

Deformation and Viewpoint Invariant Color Histograms

Justin Domke and Yiannis Aloimonos
Computer Vision Laboratory, Department of Computer Science
University of Maryland
College Park, MD 20740, USA
domke@cs.umd.edu, yiannis@cfar.umd.edu

Abstract

We develop a theoretical basis for creating color histograms that are invariant under deformation or changes in viewpoint. The gradients in different color channels weight the influence of a pixel on the histogram so as to cancel out the changes induced by deformations. Experiments show these histograms to be invariant under a variety of distortions and changes in viewpoint.

1 Introduction

Color histograms have been widely used for object recognition. Though in practice these histograms often vary slowly under changes in viewpoint, it is clear that the color histogram generated from an image surface is intimately tied up with the geometry of that surface, and the viewing position. In this paper, we suggest a method to create color histograms based on the color gradients. We will see that our method is invariant under any mapping of the surface which is *locally* affine, and thus a very wide class of viewpoint changes or deformations.

1.1 Related Work

The use of color histograms for recognition was initiated in a famous paper by Swain and Ballard [4]. This paper used simple color histograms for object recognition, with surprisingly strong results. Among the work that followed this, Funt and Finlayson [1] reduce sensitivity to lighting by considering changes among neighboring pixels, similarly to Land's Retinex theory. Stricker and Orengo [3] provide a measure of color similarity between images, storing only the dominant features of the histogram. They interpret the color histogram as a probability distribution, and then store only the central moments.

2 Derivation

Our derivation will proceed as follows. First, we will explore how small areas change in size under a given local affine transformation. Next, we will show that, given two corresponding points, the derivatives in two color channels may be used to recover the relevant

parameters of the homography, and thus the ratio of differential areas.¹ Finally, we will show that on this basis, the derivatives can be used to 'weight' pixels when constructing a color histogram to cancel out viewpoint dependent effects.

2.1 Local Affine Approximation

Given two corresponding points in two images, one can change coordinate systems such that both points lie at the origin in their respective systems. Consider this done. Assuming that the transformation between the two images is continuous, it is locally linear. Note that this is a weak assumption, satisfied by transformations such as homographies, changes in viewpoint or smooth deformations. Thus, for *infinitesimal* \mathbf{x} we can write the local affine transformation as

$$\mathbf{x} = H\mathbf{x}' \quad (1)$$

where

$$H = \begin{bmatrix} A & B \\ C & D \end{bmatrix}. \quad (2)$$

Here, \mathbf{x}' and \mathbf{x} are the points in the first and second coordinate systems, respectively.

2.2 Differential Area From Gradients

By definition, the differential area in the first coordinate system is

$$da' = dx' dy'. \quad (3)$$

We can then find the corresponding area in the second coordinate system.

$$da = \left| \frac{\partial \mathbf{x}}{\partial x'} \times \frac{\partial \mathbf{x}}{\partial y'} \right| = |AD - BC| dx' dy' \quad (4)$$

So, the ratio of differential areas is

$$\frac{da}{da'} = |AD - BC|. \quad (5)$$

2.3 Invariant Color Histograms

Denote the image intensity function in one color channel by $f(x,y) = f'(x',y')$. Using standard calculus, we can relate the image derivatives.

$$\frac{\partial f'}{\partial x'} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial x'} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial x'} \quad (6)$$

$$\frac{\partial f'}{\partial y'} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial y'} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial y'} \quad (7)$$

¹We emphasize that no correspondence is used in our technique- this is merely an artifact of the derivation.

Now, write the derivative of f with respect to x , evaluated at the origin as f'_x , and for f' similarly. Evaluating the coordinate derivatives yields

$$f'_x = f_x A + f_y C \quad (8)$$

$$f'_y = f_x B + f_y D. \quad (9)$$

Notice that, given only the derivatives of f in both images, it will not be possible to recover $|AD - BC|$, and thus the differential area. Two solutions to this problem are apparent. First, one could use higher-order derivatives. Second, one can use other color channels. Given the difficulty of accurately measuring high-order derivatives, we will pursue the second option in this work. Denote the image intensity function in a second color by g . The derivatives will obey the same constraints as above.

$$g'_x = g_x A + g_y C \quad (10)$$

$$g'_y = g_x B + g_y D \quad (11)$$

Now, given all derivatives, it is possible to solve for each of A , B , C , and D .

