

Pose Hashing with Microlens Arrays

Ian Schillebeeckx^(✉) and Robert Pless

Department of Computer Science and Engineering,
Washington University in St. Louis, St. Louis, USA
{ischillebeeckx,pless}@wustl.edu

Abstract. We design and demonstrate a passive physical object whose appearance changes to give a discrete encoding of its pose. This object is created with a microlens array that is placed on top of a black and white pattern; when viewed from a particular viewpoint, the lenses appear black or white depending on the part of the pattern that each microlens projects towards that viewpoint. We analyze different design considerations that impact the information gained from the appearance of microlens array. In addition, we introduce the process through which the discrete microlens pattern can be turned into a viewpoint and a pose estimate. We empirically evaluate factors that impact viewpoint and pose estimation accuracy. Finally, we compare the pose estimation accuracy of the microlens array to other related fiducial markers.

Keywords: Microlens array · Fiducial marker · Pose estimation

1 Introduction

Fiducial markers are critical to many visual, augmented reality and robotic tasks. Most often, fiducial markers serve to create geometric constraints that relate the known position of the marker on the object to the observed position of the marker in the image. Multiple fiducial markers, or larger patterns with several observable points can be combined to constraint the relative pose of an object.

These approaches that relate the fiducial marker to the image suffer from two shortcomings. First, the geometric constraints from a set of points on a fronto-parallel plane are known to be sensitive to noise [1–3]. Second, for any viewpoint, patterns or sets of markers that are collinear or nearly collinear lead to poorly conditioned estimators for the pose estimation problem [4, 5].

Some recent work attacks these limitations by designing fiducial markers whose appearance changes depending on the direction from which they are viewed. Images of these fiducial markers directly encode constraints on the object rotation and therefore require fewer markers and eliminate the need for non-collinear markers for pose estimation. Marker designs include lenticular arrays and microlens arrays that have a moire-pattern to encode the viewing angle and go by the name “Lentimark” [6, 7] and “Arraymark” [8, 9]. Other work uses lenticular arrays that change color depending on their viewing angle [10, 11].

This paper expands this line of work. We propose the integration of a microlens array with a random black and white pattern adhered to the back. Because these microlenses focus parallel rays onto an approximately single point of their back-plane pattern, they will appear either black or white. Using different patterns behind different microlens dots means that a given viewing direction will cause some dots to be black, some to be white, and some to be in between because the viewpoint is on the boundary of black and white regions. The major contributions of this work are:

- the design of a microlens based fiducial marker with a random black and white back-plane pattern,
- a discrete, combinatorial approach to geometric inference based on images from this fiducial marker, and
- experimental evaluation of a physical instantiation of this marker design.

2 Related Work

Fiducial markers are widely used in computer vision and augmented reality. The most common current versions are AprilTags [12] and the ARTag [13], which are often used within augmented reality libraries such as ARtoolkit [14]. These tags are black and white patterns that are defined for easy detection, and which encode an index that can be used to differentiate multiple tags that are both visible at the same time. Larger scale calibration setups include the de-facto standard approach by Zhengyou Zhang [15]. This method requires the use of a large black-and-white checkerboard pattern to estimate the pose of the camera with respect to the pattern. In all cases, however, the fiducial marker is explicitly constructed to have a similar appearance for all possible viewing angles.

Both AprilTags and ARTags require matching a pattern, so the local orientation of the fiducial marker (and not just the position) is also available as a geometric cue. The derivation of geometric constraints that take advantage of this cue has been done for rigid body pose estimation [16], and in the context of geo-location where the correspondence of image points to a 3D scene model is unknown [17].

Most related to our work are approaches that explicitly make patterns that change their appearance depending on their viewing direction. The use of microlens arrays to both capture and render a complete lightfield was suggested by Nobel Prize winner Gabriel Lippmann in 1908 [18], in the context Integral Photography. Here “Integral” is used in the sense of complete, and the microlens arrays create a light field pattern that changes as the viewer moves so as to create the perspective effects that an array of microlenses captured on film by sampling the light field. In the context of fiducial marker design, Agam Fiducials were suggested in 2000 as drawings on corrugated paper where different orientations of the paper are painted different colors so that the visible color gives a constraint on the viewing direction [19]. Next, and more recently, the “Lentimark” [6,7] and “Arraymark” [8,9] fiducial markers use lenticular arrays or microlens arrays

that change appearance for different viewpoints. Using a moire encoding pattern, this change in appearance is a bar or cross that translates relative to the rest of the marker. This beautiful approach gives both human and machine interpretable fiducial markers, but requires the fiducial to be imaged at relatively high resolution in order for the relative position of the dark spot to be measured accurately. Finally, a recent approach creates a lenticular pattern whose apparent color depends on its relative orientation to the camera [10, 11]. This avoids the need for the fiducial marker to take up many pixels, but adds the challenge of using color as a good measurement in mixed lighting environments.

