ACKNOWLEDGEMENTS:

The authors thank B. Crowley and N. Ostrofe for the preparation of this manuscript. This research was supported by the U.S. Dept. of Energy, Office of Scientific Computing, under Contract DE-AC03-76SF00098, and in part by the National Institutes of Health, under grant number AIO8427-25.

REFERENCES

[1] A.A. Amini, R.W. Curwen, R.T. Constable, and J.C. Gore, "MR physics-based snake tracking and dense deformations of tagged mr cardiac images," *AAAI Conf. Applications of Computer Vision to Medical Image Processing*, pp. 126-130, 1994.

[2] A. Amini, T. Weymouth, and R. Jain, "Using dynamic programming for solving variational problems in vision," *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp. 855-867, 1990.

[3] R.A. Brooks, "Model-based three-dimensional interpretations of two-dimensional images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 5, no. 2, pp. 140-150, 1983.

[4] C. Bustamante, "Direct observation and manipulation of single dna molecules using fluorescence microscopy," *Ann. Review of Biophysics and Biophysical Chemistry*, vol. 20, pp. 415-446, 1991.

[5] C. Bustamante, T. Houseal, D. Beach, and F.M. Maestre, "Fluorescence microscopy of the dynamics of supercoling, folding, and condensation of bacterial chromosomes, induced by acridine orange," *J. Biomolecular Structure and Dynamics*, vol. 8, pp. 643-655, 1990.

[6] J.F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 670-698, Nov. 1986.

[7] Q. Huang and C. Stockman, "Generalized tube model: Recognizing 3D elongated objects from 2D intensity images," *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 104-109, New York, June 1993.

[8] A. Huertas, W. Cole, and R. Nevatia, "Detecting runways in complex airport scenes," *Computer Vision, Graphics, and Image Processing*, vol. 51, no. 2, pp. 107-145, Aug. 1990.

[9] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int'l J. Computer Vision*, vol. 1, pp. 321-331, 1988.

[10] F. Leymarie and M.D. Levin, "Tracing deformable objects in the plane using an active contour model," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, pp. 617-635, 1993.

[11] C. Nastar and N. Ayache, "Classification of nonrigid motion in 3D images using physics-based vibration analysis," *IEEE Workshop Biomedical Image Analysis*, pp. 61-69, 1994.

[12] B. Parvin, C. Peng, W. Johnston, and M. Maestre, "Tracking of tubular objects for scientific applications," *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 295-301, 1994.

[13] A. Pentland and B. Horowitz, "Recovery of nonrigid motion and structure," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, pp. 730-742, 1991.

[14] A. Pentland and S. Sclaroff, "Closed-form solutions for physically based shape modeling and recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, pp. 715-729, 1991.

[15] R. Bellman, *Dynamic Programming*. Princeton Univ. Press, 1957.

[16] L. Song and M. Maestre, "Unhooking dynamics of u-shaped dna molecule undergoing gel electrophoresis," *J. Biomolecular Structure and Dynamics*, vol. 9, pp. 87-99, 1991.

[17] D.J. Williams and M. Shah, "A fast algorithm for active contours and curvature estimation," *Computer Vision, Graphics, and Image Processing*, vol. 55, no. 1, pp. 14-26, 1992.

Finding Waldo, or Focus of Attention Using Local Color Information

François Ennesser and Gérard Medioni

***Abstract*—We present a method to locate an "object" in a color image, or more precisely, to select a set of likely locations for the object. The model is assumed to be of known color distribution, which permits the use color-space processing. A new method is presented, which exploits more information than the previous Backprojection Algorithm of Swain and Ballard at a competitive complexity. Precisely, the new algorithm is based on matching *local histograms* with the model, instead of directly replacing pixels with a confidence that they belong to the object. We prove that a simple version of this algorithm degenerates into Backprojection in the worst case. In addition, we show how to estimate the scale of the model.**

Results are shown on pictures digitized from the famous "Where is Waldo" books. Issues concerning the optimal choice of a color space and its quantization are carefully considered and studied in this application. We also propose to use *co-occurrence histograms* to deal with cases where important color variations can be expected.

