THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Scene recovery from many randomly distributed single pixel cameras

Scene recovery from many randomly distributed single pixel cameras

R. B. Fisher
University of Edinburgh - rbf@inf.ed.ac.uk

Abstract

This paper examines two scene recovery problems arising when interpreting data from massive numbers of randomly distributed single pixel cameras. Assuming that the camera positions and orientations are approximately known, the paper shows that both distant and nearby scenes can be reconstructed, and analyzes how recovery performance varies with sensor parameters.

Keywords: distributed sensing, sensor self-organization, scene recovery

1. Introduction

Hypothesize the existence of a solid state sensor consisting of a photodiode single pixel camera, fixed lens, photovoltaic power supply and local communication system (*e.g.* radio or optical). The 'camera' measures the intensity of light arriving within a fixed angle from its optical axis. These devices are feasible with current technology and could be manufactured cheaply. Potentially, a collection of these devices could form the framework for low-cost dense-coverage survey and surveillance systems in restricted access applications. Thousands of devices could be randomly dropped from an air vehicle to cover a potential surveillance zone. Planetary spacecraft landing sites, military battlefield surveillance, police surveillance and exclusion zone surveillance are possibilities for these multi-sensor systems.

There is much research on distributed traditional camera systems, where each camera records an array of light. What has not been considered much is the self-organization and interpretation of multiple single-pixel sensors, *i.e.* photocell sensors reporting a gray level value. There are two fundamental vision problems arising in this scenario: 1) determining where the sensors are and 2) making sense of their collective observed light intensity. This paper presents results on the latter problem. **The research presented here shows that it is possible to reconstruct both nearby and distant scene structure from a collection of thousands of such sensors.**

2. Background

There is no prior art known to us on this vision problem when using single pixel sensors. For the past 5 years or so, there has started to be good work on networks of traditional cameras. When working with multiple full cameras with overlapping fields of view, standard camera calibration methods are possible using autocalibration from feature point correspondences [6], observed calibration targets [16], special viewpoints [7] or multiple views of moving objects [15]. When cameras do not overlap, then persistence of the tracked target's properties have been used to establish the topological relationships between cameras, in particular their linked entry and exit zones [13, 4, 11, 12]. By making assumptions about target motions, then metrical relationships can be used to infer the relative camera positions [3, 8, 14]. The closest known research self-organizes randomly placed single-bit sensors into a network, using peer-to-peer signal strength decay to estimate relative distances between sensors [2].

This paper is about scene reconstruction, given the sensors and their positions. If the sensors were regularly placed, then methods inspired by CT/MRI image reconstruction might be possible, but here the sensors are irregularly distributed. A space carving-like method (*e.g.* [9]) can be used for nearby scene recovery and a plenoptic-like algorithm (*e.g.* [10]) is possible for distant scene recovery, but the previous literature uses standard cameras.
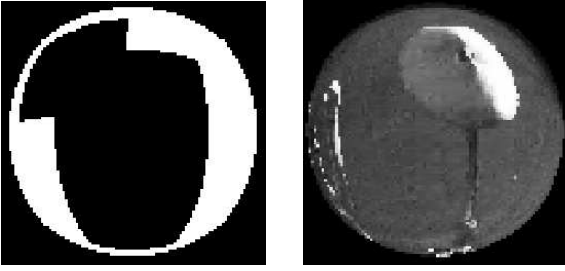
Figure 1. Binary and gray level scenes as projected onto the ground plane.

3. Sensor and Distribution Models

All sensors are identical, reporting a single gray-level value based on integrating light arriving within a cone of angle α about the sensor's optical axis. These are essentially 'raxal' cameras [5], here characterized by their position $\vec{c}$, optical axis orientation $\vec{a}$ and aperture half-angle α.

The sensor positions are random, uniformly distributed over the scene, within a bounded radius *e.g.* $r = 1$ about the scene center. We have assumed that they lie on a plane, but this is not an important point. Sensor orientations are random, uniformly distributed over the viewsphere. We ignore sensors whose optical axes are less than 0.35 radians above the ground plane as these provide little usable data.

