In Jähne et al., Mustererkennung 1996, 18. DAGM–Symposium, Springer, 1996

Color–Orientation Indexing

Udo Mahlmeister*, Harro Pahl, and Gerald Sommer

Institut für Informatik
Christian–Albrechts–Universität zu Kiel
Preußerstrasse 1–9, D–24105 Kiel, Germany
email: uhm@informatik.uni-kiel.de

Abstract. Observing the development of content based image retrieval systems hindered by the lack of efficient image representations, color histogram based indexing techniques have been used quite successfully. Though their performance strongly depends on illumination conditions being controlled, there has been only small effort to make them invariant to illumination. By introducing color–orientation histograms we present an integrated representation for color and local orientation, achieving robustness to several illumination conditions for free. Our method involves steerable filter techniques and Lab–color space conversion.

1 Introduction

At the latest since the Internet's World Wide Web has gained widespread use, the enormous gap between the importance of visual information on the net and the facilities to retrieve it from there through slow communication channels has become obvious. Not only in distributed databases visual information is handled as an appendix to symbolic information and accessed via annotations or manually edited meta–information in an unflexible way. At the same time, the amount of imagery is increasing by far more rapidly than the network bandwidth to deliver it. To protect storage, transmission, and receiving systems, besides the human consumer from congestion, new efficient access and compression structures have to be developed. Expecting the today's non–interactive techniques soon to fail this challenge, a new generation of *Visual Information Management Systems (VIMS)* has been sketched by several authors[5, 6, 11]. VIMS are intended to seamlessly integrate visual, auditory, and symbolic information in many application domains, such as Geographic Information Systems, Engineering Visualization Systems, Medical Information Systems or Education Systems. The key paradigm of VIMS is *content based image retrieval.* This includes the features *query by example*, i.e. find all images similar to this, *query by image syntax*, e.g. find all images with blue vertical stripes on top, *automatic or interactive annotation* of large databases, and, mostly underestimated, *progressive transmission* guided *by visual attention.*

In contrast to traditional image retrieval systems, VIMS are designed to find out the users interest, perhaps interactively, in order to control the process of selecting, compressing and transmitting images. Probably, the next order of magnitude in compression rates can only be achieved by considering the users interest at an early stage of the coding process. The earlier the users interests come into play in the line of transmission, the more efficient visual information can be handled. Furthermore, if transmission of a single image is no longer considered as an instant, but as a process in time, transmission costs could be saved by sorting and transmitting image regions in a "most interesting first"

* partially supported by DFG, grant So 320/1–1 and Ei 322/1–1

order, allowing the receiver to abort transmission as early as possible. Simple progressive transmission strategies are known for years but have not yet been used in the context of content based image retrieval or visual attention, perhaps due to the lack of efficient methods describing and locating regions of interest[16].

Common to all variations of content based image retrieval is the need for a representation which (1) facilitates users to express queries, i.e. constitutes the primitives for some natural language based image syntax, (2) provides measures for perceptual similarity, (3) has fast extraction algorithms (4), facilitates efficient database indexing (5), is invariant or at least robust to lighting conditions and camera/object position. We could meet the first point, if we provided a complete segmentation and identification of objects. Unfortunatly these two problems are still unsolved in computer vision because of their intractable genericity. Even if they were tackled, their complexity would be prohibitive for interactive applications. Since Swain et al.[15] showed the discrimination performance of histograms on color values in their work on "Color Indexing", there seems to be a way out of this dilemma. Without using complex spatial or geometrical information they could index into large databases with considerable match percentiles. Nevertheless, more recent work showed, that using spatial information could increase the performance of color indexing and provides the user with a more expressive image description language[13]. Besides, simple color indexing schemes seem to be very sensitive to lighting conditions, if they don't have a color constancy algorithm working in front.

In this paper we will show, that color indexing reveals to be a powerful representation when completed by the feature of local orientation. In the next section we will give an overview on some variations of color indexing. The integration of local orientation into color histogram processing is described in the third section. The experiments in section four will reveal the superior robustness of color orientation indexing in comparison to related color indexing techniques, using a testbed of images taken under a variety of conditions.