***Index Terms*—Object recognition, focus of attention, color images, color quantization, color histograms.**

I. INTRODUCTION

Many known algorithms are able to locate an object in a complex scene (see [6] for a recent overview). However, most of them require expensive processing of the entire scene (see, for example, [4]). Such algorithms do not solve the focus of attention problem, where the only place of interest to the observer lies around the object itself, allowing the greatest part of the scene to be quickly discarded.

To address the focus of attention problem, new indexing algorithms have been proposed that allow for much faster processing, assuming that the object of interest has a known color distribution (see [1]).

The original Backprojection algorithm of Swain and Ballard [1], analyzed in Section II, performs a point-by-point processing, relying only on statistical information relative to the entire scene. Since its performance depends on the color space used and its quantization, these issues are treated in detail. In Section III, we present an algorithm that exploits more information than backprojection at a competitive cost by directly matching local sub-images with the model. Since this new method still appears to be sensitive to color variations, we address this issue in Section IV.

The reliability of the new method is demonstrated by extensive testing performed on three Waldo pictures (see Fig. 1. a-c): the Beach, Castle Siege, and Pirates images. These images were obtained by scanning the printed copy of the book "Where is Waldo: The Magnificent Poster Book" by Martin Hanford (with permission), on an Optronics scanner, using three different color filters, and a scanning spot size of 400μ on the 20cm × 20cm portion of the beach image and 200μ on the 10cm × 10cm portions of the other pictures. This process is guaranteed to introduce a significant amount of halftone noise to the images. Furthermore, these images are specifically designed to fool the observer by "hiding" the target, offering instead many lures and significant occlusion. The model was a 32×80 image extracted from the Beach picture, digitized with a pixel size of 100μ (see Fig. 1. d). Satisfactory results were also obtained on digitized video images, despite variations in scale and lighting conditions.

Manuscript received June 14, 1994; revised Mar. 3, 1995.

The authors are with the Institute for Robotics and Intelligent Systems, University of Southern California, Los Angeles, CA 90089-0273; e-mail: medioni@iris.usc.edu.

IEEECS Log Number P95075.

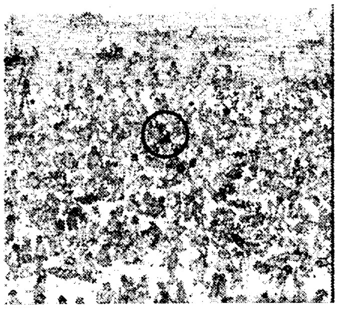


Fig. 1a. The Beach image (512 × 512).

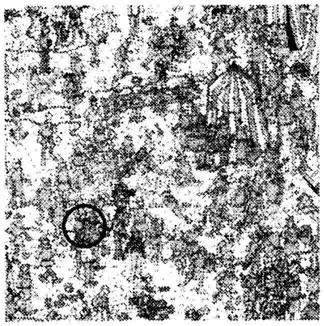


Fig. 1b. The Castle Siege image (500 × 500).

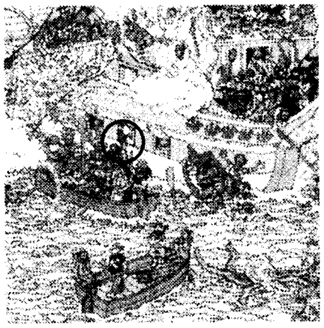


Fig. 1c. The Pirates image (500 × 500).

Fig. 1d. The Beach model (32 × 80 in high resolution).

II. THE BACKPROJECTION ALGORITHM

The novelty of Swain and Ballard's *backprojection algorithm* [1] resides in the use of *color information instead of spatial information.* Its basic features are histograms constructed on some (quantized) color space.

Let I be the histogram of the scene and M the model histogram. The algorithm first computes the "quotient" of the model histogram with the scene histogram, $M/I = \{M_i/I_i\}_{buck.i}$. This constructs a look-up table in the color space by replacing each color appearing in the image with a confidence value representing how much it is characteristic of the searched object. This histogram is then *backprojected* on the image: each pixel (x, y) of color i in the original image is replaced by its confidence value M_i/I_i. The expected locations should appear as peaks in the smoothed image.

