

0 OBJECT RECOGNITION USING MULTIDIMENSIONAL RECEPTIVE FIELD HISTOGRAMS AND ITS ROBUSTNESS TO VIEW POINT CHANGES

Bernt Schiele and James L. Crowley
GRAVIR, Institute IMAG, 46, avenue Félix Viallet,
38031 Grenoble Cedex, France

ABSTRACT

This chapter presents a technique to determine the identity of objects in a scene using multidimensional histograms of the responses of a vector of local linear neighborhood operators (receptive fields). This technique can be used to determine the most probable objects in a scene, independent of the object's position, image-plane orientation and scale.

The first part of the chapter summarizes the mathematical foundations of multidimensional Receptive Field Histograms [1] and gives a recognition example on a database of 103 objects. The second part of the chapter describes experiments to evaluate the robustness of multidimensional receptive field histograms to view point changes, using the Columbia image database [2]. In this experiment we examine the performance of different filter combinations, histogram matching functions and design parameter of the multidimensional histograms.

1. INTRODUCTION

In [1] we generalized the color histogram approach of Swain and Ballard [3] to the use of *Multiple Receptive Field Histograms*. The main idea of the approach is to calculate a multi-dimensional histogram of the response of a vector of receptive fields. An object in an image can be identified by matching the multi-dimensional histogram from a region of the image with a histogram of the sample of the object. In [1] we have shown that the technique can be used to identify objects in the presence of image plane rotation and scale.

In the color histogram approach of Swain and Ballard [3] an object in an image is identified by matching a color histogram from a region of the image with a color histogram from a sample of the object. Their technique has been shown to be remarkably robust to changes in the object's orientation, changes of the scale of the object,

partial occlusion or changes of the viewing position. Even changes in the shape of an object do not necessarily degrade the performance of their method. However, the major drawback of their method is its sensitivity to the color and intensity of the light source and color of the object to be detected.

Several authors have improved the performance of the original color histogram matching technique by introducing measures which are less sensitive to illumination changes. For example, Funt and Finlayson [4] propose the use of derivatives of the logarithms of the colors to provide color-constancy. Healey and Slater [5] have used the moments of color histograms to improve robustness to light intensity changes. Ennesser and Medioni [6, 7] have improved the performance to find a particular object in a difficult scene. Hunke, Schiele and Waibel [8, 9] have shown that normalizing the color vector by luminance provides a reliable means to detect skin color for tracking human faces. They exploit the simplicity of the approach to design a system for real-time tracking of human faces under varying conditions.

The color histogram approach is an attractive method for object recognition, because of its simplicity, speed and robustness. However, its reliance on object color and (to a lesser degree) light source intensity make it inappropriate for many recognition problems. The focus of our work has been to develop a similar technique using local descriptions of an object's shape provided by a vector of linear receptive fields [1]. For the Swain and Ballard algorithm, it can be seen that robustness to scale and rotation are provided by the use of color. Robustness to changes in viewing angle and to partial occlusion are due to the use of *histogram matching*. Thus it is natural to exploit the power of histogram matching to perform recognition based on histograms of local shape properties. The most general method to measure such properties is the use of a vector of linear local neighborhood operations, or receptive fields. In [1] we have compared sensitivity and recognition reliability for a variety of local neighborhood operations. Throughout the chapter we will use only the most successful functions.

A further advantage of the use of the histogram of receptive field vectors for recognition is that neither segmentation, nor an explicit geometric model of an object is needed. An object class is described by the histogram of its local characteristics. The experiments presented below and in [1] demonstrate that this technique can be used to discriminate arbitrary objects at different scales and orientations.

This chapter has three parts: The first part presents our generalization of the color histogram method (section 2., see also [1]). In the second part (section 6.) we give the results of an recognition experiment on a database of 103 objects. In the third part of the chapter we examine the robustness of the approach under view point changes. The experiments (section 7.) demonstrate that the technique is relatively robust to view point changes (3D-rotation). The experiments also show that we can use low resolution for each axis of a receptive field histogram and still obtain high recognition rates.

