The Geometry of Colorful, Lenticular Fiducial Markers

Anonymous 3DV submission

Paper ID 12

Abstract

Understanding the pose of an object is fundamental to a variety of visual tasks, from trajectory estimation of UAVs to object tracking for augmented reality. Fiducial markers are visual targets designed to simplify this process by being easy to detect, recognize, and track. They are often based on features that are partially invariant to lighting, pose and scale. Here we explore the opposite approach and design passive calibration patterns that explicitly change appearance as a function of pose. We propose a new, simple fiducial marker made with a lenticular array, which changes its apparent color based on the angle at which it is viewed. This allows full six degree-of-freedom pose estimation with just two markers and an optimization that fully decouples the estimate of rotation from translation. We derive the geometric constraints that these fiducial markers provide, and show improved pose estimation performance over standard markers through experiments with a physical prototype for form factors that are not well supported by standard markers (such as long skinny objects). In addition, we experimentally evaluate heuristics and optimizations that give robustness to real-world lighting variations.

1. Introduction

Many visual applications are based on tracking the pose of an object relative to a camera. Often, fiducial markers are attached to those objects to simplify this pose estimation process. These markers are designed to be easy to detect and to track, for example small spheres that look similar from any viewpoint. In this paper, we consider an alternative in fiducial marker design, creating markers whose colors change based on their relative pose to the camera.

There are many possibilities for non-lambertian materials that might support this process, from active electronics to holographic materials, but we would like to select for three properties. First, we would like something easy to create from available materials. Second, we would like a passive marker that does not require power. Third, we would like a clear geometric interpretation to the appearance.

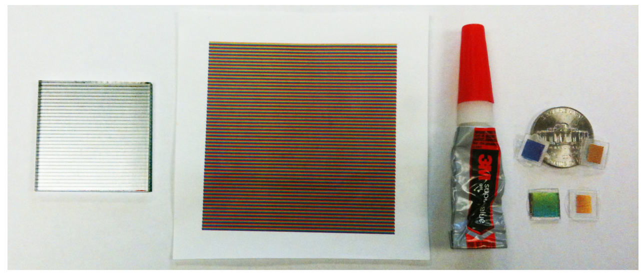


Figure 1. Gluing a lenticular plastic sheet (left) to a color pattern (center) from a standard color printer is a convenient source of fiducial markers (right) whose apparent color has a clear geometric relationship to the relative orientation to the camera. Several small markers are shown leaning against a nickel and they appear to be different colors because they are oriented differently relative to the camera. This paper discusses the design of these markers and how to use them as fiducial markers for pose estimation.

Lenticular sheets are available in hobby shops and often used to create children's toys, bookmarks, or promotional material that show an animated pattern as they turn. The optical properties of the lenticular plastic use ambient light to show different patterns when viewed from different directions. In this paper we describe how to use lenticular printing to create fiducial markers. Figure 1 illustrates the process. We design a pattern that can be printed on standard color laser printer. Adhering this to the back of a plastic lenticular sheet creates small markers whose color relates to their orientation. The contributions of this paper are:

1. The description and evaluation of a cheap way to create fiducial markers that explicitly change their apparent color based on their orientation relative to the camera.
2. The derivation of geometric constraints relating the location and color of lenticular fiducial markers to object pose.
3. Experimental demonstration of pose estimation with these fiducials showing them to be less sensitive to noise than a standard calibration pattern made of four dots arranged in a rectangle.
4. Optimizations to improve the performance of these

color varying fiducial markers under varied lighting conditions.

Current fiducial markers are based on points that are spatially spread apart in a plane or in 3D. One important application domain of colored fiducial markers is in creating pose estimation for skinny, elongated objects. In these cases, the spatial separation required to solve for pose from current fiducial markers is difficult to obtain, and results are very sensitive to noise. Because colored fiducial markers do not have this limitation, they could support novel applications of augmented reality or item tracking, especially those requiring estimates of the pose of thin objects like pencils or scalpels.

2. Related Work

Most current approaches rely on fiducial markers that have known coordinates relative to the object they are on. They use the apparent position of those markers in the image to solve for the relative pose of the object and the camera. For example, Zhang's widely used camera calibration method [24] is based on a large black-and-white checkerboard pattern, whose image is easy to interpret. Standard augmented reality libraries such as ARtoolkit [9], ARTag [5, 6] and CALTag [1] define black and white patterns designed to simplify the process of determining which fiducial marker is which, and have patterns with strong edges and corners so that line-fitting approaches can give sub-pixel image coordinates. Research in fiducial marker design also seeks markers that are easy to find at multiple scales [20, 7], or are designed to be imperceptible to humans [12, 23].

Solving from the pose estimation based on cues extend beyond the observed position of fiducial markers. In situations where the vertical direction is known for the camera (e.g. from an Inertial Measurement Unit), it is possible to solve for the pose of a camera with 2 corresponding points [10, 14]. Recent work also derives a two point solution if the observed points in the world have a known direction that projects to the image (e.g. building corners where the vertical edge is visible), and characterizes the degeneracies of these constraints [2].

A few prior works have explicitly created fiducial markers whose relative appearance depends on the direction from which they are viewed. Agam fiducials [4], take advantage of properties of 3D stairstep like structures where the vertical part of all steps are painted a different color than the horizontal parts of each step. When viewed from afar, the darkness of this pattern relates to the angle at which the stairstep is viewed. A recent patent proposes using small lenticular markers that vary in color based on the orientation from which they are viewed for the purpose of pose-estimation [11]. However, they do not describe the necessary constraints to determine pose in practice. More recently BoKodes [12] created a highly structured pattern of light projected away from one point in a scene. This structured pattern is based on thousands of small QR-codes. When a defocussed camera takes a picture of a scene, this pattern is visible, and the identity of QR-codes in view indicates the relative direction of the camera relative to the bokode marker.

