Object Recognition using Multidimensional Receptive Field Histograms

Bernt Schiele and James L. Crowley

LIFIA/GRAVIR, 46 Ave Félix Viallet, 38031 Grenoble, France

Abstract. This paper presents a technique to determine the identity of objects in a scene using 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. In this paper we describe the mathematical foundations of the technique and present the results of experiments which compare robustness and recognition rates for different local neighborhood operators and histogram similarity measurements.

1 Introduction and Motivation

Swain and Ballard [10] have developed a technique which identifies objects in an image 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. 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 color histogram approach by introducing measures which are less sensitive to illumination changes (see i.e. [5, 6, 2]).

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. 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. We have compared sensitivity and recognition reliability for a variety of local neighborhood operations, and present the results of the most successful functions below.

The first part of the paper presents our generalization of the color histogram method (section 2–4). Section 5 shows the robustness of different local neighborhood operations to additive Gaussian noise. In the second part we show the use of the histogram matching of receptive field vectors for object recognition (section 6) and experimental results (section 7).

2 Multidimensional Receptive Field Histograms

One can identify the following parameters for the *multidimensional receptive field histogram* approach:

- The choice of local property measurements (section 3),
- Measurement for the comparison of the histograms (section 4),
- Design parameters of the histograms: number of dimensions of the histogram and resolution of each axis.

The local properties should be chosen so that they are either invariant or equivariant to scale and 2D–rotation [1]. Invariant means that the local characteristics does not change with scale or 2D–rotation, while equivariant means that they vary in a uniform manner which is represented by a translation in a parameter space. Unfortunately most of the available characteristics are only scale invariant *or* 2D–rotation invariant. Therefore we use equivariant local characteristics which allow us to select an arbitrary scale and rotation (see e.g. [4, 3]). Section 3 describes the filters and normalizations which can be used.

The comparison measurement determines the separability between histograms, as we will see in the experiments described below. Different measures for the histogram comparison are introduced in section 4.

The design parameters of the histograms determine the separability between the histograms of different objects. In [8] we concluded that reducing of the resolution (number of bins per histogram axis) results in an improvement of the stability of the histograms with respect to view point changes, but also diminishes the discrimination between objects. From the experiments of [8] we concluded also that discrimination can be recovered by improving the number of histograms dimensions provided by independent local properties.

3 The local characteristics

In this section we briefly describe receptive field functions which can be used for object recognition. The calculation of local properties can be divided into the local linear point-spread function (formula (1)), and the normalization function used during measurements of local properties.

$$Img_{Mask}(x,y) = \sum_{i,j=-m,-n}^{m,n} Img(x+i, y+j) Mask(i,j) \qquad (1)$$

3.1 Filter

The first results we present are with non–equivariant filters. We have used these simple filters in our first experiments to test the power of our approach. This is followed by the description of two equivariant filter classes, Gabor filters and Gaussian derivatives.

[1] Recent results have shown that the technique is quite robust to 3D rotation. These results have been submitted to the *International Workshop on Object Representation for Computer Vision* at this conference [8]

Gradient and Laplacian Operators Our first experiments were performed with first derivative and Laplacian operators given by:

$$M_{dx} = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix} \qquad M_{dy} = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \qquad M_{lap} = \begin{pmatrix} -1 & -2 & -1 \\ -2 & 12 & -2 \\ -1 & -2 & -1 \end{pmatrix}$$

Gabor filter Gabor filters are local compact filters tuned to a spatial frequency band. Gabor filters are defined by modulating a Gaussian window with a cosine and an imaginary sine giving an even and odd filter pair. The main advantage of the Gabor filters is that one can freely choose the frequency (and therefore the scale) as well as the bandwidth of the filter.

A Gabor filter pair is compact in both space and frequency. In our experiments we have used a two–dimensional formulation of the Gabor functions proposed by Daugman [1] (in the Fourier domain):

$$G(u,v) = e^{-\pi\left((u-u_0)^2\alpha^2+(v-v_0)^2\beta^2\right)} e^{-2\pi i(x_0(u-u_0)+y_0(v-v_0))} \tag{2}$$

where (x_0, y_0) are the center coordinates of the filter, (α, β) define the width and the length, and (u_0, v_0) specify the modulation in x and y direction, which has the spatial frequency $\omega_0 = \sqrt{u_0^2 + v_0^2}$ and direction $\theta_0 = \arctan(v_0/u_0)$.

