these plots shows how bias of various estimators is affected by signal length. Each curve is the average of 150 information distance estimates, taken between signals generated by photocells separated by approximately 5° . The signal length is varied from 1359 to 43488. A good estimator is one for which the mean value changes little with the signal length. The three considered information distance estimators are: the plug-in estimator, the Miller–Madow corrected estimator and Paninski’s estimator.

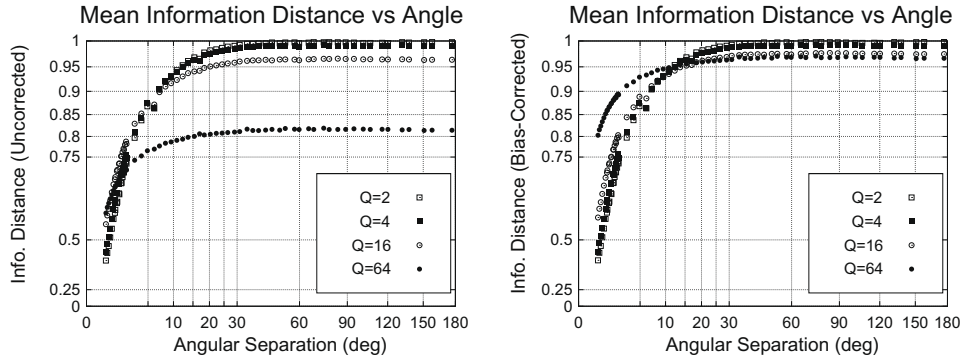


Fig. 6. Dependency of the information distance $d(i, j)$ on the angular separation θ_{ij} . The height of each plotted point is the average of 100 estimates of the information distance between signals $x(i, 1 \leq t \leq 1359)$ and $x(j, 1 \leq t \leq 1359)$. The abscissa is the angular separations between pixels i and j of the discrete camera. Each curve corresponds to a different number Q of quantization bins. The quantization bins are all taken to have weight $1/Q$. Right: Information distances computed with the first-order bias correction term of Eq. (4) removed. Left: Distances computed without correction.

$$E\{\hat{H}\} = H - \frac{Q-1}{2T} + \frac{1 - \sum_q p_x(q)}{12T^2} + O\left(\frac{1}{T^3}\right). \quad (4)$$

This bias in turn causes a bias in the information distance estimates in Eq. (3). While correcting for the first bias term $(Q-1)/2T$, is easy and results in the Miller–Madow estimator, correcting for the other terms is more delicate [30]. In addition to the plug-in and Miller–Madow estimators, we also consider here the estimator of Paninski [22].

A first consideration in the choice of estimators is the number of bins Q . Since the bias decreases slowly and is to a large extent proportional to Q , it is advantageous to re-quantize the signals from 256 graylevels to a more parsimonious representation with $Q < 256$ bins. When re-quantizing signals for estimating their information distance, we found a slight advantage in choosing bins that maximize the entropy, i.e. bins that contain equal numbers of values. All signal re-quantization in this paper is done using this scheme, also used in [21].

In order to choose a suitable estimator of the information distance, we first consider the stability of each estimator with respect to signal length. In Fig. 5, each curve is the average of 150 information distance estimates, taken between signals generated by two photocells separated by approximately 5 degrees, when the signal length varies approximately from 1350 to 43,500.⁴ The leftmost curve shows that, for $Q = 4$, all estimators behave pretty well, although the asymptotic value appears smaller than with $Q > 4$ (two plots at right). This last effect results from the information lost

in quantization. For $Q = 16$, the middle curve shows that Paninski’s estimator is stablest, but not much more than the Miller–Madow estimator, while the plug-in estimator is clearly far from its asymptotic value when the signals length is 1350. Finally, for $Q = 64$, the rightmost curve shows the trend of the previous curve greatly increased: as Q increases, the term Q^2 in the bias expression for $H(X, Y)$ or $H(X|Y)$ in Eq. (2) becomes more important and more samples are needed to reduce it. Not shown in these curves, the standard deviations of all three estimators are comparable.

From the previous result, the Miller–Madow estimator appears like a good choice, due to its relatively good performance, and to the fact that it is very easy to compute. In contrast, the Paninski estimator requires the computation of a table of same length as the signal, which renders it impractical with signals of variable length.

In order to further assess the advantage of the Miller–Madow estimator over the plug-in estimator, we show the results of other experiments performed by varying Q and the angular separation between the photocells.

Fig. 6 shows the average information distance Eq. (3) between the outputs of sensors with known angular separation. Each plot holds four curves, corresponding to $Q = 2, 4, 16$ and 64 . The left plot uses bias correction, whereas the right plot does not. The curves in the left plot are better grouped, showing that, when bias reduction is applied, the information distance estimates depend less on Q . This is important if one is interested in estimating properties of the information distance that depend on angular separation and on the ambient light-field, rather than on Q and T .

The conclusion of these experiments is that adding the first correction term in Eq. (4) to entropy estimates in Eq. (3) is justified and beneficial, especially when only short sequences are available. For

⁴ In order to obtain signals of length greater than 1359, the original length of our signal, we concatenate many signals corresponding to different sensor orientations.

the above reasons, we will use the Miller–Madow estimator in the sequel, and use $Q = 4$ bins, which is appropriate for short sequences.

One should note that correcting for bias in the entropy estimates is less important in the situation of Olsson et al. [21], where the distance measurements are embedded directly in an abstract space. Obtaining accurate information distance estimates is important in our approach because our goal is to accurately map information distance to angles.