Several additional works also create markers that explicitly look differently from different orientations. The light-field probe is created with microlens arrays that each cover a color wheel, creating a color coded lightfield that was used to understand the shape of reflective and refractive objects [20–23]. Finally, the BoKodes work uses a very small projector in the scene to transmit into the scene a collection of QR-Codes [24]. For most camera settings this appears to be a small point, but when the camera is defocused the projector image that impinges on the camera lens is visible and the relative orientation of the BoKode is inferred from which QR code is visible.

This work is most inspired by the BoKode work which creates a discrete pattern that describes an orientation constraint. But because our signal is derived from the spatial layout of microlens patterns it is not necessary to defocus the camera and therefore the orientation encoding is available under standard use cases of a camera.

3 Discrete Encoding of Viewpoint with a Microlens Array

A microlens array is plastic sheet comprised of many small lenses, called lenslets, that are arranged in a regular grid. To create visual effects with the microlens array, the thickness of the sheet is set to be the same as the focal length of the lenslet so that parallel rays are focused onto a point on the back of the sheet. Parallel rays from different directions are focused onto different points. Because light is reversible, the lenslet can also be thought of as a (dim) projector, projecting the appearance of the pattern behind the lenslet out into the world. A schematic of this is shown in Fig. 1.

Thus, the appearance of the microlens array is defined by the pattern that is underneath each microlens. When that pattern is low frequency, the appearance of each lenslet will change slowly as the array is rotated. For patterns with higher frequency, the pattern will change more quickly, and if different lenslets have different patterns the microlens array has an apparent random flickering as each lenslet changes appearance independently of the others. Figure 1 shows two example viewpoints of a microlens array with a texture of many small, randomly placed black blocks. Because this texture is high contrast, the resulting microlens appearance changes quickly for different rotations.

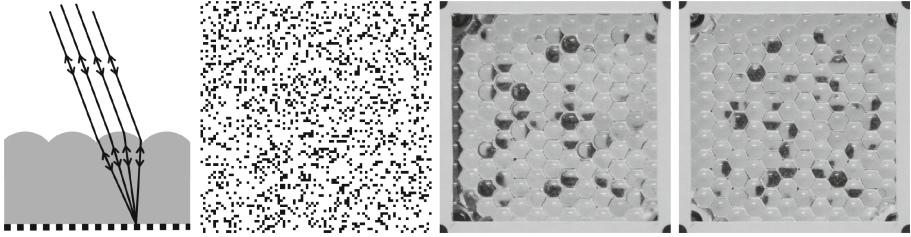


Fig. 1. A microlens array placed on top of a pattern creates an image where each lenslet magnifies a piece of the pattern below it. (Left) The microlens array is designed so that parallel rays focus at a point on the back of the array, so that point is magnified. (Left center) We show an example of a pattern with randomly placed black squares. For different viewpoints, each lenslet will focus at a different location on the pattern and thus produce the different sets of black and white appearances. (Right center and Right) We show the appearances for two different viewpoints after being transformed with a homography to make the images more easily comparable. The appearance of the array changes dramatically, and the discrete measurement of which lenslets are dark and light encodes its orientation.

By thresholding the appearance, one can limit a lenslet to 2 states to carry a bit of information. Considered together for a single viewpoint, all lenslets on the microlens array express a bit string that varies based on viewpoint. In an ideal case, a microlens array with n independent lenslets, could encode 2^n unique bit strings, and thus 2^n uniquely encoded appearances. With 12 lenslets giving binary measurements, this could suffice to encode viewpoints in the viewsphere up to 30 degrees from fronto-parallel to within 1 degree.

While this intuition of a discrete encoding of pose inspired us, random patterns may not give optimal encodings, and there is value in a non-binary classification of lenslet appearance. In the next section we consider practical approaches to choosing patterns that are most useful and consider what is the best discretization of the apparent lenslet brightness. After that we will derive a pose estimation approach and evaluate it with real images of a prototype.

4 Discretization and Entropy of Single Lenslet Measurements

A single lenslet will share the same appearance for a set of viewpoints because it is magnifying the black and white pattern directly beneath it. This magnified view constrains the orientation at which the lenslet is being viewed. To model this constraint, we consider a measurement of the intensity at the center of each lenslet, and experimentally measure the response across a set of viewpoints to create a response map. We explore thresholding the measured intensity at k intervals. The response map characterizes the apparent intensity of each lenslet when viewed from each orientation, and the state-map is the discretization of that response map in k states. Section 5 describes our measurement setup and Fig. 5 shows examples of the measured response map and discrete state maps.

