A Color Interest Operator for Landmark-based Navigation

Zachary Dodds and **Gregory D. Hager**
Department of Computer Science
Yale Univ., P.O. Box 208285
New Haven, CT 06520

Abstract

Landmark-based approaches to robot navigation require an "interest operator" to estimate the utility of a particular image region as an effective representative for a scene. This paper presents a color interest operator consisting of a weighted combination of heuristic scores. The operator selects those image regions (landmarks) likely to be found again, even under a different viewing geometry and/or different illumination conditions. These salient regions yield a robust representation for recognition of a scene. Experiments showing the reproduceability of the regions selected by this operator demonstrate its use as a hedge against environmental uncertainties.

Introduction

One important ability of natural visual systems is that they spend most of their time on "interesting" portions of their input, that is, on those aspects of an image which inform the task at hand. The Stanford Cart had one of the first artificial vision systems which looked for regions of interest from the scenes it recorded. (Moravec 1983) The Cart's "Interest Operator" sought corners and areas of high contrast in order to localize – and avoid – obstacles between it and its goal. We propose in this paper a "Color Interest Operator," also suited for robotic localization and navigation tasks.

In contrast to work in image segmentation (Liu & Yang 1994; Perez & Koch 1994; Beveridge *et al.* 1989; Panjwani & Healey 1995), feature-finding algorithms such as ours do not seek to classify each - or even most - of the pixels of an image. Instead, feature (or landmark) recognition attempts to home in on those portions of a scene which are "locally unique" and are likely to remain so under a set of possible environmental changes. To this end, a variety of approaches have been applied. For example, the geometric properties of objects in the environment have been used (Baumgartner & Skaar 1994), as well as texture (Zheng & Tsuji 1992), edge location (Huttenlocher, Leventon, & Rucklidge 1990), and dominant edge orientation (Engleson 1994). This paper presents an interest operator for determining image-based *color* landmarks suitable for navigating in an indoor environment. As such, this work comes closest to that of (Zheng & Tsuji 1992), which presents a self-navigating automobile using a combination of techniques for landmark detection outdoors. Tailored to indoor workplace environments, the approach we present differs from that work in that it offers reproduceability of landmark finding under a variety of conditions, including viewpoint geometry, lighting conditions, and, to a lesser extent, lighting composition.

In the following sections we first address what constitutes an effective, i.e. reproduceable, landmark; we follow with our choice of representation for landmarks and an overview of our feature-finding algorithm. We then consider the design-choice specifics of the algorithm in light of our goal of choosing reproduceable landmarks. We extend this consideration to the heuristics we employ to estimate the suitability of particular landmarks. Finally, we report the results of our experiments with the algorithm.

Background

Landmarks

Informally, a landmark is an object which represents an area larger than it actually occupies. A landmark simplifies our view of the landscape by representing its surroundings, as well as itself; to do so, it must be easily distinguishable from the neighboring area. That is, a landmark could be characterized as "locally unique" image region. Such a landmark does not necessarily correspond to an environmental fixture, but instead indicates a salient image-based region. Work such as (Hager & Rasmussen 1996; Engleson 1994) has explored the implications of using such image-level features as cues for navigation.

To assign a formal definition to local uniqueness, we consider a landmark to be a subset of an image, f, which by some measure is distinct from its boundary, b, so that

$$|\,\varphi(f,b)\,| < \tau \tag{1}$$

where φ is a (real-valued) function which expresses a quantifiable similarity between two image subsets, and τ is a threshhold. Geometrically, b is an immediate neighborhood or outside edge of f, excluding f itself. Intuitively, a landmark is a framed picture, whose border b surrounds a canvas, f. The function φ we will use will depend on the color properties of the image subsets f and b.

Because mobile robotics applications act within dynamic environments, they pose classes of invariance problems for a vision system. A landmark-based vision system satisfies the need for invariants by picking image patches out of a scene that the system can find again, even as other aspects of the robot's surroundings change. The fundamental goal of a landmark-finding algorithm is this reproduceability of results under a variety of environmental conditions.