3.4. Estimating angular separations from inter-signal distances

As explained earlier, our a priori knowledge of the world will be encoded in a functional relation mapping a measure of discrepancy between two signals, to the angular separation between the photocells that generated the signals. We now build this functional relation and assess its effectiveness at estimating angles.

In the previous section we detailed the computation of the distances between pixel streams, and therefore completed the construction of the angles and distances dataset \mathcal{D} (Eq. 1). From this dataset, we can finally build a constant by parts model of the expectancy of the distance knowing the angle. After verifying and, if needed enforcing, the monotonicity of this model, we invert it, obtaining a graph of angles, θ as a function of (correlation or information) distances, d :

$$\theta = \mathcal{F}(d) \quad (5)$$

Strict monotonicity has to be enforced for the correlation-based data, owing to the relatively small number of data points used for each quantized angle.

Fig. 7 shows the resulting graphs. This figure shows one of the major issues that appear when estimating the angular separation between pixels from the correlation or information distance: the graphs become very steep for large values of the distance, indicating that small changes of the distance result in large changes in the estimated angle.

Hence, a small uncertainty in the estimated distance implies a large uncertainty in the estimated angle. This problem is somewhat mitigated, in the case of the correlation distance, by limiting the abscissa to values $[0, 1/2]$.

For small distance values, on the other hand, the curves are much flatter, suggesting that small angles can be determined with greater accuracy. Both trends are particularly true for the information distance.

3.5. Experimental validation

We now assess how well angles can be estimated from the graphs obtained in the previous section. As described in Section 2.2, we

collect in each omnidirectional image the samples necessary to simulate 100 discrete linear sensors, each one having 31 photocells (100 times 31-tuples per image). For each 31-tuple of signals, we estimate the $31 \cdot 30/2$ correlation and information distances between pairs of signals and map these distance to angles using the models of Fig. 7. We assess the quality of the angle estimators by comparing these estimated angles with the known true angles.

Fig. 8(a) and (b) shows the boxplots the 100 estimated angles versus the true angles for the correlation- and information-based estimators, respectively. These plots show that the true angles are well estimated for small angles, while the spread and apparent bias of the estimators increases sharply for angles greater than five or ten degrees.

Fig. 8(c) and (d) better illustrates this trend: Fig. 8(c) plots the mean estimated angle versus the true angle. This figure shows that the estimators are essentially unbiased until five degrees and that they underestimate angles after that. Fig. 8(d) plots the mean absolute error in the estimated angles.

Fig. 8 shows how the proposed angle estimators perform on signals produced in the same conditions of environment and motion as the signals used to build the correlation and information distance-to-angle models. Since we will want to use these estimators in other situations, we must also evaluate their performance in more general conditions.

Fig. 9 shows how the proposed angle estimators perform on signals captured outdoors and indoors by the same camera setup as above. This time, the camera is held mostly horizontally, but the virtual discrete cameras still have pretty random orientations, owing to the very wide field of view. Four of the 2349 images are shown in Fig. 9(a). Again, we sampled 100 times 31-tuples of pixel-values per image and compared the true angles with the angles estimated using the correlation and information distance based estimators.

Fig. 9(b) shows the mean of the 100 estimated angles, while Fig. 9(c) shows the mean absolute error. This figure shows that, globally, the estimators actually perform better on this dataset than on the dataset acquired in the same conditions used to build the estimators.

Using the same outdoors and indoors sequence, but sampling horizontal virtual cameras allows to evaluate the performance of the angle estimators when the assumption of uniform sensor orientation is violated. Since the camera was mostly horizontal during the capture of the outdoors and indoors sequence of Fig. 9, and the virtual sensors sampled in the present experiment remained horizontal, these virtual sensors in effect have a very restricted set of orientations.

Nevertheless, Fig. 10 shows that the signals captured in these conditions allow to estimate angles with an accuracy that is less

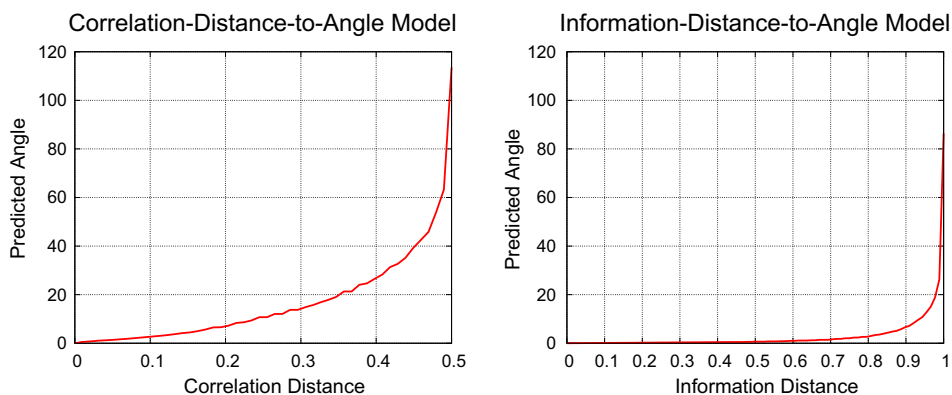


Fig. 7. Models relating correlation (left) or information distance (right) to angular separation between photocells. These models were built from simulated signals produced by the linear probe of Fig. 4(a). Signals of length $T = 135,900$, acquired indoors were used.