Lenticular arrays and their 2d counterpart, microlens arrays, have been used previously to geometrically constrain the relative orientation of an object or the relative path of a ray of light. Previous research has used microlens arrays as light field probes for Schlieren photography [21], for the reconstruction of surface geometry of transparent objects [22], and for the reconstruction of the refractive index of gases [8]. Lenticular arrays have been created to change color with changes in orientation and were used to estimate object rotation [15] with correspondences. Other research has used lenticular arrays to augment existing fiducial markers [17, 19] and microlens arrays as a fiducial marker [16, 18]. These fiducial markers, called LentiMark and ArrayMark, change appearance base on orientation. This appearance change manifests as a relative change in location of a mark that translates in reference to the rest of the fiducial marker.

We propose a set of markers made from lenticular arrays, so that the apparent color of the marker depends on their relative orientation and position to the camera.

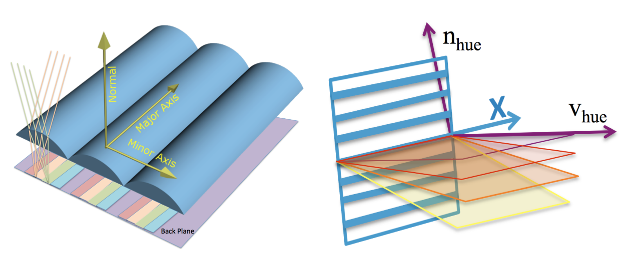
3. The Geometry of Lenticular Arrays

Lenticular arrays are usually made of plastic and have a flat-surface (the back-plane) and a front surface that is comprised of parallel cylindrical surfaces, called lenticules. The major axis of the lenticular array runs the length of the lenticules and gives the orientation of the array. The lenticules act as a partial lens focusing parallel planes of light onto the back-plane.

3.1. Viewing Angle and Hue Response

When observed from a particular direction, all the light leaving the lenticular array in a particular direction comes from a thin strip behind each lenticule. Therefore, if the backplane is made up of multiple colors that are interleaved behind each lenticule, rotating the lenticular array results in the perception of different colors. This lenticular geometry is depicted in Figure 2(a). As described in [15], we create lenticular markers by adhering blank lenticular arrays to interleaved hues and then cutting the arrays into smaller markers.

As shown in [15], the apparent color of a lenticular patch smoothly varies depending almost exclusively on the viewing direction rotated around the lenticular array's major axis. This relationship is referred to as the Hue Response



(a) (b)

Figure 2. (a) The top surface of a lenticular array is comprised of parallel cylindrical parts that focus parallel rays onto particular rows of a backplane. (b) Relative to the major axis x, any view v_{hue} of the lenticular array with a particular hue, lie on a plane denoted by its surface normal n_{hue}.

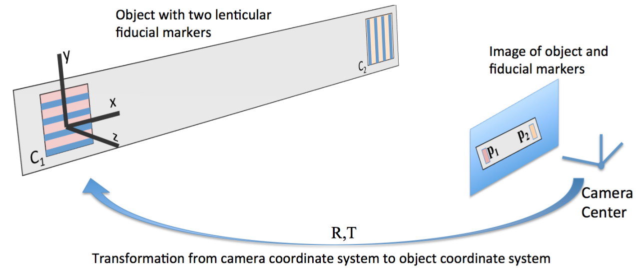


Figure 3. A pinhole camera (right) sees an image of an example flat object (left), which has two lenticular patches. The pose estimation problem asks to solve for the R, T describing the rotation and translation mapping the coordinate system of the camera to the coordinate system of the object.

Function (HRF). Relative to the major axis, any view, $\vec{v}_{hue}$, of the lenticular array with a particular hue, lies on a plane denoted by its surface normal $\vec{n}_{hue}$. Different viewed colors have differently oriented $\vec{n}_{hue}$. This relationship is depicted in Figure 2(b) for a lenticular array with a major axis $\vec{x}$ along an object's x-axis.

3.2. Lenticular Arrays as a Fiducial Marker

Given this strong relationship between hue and viewing angle, we derive a formulation for pose estimation based on observing fiducial markers composed of small lenticular arrays. Information about the pose of an object comes from both the position and the hue of the observed lenticular fiducial marker. Many configurations of lenticular fiducial markers are possible and interesting, but we first derive the constraints for two markers attached to the same plane. Because the hue of a marker changes primarily due to rotations around the major axis of the lenticular array, we orient the major axes of our lenticular markers to be perpendicular to each other so that they provide complementary information about the surface normal. The derivations of pose constraints for other configurations of two lenticular patches are similar, and we explore using more than 2 lenticular arrays to improve pose estimation accuracy.

Figure 3 shows an example of a skinny planar object with one lenticular fiducial at location C_1, and a coplanar lenticular fiducial at location C_2 which is oriented orthogonally.

4. Minimally Constrained Pose Estimation

Figure 3 also illustrates our pose estimation problem. A pinhole camera (on the right) observes an object (on the left), which contains two lenticular patches.

For simplicity, we assume that the coordinate system of the object is centered at point C_1 so that, in the coordinate system of the object, the first lenticular fiducial marker has coordinate (0,0,0). We assume the second lenticular patch is a distance d away in the direction of the x-axis of the object so C_2 has object coordinates (d,0,0). These lenticular patches are not just points, they also have an orientation, and this orientation affects the apparent color of the patch. In Figure 3, the patch at C_1 has its major axis along the x-axis of the object and the patch at C_2 is oriented along the y-axis of the object.

The pose estimation problem is to solve for the rotation matrix R and translation vector T mapping a point in the object coordinate system into a point in the camera coordinate system. In our case, the lenticular marker at C_1 in the object coordinate system will move to $RC_1 + T$, and C_2 will become $RC_2 + T$.

For lenticular patches, the question is to solve for the R, T that is consistent with p_1 being the image of $RC_1 + T$ and p_2 as the image of $RC_2 + T$, *and* both p_1 and p_2 have the correct hue for the angle at which they are being viewed. We formalize these constraints in the next section.

4.1. Pose constraints from two lenticular patches

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 on the camera, and the camera calibration is known and represented by a camera calibration matrix K. Then, if a pixel p is represented in homogeneous coordinates, it views an object that appears along a ray in space $\vec{r}$ that is defined as:

$$\vec{r} = K^{-1}p \quad (1)$$

Therefore, if we see lenticular fiducial markers at locations p_1 and p_2, then the fiducial markers must lie along rays $\vec{r_1} = K^{-1}p_1$ and $\vec{r_2} = K^{-1}p_2$.