2. MULTIDIMENSIONAL RECEPTIVE FIELD HISTOGRAMS

In [1] we did identify three main parameters of the multidimensional receptive field histogram approach:

- The choice of local property measurements or the receptive field functions. In section 3. we will describe several local characteristics based on Gaussian derivatives, which we use throughout the chapter.
- Measurement for the comparison of the histograms. In [1] we summarized that χ^2 and “intersection” are the most suitable comparison measurements. Section 4. defines them.
- Design parameter of the histograms: number of dimensions of the histogram (each axis corresponds to one local property) and resolution of each axis (number of bins per axis).

In this chapter we use only local characteristics based on Gaussian derivatives, since they are equivariant to scale and image–plane rotation. Equivariant means that they vary in a uniform manner which is represented by a translation in a parameter space. We also describe two rotation invariant characteristics, namely the magnitude of the first Gaussian derivative and the Laplace operator.

As we saw in the experiments in [1], the comparison measurement determines the separability between histograms. The two best measures (of the experiments in [1]) for the histogram comparison are introduced in section 4.. In the experiments (see section 7.) we will examine the robustness of these comparison measurements under view point changes.

The experiments described in this chapter examine the dependency of the robustness to 3D–rotation and the design parameter of the histograms. The design parameters of the histograms determine the separability between the histograms of different objects (especially the number of local characteristics, which determines the number of dimensions of the histogram). An important parameter is the histogram quantification, of number of bins, which is used to represent each histogram dimension. Reducing the number of bins results in an improvement of the stability of the histogram with respect to noise, but also diminishes the discrimination between objects.

3. LOCAL CHARACTERISTICS

This section describes receptive field functions which we use in the experiments described below. These receptive field functions are namely the first Gaussian derivative, the magnitude and direction of the first derivative and the Laplace operator. We should mention that the approach is not restricted to local characteristics based on Gaussian derivatives. Nevertheless we think that these characteristics are well suited for the object recognition task, since they are equivariant (= steerable) to scale and 2D–rotation.

The calculation of local properties can be divided into the local linear point-spread function, and the normalization function used during measurements of local

properties. As normalization function we are using the Energie normalization function, which is described in section 3.2.. This normalization has been shown to be the most robust in the presence of additive Gaussian noise [1].

3.1. LOCAL CHARACTERISTICS BASED ON GAUSSIAN DERIVATIVES

As stated in [1], it is desirable that local properties be equivariant to scale and 2D-rotation. In this chapter we are only using filters which are based on Gaussian derivatives. As described above the approach can be used with any receptive field function (see [1] for further receptive field functions as i.e. Gabor filter).

By using the Gaussian derivatives one can explicitly select the scale. This is achieved by adapting the variance σ of the derivative. Given the Gaussian distribution $G(x)$:

$$G(x, y) = e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

the first derivative in x - and y -direction is given by:

$$G_x(x, y) = -\frac{x}{\sigma^2}G(x, y) \quad (2)$$

$$G_y(x, y) = -\frac{y}{\sigma^2}G(x, y) \quad (3)$$

Therefore the derivative in the direction $v = (\cos(\alpha) \quad \sin(\alpha))^T$ is given by

$$\frac{\partial G}{\partial v} = \cos(\alpha)G_x(x, y) + \sin(\alpha)G_y(x, y) \quad (4)$$

This property of the Gaussian derivative is known as “steerability” [10]. This property can be used to calculate the first derivative in any direction α . In the following we will refer to the derivative in the direction α as Dx and in the perpendicular direction $\alpha - 90^\circ$ as Dy .