To this end, let $\mathbf{s}_1$ and $\mathbf{s}_2$ be vectors whose components represent the variable conditions in a scene. We extend (1) to claim that a landmark is reproduceable (i.e. refindable) between the two states, $\mathbf{s}_1$ and $\mathbf{s}_2$, if

$$|\,\varphi(f,b;\mathbf{s}_1)\,| < \tau_1 \Leftrightarrow |\,\varphi(f,b;\mathbf{s}_2)\,| < \tau_2 \tag{2}$$

In this work we will consider viewpoint geometry, ambient illumination intensity, and the spectral illuminant composition as the components of the environmental state vector.

Landmark Representation

We have chosen to use color as the distinguishing characteristic for landmarks. The domains we are considering – man-made, indoor environments – have properties which make color-based features a workable idea. Man-made indoor environments can often be characterized by only a few colors: those of the walls, floor, and ceiling. Highlighting that background are smaller-scale features – furnishings, decoration, or structural details – commonly with distinct color characteristics; those features provide recognizable summaries of their location.

There are a number of possible representations for landmarks based on color information. The dominant color, a mean color value, or another statistically-based measure might be used to characterize a portion of an image. We follow Swain and Ballard's work in color histogramming, originally proposed to enable fast indexing into an image database. (Swain & Ballard 1991) We represent a landmark as two histograms: one which stores the colors of the feature itself and another which stores the colors immediately around the first: f and b from equation 2.

We consider a histogram a vector of color bins, each of which contains the number of pixels which fall into that bin's color, so that a color histogram, $\mathbf{H}$, is

$$\mathbf{H} = (h_1, \cdots, h_n)^T \text{ with } h_i = \text{ \# of pixels in bin } i$$

A histogram bin represents an equivalence class of colors within the underlying color space; two histograms are comparable only if the bins used by each are the same. In this case we extend the notion of "histogram as vector" to define a histogram inner product. Given two histograms with identically defined bins, $\mathbf{H} = (h_i)^T$ and $\mathbf{G} = (g_i)^T$, their normalized histogram inner product is given by

$$\langle\, \mathbf{H}, \mathbf{G}\,\rangle = \frac{\sum_i h_i g_i}{(\sum_i h_i^2)\,(\sum_i g_i^2)} \tag{3}$$

Equation 3 can been viewed as the normalized correlation between the two histograms: 1 represents the perfect match of identical histograms, while 0 represents orthogonal, nonmatching histograms. Because none of the bins can hold a negative number of pixels, negative values for the histogram inner product can not occur. The normalized inner product in (3) will serve as the similarity function φ from (1) and (2).

Landmark-finding Algorithm

Our algorithm for finding landmarks uses a standard region-growing technique (Beveridge *et al.* 1989) where the similarity measure among subsets of an image is the inner product of their color histograms:

- Divide the image into tiles of a small, fixed size.
- Compute the color histograms of these image tiles.
- Starting with a seed tile as the current feature, include neighboring tiles if their inner product with the current feature is greater than a threshhold.
- If a tile neighboring the current feature is not sufficiently similar, add it instead to the boundary.
- Using the landmarks found, compute heuristic estimators of their suitability for use.

The algorithm can be stopped after finding the image feature containing the seed point, or continued until all of the landmarks of the image are found. In the latter case, the results of the algorithm are dependent on which tile is chosen to seed to next potential landmark. While we have found experimentally that different choice strategies do not greatly affect performance, other algorithms (Baumgartner & Skaar 1996) do eliminate this dependence on the seed choices.

Reproduceability of Landmarks

The color-histogram representation itself serves the goal of finding image landmarks under changing viewpoint geometry. Fundamentally, a histogram measures the area covered by each of the colors in a feature. By using normalized correlation, only the ratios among the bin values (area) are important for identifying a color feature. Up to the affine approximation of perspective projection, the ratio of areas of planar objects within an image remains invariant to changes in viewing angle and distance. Those image regions characterized by a single color do not even require that affine approximation be valid: they can be matched based on the reproduceability of that patch of color under different viewpoints and lighting conditions.