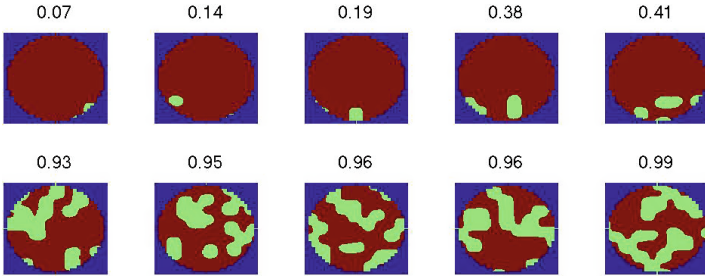


Fig. 2. The five lenslets whose discrete state map has the lowest (top row) and highest (bottom row) entropy. Green represents the views where the lenslet is in state 0, while red, state 1. The bottom lenslets are much more likely to be useful in estimating orientation; the median entropy of 90 lenslets in an array covering random dots is 0.8. (Color figure Online)

We characterize the value of a lenslet based on its entropy. The entropy H for a given lenslet and its map from viewpoint to state is:

$$H = - \sum_i^k p_i * \log_2(p_i) \tag{1}$$

where p_i is the frequency in the range (0, 1) of a lenslet being in state i for all viewpoints.

A lenslet which is in all states equiprobably has maximum entropy, and it should be easy to design a pattern to put underneath the microlens array which has this property. However, challenges in accurate printing, aligning a pattern to a lens array, and imperfectly manufactured microlenses that may be out of focus led us to use a random pattern. As a side effect, this lets us explore several questions. First, what is the entropy of a microlens array mounted on top of a random dot pattern?

Using the prototype array and texture pattern shown in Fig. 1 we calculate the entropy of each lenslet. Figure 2 shows the binary response maps and entropy measures for lenslets that are thresholded at an intensity of 100 (out of 255). We show the 5 lenslets with the smallest entropy in the first row, and the 5 lenslets with the highest entropy in the 2nd row. The median entropy for all lenslets is about 0.8 indicating that most lenslets give substantial useful constraints.

On the left of Fig. 3, we show the distribution of entropies for all 90 lenslets for optimal discretizations of 2, 3, 4, 5, and 6 states. On the right of Fig. 3, we show these optimal thresholds. This shows two interesting features. First, measurements of the microlens intensity are not binary and discrete states can be selected to give a discrete encoding that maximizes the (per lenslet) entropy. Second, when more than 2 discrete states are used, the thresholds often clump, highlighting the value of noticing when a lenslet is changing between black and white.

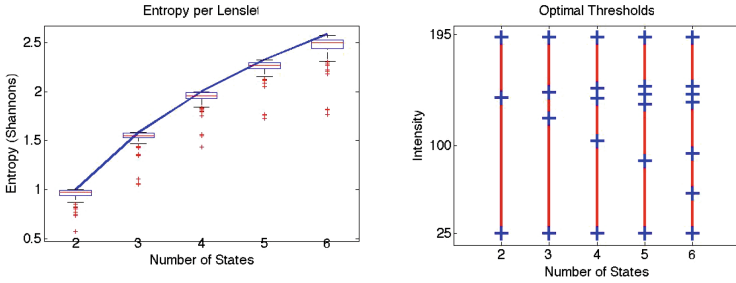


Fig. 3. We experimentally measure the entropy in discrete measurements from a microlens array covering a random pattern of small squares. Optimizing to choose discretization thresholds over all lenslets gives a distribution of the measured entropy that is close to the maximum entropy. On the right we show the optimized thresholds. This motivates the use of a random pattern as having nearly as much information as an optimal pattern.

5 Viewpoint Estimation with a Microlens Array

The state map can be used to characterize the discrete measurement for a lenslet viewed from a given viewpoint. We parameterize the viewpoint with the two-vector Θ which captures the direction of the camera rotated around the local x and y axes, and we define a_i to be the measured appearance of the lenslet. If a lenslet can exist in k states, then the lenslet appearance, a_i , is one of the first k positive integers.

We empirically determine the measurement a_i for each rotation Θ by imaging the microlens array for a grid of viewpoints with a DSLR camera. The microlens array is rotated with two programmatically controlled motors which can change the orientation of the microlens array to any rotation around the x and y axis. We scanned over the dome of viewpoints that are (30, -30) degrees from fronto parallel in 2 degree increments, yielding 677 images. We visualize the sampled viewing dome and show the motor setup in Fig. 4.