As described in Section 3.1, the image of the lenticular fiducial marker may have a different hue depending on the relative angle between the incident ray and the orientation of the marker. If the lenticular patch appears red, for example, the incident ray must lie in the plane denoted $\vec{n}_{hue}$ that has the particular orientation around the major axis where the lenticular marker appears red. $\vec{n}_{hue}$ is the cross product of the lenticular array's major axis and the viewing direction $\vec{v}_{hue}$.

Rotational Constraints

The first constraint on the rotation of the object comes from the apparent color of the first fiducial marker. The ray $\vec{r_1}$ from the camera that observes that marker must lie in the plane $\vec{n}_{hue}$ defined by the hue that the camera observes, and therefore $\vec{r_1}$ must be perpendicular to $\vec{n}_{hue}$. In the coordinate system of the object, $\vec{n}_{hue}$ is defined by the two vectors that span it: $\vec{x} \times \vec{v}_{hue}$. In practice, $\vec{n}_{hue}$ is determined via the HRF. The rotation R changes that surface normal to become $R(\vec{x} \times \vec{v}_{hue})$. So our first constraint on the rotation is that the viewing direction and the surface normal are perpendicular:

$$R(\vec{x} \times \vec{v}_{hue_1}) \cdot \vec{r}_1 = 0 \tag{2}$$

A similar constraint applies when the camera observes the other lenticular marker, except that the viewing direction is r_2, the second lenticular marker is aligned along the $\vec{y}$-axis of the object, and the observed color may be different. This leads to a second constraint:

$$R(\vec{y} \times \vec{v}_{hue_2}) \cdot \vec{r}_2 = 0 \tag{3}$$

A third constraint relies solely on the locations of the observed lenticular patches. Specifically, three rays are coplanar: the direction from the camera to the first patch $\vec{r}_1$, the direction from the camera to the second patch $\vec{r}_2$, and the displacement vector between the first and second patch $(RC_2 + T) - (RC_1 + T) = R(C_2 - C_1)$. This also leads to a geometric constraint:

$$R(C_2 - C_1) \cdot (\vec{r}_1 \times \vec{r}_2) = 0, \tag{4}$$

If the coordinate system of the object is such that C_1 is the origin, this simplifies to:

$$(RC_2) \cdot (\vec{r}_1 \times \vec{r}_2) = 0, \tag{5}$$

This provides three constraints on the rotation matrix that are independent of the estimation of the translation vector.

Translational Constraints

Once the rotation is known, we can derive three linear constraints on the translation. The first constraint is that the translation must be consistent with the observed location of the first lenticular marker. Since any point C in the object coordinate system is mapped to a location $RC + T$, the ray $\vec{r}$ viewing that point must be parallel to the vector from the origin to $RC + T$. Because we define the first lenticular marker at location C_1 to have object coordinates (0,0,0), $RC_1 + T$ simplifies to just T, and we can use the fact that the cross-product of two parallel vectors is zero to express the constraint as:

$$T \times \vec{r}_1 = \vec{0}. \tag{6}$$

The final constraint needed to estimate the translation is similar to 6, but uses the projection of the second lenticular marker:

$$(RC_2 + T) \times \vec{r}_2 = \vec{0} \tag{7}$$

These two constraints on the translation are both vector equations, so collectively these 5 constraints let us solve the pose estimation problem in Figure 3.

4.2. Optimization of Rotation and Translation

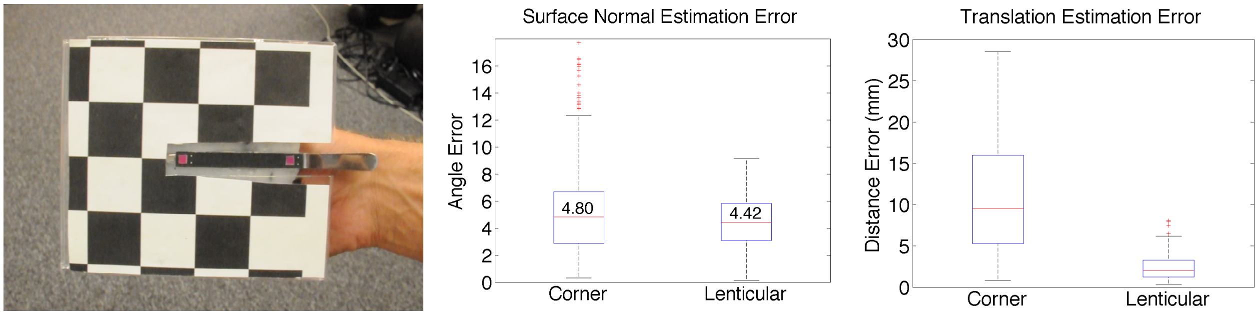
There are several possibilities to solve for the R, T using this set of constraints. In our experiments we first obtained the rotation matrix on its own by optimizing a non-linear error function over the Rodrigues vectors $\vec{\rho}$ that define a rotation matrix $R\rho$. The error function is the sum of the squared error for Equations 2, 3, and 5. We use $fminunc$ in matlab and initialize with the identity matrix.

After we have the rotation matrix R, equations 6 and 7 define a linear system of equations for T. Both error functions return solutions with zero error because they are optimizing a minimal set of constraints, as we are measuring 6 numbers (two coordinates and one hue for each of the 2 markers) to solve the 6 DOF pose estimation problem. Understandingthe configurations that lead to degenerate cases with multiple pose estimates giving zero error is an interesting problem we leave for future work. Later in this paper, we will extend this optimization routine for $n > 2$ lenticular arrays to improve pose estimation results and add robustness to varying ambient light conditions.

5. Experimental Results with Two Lenticular Markers

In this section we compare end-to-end performance of a pose estimation procedure, comparing standard fiducial markers (small points at a known location on the object) and our new lenticular fiducial markers. We build a physical prototype of the lenticular fiducial marker to estimate the pose of an object throughout a video. For the experiment, we compare accuracy for an object that is long and skinny.