The backprojection algorithm is a good focus of attention mechanism: considering an $n*n$ image and $p*p$ model and mask ($p << n$), the overall complexity of the algorithm is $O((n + p)^2 \log_2(n + p))$, where $b < (n + p)^2$ is the number of buckets in the color space.

A. Open Issues

The two most important factors affecting the efficiency of any algorithm using color space processing are the *color space* being chosen and its *quantization*.

The choice of color space has been adressed in [2], [3], [5], [12], [13]. We have tested a number of different spaces and found that linear transformations of the RGB space are quite good. 2D spaces require more care in quantization.

Proper quantization of the color space is essential. We have tested both uniform and adaptive quantization schemes. More buckets are needed in 2D spaces than in 3D spaces for equivalent results with uniform quantization. Our adaptive scheme based on clustering ([9], [14]) is trained with colors of the model and a "catch-all" bucket for non-model colors. When enough colors (>3) are present in the model, stable and meaningful clusters can be obtained. Equivalent results were obtained in 2D and 3D spaces whenever enough clean clusters could be defined.In other cases, we can use color edges, as explained later.

B. Drawbacks

The backprojection results validate the color space approach, but provide very noisy results. Furthermore, the approach does not check for the similarity in color distribution between model and target area, only their presence. Our method overcoming these problems is now presented.

III. THE LOCAL HISTOGRAMS ALGORITHM

Our algorithm for focus of attention requires the color distributions of the model and of the suspected object to be globally similar. This algorithm can only outperform backprojection and can be improved to handle important scale variations and can be efficiently implemented.

A. Overall Description

In our algorithm, we compute *local histograms*, the size and shape of which are fixed by the expected aspect of an instance. These histograms are then matched with the model.

Our matching measure is a *weighted histogram intersection* derived from the intersection method of Swain and Ballard [1]. Let L_i and M_i be the counts in bucket i for a local histogram and the model, respectively. The formula is $\sum_{buck.i} W_i \min(L_i, M_i)$,where W is a weight which can be chosen as M/I to correspond to the confidence values of Backprojection (I: histogram of the entire scene). W increases the selectivity of the method by making the measure more sensitive to some "important" colors.

B. Degeneracy Equivalence Result

Ideally, the model is rescaled to match the size of the local area it is intersected with, so that the intersection becomes maximum when $L_i = M_i$ for all buckets i: This requires the overall color distribution on the local area to be the most similar to that of the model. This way, more information is used than in the backprojection algorithm. If the instance size is underestimated, the algorithm is still likely to work, though more regions could give strong stimuli. On the other hand, when the scaling is badly overestimated, we can prove that *the degenerate Local Histograms algorithm still performs as well as Backprojection*: Even if the saturation at M_i never occurs (the object being much smaller than expected), with the weights $W_i = M_i/I_i$, the weighted intersection measure with no saturation becomes $\sum_{buck.i} W_i \min(L_i, M_i) = \sum_{buck.i} L_i(M_i/I_i)$ which is the sum of the confidence values of the colors on the area of the local histogram. This right-hand term is also the result of the backprojection algorithm when run with a rectangular window of "1" corresponding to the local area L as a smoothing mask.

C. Estimating the Scale

A slightly modified version of Local Histogramming can be used to detect instances of any size in an image. In this method, each local histogram begins with an initially small size, and is intersected with the model according to its current size. If the initial intersection is promising enough, the histogram is sequentially allowed to grow by a fixed amount in each cardinal direction, as long as this process does not decrease the normalized intersection measure, after updating and rescaling to the new size. This process goes on as long as changes are made. Then, the final area is removed from the image and memorized as a possible candidate, while a new local histogram starts to grow next to it. The process is repeated till the entire image has been covered.

The growing process is a greedy algorithm whose heuristic is the scaled intersection measure. To avoid local maxima, when growing in one direction decreases the measure moderately, a larger region can instead be checked for acceptance. Furthermore, to avoid having the object broken into pieces, each time a new local histogram begins to overlap on a previously computed region, the rectangular envelope of their union can systematically be checked for acceptance.