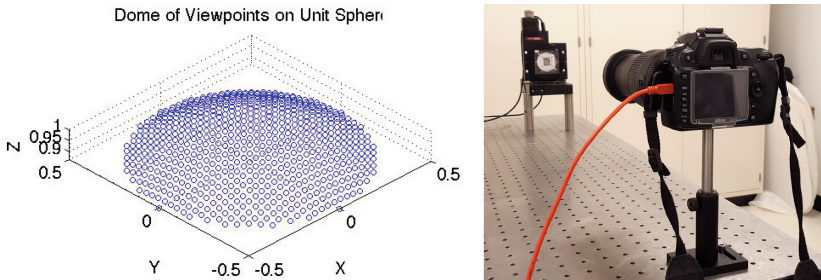


Fig. 4. The left shows the set of viewpoints from which we image the lenticular array, represented as points on a sphere. The right is a depiction of our image setup which includes a DSLR camera viewing the microlens array on a 2-axis motorized mount.

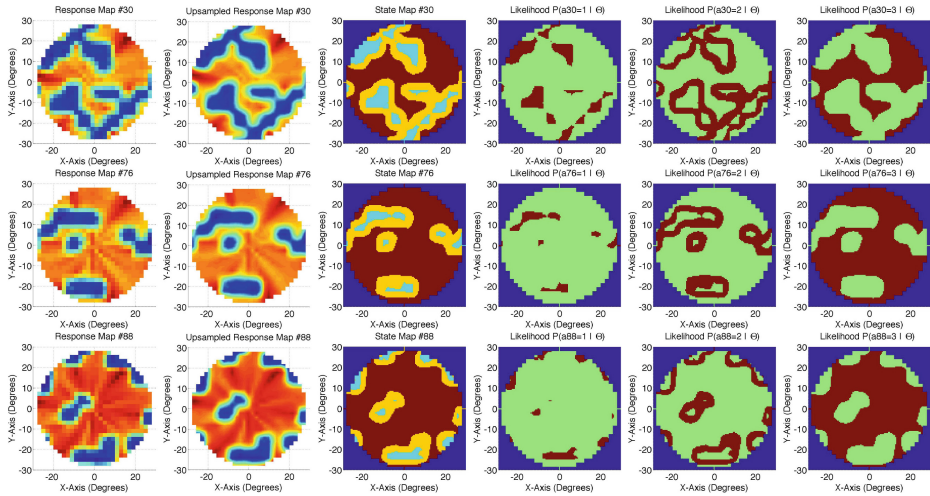


Fig. 5. For each lenslet, our calibration process characterizes the response as a function of viewpoint. Each row shows a different lenslet. From the left, the figure depicts first the raw, then the up-sampled response map. Third is the discrete state map where the response map is thresholded into one of three categories. The last three maps show the 0–1 likelihood function of the orientation as a function of each of the three possible discrete measurements.

For each image, four reference points on the corners of the microlens array are tracked and a homography is used to warp the image to a common coordinate system. Once warped, the center of each lenslet is sampled to create response maps for each lenslet. We show the response map and the responses thresholded into a trinary state map in Fig. 5.

We up-sample the response maps to have approximate measurements at every 0.5 degrees, using linear interpolation, then threshold to create the state map. The experimental section explores the performance gains for different amounts of up-sampling.

5.1 Inference

Using the microlens array, our task is to determine the 2D viewpoint orientation of camera in the reference of the microlens array. We seek to find the most probable viewpoint given the observed discrete appearance of all lenslets in the array. We employ a simple approach where black and white appearances at each lenslet vote for the most likely viewpoint.

Consider a lenslet at inference time with a state b_i . For all the Θ s in that lenslet’s statemap with state a_i that match b_i , we give one vote. Each lenslet will have a different statemap and will vote for a different set of Θ s.

To estimate the viewpoint using the entire microlens array, we choose the Θ with the maximum number of votes. In the case of ties, which normally happen

for similar Θ s, we take the average. The next section describes how to combine this with measurements of the corner points of the microlens array to determine the full pose of the microlens array.

6 Pose Estimation with Microlens Array

The pose estimation problem seeks to estimate the 3×3 rotation matrix, R , and 3×1 translation vector T , needed to transform the camera reference frame into the microlens array reference frame. As a result, R represents the direction of the surface normal of the microlens array relative to the camera's direction and T is the vector from the camera to the origin of the microlens array.