The Magnitude and Direction of the first derivative are calculated as:

$$Mag(x, y) = \sqrt{(Dx)^2 + (Dy)^2} \quad (5)$$

$$Dir(x, y) = \arctan \frac{Dy}{Dx} \quad (6)$$

The second order derivatives and the Laplace operator are calculated as:

$$G_{xx}(x, y) = \left(\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2}\right)G(x, y) \quad (7)$$

$$G_{yy}(x, y) = \left(\frac{y^2}{\sigma^4} - \frac{1}{\sigma^2}\right)G(x, y) \quad (8)$$

$$Lap(x, y) = G_{xx}(x, y) + G_{yy}(x, y) \quad (9)$$

3.2. NORMALIZATION OF THE LOCAL CHARACTERISTICS BY ENERGIE

The effects of variation in signal intensity can be removed by normalizing the inner product of a filter with a signal during convolution. Normalization should be considered from at least two points of view. The first point concerns how well the normalized convolution behaves in the presence of additive noise (see experiments in [1]). The second point concerns how the normalized convolution responds to variations in signal intensity due to differences in ambient light intensity, aperture setting or digitizer gain.

We have compared the robustness of *no* normalization to three forms of normalization: Normalization by *MaxMin*, Normalization by *Energie* and Normalization by *Variance* (see [1]).

$$Img_{ene}(x, y) = \frac{\sum_{i,j=-m,-n}^{m,n} Img(x+i, y+j)Mask(i, j)}{\sqrt{\sum_{i,j=-m,-n}^{m,n} Img(x+i, y+j)^2} \sqrt{\sum_{i,j=-m,-n}^{m,n} Mask(i, j)^2}} \quad (10)$$

In the following sections we will use only *Energy* normalization (see formula (10)) since it seems to be the most robust normalization for the considered filters in respect to additive noise. The following section shows quite satisfactory results with this normalization in the recognition experiments (see section 7.).

4. HISTOGRAM COMPARISON MEASUREMENTS

For object recognition using receptive field histograms we compare a histogram T from a database to a newly observed histogram H . Possible similarity measures can be drawn from signal processing as well as from statistics.

In [1] we concluded that the best measurements to compare histograms are the “intersection”-measurement of Swain and Ballard and the χ^2 distance, which we will introduce in the following.

4.1. χ^2 - TEST

The proper method proposed by mathematical statistics for the comparison of two histograms is the χ^2 -test. χ^2 is used here to calculate the “distance” between two histograms. We have used two different calculations for χ^2 [11]: χ_T^2 is defined, when the theoretical distribution (here T) is known exactly. Although we do not know the theoretical distribution in the general case, we have found that χ_T^2 works well in practice:

$$\chi_T^2(H, T) = \sum_{i,j} \frac{(H(i, j) - T(i, j))^2}{T(i, j)} \quad (11)$$

The second calculation χ_{TH}^2 compares two real histograms. As we know from the results of [1] the following χ_{TH}^2 gives more reliable results:

$$\chi_{TH}^2(H, T) = \sum_{i,j} \frac{(H(i, j) - T(i, j))^2}{H(i, j) + T(i, j)} \quad (12)$$

4.2. INTERSECTION

Swain and Ballard [3] used the following intersection value to compare two color-histograms:

$$\cap(H, T) = \sum_{i,j} \min(H(i, j), T(i, j)) \quad (13)$$

The advantage of this measurement is, that background pixels are explicitly neglected when they don't occur in the Model histogram $T(i, j)$. In their original work they reported the need for a sparse distribution of the colors in the histogram in order to be able to distinguish between different objects. Our experiments have verified this requirement. Unfortunately, multidimensional receptive field histograms are not generally sparse, and a more sophisticated comparison measure is required.

5. USING MULTIDIMENSIONAL RECEPTIVE FIELD HISTOGRAMS FOR OBJECT RECOGNITION

The first part of this section defines the object recognition task by the analysis of the “degrees of freedom”. The second part describes the use of multidimensional receptive field histograms for this object recognition task. This is followed by sections 6. and 7. which give experimental results of this approach.

5.1. DEGREES OF FREEDOM WITHIN THE OBJECT RECOGNITION TASK

Possible changes of the object's appearance must be considered in the object recognition task. Possible changes include:

- Rotation of the object (or the camera): we distinguish rotation in the image plane (2D rotation) and arbitrary rotation (3D rotation)
- Changes in scale
- Translation of the object (or the camera)
- Partial occlusion of the object
- Light: intensity change and direction of the light source(s)
- Noise (noise of the camera, quantization noise, blur, ...)