We track a small pair of tweezers by placing lenticular markers 55mm apart. These lenticular markers are made from the lenticular arrays detailed in [15]. Each fiducial marker is a 4mm square, comprising about 12 lenticules. For comparison, there are also 4 small fiducial markers in a rectangular pattern that are also 55mm apart in one direction, and 4 mm apart in the orthogonal direction. We determine the pose of these "corner" markers following the ARToolkit algorithms [9]. Additionally, the tweezers are



(a) (b) (c)

Figure 4. We compared lenticular fiducial markers with the state of the art "corner" method on a video (of which an example frame is shown in a). For the length of the video, the pose estimation using lenticular markers produces a better normal direction estimation (b) and position estimation (c) median with a much more compact distribution. Each box in b) contains the median value in text.

mounted to a large checkerboard so that we can use a standard toolbox [3], to give ground truth pose estimates per frame. Figure 4(a) shows this setup in a frame of the experimental video.

We calibrate our lenticular markers and compute the HRF — and the inverse HRF that maps a measured hue to a θ rotated around the major axis of a lenticular marker — through a lookup-table from the experimental measurements in a calibration video using the same setup. For each frame, we hand label the location of the lenticular patches and corners. All frames that had substantial motion blur, making fiducial dots difficult to label, were discarded. For the lenticular patches, we click each corner of the small array, and use the centroid as its location. The color of the entire area of the small lenticular patch is averaged to get the two hues needed for the lenticular marker. To get a subpixel accuracy estimate for the standard fiducial dots, we fit a Gaussian distribution over the small dot and use the mean as the dot location.

Figure 4 shows the pose estimation results for the rotation error, defined as the coplanar angular difference between the ground truth and estimated Z-axes, and translation error, defined as the euclidean distance of true position and estimated position. We show the results as boxplots with default Matlab parameters. The top and bottom blue lines of the box indicate 1st and 3rd quartiles, while the red line in the center of the box indicates the median. The whiskers indicate the minimum and maximum value considered an inlier. Outliers (seen as red crosses) lie outside 1.5 times the distance between the 1st and 3rd quantiles beyond the nearest quantile. For a normal distribution, this means outliers are outside 2.7 sigma. For Figure 4(b), the median error is written in each fiducial marker box. The median rotation (Figures 4(b)) and translation (Figures 4(c)) errors of pose estimation using the lenticular fiducial markers are less than that of the state of the art corner method. In addition, the distribution of errors is much tighter and indicates a more stable pose estimation. We find that this is due to the fact that our system is more robust to fiducial marker position noise. In Appendix C, we show pose estimation results of simulated data, where we can control noise level.

6. Over Constrained Pose Estimation

The previous experiment compared pose estimation using 2 lenticular markers against 4 standard fiducial markers. Our pose estimation method minimally constrains the 6 degrees of freedom of the pose estimation problem with the 2 hue and 4 position measurements of the lenticular markers, while the standard fiducial marker over constraint with 8 position measurements [9].

In this section, we now build on our algorithm to use more than 2 lenticular markers to improve pose estimation results. First, we show how to optimize over existing rotation and translation constraints, but with $n > 2$ markers. Second, we introduce a reprojection error refinement. In the framework of this reprojection optimization, we later introduce 2 additional color scaling variables to add robustness to different lighting environments.

6.1. Adding Additional Constraints Per Marker

A common approach to improve fiducial marker pose estimation results is to increase the number of fiducials and optimize the over constrained system of equations. As an example, only 3 corners of a square are needed to solve for pose, however, optimizing over many corners of a checkerboard drastically improves pose estimation accuracy.

Inspired by this, we show how to optimize over the arithmetic constraints presented in Section 4 for more than 2 lenticular markers. We generalize these constraints in Appendix B so that each ith marker has a local position C_i, local orientation $\vec{o}_i$, direction from camera $\vec{r}_i$, and respective direction of hue $\vec{v}^{i}_{hue}$. In the generalized form, each marker has 2 constraints on rotation, 1 using hue and 1 using relative position to another marker, and 1 constraint per marker on translation. Thus, similar to with 2 lenticular markers, we can first optimize for the rotation over all markers with the following objective function:

$$\underset{R}{\operatorname{argmin}} \sum_{i}^{n} \Big(\big(R(\vec{o}_i \times \vec{v}^{i}_{hue}) \cdot \vec{r}_i\big)^2 + \big(R(C_i - C_{i-1}) \cdot (\vec{r}_{i-1} \times \vec{r}_i)\big)^2 \Big) \quad (8)$$

and then optimize for the translation:

$$\underset{T}{\operatorname{argmin}} \sum_{i}^{n} \big((RC_i + T) \times \vec{r}_i \big)^2 \quad (9)$$

6.2. Reprojection Refinement

After using the arithmetic error to optimize for $n > 2$ lenticular markers, we could optimize over the reprojection error of all markers to refine pose estimation. Here we show the object function to simultaneously optimize for R and T over the reprojection error of hue appreance and image position:

$$\underset{R,T}{\operatorname{argmin}} \sum_{i}^{n} \frac{1}{\lambda_1} \big(h(R,T) - hue_i\big)^2 + \frac{1}{\lambda_2} \| g(R,T) - p_i \|_2^2 \quad (10)$$

where $h(R,T)$ projects a hue according to the pre-calibrated HRF given for a lenticular maker at the projected image location $g(R,T)$. Normalizing values λ_1 and λ_2 are found empirically to balance hue and position cost.

Color Calibration

The appearance given by lenticular markers can be affected by environmental lighting factors. Ambient light color may affect the hue appearance of the lenticular marker at any orientation, while glare may only change the hue appearance for certain orientations. As a result, the pre-calibrated hue/viewpoint relationship of the HRF which encodes $\vec{n}_{hue}$ does not represent the observed images and orientations in a different lighting environment. The extra information given by $n > 2$ lenticular markers, however, makes it possible to correct for these uncalibrated lighting environments.