We consider the pose estimation problem for images whose geometry is defined by a pinhole camera model. Using the standard geometric framework, we assume the origin of the camera coordinate system is centered at the pinhole, and the camera calibration is known and represented by a 3×3 calibration matrix K . According to this model, a point in the camera reference frame P is thus projected to the image location p represented in homogeneous coordinates, via K :

$$p = KP \quad (2)$$

Therefore, a reference point Q in the reference frame of the microlens array is projected at the pixel location q' on the image according to the following linear projection:

$$q' = K(RQ + T) \quad (3)$$

Because the viewpoint estimate derived in the former section is in spherical coordinates, it only gives 2 of the 3 rotation parameters (because the appearance of a microlens is invariant to rotation around the line from it to the camera). Our approach is to use the θ and ϕ estimate of the rotation of the microlens around the x and y axis from before. We construct the complete 3×3 rotation matrix R parameterized by the unknown rotation around the z -axis ψ as:

$$R = R_z(\psi)R_x(\phi)R_z(\theta) \quad (4)$$

where $R_{x,z}(w)$ are the rotation matrices to rotate around the respective axes. We denote the transformation to determine R from θ, ϕ , and ψ as $R(\theta, \phi, \psi)$.

We solve for the rotation parameter ψ and the translation vector T by using the known pixel and local locations of the 4 reference points used to rectify images of the microlens array and θ and ϕ estimated previously. Using a non-linear optimization, we solve for these four parameters via reprojection error:

$$\min_{\psi, T} \sum_i^4 \|q_i - K(R(\theta, \phi, \psi)Q_i + T)\|_2^2 \quad (5)$$

where q_i is the measured pixel location of a reference point, θ and ϕ are known and held constant, Q_i is the location of the reference point in the local

reference frame, and $\|\cdot\|_2$ denotes the euclidean norm. After optimizing for ψ , we can recover the full R matrix using $R(\theta, \phi, \psi)$.

The estimated R and T fully capture the orientation and position of the microlens array relative to the camera. In the next section, we explore the performance of using a microlens array to estimation viewpoint and pose.

7 Experimental Design and Evaluation

We explore the performance of using microlens arrays for viewpoint and full pose estimation. First, we explore how different design and environmental factors affect viewpoint estimation. Second, we explore how the number of lenslet states and the number of lenslets affect viewpoint estimation. Finally, we show pose estimation experimental results and compare these to other recent work with fiducial markers whose appearance depends on the viewpoint.

7.1 Experimental Setup

For the experiments evaluating viewpoint estimation, we test on 88 images randomly sampled from the dome of viewpoints 30 degrees from fronto parallel. These images are taken with the microlens array on the programmable motorized stage, but with Θ values that are randomly selected, and not at the same location as calibration images, but known so that there is ground-truth to compare to. We use a microlens array with a 9×10 regular grid of lenslets, and unless otherwise stated, we use all constraints from all 90 lenslets for viewpoint estimation.

In assessing the accuracy of our viewpoint estimations, we calculate the angular difference between a vector in the direction of the estimated viewpoint direction and a vector in the direction of the true viewpoint. We show summary statistics for all 88 random views as a boxplot. Each boxplot shows the median error in red, a box that shows the range from the 25 percentile to 75 percentile. Red crosses depict outliers, and have values more than 2.7σ away from the mean, where σ is computed assuming that the data are normally distributed.

7.2 Design and Environmental Factors

We first characterize the effect of design and environmental factors on viewpoint estimation. These factors affect estimation regardless of the number of states the lenslets can occupy. Therefore, to simplify this first set of experimentation, we employ binary (2-state) measurements and threshold at an intensity value of 100. This threshold reflects general observations that that “white” lenslet have intensities above 120 and “black” lenslets had intensities below 50.

We first look at performance gains from up-sampling the response map. Second, we explore the effect of the number of lenslets on viewpoint estimation. Finally, we test the microlens arrays in varying lighting environments.

Response Map Precision. We build response maps for each lenslet by sampling rotations of the microlens array in 2 degree increments. How much can we improve our rotation estimates by up-sampling these response maps? We create increasingly more up-sampled response maps and assess their orientation estimation performance. Figure 6 shows that at the initial resolution, the microlens array constrains the orientation to a median error of 0.7 degrees. Increasing the precision of the response maps to 0.5 degrees (upsampling and interpolating the response maps by a factor of 4) gives a substantial improvement, and further up-sampling the response maps has little additional benefit. We use this precision for the rest of the paper.

Number of Lenslets. Section 3 discusses the potential of combinatorial encodings of orientation, with the claim that in an ideal case 14 binary lenslets is sufficient to uniquely encode $\frac{1}{2}$ degree increments in a viewing dome of 30 degrees. This section gives an experimental evaluation of the correlation between estimation performance and number of lenslets. In this experiment, we randomly select k number of lenslets, and then perform orientation estimation. Figure 6 shows results using 20–90 lenslets in increments of 10 on the left, as well as a finer grain analysis with 10–20 lenslets in increments of 1 on the right. With 20 of the total 90 lenslets, we achieve orientation estimation accuracy with a median error below 1 degree. With fewer lenslets, the performance degrades. With less than 14 lenslets, the maximum error surpasses 10 degrees of error, anecdotally we see that sometimes in these cases there are far apart viewpoints that have very similar discrete encodings. To achieve a median viewpoint estimation error of 0.5 degrees, about 30 randomly chosen lenslets are necessary.