In our approach, changes in to scale and 2D rotation are handled by the use of steerable filters [10, 12]. Therefore we will have only one image for one object and will generalize from this image to all considered scales and 2D rotations (see experiments in [1]).

In this chapter we do consider view point changes (3D-rotation). The experiments described below examines the robustness of the approach to these rotations.

The histograms themselves are invariant with respect to translation of the image or the object, since position information is completely removed. Furthermore the histogram matching is relatively immune to minor occlusions. This was demonstrated by Swain and Ballard in the original work on color histograms [3].

Signal intensity variations are accommodated by the use of energy normalized convolution with robust filters such as Gabor filters and Gaussian derivatives.

calculated the *Mag-Lap* histogram and compared these histograms to the histogram of each image of a test-set of the same 103 objects. The images of the test-set are taken under different conditions as the database images: 20° view point change for 20 objects, 20° image plane rotation for 22 objects, 15% scale change for 30 objects and the remaining 31 objects are taken under approximately the same condition. Table 1 shows a recognition rate of 97% for the intersection measurement and a 99% recognition rate for the χ_{TH}^2 measurement. This results show the ability of the approach to recognize objects even under quite different conditions.

measure	recognition
$\cap(H, T)$	97.1
$\chi_T^2(H, T)$	96.1
$\chi_{TH}^2(H, T)$	99

Table 1. Recognition results on a database of 103 objects under different condition (15% scale change, 20° view point change and 20° image plane rotation)

Table 2 shows recognition results of the same 103 objects but with more important changes between the test-set and the database: 40° view point change for 20 objects, 40° image plane rotation for 22 objects, 30% scale change for 30 objects and approximately the same condition for the remaining 31 objects. As we can see the recognition rates for χ_{TH}^2 is still 92.3% which indicates the robustness of this measurement to changes in scale, view point and image plane rotation. The intersection measurement also gives a high recognition rate of 87.4%, even though not as good as the two χ^2 measurements.

measure	recognition
$\cap(H, T)$	87.4
$\chi_T^2(H, T)$	91.3
$\chi_{TH}^2(H, T)$	92.3

Table 2. Recognition results on a database of 103 objects under different condition (30% scale change, 40° view point change and 40° image plane rotation)

7. EXPERIMENTAL EVALUATION OF ROBUSTNESS UNDER 3D-ROTATION

This section describes experiments with the Columbia database [2] which can be seen as a benchmark for object recognition. Since this database contains 20 objects, each at 72 different viewing angles, we used this database to examine the robustness of the multidimensional receptive field histogram approach to view point changes (3D-rotation): in this experiment we examine the influence of the histogram comparison measurements χ_{TH}^2 and *intersection*, the influence of the resolution per histogram axis and the number of dimensions of the histograms.

Throughout this section we are using the following abbreviations to refer to particular combinations of filter (and the corresponding multidimensional histogram, where each axis of such a histogram corresponds to one filter):

- *Dx-Dy-Lap* : first derivative in x- and y-direction and Laplace operator (this is the only three dimensional histogram used in the experiments of this chapter)
- *Dx-Dy* : first derivative in x- and y-direction (2D-histogram)
- *Mag-Dir* : magnitude and direction of the first derivative (2D-histogram)
- *Mag-Lap* : magnitude of the first derivative and Laplace operator (2D-histogram).
Mag-Lap is the only 2D-rotation invariant filter pair in the experiments.

7.1. ROBUSTNESS TO VIEW POINT CHANGES – THE COLUMBIA DATABASE

The Columbia database can be seen as a benchmark for object recognition algorithm. This database has been created and successfully used by Murase and Nayar [2]. Also Rao and Ballard [13] database. All authors obtained 100% recognition rate. In the experiments described below we obtain also 100% recognition rate. Since the Columbia image database contains objects under controlled view point changes, we will use this database in this section to examine experimentally the robustness of our approach to view point changes.