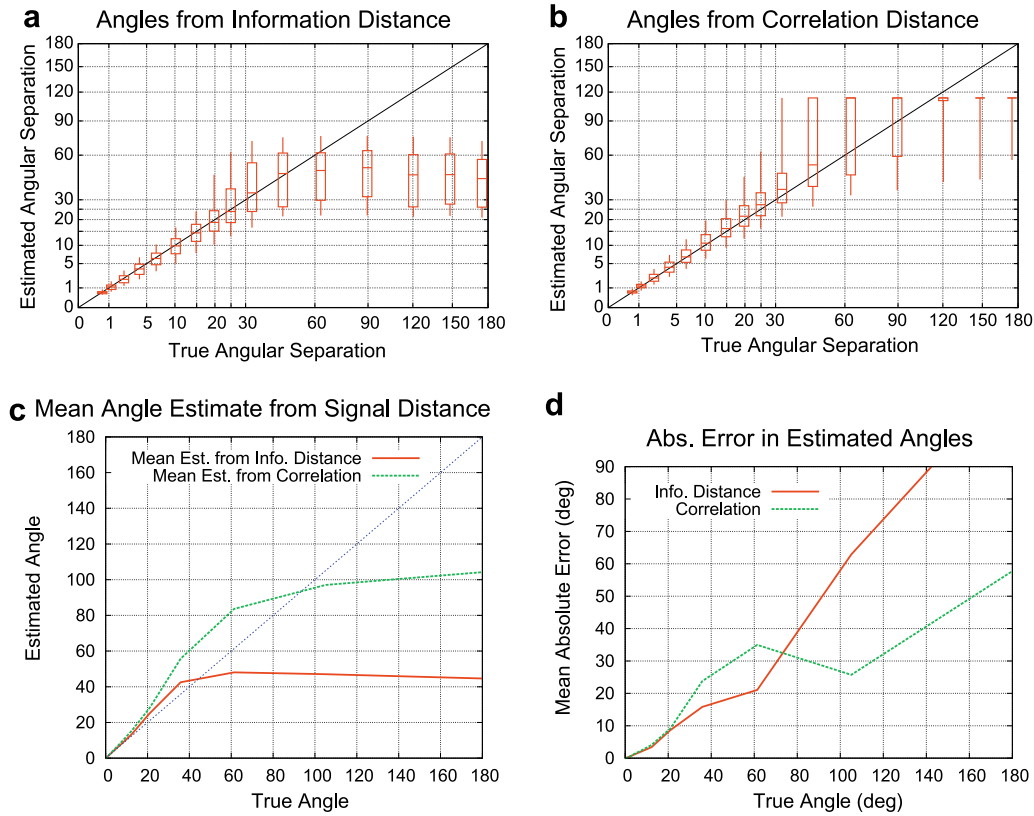
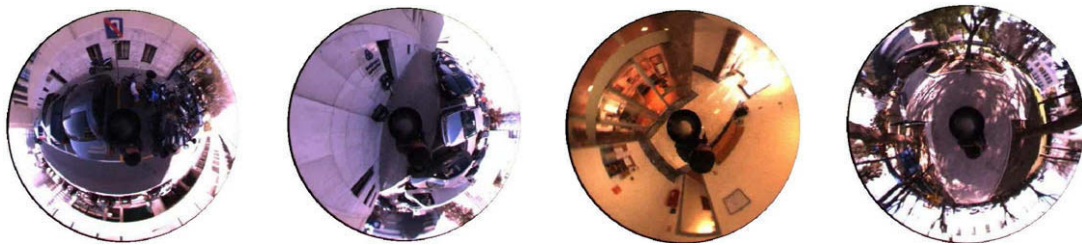


Fig. 8. Validation of the distance-to-angle models of Fig. 7. Boxplots (a) and (b) show the 5th percentile, first quartile, median, third quartile and 95th percentile of 100 angles estimated from signals of length $T = 1359$ simulated using the indoor sequence of Fig. 4(c). The angles are estimated from information distance in (a) and correlation in (b). The curves in (c) show the mean estimated angle vs. the true angle, and (d) shows the mean absolute error in the estimated angles.



(a) Four images of a sequence for testing the calibration methodology.

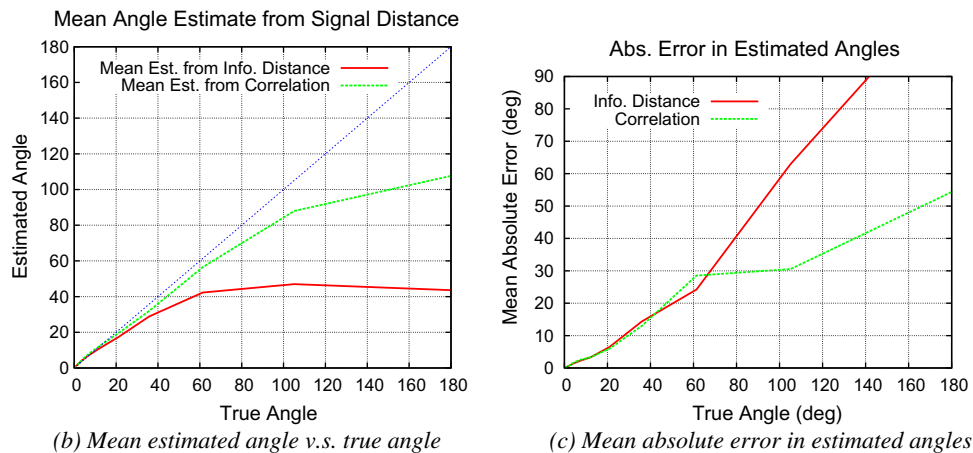


Fig. 9. (a) Four out of 2349 images of an outdoors and indoors sequence used to validate the angle estimators. (b) Mean of 100 sets of angles estimated from this sequence, plotted versus the true angle. (c) Mean absolute error in the estimated angles.