Inspired by white balancing methods, we therefore introduce 2 extra color scaling variables to the objective function of Equation 10. These two variables, s_r and s_b, scale the red and blue channels of the RGB measurement of lenticular markers to better fit the model of how lenticular markers appear given a relative orientation. The purpose of s_r and s_b is to color correct lenticular marker observations to fit the pre-calibrated HRF of a different lighting environment. The optimization in 10 can thus be updated to include the 2 new

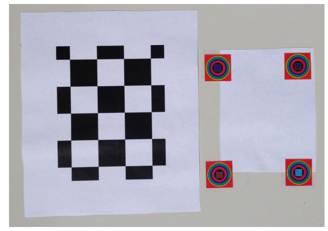


Figure 5. With a setup that mounts 4 lenticular markers on the same plane as a checkerboard, we can ground truth experiments that explore pose estimation results for more than 2 markers and color calibration routines.

color correcting variables by scaling the observed color of each lenticular marker:

$$\underset{R,T,s_r,s_b}{\operatorname{argmin}} \sum_{i}^{n} \Big(\frac{1}{\lambda_1} \big(h(R,T) - hue(M * rgb_i)\big)^2 + \frac{1}{\lambda_2} \| g(R,T) - p_i \|_2^2 \Big) \quad (11)$$

Here, rgb_i is the vector representation of the RGB measurement of a lenticular marker,

$$M = \begin{pmatrix} s_r & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & s_b \end{pmatrix}, \quad (12)$$

and $hue(..)$ is a function that coverts an RGB measurement to HSV space and returns only the hue.

This optimization function allows the multiple hue measurements of various lenticular markers to "agree" for a predicted rotation and translation under some global color calibration.

7. Experimental Results with Four Lenticular Markers

In this section we show pose estimation results when using more than 2 lenticular markers. First, we show the effect of increasing the number of markers and using the arithmetic error optimization detailed in Section 6.1. Then, we show the benefit of refining pose estimation using a reprojection optimization with and without color calibration explained in Section 6.2. Finally, we look closely at how optimizing for additional color scaling variables changes the measured hues to match the pre-calibrated HRF.

For this section, we use the setup shown in Figure 5. In this setup, we have 4 coplanar lenticular markers arranged in a rectangle, oriented as catacorner pairs with orthogonal orientations. We surround the lenticular markers with a radial hue pattern inspired by [13] to facilitate automated identification. Because the radial patterns are concentric rings, they are motion blur resistent; the radial pattern is unblurred along the direction of motion. We use concentric hues as a cue to refine centroid positioning of each lenticular marker. On the same plane as the markers, we include a checkerboard. Using the camera calibration toolbox in Matlab 2014a, the checkerboard serves to ground truth the pose of the lenticular markers. As before, we use this ground truth to also calibrate the HRF for each lenticular marker with a calibration video.

7.1. Pose Estimation with $n > 2$ Markers

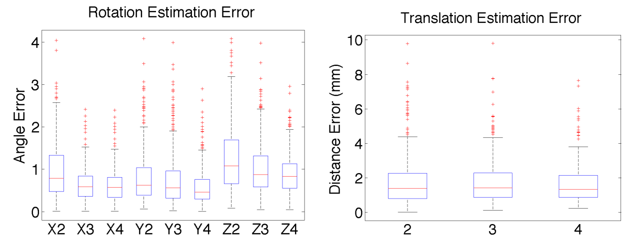
We first look at the benefit of overconstraining the pose estimation problem with more than 2 lenticular markers using the optimization of Section 6.1. In this experiment, we estimate the pose of frames from a video 3 times, each iteration adding the information from one additional lenticular marker. We estimate pose for the same video frames used to calibrate each marker's HRF in order to avoid any deleterious effects from different light environments. The video has 700 frames and captures the minimum and maximum relative orientations possible for the lenticular markers. As before with the physical prototype, we show the summary rotation and translation errors as boxplots.

In Figure 6, we show pose estimation performance as more lenticular markers are progressively used. For both the rotation and translation results, we group by trials using 2, 3, and 4 markers. Figure 6(a) shows the coplanar angular error of the individual local axes of the plane. For example, X2 indicates the error of the x-axis estimation when using 2 markers. Results show that additional lenticular markers reduces rotation error for each axis. The accuracy for the z-axis or the surface normal is especially improved. In addition, the inlier extremes and quantiles tend to become tighter, indicating a higher precision. The improvement in performance is due to the extra angular and positional information gained from adding markers across the plane.

Figure 6(b) shows the euclidian distance error of position estimation. We see slight gains in having 4 markers versus 2 or 3. The translation performance is already under 2 mm of error, so there is little room for improvement.

7.2. Reprojection and Color Calibration Refinement

Next, we show that optimizing the reprojection error can refine pose estimation results and increase rotation and translation accuracy. For this experiment, we use 3 videos, each containing around 300 frames, taken in lighting conditions different from the calibration video used to determine the HRF. The calibration video was taken under flourescent lights away from any window. Two test videos were taken outside: one in full direct sun ("sunny") and one on a completely cloudy day during the early afternoon ("cloudy"). The third video ("office") was taken indoors without lights turned on, but next to a window early in the afternoon. For



(a) (b)

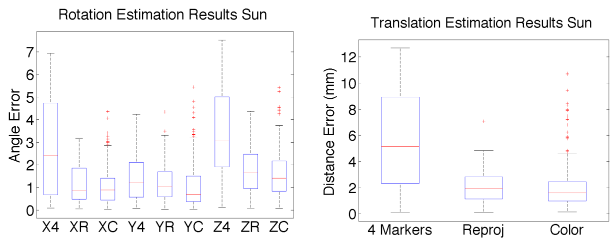
Figure 6. Although two lenticular markers are sufficient to constrain the pose estimation problem, using more lenticular markers improves rotation performance. Here we show the performance effects of increasing the number of lenticular markers in a) rotation estimation and b) translation estimation.

each video, we compare the pose estimation results 3 times:

- using the arithmetic optimization of Section 6.1 using 4 markers,
- intializing off of this pose estimation and doing the reprojection optimization of Section 6.2 using 4 markers,
- and using the extended reprojection optimization that solves for 2 additional color scaling variables.