Figure 2. The 20 objects of the Columbia database



Figure 3. 9 of the 72 3D-rotations of one object of the Columbia database

The Columbia database contains 20 objects (see figure 2) seen from 72 different view angles, where each is 5° apart from the following (see figure 3). Therefore the database contains $20 \times 72 = 1440$ images. Typically the half of these images is used as database (or learning set) and the remaining half as test set. Therefore the angle between the images in the database is $\Delta\alpha = 10^\circ$.

In the experiment described in this section we varied this angle: $\Delta\alpha = 10^\circ, 15^\circ, 20^\circ, 30^\circ, 40^\circ, 45^\circ, 60^\circ, 90^\circ$. Doing so we examine the robustness of the approach to view point changes. Besides $\Delta\alpha$ we varied the following parameter of the approach:

- resolution of each histogram axis
- we use four filter combinations (*Dx-Dy-Lap*, *Dx-Dy*, *Mag-Dir*, *Mag-Lap*)

- histogram comparison measurement (χ_{TH}^2 and intersection)

Resolution	σ	Dx-Dy	Mag-Dir	Mag-Lap	Dx-Dy-Lap
64	1.5	100	99.72	99.72	
32	1.5	100	99.72	99.86	
16	1.5	100	99.31	99.72	100
8	1.5	99.72	98.47	98.06	100
4	1.5	99.31	94.17	81.67	99.31
2	1.5	77.5	29.86	32.5	94.86

Table 3. χ_{TH}^2 for 3D-rotation

Resolution	σ	Dx-Dy	Mag-Dir	Mag-Lap	Dx-Dy-Lap
64	1.5	100	99.72	99.58	
32	1.5	100	99.72	99.31	
16	1.5	99.86	99.58	98.33	100
8	1.5	99.44	97.78	90.14	99.72
4	1.5	93.19	84.31	63.19	96.53
2	1.5	75.14	29.58	32.22	93.47

Table 4. Intersection for 3D-rotation

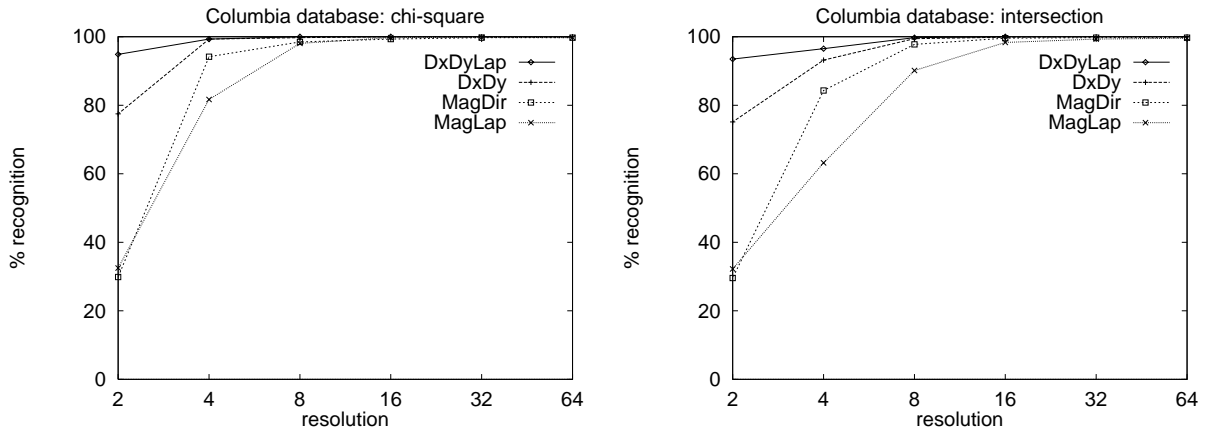
Figure 4. Columbia database: dependency of the recognition rate from the resolution of the histograms. Left: χ_{TH}^2 . Right: intersection

Table 3 and table 4 show results for different filters and different resolution ($\Delta\alpha = 10^\circ$ constant). Table 3 gives results with χ_{TH}^2 and table 4 gives results with intersection as comparison measurement. Figure 4 visualizes these results. In this experiment we can observe slightly higher recognition rates for χ_{TH}^2 than for intersection. Secondly a resolution of 8 for *Dx-Dy-Lap* and a resolution of 16 for *Dx-Dy* is sufficient to get a recognition rate of 100%.