To design a Gabor filter, we follow a method proposed by Westelius [11] to choose the standard deviation α and the spatial frequency ω_0. These two parameters determine the size and bandwidth of the filter.

Gaussian derivatives 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 $f(x,y)$ we obtain the first derivative in x-direction:

$$f(x,y) = e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad \frac{\partial f(x,y)}{\partial x} = -\frac{x}{\sigma^2} f(x,y)$$

3.2 Normalization

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 section 5). 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 correlation with *no* normalization and with two other forms of normalization.

Normalization by energy Dividing by neighborhood energy removes variations in signal strength which may be due to light source intensity variation, and thus provide a filter output vector histogram which is invariant to illumination intensity. Energy normalization also turns out to be the most robust in respect to additive Gaussian noise. Therefore we have used energy normalization in most of our experiments.

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

Normalization by mean and variance By Variance normalization we refer to subtracting the mean of each neighborhood and then dividing by the variance of the neighborhood. Variance normalization is relatively sensitive to additive Gaussian noise. This makes *Variance*–normalization unusable in our context.

$$Img_{var}(x,y) = \frac{\sum_{i,j}(Img(x+i,y+j) - \overline{Img(x,y)})Mask(i,j)}{\sqrt{\sum_{i,j}(Img(x+i,y+j) - \overline{Img(x,y)})^2}\sqrt{\sum_{i,j} Mask(i,j)^2}}$$

with $\overline{Img(x,y)} = \frac{1}{(2m+1)(2n+1)} \sum_{i,j=-m,-n}^{m,n} Img(x+i,y+j)$.

4 Histogram Comparison

This section describes possible measurements for comparing histograms. The analysis of these measurements is important, since the "intersection"–measurement, used by Swain and Ballard [10], has limitations for the use for multidimensional receptive field histograms. For object recognition using receptive field histograms we compare a histogram T from a database to a newly observed histogram H.

Sum of squared distances The sum of squared differences (SSD) is commonly used in signal processing:

$$SSD(H,T) = \Sigma_{i,j}(H(i,j) - T(i,j))^2 \tag{3}$$

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 [7]: χ^2_T 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^2_T works well in practice:

$$\chi^2_T(H,T) = \Sigma_{i,j}\frac{(H(i,j) - T(i,j))^2}{T(i,j)} \tag{4}$$

The second calculation χ^2_{TH} compares two real histograms. χ^2_{TH} also gives good results. For the moment it is not clear which of the two χ^2 measurements is more reliable:

$$\chi^2_{TH}(H,T) = \Sigma_{i,j}\frac{(H(i,j) - T(i,j))^2}{H(i,j) + T(i,j)} \tag{5}$$

Intersection Swain and Ballard [10] used the following intersection value to compare two color–histograms:

$$\cap(H,T) = \Sigma_{i,j} \min(H(i,j), T(i,j)) \tag{6}$$

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.

Bayes Rule The last section below considers the use of Bayes rule to determine for each pixel or set of pixels, the probability that it is the projection of a part of a specified object. In [9] we have introduced the following formula:

$$p(O_n|\bigwedge_k M_k) = \frac{\prod_k p(M_k|O_n)p(O_n)}{\sum_n \prod_k p(M_k|O_n)p(O_n)} \tag{7}$$

with $p(O_n)$ the a priori probability of the object O_n, and $p(M_k|O_n)$ is the probability density function of object O_n, which can be directly derived from the histogram of object O_n. This formula can be used to determine for each subregion of an image the probability of the occurrence of each object O_n only based on the multidimensional receptive field histograms of each object (see for details and recognition results [9]).

5 Robustness to additive Gaussian noise

In this section we report the results of an experiment which was designed to determine how sensitive the different combinations of filter and normalization are in respect to additive Gaussian noise. For this experiment we used 8 artificial images. We will summarize the results for one image, which we call *Sin* which contains a sine–curve with the wavelength of 45 pixels.

Figure 1 show the results. To the *Sin*–image we added Gaussian noise with variance $\sigma = 1, 2, 3, \ldots, 20$ (abscissa in the diagrams). We store the histogram of the initial image (which is equivalent to $\sigma = 0$). This histogram is then compared (by using χ^2_{TH} as distance measurement) to the histograms with additive Gaussian noise. This distance correspond to the ordinate of the diagrams.