$$\begin{aligned} A &= \frac{f_y g'_x - g_y f'_x}{f_x g_y - g_x f_y}, & B &= \frac{f_y g'_y - g_y f'_y}{f_x g_y - g_x f_y} \\ C &= \frac{f_x g'_x - g_x f'_x}{f_x g_y - g_x f_y}, & D &= \frac{f_x g'_y - g_x f'_y}{f_x g_y - g_x f_y} \end{aligned} \quad (12)$$

Substituting these expressions into Equation 5 and simplifying gives us

$$\frac{da}{da'} = |AD - CB| = \frac{|f'_x g'_y - f'_y g'_x|}{|f_x g_y - f_y g_x|}. \quad (13)$$

We can now see a direct relationship between the color derivatives, and the differential areas.

$$da |f_x g_y - f_y g_x| = da' |f'_x g'_y - f'_y g'_x| \quad (14)$$

This is the key relationship that we will use to construct invariant color histograms. Notice that one can consider the gradients of the two color channels as vectors. Then, Equation 14 says that the differential area, times the norm of the cross product of the color gradients is invariant.

If we take integrals over the areas in both images with a given color, these integrals will be equal, and so no correspondence is required. If the color of pixel \mathbf{x} is some integer x_c , then,

$$\int_{\mathbf{x}:x_c=c} |f_x(\mathbf{x})g_y(\mathbf{x}) - f_y(\mathbf{x})g_x(\mathbf{x})| da = \int_{\mathbf{x}':x'_c=c} |f'_x(\mathbf{x}')g'_y(\mathbf{x}') - f'_y(\mathbf{x}')g'_x(\mathbf{x}')| da'. \quad (15)$$

Intuitively, the contribution in the integral in Equation 15 will be largest when there are significant gradient magnitudes in both color channels, and these gradients are in different directions.

3 Algorithm

Using Equation 14, we can create a very simple algorithm to create deformation invariant color histograms. In our notation, the histogram component for color c , h_c , in a *traditional* color histogram is

$$h_c = \sum_{\mathbf{s}:s_c=c} 1. \quad (16)$$

The only difference with our method is that, rather than each pixel getting one 'vote', the influence of each pixel is weighted by a function of the derivatives.

$$h_c = \sum_{\mathbf{s}:s_c=c} w(\mathbf{s}) \quad (17)$$

$$w(\mathbf{s}) = |f_x(\mathbf{s})g_y(\mathbf{s}) - f_y(\mathbf{s})g_x(\mathbf{s})| \quad (18)$$

4 Implementation

In the experiments in this paper, we used 5 color bins in each of the three color channels, resulting in a total of 125 bins. Using such a coarse sampling of the color space is principally for display purposes. In our experiments, it was possible to accurately recover histograms with many more colors.

This method requires derivatives in two color channels, f and g . Here, we have used red as one channel, and the average of green and blue for the other. Derivatives are taken in the simplest possible way, though convolution with the filters

$$\begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}. \quad (19)$$

Experimentation with more complex methods of measuring the derivative have shown the above filters to be preferred.

5 Experiments

In all the synthetic experiments shown in this paper, the initial images are first processed with a low-pass filter to make derivatives more reliable. This filter is applied once, before warping, and so no similar process can be applied with real images. See Section 6 for further discussion of this issue.

As a first example, Figure 1 shows the computation of an invariant histogram, and compares it to a traditional histogram. We emphasize that the only difference in the computation of the two histograms is as contrasted in Equations 16 and 17.

In our experiments, we seek to answer two questions. First, how do invariant histograms change under deformation or change in viewpoint? We will see that invariant histograms are nearly unchanging under distortions or viewpoint changes that lead to large changes in traditional histograms. Secondly, how do invariant histograms change under true changes in the scene? If the invariance was gained at the price of simply being