Light Environments. Related work uses hue to encode orientation with lenticular arrays [10], and one motivation of this work is use discrete measurements of black and white patterns to avoid problems that arise with varying lighting environments. We test the sensitivity of binary microlens arrays to different lighting environments and exposure settings by exploring 3 different lighting conditions. The first lighting environment is inside under overhead lights. This is the lighting environment used to generate the response maps for all lenslets and is used for all other of the experiments in this paper. The second lighting environment is similar to the first, but with an additional, strong white directional light. The third lighting environment has the scene lit entirely by 2 blue directional lights. In Fig. 7, we show the estimation results of these 3 different lighting environments. The common office environment with overhead lights achieves the lowest error with a median error of 0.3 degrees. However, even with the very extreme lighting environment of only blue directional light, the microlens array is able to estimation orientation with a median error of 0.6 degrees. This experiment suggests that even extreme lighting environments have a minimal effect on the binary encoding of orientation giving by the microlens arrays.

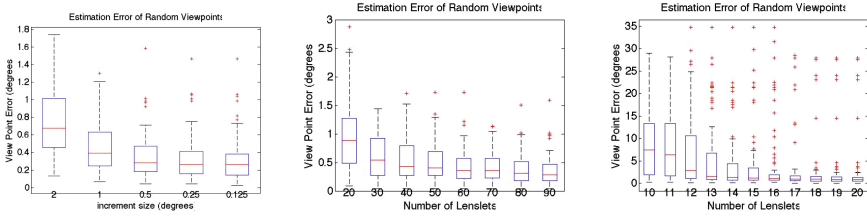


Fig. 6. (Left) Orientation estimation accuracy, as a function of the angular spacing of the response maps. The original measurements have an angular spacing of 2 degrees; all other data is based on up-sampling and interpolating this response map. (Center and Right) As fewer lenslet provide orientation cues, estimation accuracy goes down.

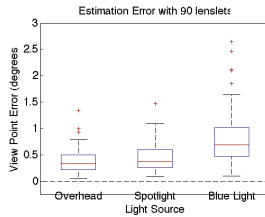


Fig. 7. Orientation estimation error as a function of lighting environment. Overhead lights match the calibration environment, and adding a spotlight light source has minimal impact. Lighting the scene with strong blue lights increases the median estimation error by about half a degree.

7.3 Measurement Discretization and Lenslet Selection

Here we explore choices driven by the entropy in the discrete state space for each lenslet. First we experiment with the number of discretized states using all lenslets, and then we determine the effect of using the most informative lenslets.

Number of States. In Sect. 4, we showed that individual lenslets with more states had a larger maximum entropy, and that by optimizing the threshold value allows a random texture to create appearances that have entropies approaching the maximum. In this section, we validate whether there is a corresponding improvement in viewpoint estimation by optimizing for high entropy. From the same response maps, we create state maps with 2,3,4,5, and 6 states. We use the same optimal thresholds for these state maps as described in Fig. 2. We also create a second binary statemap, but using a threshold of 100, as used in Sect. 7.2. To differentiate between the two binary conditions, we label one “2_opt” to indicate use of a threshold that maximizes entropy. With each choice of our discretization we get different state maps, and we use these to estimate the viewpoint. We report error versus the known true viewpoint.

The results of using all 90 lenslets are shown on the left in Fig. 8. In all cases, the median error is less than 0.5 degrees. There is no discernible trend, except

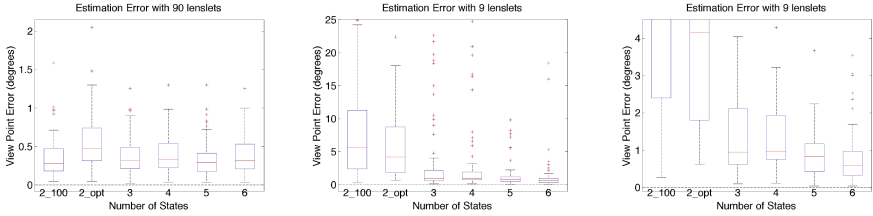


Fig. 8. The effect of the number and choice of lenslets on orientation estimation. Left, using all 90 lenslets, the median orientation error is less than 0.5 degrees across all discretization choices. The middle and right show the same plot at two scales. They highlight that using the 9 lenslets with highest entropy gives plausible results, and when using fewer lenslets it is especially important to go beyond a binary classification.

that a binary threshold at the entropy maximizing cutoff is worse than a default threshold of 100, perhaps because 100 was hand-chosen to be as far as possible from the appearance of completely white and completely black lenslets.