4. Distant Scene Recovery

Distant scene recovery is based on the realisation that only sensor orientation matters - not sensor location on the ground plane. In the text below, we first model the sensor observation and then show how to invert the process by back-projection, much as in a simplified version of MRI or CT back-projection or plenoptic view recovery [10].

The ith sensor is at position $\vec{c}_i$ and its optical axis $\vec{a}_i$ (a unit vector). Each sensor integrates light from rays within angle α of its optical axis. We assume a sensor response function $f_\alpha(\theta)$ that weights the ray's contribution as a function of the angle θ away from the optical axis, up to the maximum aperture α. $f_\alpha(\theta) = 0$ for $\theta > \alpha$. Define the set of all unit vector rays $\vec{r}$ within angle α of direction $\vec{x}$: $T_\alpha(\vec{x}) = \{\vec{r} \mid \vec{r} \cdot \vec{x} \geq cos(\alpha)\}$. Let $f_\alpha(\theta)$ satisfy $\int_{\vec{r} \in T_\alpha(\vec{x})} f_\alpha(cos^{-1}(\vec{x} \cdot \vec{r})) = 1$. With this aperture function, each sensor i observes intensity $V_i = \int_{\vec{r} \in T_\alpha(\vec{a}_i)} I_{\vec{r}} f_\alpha(cos^{-1}(\vec{r} \cdot \vec{a}_i))$ where $I_{\vec{r}}$ is the intensity of ray $\vec{r}$ from the background to the sensor.

From the set $\{V_i\}$, estimate $\hat{I}_{\vec{r}}$ for each direction $\vec{r}$. There might be multiple sensors that have observed this direction, and also some directions without an observation. The intensity in direction $\vec{r}$ is computed from all sensors whose axis vectors are within angle α of the observed direction $\vec{r}$. Define the set of all sensors that see direction $\vec{r}$: $S(\vec{r}) = \{i \mid \vec{a}_i \cdot \vec{r} \geq cos(\alpha)\}$. The estimated intensity is:

$$\hat{I}_{\vec{r}} = \frac{\sum_{i \in S} V_i f_\alpha^*(cos^{-1}(\vec{a}_i \cdot \vec{r}))}{\sum_{i \in S} f_\alpha^*(cos^{-1}(\vec{a}_i \cdot \vec{r}))}$$

where $f_\alpha^*(f_\alpha(x)) = 1$. In the experiments below we use the constant sensor function $f_\alpha(x) = \frac{1}{\pi \alpha^2}$ (α in radians here).

4.1. Experiments

Figure 1 shows a binary and a gray-level scene. The binary scene has two overlapping black rectangles on a white background. The gray-level scene is of an illuminated apple segment skewered by a stick. The viewpoint is a little unusual - here the image is distant, so it lies on a hemisphere at infinity. To display this, we have rendered a scaled orthographic projection onto the ground plane. Figure 2 shows the reconstructed images, using $N = 6000$ sensors, with aperture cone angle $\alpha = 2$ degrees. We can clearly see that reconstruction is possible.

We varied the number of sensors from 1000 - 10000 in steps of 1000, and aperture α for 2, 3 and 4 degrees. Figure 3 shows the percentage of distant scene not observed. The mean (1.1) and standard deviation (23) of the reconstructed pixel value errors are largely constant above 2000 sensors.

From elementary geometry, we can show the area of the unit hemisphere observed by a sensor with aperture angle α (in radians) is $2\pi(1 - cos(\alpha)) \doteq \pi\alpha^2$ for small α. Therefore, the percentage p of the hemisphere above 0.35 radians observed by each sensor is $p_\alpha = \frac{2\pi(1 - cos(\alpha))}{2\pi(1 - cos(\frac{\pi}{2} - 0.35))} \doteq 0.76\alpha^2$. Since sensor placement is random, the average hemisphere coverage
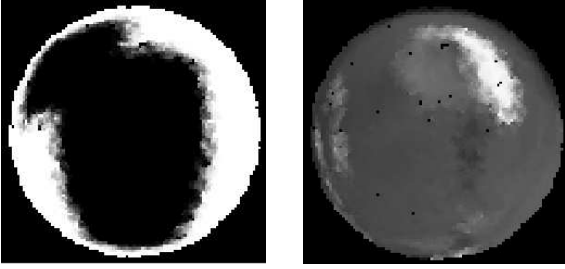
Figure 2. Reconstruction with $N = 6000$ sensors and aperture cone angle $\alpha = 2$. The black dots are where no reconstruction occurred.
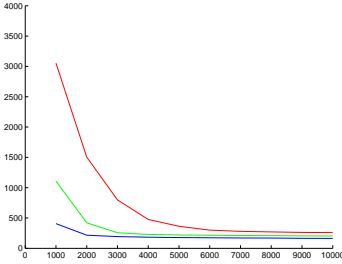