The most interesting observation in table 3 and table 4 is that the 3D histogram *Dx-Dy-Lap* always gives higher recognition rates than the 2D histograms, the more

significant the lower the resolution. This results in a recognition rate of 94.86% for the 3D histogram *Dx-Dy-Lap* with a resolution of 2 cells per histogram axis (which corresponds to only 8 cells for the whole 3D histogram). This result indicates that the adding of independent dimensions (as the *Lap*-dimension) to the histogram increases the ability to discriminate between objects.

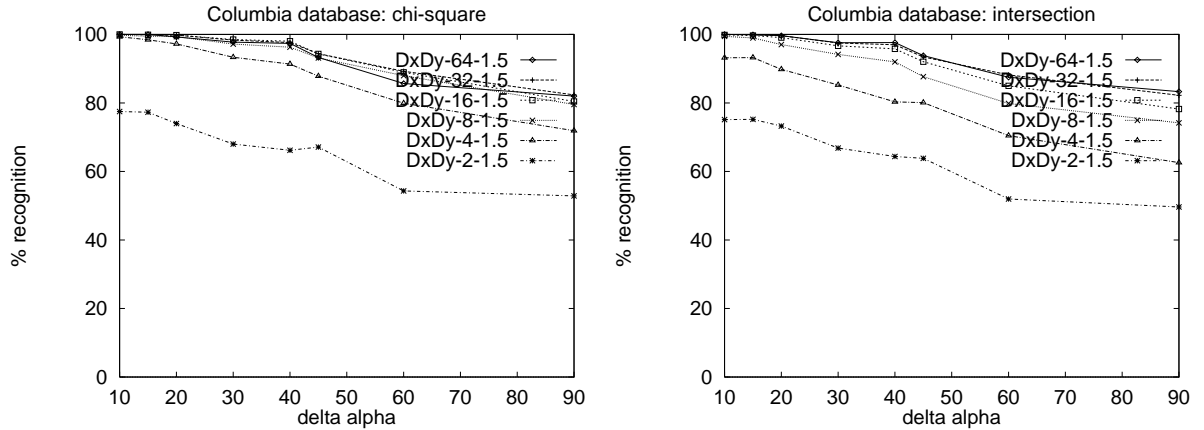


Figure 5. Columbia database: The 2D histogram *Dx-Dy*. Relation between recognition rate and $\Delta\alpha$ (see text).