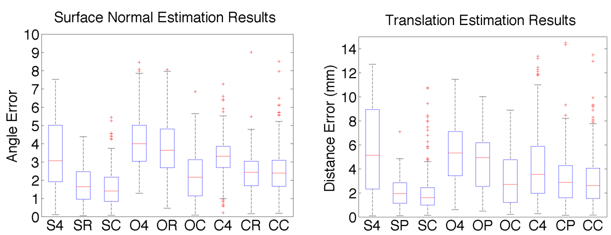
To start, we show results for the test data in full sun, shown in Figure 7. In Figure 7(a) we analyze the rotation error. Similar to before, we show the rotation and translation error grouped by each local axes and vary the optimization method. As examples, X4, XR, and XC show the rotation error for the arithmetic optimization, the reprojection optimization, and the color calibration optimization for the x-axis, respectively. We can see that for each axis, reprojection substantially improves precision and accuracy. However, by adding color calibration, we can further ameliorate rotation estimation: rotation estimation error drops down to about 1-1.5 degrees of median error, with most errors being less than 2 degrees. The same improvements can be seen in translation estimation results shown in Figure 7(b). Using reprojection refinement improves translation precision and accuracy substantially, with additional (smaller) improvements by color calibrating as well. The median translation error for color calibration is 1.6 mm versus 1.9 with only reprojection optimization.

Now, we show similar pose estimation results in Figure 8, but for the cloudy and office datasets as well. In the interest of space, we only present the z-axis error for rotation results. We denote the different datasets by S for sunny, O for Office, and C for Cloudy and the 3 optimization methods as 4,R, and C. For both rotatation and translation results across all datasets, we see an improvement over arithmetic optimization with reprojection refinement, with the highest accuracy and precision by optimizing for additional color scaling variables. Across a variety of light environments



(a) (b)

Figure 7. Optimizing for reprojection error and color calibration variables mitigates the negative effects of a lighting environment very different from calibration, such as full sun shown here. The a) rotation error reduces for each axis and b) the translation error reduces with the reprojection color calibration routine.



(a) (b)

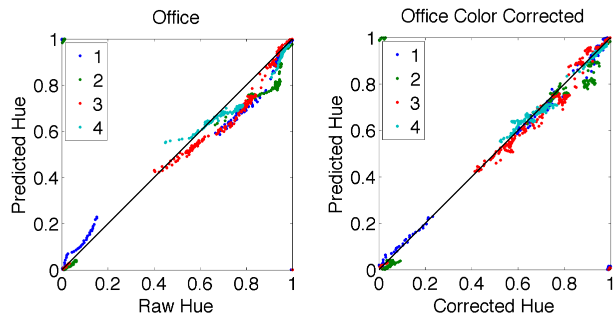
Figure 8. Despite these varying lighting conditions, reprojection optimization with color calibration produces results with very high accuracy and precision in a) surface normal estimation and b) translation estimation. Sunny, Office, and Cloudy datasets are denoted by S, O, and C respectively and 4, P, and C correspond to the three optimization routines discussed in Section 7.2.

using our color calibration optimization, we see the median rotation error to be around 2 degrees of error and median translation error less than 3 mm.

By refining the pose estimation with a reprojection optimization, we see pose estimation results improve. The reprojection objective function allows "wiggle room" in the observed position and hue of the lenticular markers versus what would be modeled from a pinhole camera. In effect, this reprojection objective makes the algorithm robust against both position and hue measurement noise for individual markers. Therefore, any errors in automatically finding the position of a marker is mitigated, as well as some lighting effects for specific angels such as glare. The two color scaling variables, on the other hand, allow a color calibration across all markers that adjusts individual frames towards different global lighting environments.

7.3. Color Calibration Analysis

In this final section, we look more closely at color calibration effect. The main challenge with different lighting environments is that the HRF no longer becomes a precise



(a) (b)

Figure 9. For a video in an office next to a window, we show color calibration results. On the left, we see the observed hue differing from the hue predicted by the HRF for the known orientations. On the right, we see that color corrected hues match the predicted hue and thus provides robustness for pose estimation in varying lighting environments.

representation of the hue/orientation relationship. Rather, in a new lightning environment, a given orientatation will result in a different hue than anticipated by the calibrated HRF. Indeed, any global light changes should affect all the markers. Therefore, a global color calibration can transform the observed hues of all the markers in an image to hues that match the HRF, and thus the lighting environment present during calibration of the HRF.

For the office data set, we analyze the observed hue and the calibrated hue versus the anticipated hue for each marker given the ground truth orientation and position in Figure 9. On the left, we see the observed hue versus predicted hue before any color correction. On the right, we see the corrected hue (via the color calibration optimization) versus the predicted hue. In the ideal case, when the observed or corrected hues match the HRF predictions perfectly, all points would be along the diagonal black line. However, as seen in Figure 9(a), all markers have systematic errors in matching the hue predictions before color correction. In particular, there are systematic color warps with hues above 0.5 shifted towards 0.8, and colors below 0.2 shifted higher.

With color correction, however, we see these lighting effects mitigated. In Figure 9(b), we see most of the points along the black line and thus the color corrected hues match the hues predicted by the pre-calibrated HRF. We show similar trends for the full sun and cloudy videos in Appendix C. The 2 optimized color scaling variables succeed in recovering the global lighting environment present during calibration of the HRF.

8. Conclusion

This work introduces a novel fiducial marker that can be made from easily accessible materials and commodity tools. This fiducial marker explicitly changes appearance due to viewpoint and two markers are sufficient to completely constrain the pose of an object. For objects that are long and skinny, these color based markers are more accurate than pose estimation based entirely on the tracking of the position of nearby fiducial points. In addition, by adding more markers and optimizing for a color calibration per frame, we can mitigate the negative effects of lighting environments different from the calibration environment, improving pose estimations.