Figure 3. Pixels of scene image not recovered. All curves are for 1000-10000 sensors (horizontal axis) for apertures $\alpha = 2$ (top curves), $\alpha = 3$ (middle curves) and $\alpha = 4$ (bottom curves).

by N sensors is $1 - (1 - p)^N$. For the values of α used, the estimated minimum number of sensors needed to obtain 99.9% coverage is:

α (degrees)	1	2	3	4	5
N	29834	7456	3313	1864	1192

The curves in Figure 3 shows that this is a reasonable estimate as the percentage of non-observed scene is largely stabilized when more sensors than the threshold number given in the table are used.

5. Nearby Scene Recovery

The general principle behind recovery of nearby scenes is like a mixture of stereo triangulation and CT/MRI reconstruction. The algorithm presented below is a variant of the space carving approach [9]. The stereo aspect is straightforward: since we know where the sensors are, we could intersect matching rays to deduce the 3D target point's position (if we knew the corresponding rays). In standard feature-based stereo algorithms, the camera epipolar geometry and local intensity descriptions constrain the potential matches between feature points (*e.g.* those found by the Harris corner finder). Unfortunately, we do not know the corresponding rays. All we have initially are single intensity samples which are highly ambiguous.

The algorithm works in two stages: 1) Hypothesize the background, using the method of Section 4. 2) Hypothesize nearby surfaces by a space-carving process. The second stage is described here.

5.1. Volumetric Recovery

As we initially have no hypotheses about which rays observe foreground surfaces and which observe background, we effectively intersect all rays. We do this using a form of space carving algorithm, over a voxel space that represents where surface points are in the 3D space over the sensors.

To hypothesize the surface efficiently, space near to the ground plane is partitioned into an array of $W \times W \times H$ voxels. The reconstruction algorithm is based on back-projection of the rays into the voxels. The core of the algorithm uses this array for counting the number of sensor observations that pass through each voxel. Thus, rather than having to do an $O(N^2)$ intersection algorithm for N sensors, we need only an $O(HN)$ vote casting algorithm. As N is big and H small, this is more efficient.

If there is an observable surface lying in a voxel, then that surface patch would be observed by any sensor ray that passes through the voxel. The observed intensity of each of the sensor rays will be the same (if we ignored noise and the actual

3

Figure 4. ... in front of a distant scene (a) and its reconstruction ... the side ...
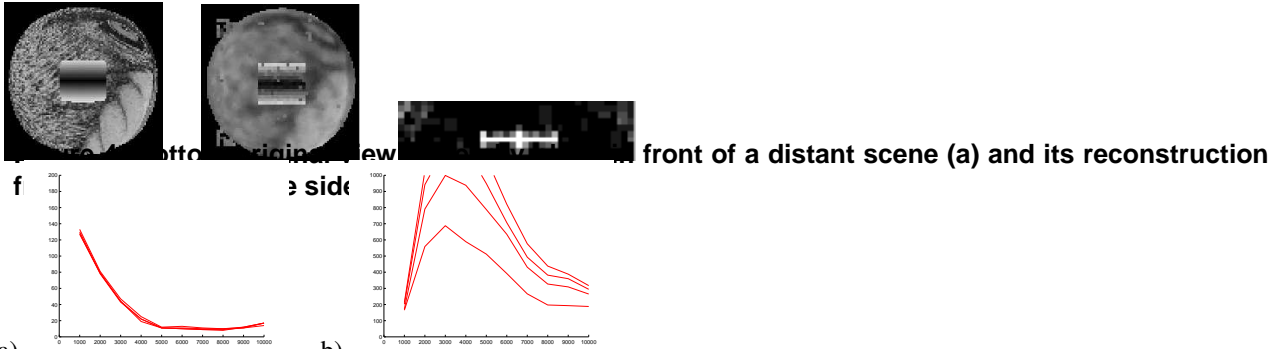
Figure 5. Variation of a) number of true voxels missed, b) number of false voxels created. The top curves are for $\alpha = 4$, the middle for $\alpha = 3, 3$ and the bottom for $\alpha = 1$.