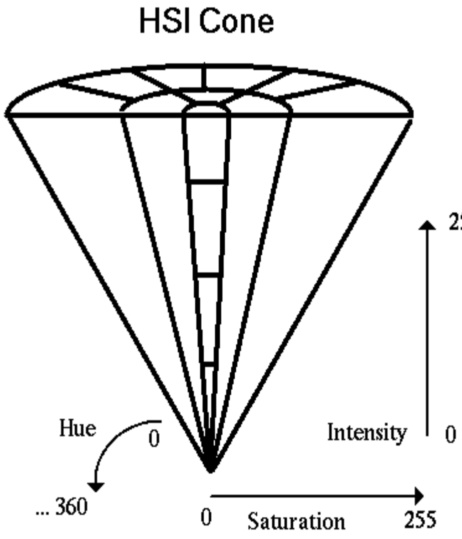

Figure 1: **The HSI Cone** The divisions within this cross-sectional view represent the bins of response values considered equivalent for the purposes of landmark recognition.

As modeled in (Funt 1995), a change in ambient lighting intensity corresponds to a linear variation in pixel values along all three axes of RGB space. In the Hue-Saturation-Intensity color space, however, such a change corresponds only to a variation of the vertical intensity axis. This explicit representation of pixel intensity makes HSI space a natural choice for our representation of color; HSI space has been considered for image segmentation purposes (Liu & Yang 1994; Perez & Koch 1994) and for landmark recognition (Zheng & Tsuji 1992). Its often-cited shortcoming are its singularities at the vertex, where hue and saturation are undefined, and along the intensity axis, where hue is undefined. (Liu & Yang 1994)

Dividing up HSI space into bins along lines of hue and saturation would necessarily group all of the greyscale pixels into a single bin. Such a scheme runs counter to our intuitive segmentation of a scene. In reality, most lighting changes we would expect a landmark-finding system to handle do not span such extreme variations in intensity. We compromise by dividing HSI space into finer distinctions of hue when hue is most meaningful and into finer distinctions of intensity when hue is less meaningful. Figure 1 shows a top-down and a cross-sectional view of our method of assigning bins. In our system, we use twelve bins at high saturation, corresponding to hue angles of 30 degrees each, and six bins at lower saturation, corresponding to hue angles of 60 degrees each. The bin containing the lowest-saturation pixels, the greyscale values, was divided into four intensity bins. The divisions along the saturation axis were drawn at 30% and 8% of maximum saturation: these low figures reflect that most objects in the test environments used were relatively unsaturated. To deal with the singularity at the vertex of the cone, a low-intensity bin replaces the portions of the color bins near the vertex.

Thus, the bin divisions for our experiments occur at

$$\begin{aligned}
hue &= 30n^\circ \;\; (n = 1, ..., 12;\; sat > 30\%) \\
hue &= 60n^\circ \;\; (n = 1, ..., 6;\; sat > 8\%) \\
int &= 64n \;\; (n = 1, ..., 3;\; sat < 8\%) \\
red, green, blue &< 40 \;\; (out\ of\ 255)
\end{aligned}$$

where *int* refers to intensity, *sat* refers to saturation, and *hue* refers to the hue "axis" of HSI space. The last bin is considered a part of the low-saturation, low-intensity bin, rather than a separate color class.

The remaining environmental factor we wish to consider is that of illuminant composition. If we restrict our attention to single-colored illuminants and landmarks, then changes in illuminant color will correspond to changes in the angle formed by the perceived color and the intensity axis of the HSI cone. It is that angle which specifies the bin to which a pixel will be assigned by the above equations. We do not use a color constancy algorithm, but even without modeling the illuminant color, the histogramming approach presented above uses bins large enough to absorb small changes in illuminant composition. Aliasing effects inherent in all histogramming techniques cause certain colors (those near bin boundaries) to be more sensitive to illumination changes than others.

Creating A Color Interest Operator

With a procedure in place for dividing an image into potential landmarks, we introduce a group of heuristics with which we can evaluate the landmarks found. Suitable landmarks will be those which are refindable under the environmental changes mentioned in the previous section. A linear combination of the following heuristics produces a real-valued score for each distinct region of an image.