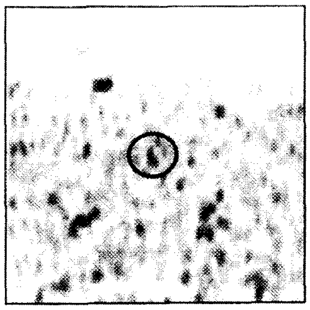
D. Complexity

We assume that the number of buckets b is no larger than n^2. Our implementation makes full use of the overlap between neighboring regions, replacing the computation of a $p*p$ histogram and an intersection of b buckets with an update using only $2p$ pixels and buckets. So the overall complexity is in $O((n + p)^2p)$. This should be compared to the backprojection method, which was in $O((n + p)^2\log_2(n + p))$: As long as the scene is large compared to the mask (precisely, $n > (2^p - p)$ for similar factors), the backprojection algorithm offer no advantage over local histogramming. This is the case in most interesting search problems.[5]

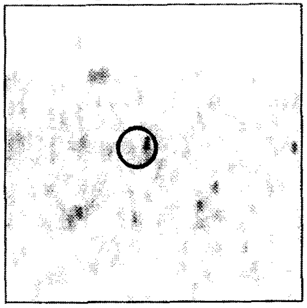
E. Results

Both backprojection and the Local histograms algorithm with weights $1/I$ and M/I have been implemented. A variety of color spaces have been tested with different quantization schemes.

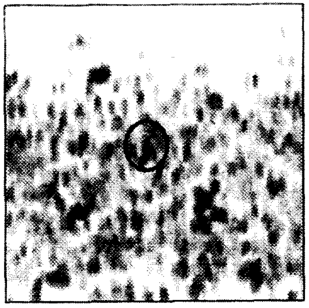
The Beach image gave correct results with uniform quantization, as shown on Fig. 2. Only local histogramming was able to find Waldo as its strongest peak, with weights M/I or $1/I$, the latter giving cleaner results. Other models and quantization schemes were also tested successfully.



(a) Backprojection



(b) W=1/I



(c) W=M/I

Fig. 2. Beach image results (RGB space, eight buckets per axis)

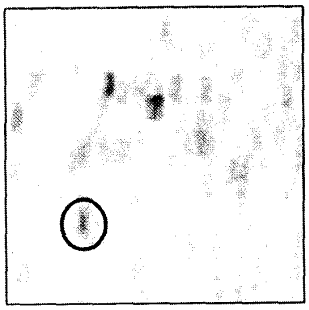
The Castle Siege image, chosen for the small size of the target, gave good results with Local Histogramming and any weighting, provided that appropriate quantization parameters were used. Even wrong scale estimates provided good results, although some selectivity was lost. Again, Backprojection was much less efficient. Fig. 3 shows these results.

Overall, any of the previous algorithms was able to find Waldo as one of its strongest peaks in each of the above pictures, with a uniform quantization using eight to 16 buckets per axis in 3D spaces, and 16 to 32 buckets in 2D spaces.

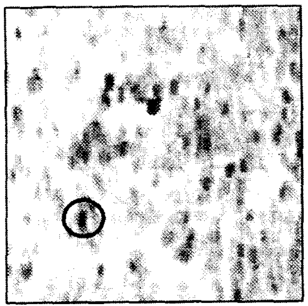
However, both methods are very sensitive to color variations, such as the Pirates image where Waldo is quite dark. The following scheme allows us to become robust to color variations.

---

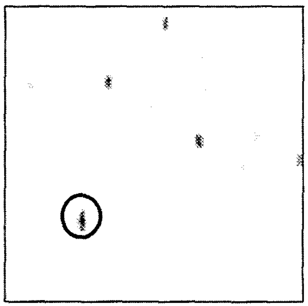
[5] The computational burden of Local Histogramming can be further reduced in practice by skipping histogram calculations when the measure is low and using a high-resolution input only when it becomes high.