sensor aperture) as they are observing the same surface. Note that: 1) many rays won't intersect any of the surface voxels because of their random orientations, 2) each surface voxel could be observed by many nearby rays and 3) each ray will be averaging different surface patch intensities because of the aperture α.

1. Find the set S_j of all rays that observe the voxel j at position $\vec{v}_j$. Given a sensor position $\vec{c}_i$ and axis direction $\vec{a}_i$, the voxel is observed if $(\vec{v} - \vec{c}_i) \cdot \vec{a}_i \geq cos(\alpha)$. This set is found by projecting each sensor ray through the voxel array and then examining the sets S_j.
2. Hypothesize the presence of a surface and its lightness based on S_j. If there are 0 or 1 values in S_j, then there is no evidence that there is a surface at voxel j.
3. If there is a surface corresponding to the sensors in S_j, then the sensors should observe approximately the same intensity. We compute the standard deviation of the observed intensities in S_j and only accept this as a surface voxel if the standard deviation is below a threshold. The mean of the observed intensities is the estimated surface intensity.
4. To reduce spurious surface detections, we eliminate voxels that only have 0 or 1 face-connected neighbors.

5.2. Experiments

Figure 4a shows a scene with a surface (the rectangle with a grating reflectance pattern) parallel to the ground plane against a distant background (a portion of the 'mandrill' image). The viewpoint is the same as before. Figure 4b shows the corresponding reconstructed image and Figure 4c shows the reconstructed space projected onto the side of the voxel cube. Here $N = 10000$ sensors, with aperture cone angle $\alpha = 3$ degrees. We used $W \times W \times H = 40 \times 40 \times 10$ voxels covering 80% of the sensor field. We can clearly see that reconstruction is possible, and there are few spurious points inferred.

We varied the number of sensors from 1000 - 10000 in steps of 1000, and also the aperture α for 2, 3 and 4 degrees. Figure 5 shows some measurements that help characterize this distributed sensor.

Figure 5a shows the number of surface voxels unrecovered by the process. As the number of sensors increases, more of the surface is found, until the number of spurious observations in each voxel starts to increase, which causes the voxel to be rejected. Figure 5b shows the number of spurious surface voxels infered. Here, the peak in the middle occurs because there are always spurious correspondences. For low numbers of sensors, few spurious correspondences occur. For large numbers of sensors, many spurious correspondences occur, but the variations in gray levels in the correspondences becomes more more extreme so the false voxels are rejected by the consistency checking. In between, more correspondences occur, but not enough at each voxel to eliminate them. The mean intensity (2) and depth reconstruction (0) errors stablize when using more than 3000 sensors. The different sensor aperture sizes explored here had little effect on the results, probably because of the averaging effects over the whole sensor's field of view.

Three types of noise were considered: a) error in the estimated optical axis direction, b) error in the estimated ground plane position of the sensor and c) error in the measured intensity. In these experiments, we fixed the number of sensors $N = 10000$ and sensor aperture $\alpha = 3$. We observed the mean and standard deviation of the gray level and depth reconstruction, the number of surface voxels not recovered and the number of false surface voxels created.

Figure 6 shows the number of surface voxels that were not recovered as the standard deviation of noise in the a) sensor axis orientation, b) sensor position increases. The effect of the introduced noise is that corresponding sensor rays will no longer intersect thus the increase in missed voxels. The curves for intensity noise error are similar. Altogether there are about 600 true surface voxels with up to 250 voxels lost as noise increases.
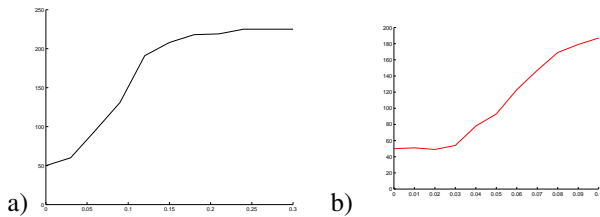
4

Figure 6. Number of surface voxels unrecovered as a function of a) optical axis orientation error, b) sensor position noise.