The **size** of the feature – a typical image segmentation by the algorithm above yields many small and single-tile "features," mostly occurring between larger regions. In an image segmentation algorithm these smaller regions would be reconsidered and either added to one of their neighbors or made into a segment of their own. Because a landmark-finding algorithm does not seek to classify each pixel in an image, regions smaller than a threshhold are discarded. On the other hand, too-large regions are presumed to be background and are also disregarded. Note that this does introduce a dependence on viewpoint: imaging locations too close or too far from a given landmark will not result in its identification. Some dependence on viewpoint is inevitable with any single-scale representation of the environment, such as ours.

The **color-distinctiveness** of a feature with respect to the whole image is a measure of how different a given landmark's color composition is from the rest of the image. As the landmark is a part of the whole image, the inner product between its histogram and the image's cannot be zero, but it does provide a means to compare how distinctive different features are. If the region contains colors which are relatively rare in the whole image, its inner product with that image will be lower than others'. By favoring features with low scores on this heuristic, we are ensuring that locally unique image patches are those chosen to act as landmarks.

Another heuristic based on the distinctiveness of a feature's colors is the **distinctive range** of that feature. A landmark's distinctive range is the image-based distance from it to the closest feature which has a highly correlated histogram. Larger distinctive ranges indicate wider areas in which a feature is unique in its color structure. For example, a fire extinguisher has a large distinctive range and will be easily recognizable provided there are no other red objects around. Hanging next to a red exit sign, however, its range and its reproduceability based on color information alone are much smaller.

The **edges** between color patches have been used to make Swain and Ballard's color-histogram indexing technique more robust to differing illuminant intensities. (Funt 1995) We adapt this idea by considering the relationship between a landmark and its boundary, or frame. By definition, a landmark's frame must differ in its color composition from the its body. However, landmarks with weak color edges, i.e. with frames only slightly distinct from their enclosed features, will be more susceptible to spill into their frames with small changes of lighting or viewpoint occur. This **frame distinctiveness** heuristic ranks an image's features according to the inner product between their frames and their bodies; lower inner products correspond to greater color distinctions and, thus, better-isolated regions.

Whether its histogram is unimodal or multimodal, the **average saturation** of a feature indicates the strength of the color(s) in that feature: a low average saturation indicates that the feature is "washed out," i.e. that the color(s) are close to greyscale shades. For low-saturation features the composition of the light of the scene plays a relatively larger role in determining their hue, so they would be expected to be less reproduceable under varying lighting conditions. Further, if the saturation is low enough that the feature is largely represented by the greyscale histogram bins, varying intensities of light will change the histogram substantially. Thus, we expect that features with higher saturation will appear more reliably as lighting conditions change.

The **circularity** of a landmark is, informally, a measure of how compactly its body of pixels fits within its perimeter. We estimate circularity by computing the ratio of the length of a feature's perimeter with the circumference of a circle enclosing equal area. Though the applicability of this heuristic depends on the regularity of objects one expects in the environment, experimental results suggest that those landmarks which are very far from circular, e.g. those not convex or very thin, are less likely to remain stable under different viewing conditions. In particular, non-convex shapes are often meandering patches of background. The circularity heuristic penalizes such regions and reinforces those which have less unusual shapes.

Heuristic	**Weight**
Dinstinctive Range	3.0
Saturation	2.5
Color Distinctiveness	1.0
Frame Distinctiveness	1.0
Circularity	0.8
Size	0.5

Figure 2: **Heuristic Weights for Representing the Environment** The relative importance of the individual heuristics for finding a landmark suitable for representing the environment. Such a landmark should be reproduceable by the system under different conditions, e.g. changes in lighting contrast, intensity, and composition.

With a given set of weights, such as those in Figure 2 these heuristics comprise a color interest operator. That is, for a particular image region the operator supplies a score which indicates its "interest" relative to the other regions in the image. The relative weight of each heuristic was tested experimentally, using the reproduceability of the three most salient image regions to tune the parameters over a variety of images; about a 20% change in these weights' values leave the ranking of the top six landmarks in our test scenes unaffected. The order of the weights' importance reflects the goal that landmarks be reproduceable under a variety of conditions. Size and circularity, the two least important heuristics, are also the two which are directly affected by changes in camera orientation and position. On the other hand, saturation and the distinctiveness measures, when high, ensure that changes in lighting conditions will not dominate the camera response of an image region.