(a) W=1/I, wrong scale = 2, 16 buckets.



(b) W=1/I, 8 buckets



(c) W=M/I, 16 buckets

Fig. 3. Castle image results (RGB space).

IV. TOLERANCE TO COLOR VARIATIONS

A. Co-occurrence Histograms

The color variation problem becomes critical when we need to rely heavily on color information, so that we need to distinguish between close hues. Unfortunately, due to possible changes in illumination, shadow, etc., the color of a given object may vary, creating, at best, a shift in the color space. A solution is to use improved features, by replacing the pixels considered in the classical histograms by a given pattern of points (say two to four points), the color of which would be read simultaneously to be stored as one information in the histogram: In this way, some spatial information is added, allowing us to relax our constraints on color discrimination. Precisely, instead of checking if the color of the pixel in position (x, y) is C, we now check if, for some ordered set of $\{(a_i, b_i)\}_{i=1...n}$, the colors of the pixels at $(x + a_i, y + b_i)$ are some ordered set of $\{C_i\}_{i=1...n}$ (e.g., a point is white, and its neighbor to the right is blue). Note that this is equivalent to instantiating the color variations between the chosen points, meaning that color edges are now distinguished, but such histograms will be more sensitive to rotation and scale variations.

This method multiplies the complexity of the algorithm and the dimension of the color space by the number of points used. Yet any color-using algorithm can still be applied in the same way, by just replacing the original k-bands color image by a $n \times k$-bands image, where the colors of the chosen set of n neighboring points have simply been overlapped. We refer to this technique as co-occurrence histogramming.

B. Results

The co-occurrence method has been tested with two and four points. For increased robustness with respect to scale, the histograms values were actually taken as a sum of the values obtained on two different sets of pixels, corresponding to two different distances between neighbors.

The Beach Image: Compared to the results of Backprojection without the co-occurence method, the gains that are obtained are significant in 3D color space while the results in 2D color space varies. We also investigated the influence of the parameters (pattern of points and size of the pattern). The optimal pattern of points to use depends on the color-edge characteristics of the model (breadth and orientation of the color strips, etc.), and the output peaks will basically retain those characteristics, making the method more selective. Averaging the histograms obtained for different sizes of the pattern makes this method robust to scale variation. The increased dimension of the filtering space provided by this method allows us to obtain a number of buckets that is sufficient to achieve good selectivity, even when 2D color spaces and coarse color quantization is used.

The Pirates image, chosen for its many lures, was impossible to treat with the plain method, because Waldo appeared much darker than in the model. A very coarse quantization was necessary for Waldo to create a peak, but then the selectivity was too poor. Even plain adaptive quantization was fooled by the lures. Co-occurrence histograms were the only way to combine color tolerance with sufficient selectivity.

In order to compare the simple and co-occurrence methods, we also searched the "Pirates" model in the "Pirates" image with many possible methods and parameters, tuning the quantization so that the same number of buckets was distinguished in any space: Since the colors were perfect, 2D and 3D spaces gave equivalent results. But the plain method was still fooled by the lures, while co-occurrence histogramming was definitely more selective. By varying the scale estimate in Local Histogramming, we verified that the co-occurrence method was not too sensitive to scale variations. Some results are shown in Fig. 4.

Real-world tests: Our algorithm was also successfully tested on video images from the outdoor world. We tried to locate various models in sequences of about 100 frames. The models were, of course, easier to locate than Waldo, but there were slight changes in orientation and illumination, important scale variations (1 to 4 ratio), as well as partial occlusions, and possible absence of the target object.

V. CONCLUSION

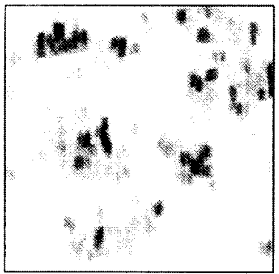
We have presented a new algorithm for focus of attention in color images based on local histogramming. It can be implemented at a relatively low computational cost and can be speeded up in many ways. It exploits more information than the former backprojection algorithm of [1]. Excellent results were obtained on complex pictures, using only low resolution images. Furthermore, our method can be modified to handle large scale variations at a low cost, by adaptively constructing the local histograms.