2 Color Indexing

Swain and Ballards "Color Indexing" method has been modified several times[3, 2] to compensate for some of its weaknesses, though the strength of its simplicity is still valid. Color Indexing is based on simple operations with color histograms, i.e. histograms on pixel values of an image in some color space (RGB, Luv, Lab):

Because color histograms are computed on images without prior figure–ground segmentation they are quite efficient low–level representations. Additionally they are robust to occlusion and small changes in view. Unfortunately, they depend strongly on intensity and color of lighting. Color histograms can handle two major problems important to VIMS, depending on the operation performed on them: object *localization* is achieved by histogram backprojection and *identification* by histogram intersection. In this section both methods are analyzed with respect to invariances and robustness.

2.1 Histogram Intersection

To solve the key problem of indexing, "given an image, show me the best matching models" a similarity measure between images for producing a similarity ranking must be provided. Histogram intersection yields such a measure by computing, how many of the pixels in the model are found in the image. For an image histogram I and a model histogram M, the match value $H(I, M)$ is calculated as the intersection with the image histogram I

normalized by the number of pixels in the model[15]:

$$H(I, M) := \sum_{i=0}^{N-1} \min \{I_i, M_i\} \tag{1}$$

where N is the number of color bins. Each histogram bin contains the frequency of a color which is a measure for the area occupied by that color. Obviously, intersection values are rotation and shift invariant, because histogram counts don't change under these transformations. Normalizing $H(I, M)$ by the number of pixels in the model does not make the intersection value in eq. (1) invariant to scaling. Models would have to be scaled according to another scale/distance cue before histograming.

Recently it has been proposed to match the first few moments and central moments of histograms[13]. The goal is to reduce the index dimensionality and to make the matching more robust to any kind of distortions.

2.2 Histogram Backprojection

The answer to "Where in the image are the colors that belong to a given model?" is found by generating a confidence image a and determining the locations of its maxima. The confidence measure R_i backprojected to a for each input pixel with color i is defined by:

$$R_i := \min \left\{ \frac{M_i}{I_i}, 1 \right\} \tag{2}$$

This ratio emphasizes locations whose color is highly present in the model and rarely present in the image. On the other hand, locations are suppressed whose colors are rarely seen in the model but very often seen in the image. The pure backprojection image is quite noisy, so it has to be smoothed before maximum detection.

Computing histograms locally, in a sliding window manner, a local confidence measure can also be derived by sliding intersection with the model histogram[2]. In fact, histogram backprojection has been shown to be a special case of this more general method.

2.3 Color Constancy

In general, lighting cannot be controlled in natural environments. While the apparent color of object surfaces is strongly affected by illumination, an efficient color constancy mechanism is crucial to recognizing objects by color cues.

Modeling intensity and color of illumination as offset in color space[8], a simplification anyway, a change in illumination will result in a shift in histogram space. This shift cannot be detected and compensated reliably. Histogram based methods therefore cannot be invariant to illumination. This fact was reported in the literature[15, 3] and is also confirmed by our experiments. For future work, Swain et al. suggested to switch a global color constancy algorithm (e.g. [9]) in front of histograming. Because the complexity of algorithms could exceed the complexity of indexing, we agree with Funt et al.[3], as this step would destroy the overall elegance and efficiency of "Color Indexing".

Small changes in illumination can be absorbed naturally by coarse quantization at the expense of discrimination performance. Though Swain et al. reported insensitivity to quantization parameters[15], we and also Stricker et al.[14] found that quantization and the choice of a proper color space is critical for color constancy in pixel based algorithms.

To cope also with spatially varying illumination, Retinex–like algorithms[4] based on linear filtering[8, 3] have proved to be better suited for indexing. Basically, these algorithms extract the color value ratio between adjacent color patches by linear filtering in a proper color space, assuming the ratios invariant to illumination intensity and color. The method is also biologically motivated by color opponent receptive fields found in human's visual cortex.

2.4 Spatial Information

The success of color histogram based methods is primarily due to the fact that they don't use cues except color, renouncing completely from spatial and geometrical features and their inherent complexity. It has been argued, that spatial features are scale specific thus making scale (and distance) invariant processing more difficult.

Nevertheless, Stricker et al.[13] realized that weak global spatial information improves the performance of color indexing. They used a pseudo segmentation of images into five fuzzy regions, matching each region separately. They also introduced the use of moments and central moments directly on pixels rather than histograms to describe color distributions.

A very important family of cues has been completely ignored by the color indexing community: local geometrical features. In the paradigm of "local geometry", initiated by Koenderink[7], the geometrical properties of an image point are described by local jets i.e. sets of directional spatial derivatives. Completed with steerable filter techniques, and assisted by differential geometry and Lie group theory, more complex structures may be described in a generic framework[10]. Though this is a very flexible means of description, it assumes precise selection and localization of points, which is a very difficult task at small scales.