6. Discussion and Conclusions

Sections 4 and 5 have shown that it is possible to recover both distant and nearby scene structure from large numbers of randomly distributed single pixel sensors.

Recovery requires estimates of the position and orientation of the sensor. We will report on how they can be estimated in another publication, but a process based on temporal signal correlation is producing preliminary results that suggest that pose recovery is feasible.

As for the results presented here, further effort could be put into reducing the number of spurious voxel recoveries. In particular, using probabilistic voting [1] and exploiting surface continuity may allow refining both the foreground target and background outlier voxels. One of the next steps would be to experiment with real data, but we have not yet found a practical way to acquire 10000 randomly distributed observations (*e.g.* using a traditional camera) where the sensor location is known.

Acknowledgements

References

[1] R. Bhotika, D. J. Fleet, K. N. Kutulakos, "A Probabilistic Theory of Occupancy and Emptiness", Proc. Eur. Conf. on Computer Vision Vol III, pp 112 - 130, 2002.

[2] L. Doherty, L. El Ghaoui, K. S. J. Pister. "Convex Position Estimation in Wireless Sensor Networks", Infocom 2001, Anchorage, AK, April 2001.

[3] R. B. Fisher. "Self-organization of randomly placed sensors", Proc. Eur. Conf. on Comp. Vis., vol 4, pp 146-160, 2002.

[4] W. E. L. Grimson, K. Tieu, G. Dalley, "Inference of non-overlapping camera network topology by measuring statistical dependence", Proc. Int. Conf. Comp. Vis., 2005.

[5] M. D. Grossberg, S. K. Nayar "A General Imaging Model and a Method for Finding its Parameters", Proc. of IEEE Int. Conf. on Comp. Vis., Vancouver, Canada, July 2001.

[6] R. Hartley, A. Zisserman. "Multiple view geometry in computer vision". Cambridge Univ. Press, Cambridge, 2000.

[7] H. Ishiguro. "Distributed Vision System: A Perceptual Information Infrastructure for Robotic Navigation", Proc. Int. Joint Conf. on Artif. Intel. 36-41, 1997.

[8] O. Javed, K. Shafique, M. Shah. "KNIGHT: A real-time surveillance system for multiple overlapping and non-overlapping cameras", Proc. Int. Conf. on Multimedia and Expo., July 2003.

[9] K. N. Kutulakos, S. M. Seitz, "A Theory of Shape by Space Carving", Int. J. Computer Vision, Vol 38, No. 3, pp. 199-218, July-August 2000.

[10] L. McMillan, G. Bishop. "Plenoptic Modeling: An Image-Based Rendering Approach", Proc. SIGGraph pp 39-46, 1995.

[11] D. Makris, T. Ellis, J. Black. "Bridging the gap between cameras", Proc. Comp. Vis. Pat. Recog, June 2004.

[12] D. Marinakis, G. Dudek, D. Fleet. "Learning sensor network topology through Monte Carlo expectation maximization", Proc. Int. Conf. on Robot. and Autom., 2005.

[13] A. Nakazawa, H. Kato, S. Inokuchi. "Human Tracking Using Distributed Vision Systems", Proc. Int. Conf. on Pattern Recog., Vol 1, pp 593-596, 1998.

[14] A. Rahimi, B. Dunagan, T. Darrell. "Simultaneous Calibration and Tracking with a Network of Non-Overlapping Sensors", Proc. Comp. Vis. and Pat. Recog., 2004.

[15] S. N. Sinha, M. Pollefeys, L. McMillan. "Camera network calibration from dynamic silhouettes", Proc. IEEE Conf. on Comp. Vis. and Pat. Recog., Washington DC, Volume 1 pp. 195-202, 2004.

[16] Z. Zhang. "Flexible camera calibration by viewing a plane from unknown orientations", Proc. Int. Conf. Comp. Vis., Greece, pp 666-673, 1999.