Using Minimal Lenslets. With 90 lenslets, there is a wealth of information. In order to test the limits of a the number of lenslets, we tested the viewpoint estimation with the same conditions of the previous section, but only using 9 lenslets. The 9 lenslets used were the lenslets with the highest entropy state maps. The middle of Fig. 8 shows these results. With fewer lenslets, and less information, the trend suggests that more states results in better estimates. In addition, using an optimal threshold with 2 states shows advantages with few lenslets. The left of Fig. 8 shows results that magnifies the y axis in order to better analyze estimation results for state maps with more than 2 states. Having more than 2 states results in median errors less than 1 degree. A statemap with 6 states results in the best performance, with 0.6 degrees of median error. These empirical results corroborate that, when a limited number of lenslets can be used, using more states results in better viewpoint estimation performance.

7.4 Pose Estimation

The unique appearance of a microlens array results in low error viewpoint estimations. It is therefore useful to use microlens arrays for pose estimation as well. In the next section, we determine the poses of a new set of images. In these images, we place the microlens array ≈ 1 m away and rotate it around the vertical axis from fronto-parallel to 30 degrees in 1 degree increments. We compare pose estimation to the standard fiducial marker, ARToolkit. After, we use these results to compare against other published results of a similar experiment.

Direct Comparison to ARToolkit. In Fig. 9, we show pose estimation results of using the microlens array. As a comparison, we also include results of using just the 4 reference points as used by ARToolkit [14] to estimate pose. In the

left 2 plots of Fig. 9, we show the rotation estimation errors per axis using the microlens array (leftmost) and using just 4 reference points (left-center). The rotation error is defined as the angular difference between the unit axes using the true and estimated rotations. The translation error is defined as the euclidean distance between the true and estimated position. For each plot, the title shows the median errors over all frames for the axes in x, y, z order.

For all views, the microlens array is able to determine rotations accurately. In contrast, the standard “4 corner method” suffers from the well understood ambiguity of points on a fronto-parallel plane [1–3]. Since the microlens array gives orientation cues directly from the viewpoint estimation, our method does not suffer from this ambiguity.

In the right 2 plots of Fig. 9, we show the translation estimation errors per axis using the microlens array (right-center) and using just 4 reference points (rightmost). For both methods, there is some systematic error in the Z-axis while the X and Y axes have almost no error. For the ARToolkit method, the error could be a consequence of error in the rotation estimation. The images were taken at a long focal length (≈ 300 mm), so the systematic errors may also be a consequence of errors in calibrating the K matrix.

Indirect Comparison Against Related Work. In the final experiment, we share a comparison of our rotation estimation results from above with the results reported in work that uses “Chromo-coded Markers” [10], the “Lentimark” [6], and the “Arraymark” [8]. All three papers perform a similar rotation estimation experiment that rotates the object around a single axis. In addition, all three papers also report the rotation error for the x,y, and z local axes. Table 1 summaries results reported for the “Chromo-coded Markers”, “Lentimark”, and “Arraymark” against our experiments done above in this section. The previous works and our method have comparably low rotation errors. However, our microlens array has slightly superior rotation estimation error for estimating the z axes.

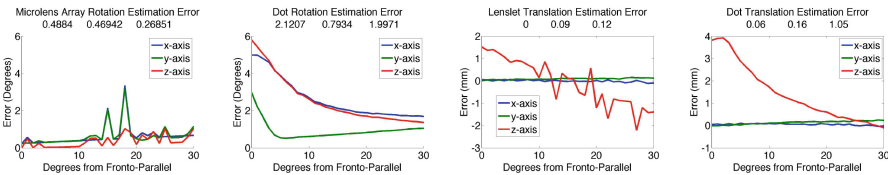


Fig. 9. We compare the rotation and translation estimation accuracy of our result with the method employed by the popular ARToolkit. In contrast to most position based fiducial markers like those used with the ARToolkit, our fiducial marker does not suffer from the well-known ambiguities of points that lie on a plane near fronto-normal

Table 1. The microlens array has similar or better accuracy than other work based on fiducial markers whose appearance is viewpoint dependent.