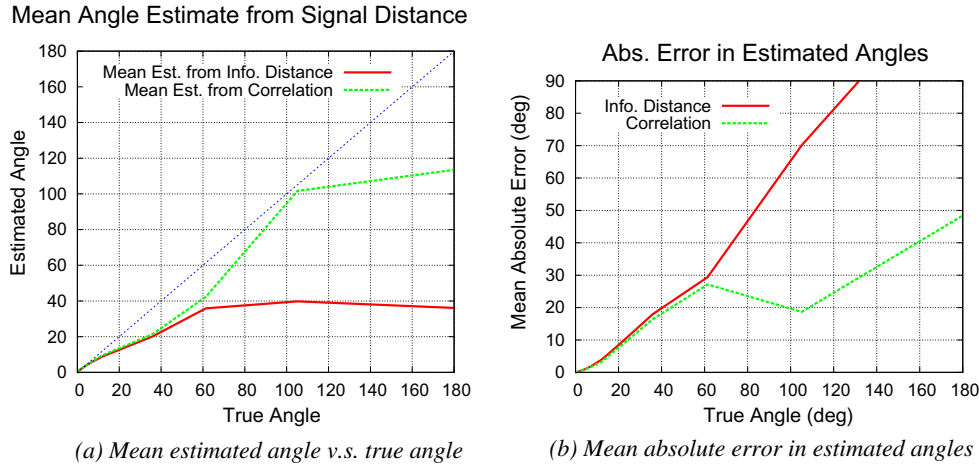


Fig. 10. Results of the angle estimators estimated from the sequence of Fig. 9, but with the virtual linear sensor kept mostly horizontal. (a) Mean of 100 sets of angles, plotted versus the true angle. (b) Mean absolute error in the estimated angles.

good than that of Fig. 9, but not much less good. One should note that the mean estimated angle is always smaller than the true angle. This seems to indicate that the correlation is greater between signals obtained by sampling along the horizon.

Finally we go a step further in testing the generalization ability of the proposed angle estimators. We now use different sensor, point-and-shoot Olympus Stylus 300 camera that can acquire 160×120 videos at ~ 15 frames per second. Multiple sequences of typically 1050 images (70 s) were taken indoors and outdoors, while the camera was moving, predominantly horizontally and forward, but also with different orientations. The horizontal and vertical fields of view cover 50 and 35 degrees, respectively. These sequences were concatenated, forming a single sequence of 22,822 images. Fig. 11(a) shows four of the images in the sequence. Obviously, this sequence does not allow to simulate a sensor with 180 degrees of field of view. In this experiment, we use a linear sensor consisting of 31 photocells spanning 35 degrees. The first two pixels are separated by 0.5° , the separation between consecutive photocells increases geometrically, so that the 31st photocell is 35 degrees away from the first.

Fig. 11(b) and (c) shows the precision and accuracy of the estimated angles. This figure shows that both estimators slightly overestimate angles, possibly due to the high texture contents that is often present in the images. The estimator based on information distance performs somewhat better than that based on correlation.

4. Calibrating a discrete camera

Having seen how to estimate the angle between two pixels from the correlation or the information distance between their respective data streams, we now estimate the whole sensor geometry from angle estimates. In this section, we produce sequences of pixel signals in the same conditions as in Section 2.2, except that the sensor shape is different. The distance between all pairs of signals are then estimated, and the angular separation between the pixels are estimated using Section 3.4. Finally, these angle estimates are embedded in the sphere using the algorithms defined in this section.

4.1. Embedding points in the sphere

The last step in order to solve the discrete camera calibration problem will be taken by solving the problem:

Problem 1 (Spherical embedding problem). Given angle estimates θ_{ij} , $1 \leq i, j \leq N$, find points X_i on the unit sphere, separated by angles approximately equal to θ_{ij} , i.e. $X_i^T X_j \simeq \cos \theta_{ij}$, for all i, j .

This problem can be reduced to the classical problem of distance geometry [5].

Problem 2 (Euclidean embedding problem). Given distance estimates D_{ij} , $1 \leq i, j \leq N$, find points Y_i in a metric vector space, such that, for all i, j , $\|Y_i - Y_j\| \simeq D_{ij}$.

Indeed, by defining an extra point $Y_0 = (0, 0, 0)$, and distances $D_{ij} = \sqrt{2 - 2 \cos \theta_{ij}}$, the mapping of the first problem to the second is immediate.

4.2. Factorization method

Solutions to both problems (with exact equality, rather than approximate) were published by 1935 [29].⁵ Schoenberg's Theorem 2 [29] states that if the matrix C with terms $C_{ij} = \cos \theta_{ij}$ is positive semidefinite with rank $r \geq 1$, then there exist points on the unit $(r - 1)$ -dimensional sphere that verify $X_i^T X_j = C_{ij}$ for all i, j . This result directly suggests the following method for embedding points in the 2-sphere:

1. Build the matrix C with terms $C_{ij} = \cos \theta_{ij}$, $1 \leq i, j \leq N$.
2. Compute, using the SVD decomposition, the rank-3 approximation $\tilde{C} = U U^T$ of C , where U is $N \times 3$. In essence, this involves square-rooting the three largest singular values of C and using them to scale the first three components of the left singular values.⁶
3. Define $X_i = (U_{i1}, U_{i2}, U_{i3}) / \|(U_{i1}, U_{i2}, U_{i3})\|$.