References

[1] B. Atcheson, F. Heide, and W. Heidrich. Caltag: High precision fiducial markers for camera calibration. In *International Workshop on Vision, Modeling and Visualization*, volume 10, pages 41–48. Citeseer, 2010. 2

[2] M. Bansal and K. Daniilidis. Geometric urban geo-localization. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 3978–3985. IEEE, 2014. 2

[3] J.-Y. Bouget. Camera calibration toolbox for matlab. http://vision.caltech.edu/bouguetj/calib_doc/index.html. Accessed: 2013-10-09. 5

[4] A. M. Bruckstein, R. J. Holt, T. S. Huang, and A. N. Netravali. New devices for 3d pose estimation: Mantis eyes, agam paintings, sundials, and other space fiducials. *International Journal of Computer Vision*, 39(2):131–139, 2000. 2

[5] M. Fiala. Artag, a fiducial marker system using digital techniques. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2005. 2

[6] M. Fiala. Designing highly reliable fiducial markers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(7):1317–1324, July 2010. 2

[7] A. Herout, M. Zacharias, M. Dubská, and J. Havel. Fractal marker fields: no more scale limitations for fiduciary markers. In *Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on*, pages 285–286. IEEE, 2012. 2

[8] Y. Ji, K. Ye, and J. Yu. Reconstructing gas flows using light-path approximation. In *International Conference on Computational Photography (ICCP)*, 2013. 2

[9] H. Kato and M. Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *IEEE and ACM International Workshop on Augmented Reality*, 1999. 2, 4, 5

[10] Z. Kukelova, M. Bujnak, and T. Pajdla. Closed-form solutions to minimal absolute pose problems with known vertical direction. In *Computer vision–ACCV 2010*, pages 216–229. Springer, 2011. 2

[11] E. J. Larsen. Tracking head position and orientation. US Patent 8922644, 2014. 2

[12] A. Mohan, G. Woo, S. Hiura, Q. Smithwick, and R. Raskar. Bokode: imperceptible visual tags for camera based interaction from a distance. *ACM Transactions on Graphics (TOG)*, 28(3):98, 2009. 2

[13] M. Prasad, B. Chandran, and M. Brown. A motion blur resilient fiducial for quadcopter imaging. In *Proc. IEEE Workshop on Applications of Computer Vision (WACV)*, 2015. 6

[14] O. Saurer, F. Fraundorfer, and M. Pollefeys. Homography based visual odometry with known vertical direction and weak manhattan world assumption. In *Vicomor Workshop at IROS*, volume 2012, 2012. 2

[15] I. Schillebeeckx and R. Pless. Structured light field design for correspondence free rotation estimation. In *International Conference on Computational Photography (ICCP)*, 2015. 2, 4

[16] H. Tanaka, Y. Sumi, and Y. Matsumoto. A high-accuracy visual marker based on a microlens array. In *IEEE/RSJ International Conference on Intelligent Robotics and Systems*, 2012. 2

[17] H. Tanaka, Y. Sumi, and Y. Matsumoto. A visual marker for precise pose estimation based on lenticular lenses. In *IEEE International Conference on Robotics and Automation*, 2012. 2

[18] H. Tanaka, Y. Sumi, and Y. Matsumoto. Further stabilization of a microlens-array-based fiducial marker. In *IEEE International Symposium on Mixed and Augmented Reality*, 2013. 2

[19] H. Tanaka, Y. Sumi, and Y. Matsumoto. A solution to pose ambiguity of visual markers using moire patterns. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014. 2

[20] K. Tateno, I. Kitahara, and Y. Ohta. A nested marker for augmented reality. In *Virtual Reality Conference, 2007. VR'07. IEEE*, pages 259–262. IEEE, 2007. 2

[21] G. Wetzstein, R. Raskar, and W. Heidrich. Hand-Held Schlieren Photography with Light Field Probes. In *International Conference on Computational Photography (ICCP)*, 2011. 2

[22] G. Wetzstein, D. Roodnick, W. Heidrich, and R. Ramesh. Refractive shape from light field distortion. In *Proc. IEEE International Conference on Computer Vision*, 2011. 2

[23] G. Woo, A. Lippman, and R. Raskar. Vrcodes: Unobtrusive and active visual codes for interaction by exploiting rolling shutter. In *Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on*, pages 59–64. IEEE, 2012. 2

[24] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *Proc. IEEE International Conference on Computer Vision*, pages 666–673, 1999. 2

A. Pose Estimation Simulation of a Skinny Object

To test the impact of noise on pose estimation using lenticular markers, we simulated the physical implementation of Section 5. We randomly generate locations for the simulated object by defining translations uniformly distributed in a box between 0.5 and 2 meters in front of the camera. Rotations are generated randomly with the constraint that the angle between the surface normal of the object and the z-axis of the camera is less than $35°$.

For each object location we project the locations of the standard fiducial markers and our lenticular fiducial markers to get image positions of simulated markers. To simulate the color of the lenticular fiducial marker, we assumed that the printing process created no artifacts and implemented a simple ray tracer to model the optical effects at each lenticule (including the failure of the elliptical lenticular lens to perfectly focus parallel rays) in order to compute the hue.

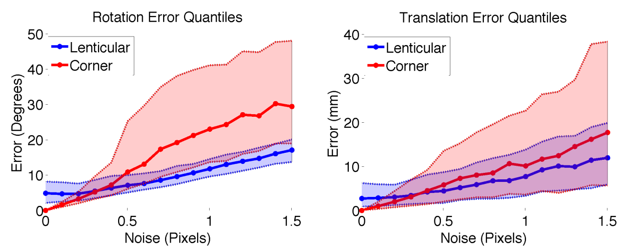
With this setup, we model noise to hue measurements by adding noise in the range $[-0.01, 0.01]$ to capture unmodeled effects that might come from, for example, glare off the lenticular array. We model location error as a 2D Gaussian whose standard deviation is a multiple of 0.1 pixels. Figure 10 shows the relative error in pose estimation as this noise increases. Rotation error is measured as the angular error between the surface normal measured from the simulated object position and the surface normal returned by the pose estimation process. Translation error is measured as the Euclidean distance between the simulated center of the object and the center of the object as estimated by the pose estimation process.