	Lentimark	Chromo-coded marker	Arraymark	Microlens array
x	0.372	0.910	0.63	0.488
y	1.359	0.703	0.47	0.469
z	0.324	1.222	0.61	0.269

8 Conclusion

We present a novel type of fiducial marker whose appearance is designed to give a combinatorial encoding of its orientation. This combinatorial encoding has the advantage over previous approaches that it can take on different form factors (including those where lenslets are disjointly spread across a plane), and it is insensitive to variations in the color of the scene lighting. We derive an approach to solve for the rotation of this new micro-lens based fiducial marker show that it improves pose estimation when combined with standard location based fiducials.

References

1. Uematsu, Y., Saito, H.: Improvement of accuracy for 2d marker-based tracking using particle filter. In: 17th International Conference on Artificial Reality and Telexistence, pp. 183–189. IEEE (2007)
2. Abawi, D.F., Bienwald, J., Dorner, R.: Accuracy in optical tracking with fiducial markers: an accuracy function for artoolkit. In: Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 260–261. IEEE Computer Society (2004)
3. Pentenrieder, K., Meier, P., Klinker, G., et al.: Analysis of tracking accuracy for single-camera square-marker-based tracking. In: Proc. Dritter Workshop Virtuelle und Erweiterte Realitt der GIFachgruppe VR/AR, Koblenz, Germany (2006)
4. Haralick, R.M., Shapiro, L.G.: Computer and Robot Vision. Addison-Wesley Longman Publishing Co., Inc., Boston (1991)
5. Fitzpatrick, J.M., West, J.B.: The distribution of target registration error in rigid-body point-based registration. IEEE Trans. Med. Imaging **20**(9), 917–927 (2001)
6. Tanaka, H., Sumi, Y., Matsumoto, Y.: A visual marker for precise pose estimation based on lenticular lenses. In: IEEE International Conference on Robotics and Automation (2012)
7. Tanaka, H., Sumi, Y., Matsumoto, Y.: A solution to pose ambiguity of visual markers using moire patterns. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (2014)
8. Tanaka, H., Sumi, Y., Matsumoto, Y.: A high-accuracy visual marker based on a microlens array. In: IEEE/RSJ International Conference on Intelligent Robotics and Systems (2012)
9. Tanaka, H., Sumi, Y., Matsumoto, Y.: Further stabilization of a microlens-array-based fiducial marker. In: IEEE International Symposium on Mixed and Augmented Reality (2013)

10. Schillebeeckx, I., Little, J., Kelly, B., Pless, R.: The geometry of colorful, lenticular fiducial markers. In: International Conference on 3D Vision (2015)
11. Schillebeeckx, I., Pless, R.: Structured light field design for correspondence free rotation estimation. In: International Conference on Computational Photography (ICCP) (2015)
12. Olson, E.: AprilTag: a robust and flexible visual fiducial system. In: 2011 IEEE International Conference on Robotics and Automation (ICRA), pp. 3400–3407. IEEE (2011)
13. Fiala, M.: Artag, a fiducial marker system using digital techniques. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2005)
14. Kato, H., Billingham, M.: Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In: IEEE and ACM International Workshop on Augmented Reality (1999)
15. Zhang, Z.: Flexible camera calibration by viewing a plane from unknown orientations. In: Proceedings of IEEE International Conference on Computer Vision, pp. 666–673 (1999)
16. Liu, X., Cevikalp, H., Fitzpatrick, J.M.: Marker orientation in fiducial registration. In: 2003 International Society for Optics and Photonics on Medical Imaging, pp. 1176–1185 (2003)
17. Bansal, M., Daniilidis, K.: Geometric urban geo-localization. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3978–3985. IEEE (2014)
18. Lippmann, G.: La photographie integrale. *Comptes-Rendus* (1908)
19. Bruckstein, A.M., Holt, R.J., Huang, T.S., Netravali, A.N.: New devices for 3d pose estimation: mantis eyes, agam paintings, sundials, and other space fiducials. *Int. J. Comput. Vis.* **39**(2), 131–139 (2000)
20. Wetzstein, G., Roodnick, D., Heidrich, W., Raskar, R.: Refractive shape from light field distortion. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 1180–1186. IEEE (2011)
21. Wetzstein, G., Heidrich, W., Raskar, R.: Computational schlieren photography with light field probes. *Int. J. Comput. Vis.* **110**(2), 113–127 (2014)
22. Ji, Y., Ye, K., Yu, J.: Reconstructing gas flows using light-path approximation. In: International Conference on Computational Photography (ICCP) (2013)
23. Yang, W., Ji, Y., Lin, H., Yang, Y., Bing Kang, S., Yu, J.: Ambient occlusion via compressive visibility estimation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, June 2015
24. Mohan, A., Woo, G., Hiura, S., Smithwick, Q., Raskar, R.: Bokode: imperceptible visual tags for camera based interaction from a distance. *ACM Trans. Graph. (TOG)* **28**(3), 98 (2009)