In the absence of noise, this very simple algorithm gives the exact solution, up to a unitary transformation. When there is noise, however, it is not optimal in many ways. In particular, it does not take into account the structure of the approximately multiplicative noise in the angles θ_{ij} .

In practice, this algorithm is very sensitive to noise: Fig. 12 shows the results of the reconstruction, when the pixels lie on a triangular grid of the image plane, scaled so that the angular diameter of the whole discrete camera is 5° (left), 10° (middle) or 20° (right). Please note that the lines connecting the photocell positions serve only to better illustrate the correspondence.

⁵ Schoenberg cites [15,19] as, respectively, previous solutions to Problems 1 and 2. These are the earliest references we are aware of.

⁶ Using Matlab/Octave notation: `[u, s, v] = svd(C); U = u(:, 1:3) * sqrt(s(1:3, 1:3))`.



(a) Four images from a sequence of 22822 for testing the calibration methodology.

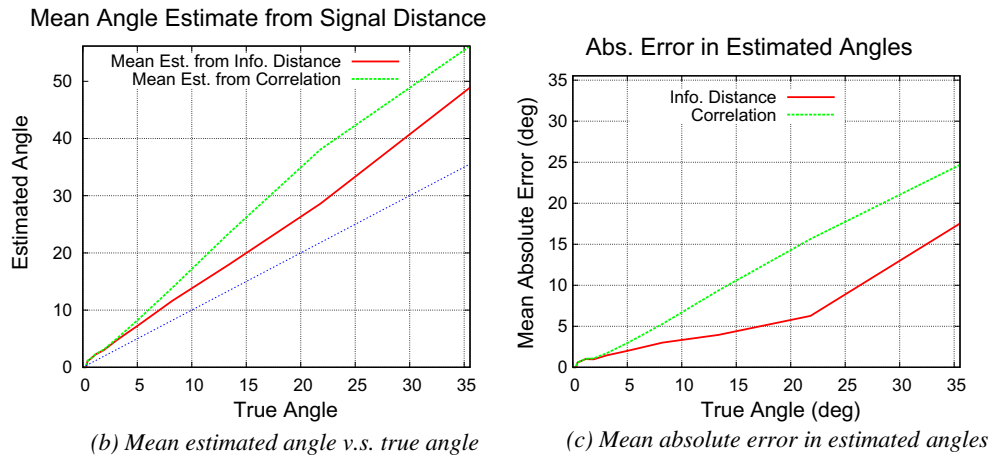


Fig. 11. (a) Images 300, 8400, 11,300 and 13,100, in a set of 22,822 images acquired with an off-the-shelf digital camera. (b) Mean of 100 sets of angles, plotted versus the true angle. (c) Mean absolute error in the estimated angles.

The reconstructed grids shown here are typical of what researchers reproducing our method may encounter. It is easy to see that for small-diameter discrete cameras, the estimated shape is pretty well determined by the information distance between the pixel streams, whereas, when the angular diameter increases, the noise in the estimated angles gradually overwhelms the embedding algorithm. It is common for the factorization method to break down with sensors of 20° diameter or less [11].

A different method is thus needed if we wish to obtain better calibration results. It is easy to verify that this situation is not directly tractable by variable-error factorization methods used in computer vision, such as [14]. In the next section, we will define and validate a more performant algorithm.

4.3. Robust nonlinear embedding method

Noting that the error in the estimated angles is approximately proportional to the actual angle suggests that the embedding method should weigh less heavily large angular estimates.

One such method is Sammon's algorithm [27], which we adapt to spherical – rather than Euclidean – embedding. In this paper, we minimize the weighted sum of the differences of the angle estimates with the internal products of directions:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \sum_{i,j} w_{i,j} (X_i^\top X_j - C_{ij})^2, \quad (6)$$

where $\mathbf{X} = \{X_1, X_2, \dots, X_N\}$ and

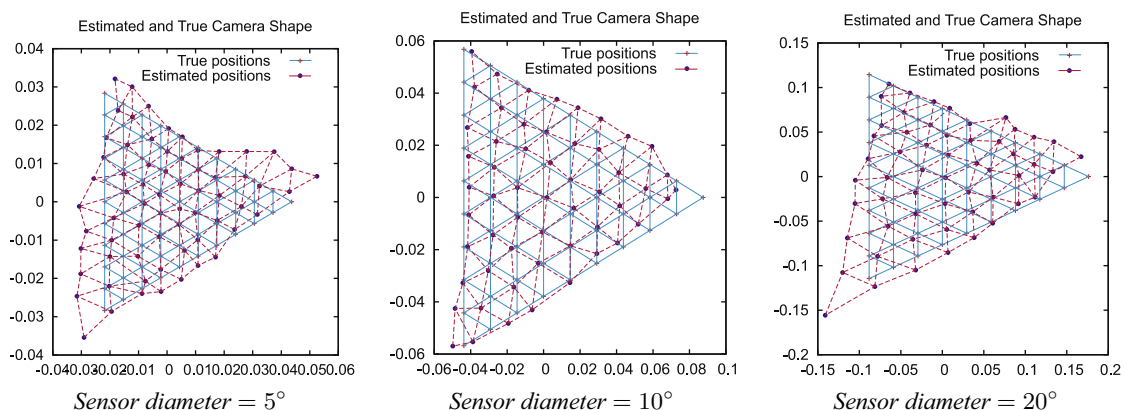


Fig. 12. True and estimated pixel layouts of a discrete camera consisting of photocells lying on a triangular grid. To estimate the layout, each pixel was sampled 1359 times. The information distance between pairs of pixels was estimated from these values, and converted to angular estimates using our non-parametric model. The angular distances were then embedded in the sphere. For visualization, they are aligned by the usual procrustes method, mapped to the plane by projective mapping with unit focal length, and line segments indicate the original pixel neighborhood relations.