In the following we will outline our strategically different approach to deal with local geometry, which naturally integrates color cues. Instead of using the vector of local jets to represent the geometry of a single point, we consider the distribution of local orientations in its neighborhood. The angle of local orientations is computed as the center of mass of the local energies at a set of orientations, exploiting the properties of steerable filters. While this method has been used to grey level images only[1], we successfully applied it to Lab–color images, providing a smart representation of color and local orientation cues[8].

3 Color–Orientation Histograms

The algorithm to compute *color–orientation histograms (COH)* is as follows:

1. Convert RGB–color images coming from the camera to Lab color space.
2. Apply a set of Andersson basis filters to each color channel in order to extract local energy at four orientations $\theta = 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$.
3. Compute the local orientation argument $\varphi_\ell(x, y)$ and amplitude $a_\ell(x, y)$ images for each color channel $\ell \in \{\mathrm{L}, \mathrm{a}, \mathrm{b}\}$.
4. Quantize the joint Lab–argument–amplitude space into a set of color indices.
5. Compute the histogram.

The CIE–Lab color space[12] was designed to realize a perceptual distance measure with simple Euclidean metric. This property facilitates the construction of a measure for perceptual similarity between color images, as claimed in point (2) of section 1. Also, the non–linearity of Lab converts multiplicative signal components due to illumination into offsets which can be suppressed by bandpass filtering, thereby achieving color constancy. The orientation steerable filters used in step 2 were designed by Knutsson and

Andersson[1]. Their application to Lab–images is described in more detail in an earlier paper[8]. By computing the center of mass from four orientation responses, we obtain the argument and magnitude of local orientation, i.e the angle and contrast of color edges in the image. Before histograming we need to quantize each $6D$–feature vector into a color orientation index. The vector quantizer design is guided by the following assumptions: (a) the orientation of color edges is of higher interest than their intensity and (b) luminance orientation is more important than chrominance orientation (c) orientation of edges below a certain strength should be ignored.

Therefore, in each color channel orientation is considered only, if energy exceeds a certain channel specific relevance threshold. Due to color constancy, we need not adapt these thresholds to images. If luminance energy of a pixel falls below this threshold, chrominance orientation is discarded and this pixel is excluded from histogram calculation. The orientation range $[0, \pi[$ of luminance and

4 Results

The following experiments are designed to compare the performance of COHs with Swain's simple RGB–histograms[15], and Funt's histograms[3] when used with the standard intersection and backprojection algorithm. The aim of our tests is to demonstrate the superior robustness of COHs to variations in several illumination parameters: color, intensity, and quality (shading etc.). To avoid combinatorial explosion due to variations of illumination parameters, we restrict our tests to a rather small model image database. Because of the fact that it is impossible to control illumination parameters in real world applications, experiments to demonstrate discrimination performance without varying illumination are of questionable use. Even simulating illumination with known spectral reflectance distributions[3] cannot substitute real illumination experiments, because effects on the A/D–conversion, e.g. saturation effects, are not considered appropriately.

The test images we use in our experiments contain bottles and cans with beverages of several brands in upright pose (see fig. 1). To compare the matching performance of the three histogram types, the rankings and tolerances for several illuminations are listed pairwise in tables 1 and 2. A Rank n says that the correct model corresponds to the nth highest match value. The tolerance is calculated as the difference between the correct match value and the second best match value. For robust matching high tolerance values are to be attained. The overall matching performance is given as the average match percentile (see [15] for a definition), denoted by "amp" at the bottom of tbls. 1 and 2. Also, the average tolerances, denoted by "at", are listed in the tables. The center frequencies of Andersson filters are set to 64 cycles/image Fig. 1,r. shows a quantized local orientation map.

Swain's RGB–histograms have not been tested under colored illumination for they are far from color constant. Comparing the ranks `green` and `blue` in tbl. 1 and 2, respectively, even Funt's histograms can not be considered color constant, though this has been claimed to be the main feature of his representation. With COHs, a perfect match cannot be achieved either, but correct models are recognized as at least second best matches. Neither too dim nor too bright illumination affects COH's rankings considerably, as indicated by columns `dark` and `sat` , though histograms of the original RGB–images are rather distorted, holding 1/3 of pixels in the last bin of each channel.