After the region-growing algorithm identifies an image's salient regions (typically taking 1-3 seconds, including the HSI conversion), finding the edge distinctiveness, circularity, and color distinctiveness of each requires time proportional to the number of features, which can be bounded by adjusting the threshhold on admissable feature size. The distinctive range requires time proportional to the square of the number of features in the image, and is the most expensive of the heuristics to compute. All of these heuristics are computed during the initial region-growing algorithm; none of them require more than the one pass that the algorithm performs over the image.

Experiments

In order to test the landmark-finding algorithm, we ran it on several scenes under a variety of different lighting conditions. The algorithm was implemented on a Sparc20 using a K2TV300 color framegrabber and Sony XC999 CCD color camera.

In the first test, we varied the environmental conditions under which scenes were viewed. A landmark, chosen as the "best" by the heuristics from one viewpoint and lighting state, was stored as a histogram and considered the canonical representation of that landmark. Then, as the viewing angle, the distance to the landmark, the lighting intensity and the lighting composition was altered, the new landmarks found in each image were compared with the stored, canonical image. If the match score, i.e. the normalized correlation between the stored feature and the best-matching feature, was greater than our threshhold of 0.95, and if the best-matching feature corresponded to the same physical landmark, the recognition was considered a success.

The first landmark considered is a set of reddish/brown mountains (from a wall-mounted poster), surrounded by a blue and black background. The second is an orange toolbox amid a more cluttered scene. The landmarks were originally found from a distance of about two meters, with the histograms stored for future comparisons. Both had maximal distinctive range, high saturation, and high color-distinctiveness with respect to the whole image, though the scenes do contain other small image patches of the same color, also highlighted as white in Figure 3.

The variety of conditions under which these landmarks are reproduceable is shown in Figure 4. The upper limit in distance came from the landmark becoming too small for the algorithm to pursue, while the lower limit arose as additional details became apparent and the landmark split into two distinct features, yielding no one best match. As the angle between the camera's optical axis and the poster increased, the specularity of the poster's glass became more important until it subsumed the original landmark at about 50°. The intensity range is bounded by two basic factors: the resolution of the camera sensors and the extent that the illuminant's color affects the perceived feature color. Because the light was close to white in both cases, the latter factor did not have a noticeable effect. Because the camera saturates at 255, however, all pixels become white as intensity increases beyond that value. At low intensity, the feature contained enough unsaturated pixels to bleed into the surrounding black frame.

In all cases within the above limits the correct feature was the best match for the stored landmark In general, however, this best-match property depends

Figure 3: **Two Canonical Scenes.** The higest-scoring features are shown as white in these scenes. In the top scene (#1), the reddish-brown mountains from a wall poster are the top feature. In the bottom scene (#2), an orange tool box stands out as the most salient image region.

Scene	**Parameter**	**Low**	**High**	**Res.**
#1	Distance	90cm	350cm	10cm
#1	Angle	0°	50°	5°
#1	Intensity	45	255	10
#2	Distance	120cm	410cm	10cm
#2	Angle	0°	40°	5°
#2	Intensity	45	255	10

Figure 4: **Limits of Reproduceability** The ranges under which the canonical landmark is refound as the best match with score greater than a threshhold (0.95). The last column is the measurement resolution used for each environmental parameter tested.

on the number of distractors in the scene of similar color composition. For scenes with many distractors, additional comparisons can help distingush landmarks from one another. For example, the match score for the *frames* can also be used to distinguish among a number of closely matched landmarks. From the start, the color interest operator places a premium on uniqueness in the scene, through the distinctive range, so that many easily-confusable patches are unlikely to receive high "interest" scores.