Figure 10(a) show the results for different noise levels. At zero added noise, the standard approach based on tracking corners is perfect (because there is no error in the simulated point locations) while our system has noise introduced in the estimation of the hue. However, for all but the smallest noise levels, the lenticular fiducial marker has lower median error and more consistent error magnitudes for both translation and rotation.

B. Generalized Geometric Constraints for Pose Estimation Per Lenticular Marker

We solve for the same pose estimation problem of a pinhole camera that views the lenticular markers shown in Section 4, but generalize to accomidate $n > 2$ lenticular markers.

In the local coordinate system of an object, each marker i has a local position C_i and local orientation $\vec{o}_i$ in reference to a local coordinate system of the lenticular markers. The pose estimation problem then solves for the rotation matrix R and translation vector T mapping a point in the object coordinate system into a point in the camera coor-



(a) (b)

Figure 10. By using lenticular fiducial markers instead of the typical 4 corner markers, we achieve ameliorated pose estimations for rotation (a) and translation (b). In both figures, we increase the amount of error simulated along the x-axis, and show the median value for each simulation surrounded by the 1st and 3rd quantile.

dinate system. Therefore, each lenticular marker at local position C_i will move to $RC_i + T$ in the camera coordinate system. For geometric constraints, we relate the direction of viewing the markers and their observed hues to a single R and T that describes this movement.

Rotation Constraints

Each lenticular marker gives a constraint related to its observed hue. Similar to Equation 2, the observed direction of the marker $\vec{r}_i$ must be perpendicular to the rotated plane $\vec{v}^i_{hue}$ spanned by the marker's orientation $\vec{o}_i$ and the local direction corresponding to the observed hue $\vec{v}^i_{hue}$:

$$R(\vec{o}_i \times \vec{v}^i_{hue}) \cdot \vec{r}_i = 0 \tag{13}$$

In addition, each marker gives a constraint related to its relative position to another marker. Generalizing Equation 4, the directions from the camera to two lenticular markers $\vec{r}_i$ and $\vec{r}_{i-1}$ and the relative displacement between the two markers should be coplanar:

$$R(C_i - C_{i-1}) \cdot (\vec{r}_{i-1} \times \vec{r}_i) = 0, \tag{14}$$

For the origin marker, this relative position constraint involves the position of the origin marker and the n_{th} marker:

$$R(C_i - C_0) \cdot (\vec{r}_0 \times \vec{r}_i) = 0, \tag{15}$$

For $n > 2$ markers, each marker gives 2 geometric constraints on R. Therefore, there are a total of $2n$ constraints to solve for R. For example, for 3 markers there are twice as many contraints as free variables in R.

Translational Constraints

With the rotation known, we can now solve for the translation. We extend Equation 7 for an arbitrary number of lenticular markers. In this constraint, the direction of the pixel location of the lenticular marker $\vec{r}_i$ should be parallel to the direction of the 3d location of the marker in the camera's coordinate space:

$$(RC_i + T) \times \vec{r}_i = \vec{0} \tag{16}$$

This constraint is a vector equation, so each marker gives 3 constraints. With n markers, we thus have $3n$ constraints to solve for T. Therefore, in total we have $5n$ constraints with n markers to solve for the 6 degrees of freedom of the pose estimation problem.

C. Additional Color Calibration Results

In Figure 11, we show color calibration results of all datasets with different lighting environments than the environment used to calibrate each marker's HRF. In the left column, we show the observed hue versus the predicted hue based on the known location and orientation of each marker. In the right column, we show the better hue predictions after color calibration. For all three datasets, we see that points fall closer to the diagonal line, indicating color corrected hues that better match the hue/orientation relationship of the pre-calibrated HRF.

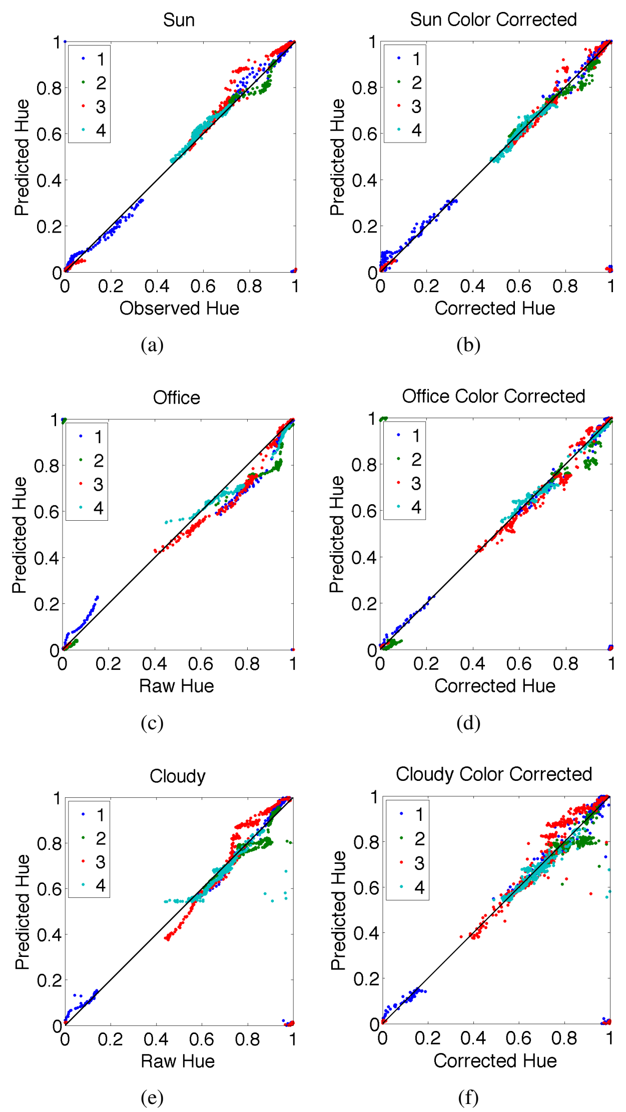


Figure 11. For three different lighting environments, we show color calibration results. In the left column, we see the observed hue differing from the hue predicted by the HRF for the known orientations. In the right column, we see that color corrected hues match the predicted hue and thus provides robustness for pose estimation in varying lighting environments.