The quality of light is also a parameter which cannot be controlled in real world. Introducing the illumination `shady` , we tried to produce sharp shades by simply reducing diffuse reflections. Even under these circumstances which induce strong non–linear distortions into RGB–histograms, COHs performed well in comparison to Funt's and Swain's

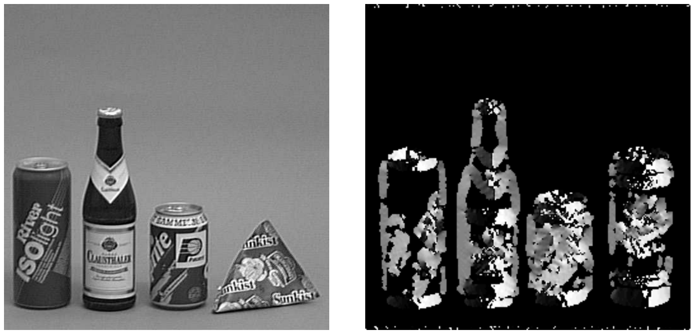

Fig. 1. Left: A typical test image, Original resolution 512×512, 3×8bit RGB, Camera: Sony 77 at Spectral sensitivity 3600 K, gain 0 dB. The model images are not cut from the original images but shot in a separate session with illumination similar to `std` –illumination. Right: quantized local orientation map. $[0, \pi] \rightarrow [\text{black}, \text{white}]$

histograms. By leaving the luminance component out of account, this effect could be further enhanced, because shades induced by white light mainly influence the luminance of histograms.

Color–orientation histograms						
model	`dark`	`std`	`shady`	`green`	`blue`	`sat`
Iso	1 .126	1 .289	1 .275	1 .072	1 .191	1 .239
Claust	1 .030	1 .084	1 .012	2 .048	2 .001	1 .024
Sprite	1 .101	1 .165	2 .013	1 .018	1 .110	1 .055
SunOr	1 .195	1 .245	1 .270	1 .182	1 .140	1 .191
Kings	1 .079	1 .132	1 .071	2 .025	1 .093	1 .065
Holsten	1 .051	1 .078	1 .048	2 .020	1 .025	1 .058
7Up	1 .097	1 .137	1 .124	1 .210	1 .057	1 .086
Coke	1 .085	1 .106	1 .082	1 .210	1 .055	1 .132
SunZi	1 .184	1 .167	1 .043	1 .052	1 .121	1 .194
Punika	1 .085	1 .164	1 .120	2 .015	1 .029	1 .097
amp	100.0%	100.0%	98.9%	95.6%	98.9%	100.0%
at	0.103	0.157	0.106	0.078	0.082	0.114

Table 1. Intersection matching ranks (first column) and tolerance values (second column) for 10 objects (names on the left) under 6 illumination conditions (`std` = standard, `sat` = saturated) using **COHs**. Relevance threshold at 0.5. Quantization intervals: 16 for L–, 8 for a– and b–channel. The average matching percentile (amp) and average tolerance (at)

In a last experiment, we investigated histogram backprojection in COHs in order to localize bottles in test images by simple maximum detection. In tbl. 3, position ranks for COHs and Swain's histograms are compiled. Position rank n indicates, that the correct position of the model is localized at the nth search iteration. After each step, the detected region is suppressed to drop it from the maximum detection of the next step. Consistent with Funt's results, we found that our histogram works quite well under all illumination condition, except for slight degradations with standard(`std`) illumination.

Funt's histograms						
model	`dark`	`std`	`shady`	`red`	`green`	`blue`
Iso	1	1	1	1	1	1
Claust	8	2	5	8	6	8
Sprite	1	1	1	1	1	1
SunOr	1	1	1	1	2	3
Kings	2	1	1	2	3	4
Holsten	5	1	2	5	4	5
7Up	2	1	1	6	2	5
Coke	6	3	3	3	3	4
SunZi	3	2	4	4	3	4
Punika	9	1	2	8	5	8
amp	71%	96%	88%	74%	73%	60%

Swain's RGB–histograms			
model	`dark`	`std`	`shady`
Iso	1 .350	1 .474	1 .106
Claust	7 .067	1 .084	1 .030
Sprite	1 .186	1 .331	1 .171
SunOr	2 .019	3 .019	2 .000
Kings	1 .028	1 .329	1 .010
Holsten	2 .044	1 .106	1 .007
7Up	2 .042	1 .227	1 .221
Coke	1 .020	1 .175	1 .133
SunZi	2 .086	2 .016	3 .047
Punika	6 .032	1 .353	2 .037
amp	83.3%	97.8%	95.6%
at	0.874	0.211	0.076

Table 2. a) Intersection matching ranks for 10 objects (names on the left) under 6 illumination conditions (`std` = standard) using **Funt et al. histograms** from Laplacian filtered RGB–images. Quantization intervals per channel: 16. In the last row the average matching percentiles (amp) are listed. b) Same under 3 different illuminations (`std` = standard) using simple **RGB histograms**.