$$w_{ij} = \begin{cases} \max \left\{ 0, \frac{1}{1-c_{ij}} - \frac{1}{1-c_0} \right\} & \text{if } c_{ij} \neq 1 \\ \frac{1}{\eta} & \text{otherwise.} \end{cases}$$

Algorithm 2. Spherical embedding

Input: Distance matrix Θ with terms θ_{ij} , $1 \leq i, j \leq N$, $\theta_{ij} \in [0, \pi]$

Output: Embedded points X_i , $1 \leq i \leq N$ on the unit sphere.

1. Compute the minimum spanning tree \mathcal{T} of Θ .
 2. Select the triplet (i, j, k) of points such that $(i, j) \in \mathcal{T}$, $(i, k) \in \mathcal{T}$ and the spherical angle $\hat{j}ik$ is maximized.
 3. Choose X_i, X_j and X_k s.t. $X_i^\top X_j = \cos(\theta_{ij})$, $X_i^\top X_k = \cos(\theta_{ik})$ and $X_j^\top X_k = \cos(\theta_{jk})$.
 4. Define the set of already embedded points $\mathcal{B} = \{i, j, k\}$.
 5. Define the set \mathcal{N} of neighbors of \mathcal{B} that are not in \mathcal{B} : $\mathcal{N} = \{i \mid \exists j \in \mathcal{B}, (i, j) \in \mathcal{T}\}$.
 6. For each element i in \mathcal{N} , taken from the closest to \mathcal{B} to the furthest:
 7. Find $j \in \mathcal{B}$ s.t. $(i, j) \in \mathcal{T}$ and $k \in \mathcal{B}$ that maximizes the spherical angle $\hat{j}ik$.
 8. Triangulate X_i from θ_{ij} and θ_{ik} (disambiguate using other points of \mathcal{B}).
 9. Update $\mathcal{N} := \mathcal{N} \setminus \{i\}$, $\mathcal{B} := \mathcal{B} \cup \{i\}$.
 10. End while
 11. If $\mathcal{B} \neq \{1 \dots N\}$, then goto 5.
-

To reflect the fact that big angles are less well estimated, we set $c_0 = 0.9$, so that estimates greater than $\arccos(0.9) \simeq 25^\circ$ be ignored. The other parameter, η is set to 1, allowing the points X_i to stray a little bit away from the unit sphere. Our implementation is inspired by the second-order iterative method of Cawley and Talbot (<http://theoval.sys.uea.ac.uk/~gcc/matlab/default.html>).

For initialization, we modify the algorithm by Lee et al. [18], which operates on the plane, so that it operates on the sphere.

Combining Sammon's algorithm to that of Lee, Slagle and Blum is not an unusual practice for embedding in Euclidean space, e.g. [2].

In our case, the algorithm is adapted to embed in the unit sphere an $N \times N$ distance matrix with terms θ_{ij} , $1 \leq i, j \leq N$, $\theta_{ij} \in [0, \pi]$. The algorithm is detailed in Algorithm 2. It is easy to see that this algorithm favors using small angle estimates and ignoring larger ones: first, in step 1, the edges that are part of the minimum spanning tree are by nature short (for example, the nearest-neighbor graph is a subgraph of the minimum spanning tree). Then, when adding a point to the set \mathcal{B} of already embedded points (step 7), this is done by triangulating with two points of \mathcal{B} that are necessarily close to the added point. Also note that the angle $\hat{j}ik$ in step 7 cannot be greater than $\pi/3$, by definition

of the minimum spanning tree, and the triangle used to localize the newly added point is far from a degenerate configuration. In addition to the useful properties just mentioned, this algorithm is simple to implement.

4.4. Testing the generality of the method

We now evaluate the results of the complete calibration methodology, and in particular the robust embedding algorithm on data produced by the angle-estimating method of Section 3.4. In this section we test two main cases: (i) calibrating non-planar discrete cameras having enlarged fields-of-view, using the same image acquisition device described in Section 2.1, and (ii) calibrating a rectangular discrete camera, but based on a different image acquisition device, namely a (consumer) digital camera.

4.4.1. Calibration from an indoors-and-outdoors sequence

The main objective here is testing the calibration of large fields-of-view while using calibration images that have completely different illuminations and textures (as compared to indoors image datasets). For this purpose, we produce sequences of pixel signals in the same conditions as previously, using the outdoors and indoors sequence shown in Fig. 9(a), except that the discrete-sensor shape is different. The information and correlation distances between pixels is then estimated from these signals, the angular separation between the pixels is estimated using Section 3.4, and the embedding method of Section 4.1 is applied to these angle estimates.

Fig. 13 shows the results of our calibration method on sensors covering more than a hemisphere, which thus cannot be embedded in a plane without significant distortion. It should be noted that, although the true sensor is each time more than hemispheric, the estimated calibration is in both cases smaller. This shrinkage is a known effect of some embedding algorithms, which we could attempt to correct.