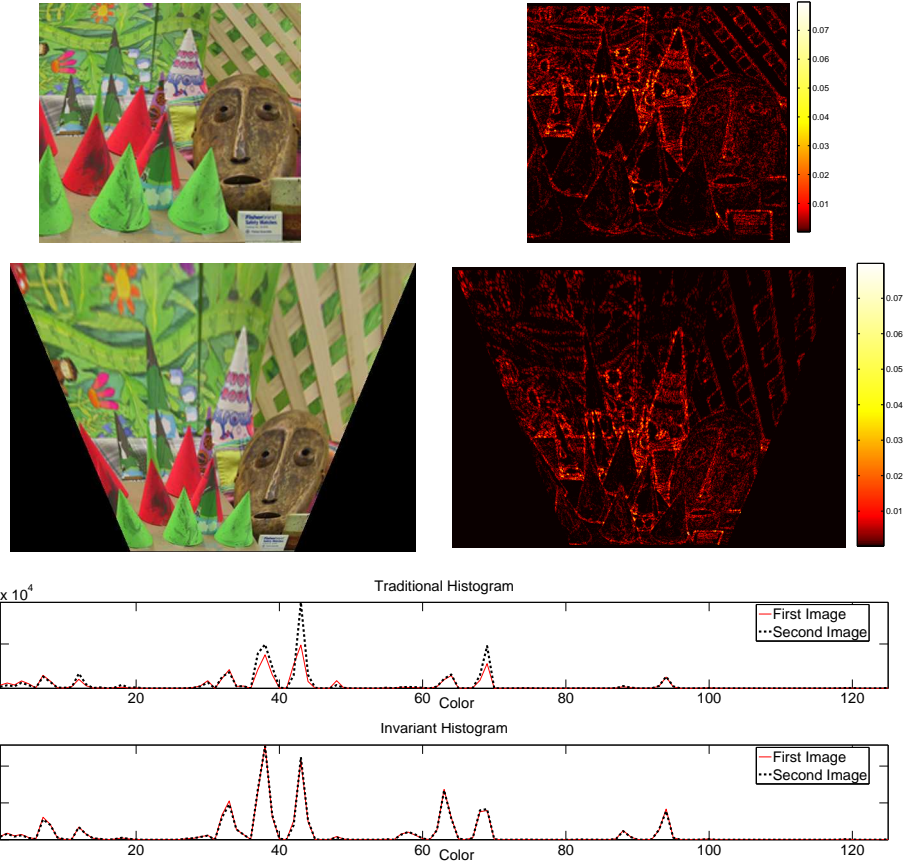


Figure 1: First row: initial image, and $h(\cdot)$ mapped over it. Second row: image warped under a homography, and $h(\cdot)$ mapped over it. Third row: traditional (unweighted) color histogram. Fourth row: Invariant color histogram.

insensitive to the images, we would not have a useful technique. We will see that this is not the case- under true changes in the scene, invariant histograms appear to have similar sensitivity as traditional histograms.

To investigate changes under deformation, we warped an image under a range of radial distortions, and computed histograms for each resulting image. To compute the difference of two histograms we used the sum of squared differences. To make the comparison between traditional and invariant histograms meaningful despite traditional histograms having vastly higher magnitudes, the error is scaled by the sum of squares of the first histogram.

$$\text{difference}(h^A, h^B) = \frac{\sum_c (h_c^A - h_c^B)^2}{\sum_c (h_c^A)^2} \quad (20)$$

Results are shown in Figure 2. We can observe that though traditional histograms increase in difference with larger distortions, the invariant histograms are essentially un-