The issues that arise in indexing via color histograms, such as the choice of a color space and its quantization, were also considered. An adaptive quantization trained on the model is an appropriate quantization choice for a focus of attention, but the optimal color space depends on the application.

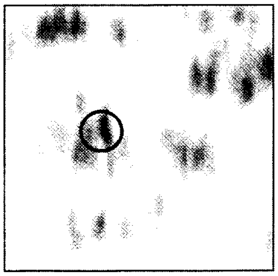
The main problem with such algorithms is that in general, color fidelity can not be expected. Our solution to that problem is to use co-occurrence (edge-based) histograms. This increases the dimension of the space used for discrimination, by incorporating some spatial knowledge. This allows the algorithm to perform well with reduced color information, but is computationally expensive. Our method can combine good selectivity with coarse quantization in 2D color representations, providing good robustness with respect to color variations.

(a) W=M/I, rb space, 24 buckets.

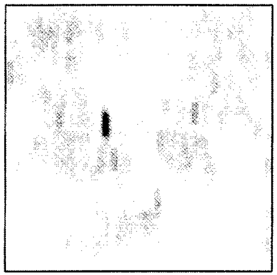
(b) W=M/I, RGB space, 8 buckets



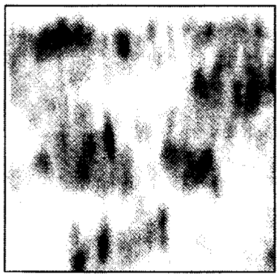
(c) W=1/I, RGB space, 8 buckets, wrong scale = 1/2



(d) W=1/I, gb space, 3 buckets.



(e) W=1/I, RGB space, 2 buckets



(f) Backprojection, RGB space, 2 buckets

Fig. 4. Pirate image results (perfect model, (a), (b), and (c) basic histogramming methods, (c), (d), and (e) co-occurrence histoprogramming with four pixels.

ACKNOWLEDGMENTS

We would like to thank the editors of the Waldo books, Little Brown & Co., for extending permission to use the images presented in this paper.

This research was supported by a National Science Foundation Grant under award No. IRI-9024369.

REFERENCES

[1] M.J. Swain and D.H. Ballard, "Indexing via color histograms," *Proc. ICCV* 90, pp. 390-393.

[2] W. K. Pratt, *Digital Image Processing*. John Wiley & Sons, 1978.

[3] C.L. Novak and S.A. Shafer, "Color vision," *Encyclopedia of Artificial Intelligence*, pp. 192-202, John Wiley & Sons, 1992.

[4] T.F. Syeda-Mahmood, "Data and model-driven selection using color regions," *Proc. ECCV* 92, pp. 115-123.

[5] G.E. Healey, "Segmenting images using normalized colors," *Physics-Based Vision: Principles and Practice—Color*, p. 166, 1992.

[6] W.E.L. Grimson, *Object Recognition by Computer—The Role of Geometric Constraints*. MIT Press, 1990.

[7] F. Stein and G. Medioni, "Recognition of 3D objects from 2D groupings," *Image Understanding Workshop*, San Diego, Calif., Jan. 1992.

[8] T. Pavlidis, *Algorithms for Graphics and Image Processing*. Computer Science Press, 1982.

[9] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*. John Wiley & Sons, 1973.

[10] A.K. Jain, *Fundamentals of Digital Image Processing*. New York: Prentice Hall, 1989.

[11] G. Finlayson, M. Drew, and B. Funt, "Diagonal transforms suffice for color constancy," *Proc. IEEE Int'l Conf. Computer Vision*, Berlin, 1993.

[12] D. Ballard and C. Brown, *Computer Vision*. Academic Press, 1982.

[13] Y. Ohta, T. Kanade, and T. Sakai, "Color-information for region segmentation, *CGIP*, vol. 13, p. 222.

[14] G.H. Bell and D.J. Hall, "ISODATA, A novel method of data analysis and pattern classification," NTIS Report AD 699616, 1965.