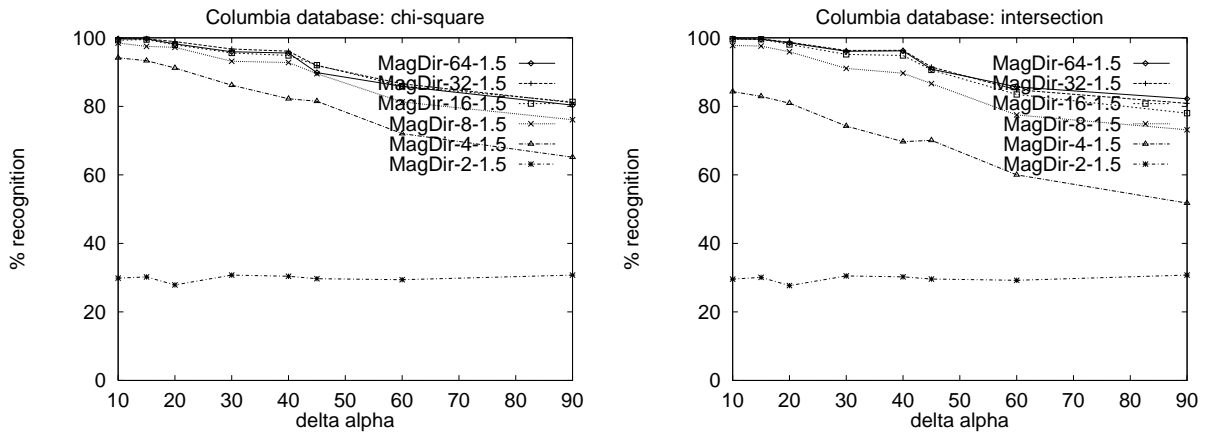


Figure 6. Columbia database: The 2D histogram *Mag-Dir*. Relation between recognition rate and $\Delta\alpha$ (see text).

The figures 5 through 8 show recognition rates depending on $\Delta\alpha$, the resolution and the different filter combinations. The performance of the different 2D histograms is very similar (one can rank *Dx-Dy* first, *Mag-Dir* second and *Mag-Lap* third). On the other hand the 3D histogram *Dx-Dy-Lap* gives high recognition rates even for large $\Delta\alpha$ and low resolutions. Therefore we think we will achieve even higher recognition rates by adding dimensions to the histograms. The increase of memory can be probably compensated by the decrease of the resolution.

In these figures 5 through 8 we can observe a graceful degradation of the recognition rate, which indicates the desired robustness of the approach to view point changes.

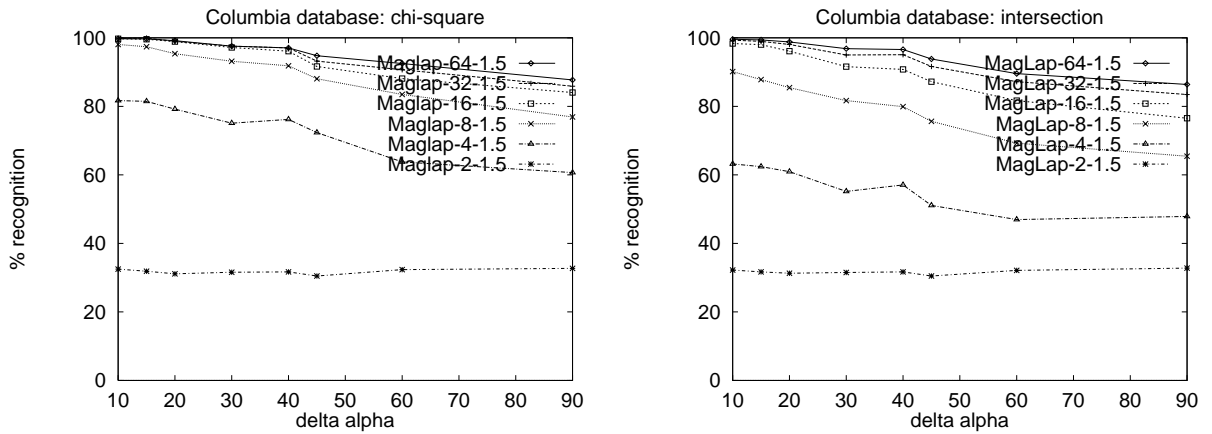


Figure 7. Columbia database: The 2D histogram *Mag-Lap*. Relation between recognition rate and $\Delta\alpha$ (see text).