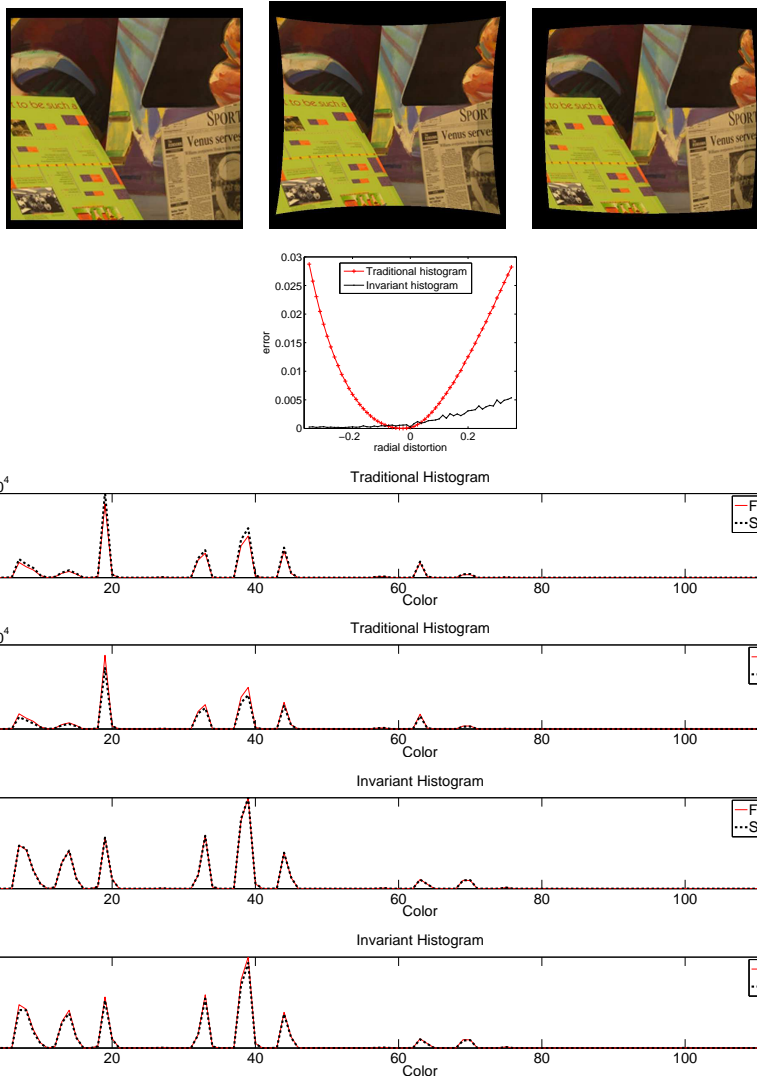


Figure 2: Histograms under radial distortion. First row: The first image, and the result of warping it under the minimum and maximum distortion coefficients. Second row: The sum of absolute differences of the first image and each of the warped image. Bottom rows show traditional and invariant histograms comparing images as labeled.

changed. Since the image we have used are taken from a stereo sequence [2], we can compare how histograms change under small change in structure. Here, the actual surfaces in view are changing, so invariance is neither expected nor desired. In Figure 2 we compare the histograms resulting from images taken with a continually increasing viewpoint. Here, the invariant histogram has similar performance to the traditional histogram. Notice particularly that despite the small change in viewpoint, the differences for the invariant histogram quickly become larger than those resulting from the major distortions

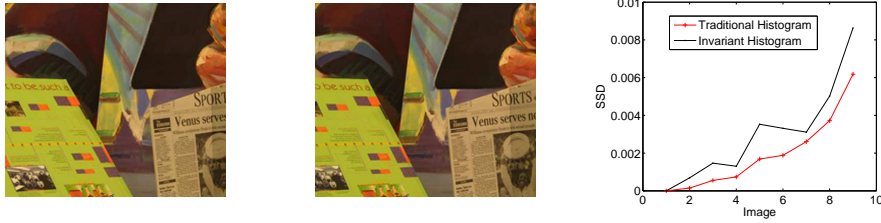


Figure 3: Effects of changing scene. Left: First image. Center: Final image. Right: Histogram differences computed with respect to the first image in the scene. Compare to Figure 2.

in Figure 2. This suggests that invariance with respect to deformation is possible without giving up sensitivity to true changes in the scene.

In Figure 4 we see the behavior of histograms under a range of projective transformation. There, the first image was warped under a range of homographies, $H = \lambda H_1 + (1 - \lambda)H_2$ where $0 \leq \lambda \leq 1$. We should note that for $\lambda = .5$, $H = I$. We see again here that the invariant histogram remains very similar despite the distortions.

Finally, Figure 5 shows traditional and invariant histograms derived from a pair of real images. The errors in the invariant histogram are likely due to the fact that that this technique accounts for deformation changes only- no attempt is made to account for changes due to lighting conditions, or image noise.

6 Conclusions

In this paper we have seen that, given two corresponding points in two images, if the gradients in two color channels were available, the local affine transformation could be recovered. On this basis, it is possible to construct a measure using the gradients which cancels out the changes in image area resulting from viewpoint. Using this measure, we can integrate over areas in two images having a given color. The resulting value will be invariant to any transformation which is locally affine, and hence essentially any smooth transformation.