4.4.2. Calibration using a different image sensor

Fig. 14 shows how our method applies to signals produced by a different sensor from the one used to build the distance-to-angle models, namely an Olympus Stylus 300 camera. An 8-by-8 square grid pixels spanning 34 degrees was sampled along a 22,822 image sequence taken indoors and outdoors. From this sequence, the estimated angles were generally greater than the true angles, which explains the absence of shrinkage. The higher angle estimates were possibly due to higher texture contents of the sequence. The estimated angles were also fairly noisy, possibly due to the sequence length, and we surmise that longer sequences would yield better results.

These results represent typical results that researchers reproducing our method may encounter.

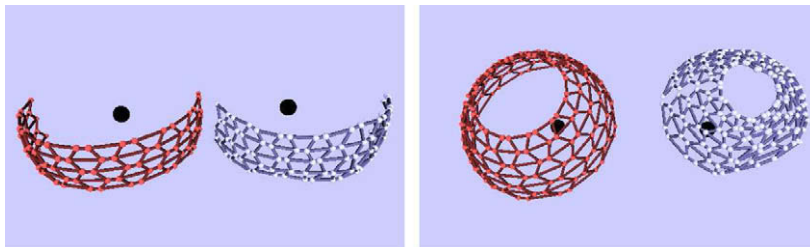


Fig. 13. Calibrations of two different sensors covering more than one hemisphere. On the left, a band-like sensor consisting of 85 photocells, calibrated from correlations (estimated: smaller, true: bigger). On the right, a discrete camera covering more than $180^\circ \times 360^\circ$, of 168 photocells, calibrated from the information distance (estimated: smaller, true: bigger). Each ball represents a photocell except the big black balls, representing the optical center.

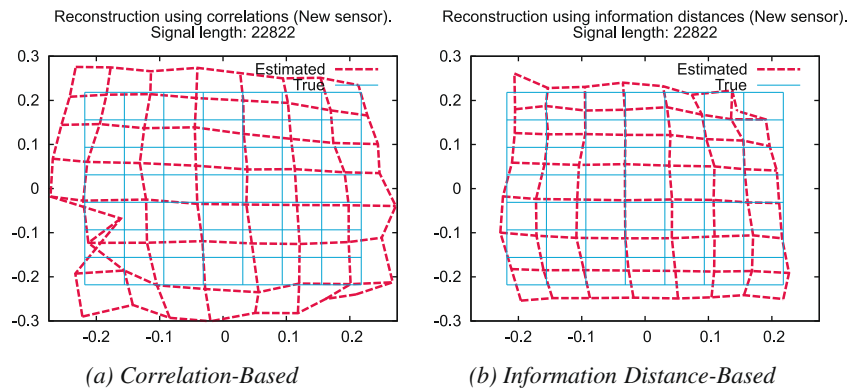


Fig. 14. Reconstructed and true pixel layouts of a discrete camera consisting of photocells lying on a rectangular grid (a and b). The sensor used differs from that with which the models of Fig. 6 were built. The reconstructions are obtained by first estimating the pairwise angular distances, then embedding the angles in the sphere (see text). For visualization, the reconstructions are aligned by the usual procrustes method, mapped to the plane by projective mapping with unit focal length. Added line segments show the true pixel neighborhood relations. Plot (a) is obtained from the correlation distance, and plot (b) from the information distance.

5. Discussion

This paper addressed the problem of determining the geometry of a set of photocells in a very general setting. We have shown that a discrete camera can be calibrated to a large extent, using just two pieces of data: a table relating information distances to angles; and a long enough signal produced by the camera. The main assumptions are that the camera motion directs each pixel uniformly in all directions, and that the environments in which the table is built and the one in which the calibrated sensor evolves are statistically similar.

Our algorithm proceeds in three main computational steps. We first estimate the information distance between each pair of signals. Then, using a pre-built table relating distances to angles, we transform the information distances into estimated angles. In the final step we embed the angles in a sphere, with a method that takes into account the noise introduced by the estimation process. It should be noted that the first and third steps are in themselves open problems that are the object of much research [1,3,5]. In particular we adapted two embedding methods [18,27] and obtained an algorithm adapted for the approximately multiplicative noise in the estimated angles. Experiments on simulated sensors assess the validity of the proposed approach.

As said above, we assumed an isotropic light-field, which can be easily produced manually moving the camera in any environment with enough visual stimuli. Dropping this assumption might allow to calibrate, for example, a sensor that is moving mostly in one direction, e.g. the eyes of a fly that mostly flies forward. It could also be interesting to generalize the proposed method to calibrate a discrete camera with a non-central calibration model, e.g. an earthworm with eyes all over the body. This poses an immediate challenge, because describing the geometric relation between two non-intersecting rays requires four parameters, while a single angle describes the relation between two intersecting rays. Future work will address this open problems.

Acknowledgments

This work was supported by TYZX, Inc, by the Portuguese FCT POS_C program that includes FEDER funds, and by the EU-project URUS FP6-EU-IST-045 062.

References

[1] J. Beirlant, E. Dudewicz, L. Györfi, E. van der Meulen, Nonparametric entropy estimation: an overview, *International Journal of Mathematical and Statistical Sciences* 6 (1997) 17–39.

[2] G. Biswas, A.K. Jain, R.C. Dubes, Evaluation of projection algorithms, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 3 (6) (1981) 701–708.