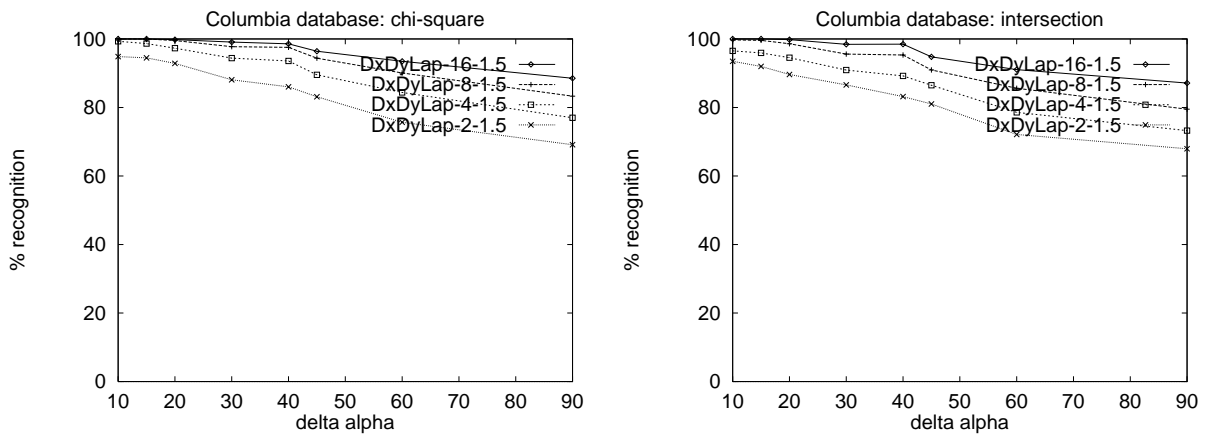


Figure 8. Columbia database: The 3D histogram *Dx-Dy-Lap*. Relation between recognition rate and $\Delta\alpha$ (see text).

8. CONCLUSION

In [1] we have shown how the color histogram matching technique of Swain and Ballard can be generalized to use vectors of local image properties measured by normalized convolution with local receptive fields. We have found that this technique present a fast and robust method to determine if a specified object is present in an image of a scene. This method can be used with any local filter as i.e. Gaussian derivatives and Gabor filters.

In the present chapter we have demonstrated that the approach is relatively robust to view point changes (graceful degradation of the recognition rates). The approach can be made more robust by increasing the number of dimensions of the histograms. We also showed that the performance of the approach is still high even with a small number of bins per histogram axis. Therefore the increase of memory for multiple dimensions can be compensated by the decrease of the bins per axis. Our experiments have also demonstrated that the χ^2 test provides the best ability to discriminate between objects.

REFERENCES

1. B. Schiele and J. L. Crowley. Object recognition using multidimensional receptive field histograms. In *ECCV'96, Fourth European Conference on Computer Vision*, 14–16 April 1996.
2. H. Murase and S. K. Nayar. Visual learning and recognition of 3d objects from appearance. *International Journal of Computer Vision*, 14:5–24, 1995.
3. M.J. Swain and D.H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.
4. B. V. Funt and G. D. Finlayson. Color constant color indexing. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 17(5):522–529, 1995.
5. G. Healey and D. Slater. Using illumination invariant color histogram descriptors for recognition. In *International Conference on Computer Vision and Pattern Recognition*, pages 355–360, 1994.
6. F. Ennesser and G. Medioni. Finding waldo, or focus of attention using local color information. In *International Conference on Computer Vision and Pattern Recognition*, pages 711–712, 1993.
7. F. Ennesser and G. Medioni. Finding waldo, or focus of attention using local color information. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 17(8):805–809, 1995.
8. M. Hunke. Locating and tracking of human faces with neural networks. Technical Report CMU-CS-94-155, Carnegie Mellon University, August 1994.
9. B. Schiele and A. Waibel. Gaze-tracking based on face-color. In *IWAFGR 95, International Workshop on Automatic Face-and Gesture-Recognition*, pages 344–349, June 1995.
10. W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 13(9):891–906, 1991.
11. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 2nd edition, 1992.
12. L.M.J. Florack, B.M. ter Haar Romeny, J.J. Koenderink, and M.A. Viergever. General intensity tranformations and second order invariants. In *SCIA'91 Proceedings of the 7th Scandinavien Conference on Image Analysis*, pages 338–345, 1991.
13. R. P. N. Rao and D. H. Ballard. Object indexing using an iconic sparse distributed memory. In *ICCV'95 Fifth International Conference on Computer Vision*, pages 24–31, 1995.