Future work should further address practical issues with finding these histograms in real images. One could incorporate techniques to overcome color changes due to lighting [1]. Another issue is the measure of 'difference' between two invariant histograms. Images, and specifically gradient measurements are inevitably corrupted by noise. If many measurements are taken the noise is likely to average out, but this will not occur for few measurements. A close observation of some of our results suggests that errors result in the histogram precisely at those points that have the smallest components in the traditional histogram, and thus *at those points derived from the fewest image measurements*. Thus, when measuring the 'distance' between two invariant histograms, it would be natural use the number of pixels having each value, to weight the importance of each component. This could be naturally phrased in a probabilistic view.

The technique developed in this paper will only work so far as the mathematical model of the problem is accurate. We emphasize that this is often not the case. The method is

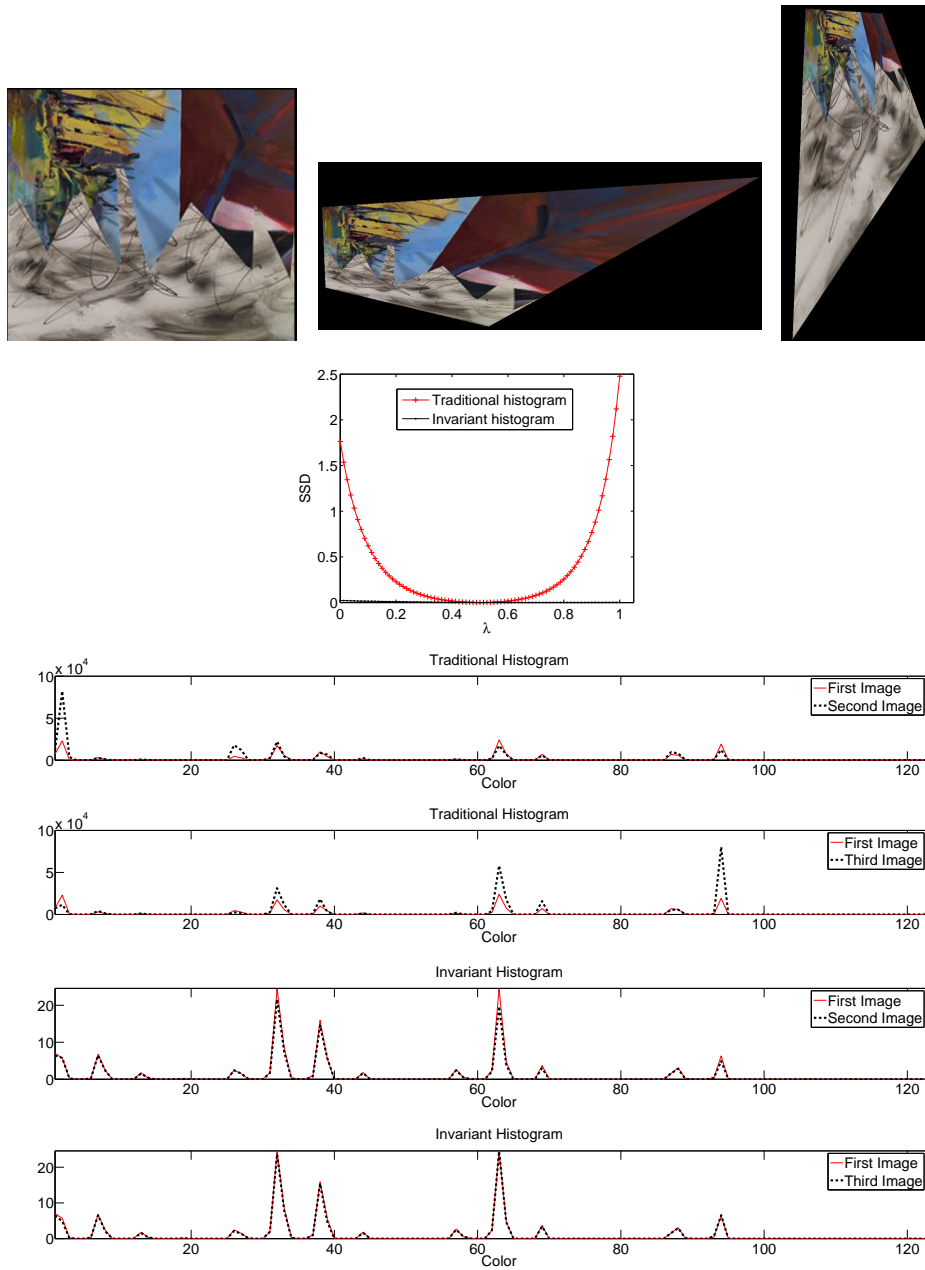


Figure 4: Histograms under projective transformations. First row: The first image, and the result of warping it under H_1 and H_2 . Second row: The sum of absolute differences of the first image and each of the warped image. Bottom rows show traditional and invariant histograms comparing images as labeled.