[3] P. Biswas, T.-C. Liang, K.-C. Toh, Y. Ye, T.-C. Wang, Semidefinite programming approaches for sensor network localization with noisy distance measurements, *IEEE Transactions on Automation Science and Engineering* 3 (4) (2006) 360–371.

[4] J.P. Crutchfield, Information and its metric, in: L. Lam, H.C. Morris (Eds.), *Nonlinear Structures in Physical Systems – Pattern Formation, Chaos and Waves*, Springer-Verlag, 1990, pp. 119–130.

[5] Jon Dattorro, *Convex Optimization & Euclidean Distance Geometry*, Meboo Publishing, 2005.

[6] R. Fergus, A. Torralba, W.T. Freeman, Random lens imaging, Technical Report MIT CSAIL TR 2006-058, Massachusetts Institute of Technology, 2006.

[7] W.T. Freeman, E.C. Pasztor, O.T. Carmichael, Learning low-level vision, *International Journal of Computer Vision* 40 (1) (2000) 25–47.

[8] C. Geyer, K. Daniilidis, A unifying theory for central panoramic systems and practical applications, *Proc. ECCV*, vol. II, Springer-Verlag, 2000, pp. 445–461.

[9] E. Grossmann, J.A. Gaspar, F. Orabona, Calibration from statistical properties of the visual world, in: *European Conference on Computer Vision*, 2008.

[10] E. Grossmann, E.-J. Lee, P. Hislop, D. Nistér, H. Stewénus, Are two rotational flows sufficient to calibrate a smooth non-parametric sensor?, *Proceedings of IEEE CVPR* (2006).

[11] E. Grossmann, F. Orabona, J.A. Gaspar, Discrete camera calibration from the information distance between pixel streams, in: *Proceedings of Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, OMNIVIS*, 2007.

[12] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[13] A. Hyvärinen, J. Hurri, P.O. Hoyer, Natural image statistics – a probabilistic approach to early computational vision, in: *Computational Imaging and Vision Series*, vol. 39, 2009. Available from: <<http://www.naturalimagestatistics.net/>>.

[14] M. Irani, P. Anandan, Factorization with uncertainty, in: *European Conference on Computer Vision*, 2000.

[15] L. Klanfer, Metrische charakterisierung der kugel, in: *Ergebnisse eines mathematischen kolloquiums*, vol. 4, pp. 43–45, 1933.

[16] I. Kohler, Experiments with goggles, *Scientific American* 206 (1962) 62–72.

[17] W.J. Krzanowski, *Principles of Multivariate Analysis: A User's Perspective*, Clarendon Press, 1988 (Statistical Science Series).

[18] R.C.T. Lee, J.R. Slagle, H. Blum, A triangulation method for the sequential mapping of points from n -space to two-space, *IEEE Transactions on Computers* 26 (3) (1977) 288–292.

[19] K. Menger, New foundations of euclidean geometry, *American Journal of Mathematics* 53 (1931) 721–745.

[20] D. Nistér, H. Stewénus, E. Grossmann, Non-parametric self-calibration, in: *Proceedings of ICCV*, 2005.

[21] L. Olsson, C.L. Nehaniv, D. Polani, Sensory channel grouping and structure from uninterpreted sensor data, in: *NASA/NoD Conference on Evolvable Hardware*, 2004.

[22] L. Paninski, Estimation of entropy and mutual information, *Neural Computation* 15 (2003) 1191–1254.

[23] D. Pierce, B. Kuipers, Map learning with uninterpreted sensors and effectors, *Artificial Intelligence Journal* 92 (1997) 169–229.

[24] B. Potetz, T.S. Lee, Scaling laws in natural scenes and the inference of 3d shape, in: *NIPS – Advances in Neural Information Processing Systems*, MIT Press, 2006, pp. 1089–1096.

[25] S. Ramalingam, P. Sturm, S. Lodha, Towards complete generic camera calibration, in: *Proceedings of CVPR*, vol. 1, 2005, pp. 1093–1098.

[26] R. Roy, Spectral analysis for a random process on the sphere, *Annals of the Institute of Statistical Mathematics* 28 (1) (1976).

- [27] J.W. Sammon Jr., A nonlinear mapping for data structure analysis, *IEEE Transactions on Computers* C-18 (1969) 401–409.
- [28] G. Schmidt, D.T. Moore, Tapered gradient index microlenses for compound lens arrays, in: *Proceedings of SPIE*, vol. 6342, International Optical Design Conference, 2006.
- [29] I.J. Schoenberg, Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espaces distanciés vectoriellement applicable sur l'espace de Hilbert", *Annals of Mathematics* 36 (3) (1935) 724–732.
- [30] T. Schuermann, Bias analysis in entropy estimation, *Journal of Physics A: Mathematical and General* 37 (2004) L295–L301 (arXiv:cond-mat/0403192v3).
- [31] F.W. Snyder, N.H. Pronko, *Vision with Spatial Inversion*, University of Wichita Press, 1952.
- [32] G.M. Stratton, Some preliminary experiments on vision without inversion of the retinal image, *Psychological Review* 3 (6) (1896) 611–617.
- [33] A. Torralba, A. Oliva, Statistics of natural image categories, *Network: Computation in Neural Systems* 14 (2003) 391–412.
- [34] R. Tsai, An efficient and accurate camera calibration technique for 3D machine vision, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 1986.
- [35] Y.N. Wu, S.-C. Zhu, C.-E. Guo, From information scaling of natural images to regimes of statistical models, Technical Report 2004010111, Department of Statistics, UCLA, 2004.