Method	**Rank**	`dark`	`std`	`sat`	`red`	`green`	`blue`	**Total**
	$= 1$	16	16	17	12	15	15	91
COH	$= 2$	1	1	0	3	1	1	7
	≥ 3				2	1	1	4
	$= 1$	12	17	15	4	7	9	64
Swain	$= 2$	4		2	1	3	1	11
	≥ 3	1			12	7	7	27

Table 3. Distribution of position ranks for COHs and Swain's histograms.

5 Conclusions

Introducing color orientation histograms (COHs), we demonstrated that color indexing techniques benefit twice from the use of local orientation as an additional feature. First, the significance of local orientation for sparse image representations improves the discrimination performance of both localization and identification methods based on color histograms. Secondly, illumination invariant cues are present only at color edges, the sources of local orientation in the image. We showed the superior robustness of COHs to changes in illumination color, intensity, and quality. To describe local geometry, we proposed to use local orientation gathered in the neighborhood of a point via histograming instead of relying on the information from inaccurately positioned local jets. Also, we showed that splitting up the color signal into luminance and chromaticities components by using the Lab–color space facilitates adaptation to the role of each color component in image formation, while providing a coherent measure for perceptual similarity.

The additional computational complexity induced by COHs is dominated by linear filtering, which could be parallelized effortlessly. This makes the goal attainable, to use COHs not only for VIMS but also for localization tasks in real–time robotic applications. Tracking human faces with the stereo camera head will be our next step in this promising direction.

References

1. M.T. Andersson. *Controllable multidimensional filters and models in low level computer vision.* PhD thesis, Linköping University, 1992.
2. F. Ennesser and G. Medioni. Finding waldo, or focus of attention using local color information. *IEEE Trans. PAMI*, 17(8):805–809, 1995.
3. B. V. Funt and G. D. Finlayson. Color constant color indexing. *IEEE Trans. PAMI*, 17(5):522–529, 1995.
4. B. K. P. Horn. Determining lightness from an image. *Computer Graphics and Image Processing*, pages 277–299, 1974.
5. R. Jain. NSF–ARPA workshop on visual information mangement systems. Workshop report, Univ. of Michigan, Ann Arbor, 1993.
6. R. Jain, A. P. Pentland, and D. Petkovic. NSF–ARPA workshop on visual information mangement systems. Workshop report, Univ. of California at San Diego, 1995.
7. J. J. Koenderink. The structure of images. *Biol. Cybern.*, 50:363–370, 1984.
8. U. Mahlmeister, B. Schmidt, and G. Sommer. Preattentive colour features by steerable filters. In G. Sagerer et al, editor, *17. DAGM*, pages 464–472. Springer, 1995.
9. L. T. Maloney and B. A. Wandell. Color constancy: a method for recovering surface spectral reflectance. *Opt. Soc. Am. A*, 3:29–33, 1986.
10. M. Michaelis and G. Sommer. Junction classification by multiple orientation detection. In J.-O. Eklundh, editor, *ECCV*, volume I, pages 101–108. Springer, 1994.
11. R. W. Picard. Light–years from Lena: Video and image libraries of the future. Tech. report, M.I.T. Media Labaratory Perceptual Computing Section, 1995.
12. W. F. Schreiber. *Fundamentals of Electronic Imaging Systems.* Springer, 3rd edition, 1993.
13. M. Stricker and A. Dimai. Color indexing with weak spatial constraints. In *Storage and Retrieval for Image and Video Databases IV*, volume 2670 of *SPIE Proceedings Series*, Feb. 1996.
14. M. Stricker and M. Orengo. Similarity of color images. In *Storage and Retrieval for Image and Video Databases III*, volume 2420 of *SPIE Proceedings Series*, pages 381–392, Feb. 1995.
15. M. J. Swain and D. H. Ballard. Color indexing. *Internat. Journ. of Computer Vision*, 7(1):11–32, 1991.
16. H. Zabrodsky and S. Peleg. Attentive transmission. *Journ. of Visual Communication and Image Representation*, 1(2):189–198, 1990.