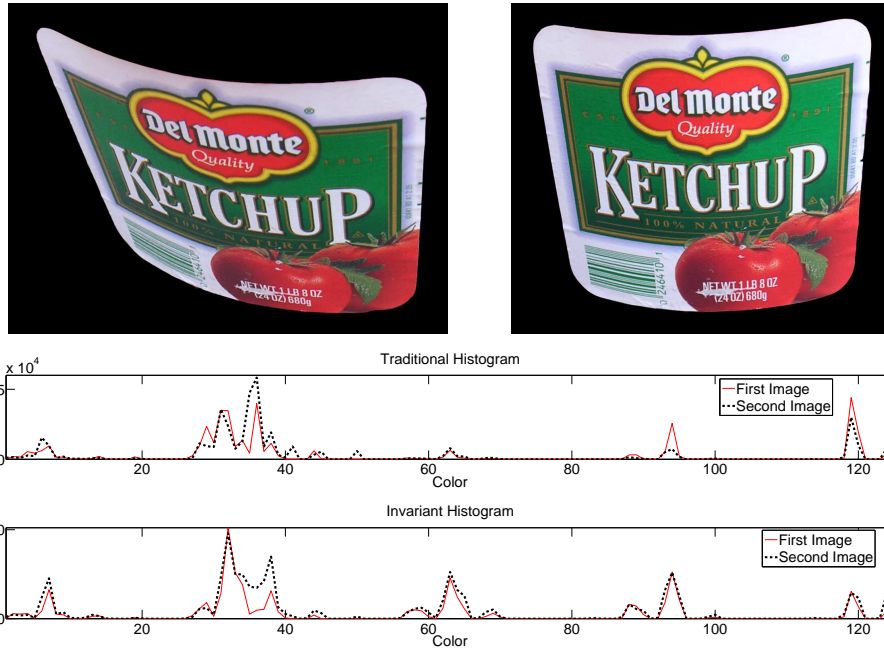


Figure 5: Traditional and Invariant histograms derived from segmented regions of two real images.

invariant only to pure deformations of the scene. In real images, however, other effects will often be present. These include changes due to reflectance, and illumination, for which the literature offers several potential solutions. Perhaps more problematic is the fact that filtering is present in the imaging process. This means that the signals in two images are not just transformed version of each other- different signals are present.

Intuitively, our model of the imaging process suggests that an image of a physical object is simply the projection of the intensity function onto the image plane. Crucially, our model states that if the scale of the observed object is halved, all derivatives will be doubled. However, if a real image is taken of a rich texture from far away, rather than resulting in high derivatives, the image will often be near a constant color, with very *low* derivatives. If a technique is to work for arbitrary objects and scales, it will be necessary to properly account for the real behavior of images under scale changes.

The present work is extendable in two major directions. First one could use measurements other than gradients to do the 'balancing' required to create invariance. Using other measurements here could significantly increase robustness of image effects. Second, one could create histograms of quantities other than color. By using histograms of more sophisticated filters, it is likely that recognition of regions could be made significantly more powerful.

References

- [1] Brian V. Funt and Graham D. Finlayson. Color constant color indexing. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(5):522–529, 1995.
- [2] Daniel Scharstein and Richard Szeliski. High-accuracy stereo depth maps using structured light. In *CVPR (1)*, pages 195–202, 2003.
- [3] Markus A. Stricker and Markus Orengo. Similarity of color images. In *Storage and Retrieval for Image and Video Databases (SPIE)*, pages 381–392, 1995.
- [4] Michael J. Swain and Dana H. Ballard. Color indexing. *Int. J. Comput. Vision*, 7(1):11–32, 1991.