To test the reproduceability of landmarks under illuminants of different compositions, the same scene was viewed under five lights: halogen, fluorescent, incandescent, red-filtered, and blue-filtered. The halogen light was considered the canonical composition; match scores were then taken with the other four lights. Note that a color constancy algorithm was not used in this experiment; pre-processing the image to appear as if illuminated by a standard illuminant, (Forsyth 1990), would further improve the results. The three regions considered were (1) the reddish mountains, (2) a blue portion of the poster beside them, (3) a low-saturation, off-white patch (one of the image regions deemed least suitable for use as a landmark). The results, shown in Figure 5, demonstrate the sensitivity of the interest operator to large changes in illumination composition. Under the colored illuminants, all three features showed greatly reduced match scores, though the green and red illuminants respectively matched the color of the bluish and reddish landmarks enough to maintain some degree of match. The results also reinforce the importance of saturation within the interest operator. The difference between the fluorescent and halogen lights sufficed to change the pixels of the low-saturation patch to the point of unrecognizability, while the match values between the three uncolored lights are high enough to apply a high threshhold for accepting matching features – a threshhold also applicable to large variations of illumination intensity and camera viewpoint.

Region	**Hal.**	**Flo.**	**Inc.**	**Red**	**Green**
1	0.99	0.99	0.94	0.62	0.00
2	0.99	0.98	0.99	0.00	0.98
3	0.99	0.06	0.98	0.01	0.00

Figure 5: **Match Scores for different lights** The values represent the histogram inner products of three landmarks found under halogen (the canonical illuminant), fluorescent, incandescent, red-filtered, and green-filtered lights.

Conclusion

This work has presented an operator which locates landmarks suitable for representing an environment even when changes in that environment are possible. The sensitivity of this color interest operator to the camera's viewing geometry, the ambient illumination contrast, and the illuminant composition have been considered.

In general, indoor environments often contain many potential color landmarks. Fire extinguishers and exit signs are usually a bright patch of red against more modestly colored backgrounds. This is not to say that such features are dense in any environment. Certainly there are uniform or uniformly-textured areas without strong features. Many environments do, however, contain sufficient variety of landmarks to support a feature-based navigation system. We are extending this work with a simple agent which uses salient image regions to interact with its environment.

Acknowledgments This research was supported by ARPA grant N00014-93-1-1235, Army DURIP grant DAAH04-95-1-0058, by National Science Foundation grant IRI-9420982, and by funds provided by Yale University.

References

Baumgartner, E., and Skaar, S. 1994. An autonomous vision-based mobile robot. *IEEE Trans. on Automatic Control* 39(3):493–502.

Baumgartner, E., and Skaar, S. 1996. Region competition: Unifying snakes, regiongrowing, and bayes/mdl for multiband image segmentation. *IEEE Trans. on Pat. An. and Machine Intel.* 18(9):884–900.

Beveridge, J.; Griffith, J.; Kohler, R.; Hanson, A.; and Riseman, E. 1989. Segmenting images using localized histograms and region merging. *Int'l Journal of Computer Vision* 2(1):311–347.

Engleson, S. 1994. *Passive Map Learning and Visual Place Recognition*. Ph.D. Dissertation, Yale U.

Forsyth, D. 1990. A novel algorithm for color constancy. *Int'l Journal of Computer Vision* 5(1):5–36.

Funt, B. 1995. Color-constant color histograms. *IEEE Trans. on Pat. An. and Machine Intel.* 17(5):522–529.

Hager, G. D., and Rasmussen, C. 1996. Robot navigation using image sequences. In *Proceedings, AAAI*, 938–943.

Huttenlocher, D.; Leventon, M.; and Rucklidge, W. 1990. *Algorithmic Foundations of Robotics*. A. K. Peters. 85–96.

Liu, J., and Yang, Y. 1994. Multiresolution color image segmentation. *IEEE Trans. on Pat. An. and Machine Intel.* 16(7):689–700.

Moravec, H. 1983. The stanford cart and the CMU rover. *Proceedings of the IEEE* 71(7):872–883.

Panjwani, D., and Healey, G. 1995. Markov random field models for unsupervised segmentation of textured color images. *IEEE Trans. on Pat. An. and Machine Intel.* 17(10):939–954.

Perez, F., and Koch, C. 1994. Toward color image segmentation in analog vlsi: Algorithm and hardware. *Int'l Journal of Computer Vision* 12(1):17–42.

Swain, M., and Ballard, D. 1991. Color indexing. *Int'l Journal of Computer Vision* 7(1):11–32.

Zheng, J., and Tsuji, S. 1992. Panoramic representation for route recognition by a mobile robot. *Int'l Journal of Computer Vision* 9(1):55–76.