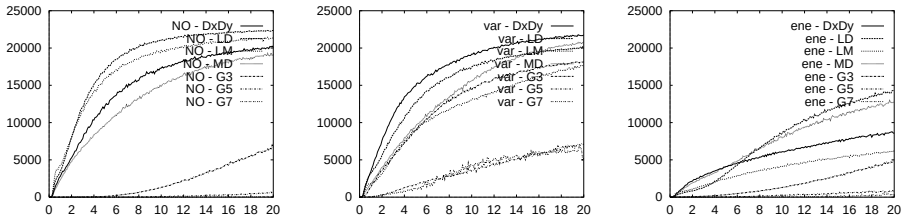

Fig. 1. *Sin:* Left: Robustness with no normalization. Middle: Robustness with Variance normalization. Right Robustness with Energy normalization

In these diagrams we look mainly at the relative behavior of the different filter normalization combinations, rather than at the absolute value of the χ^2_{TH} distance between the images (which depends strongly on the design parameter of the histograms). In this experiment we used seven different pairs of filters (see section 3): *DxDy* for M_{dx} and M_{dy}, *LD* for M_{lap} and Direction of the first derivative, *LM* for M_{lap} and Magnitude of the first derivative, *MD* for Magnitude and Direction of the first derivative, *G3* for Gabor filter with wavelength of 2.8 pixel (7×7 window) in x and in y direction, *G5* for Gabor filter with wavelength

of 5.7 pixel (15×15 window) in x and in y direction, *G7* for Gabor filter with wavelength of 11.3 pixel (30×30 window) in x and in y direction.

The first statement we can make is, that the Gabor filters are much more robust to additive Gaussian noise than the other filters (e.g. figure 1). This is not surprising, since the Gabor filters are known to be robust to additive Gaussian noise (one part of the Gabor function is a Gaussian smoothing function). Only in the case of the *Variance* normalization do Gabor filters fail to behave properly (see figure 1). The second statement that we can make is in relation to the different normalizations: *no* normalization behave rather nicely (figure 1). The *Variance* normalization on the other hand disturbs the nice behavior of the Gabor filters (figure 1). But the best normalization for all of the filters is the *Energy* normalization (figure 1).

In the following sections we will use only *Energy* normalization 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).

6 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. Section 7 gives experimental results of this approach.

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:

- Changes in scale
- Rotation of the object (or the camera): we distinguish rotation in the image plane (2D rotation) and arbitrary rotation (3D rotation)
- 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 [4, 3]. 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 section 7).

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 [10].

Signal intensity variations are accommodated by the use of energy normalized convolution with robust filters such as Gabor filters. For simplicity, our first experiments were based on simple mask operators as introduced in section 3 which are not necessarily invariant to light intensity changes.

To test robustness in relation to noise we completed a series of experiments with artificial and real images, where we added Gaussian noise. The impact on the histograms (measured with an appropriate distance measure) are shown in section 5.

In this article we do not consider the other degrees of freedom mentioned above: 3D rotation and light direction. In [8] we examined the robustness of the approach to image–plane rotation and view point changes (3D-rotation).

Application for Object Recognition The system we describe here is only an initial experiment to demonstrate the capabilities of the approach for object recognition. Further investigation must be performed in the use the multidimensional receptive filed histograms in a more thorough manner.

In this experimental version of the system, the database consists of histograms of each object at a set of scales and 2D orientations. A new histogram of an observed object is compared to each histogram of the database to find the closest match.

7 Experimental results

This section describes three experiments with the use of multidimensional receptive field histogram for object recognition: in the first experiments we consider scale, in the second we consider scale and image–plane rotation. In the last experiment we generalize from one single view of an object to 5 different scales.

7.1 Scale Experiment

In this section we report results from a recognition experiment with different scales of objects. We employed two series of images of 31 objects (see figure 2) at 6 different scales (approximate difference between each scale is 10%, see figure 2). The total number of images is therefore $2 \times 31 \times 6 = 372$. The first series have been used to calculate the histogram database and the second series have been used as test–set.

As mentioned above (section 6) we have different parameters in the *multidimensional receptive field histogram* approach. In this experiment we varied the local properties and the histogram comparison measurement The design parameter of the histograms have been fixed (2–dimensional with resolution of 32 cells per axis, for variation of the design parameters see [8]). For local properties, we used the same pairs of filters as in section 5. All experiments were performed with only two filters, as a minimal limiting case (in [8] we showed that recognition rates can be improved by increasing the number of local properties measured at each pixel).

Table 1 shows the recognition rate for different filter–pairs and different histogram comparison measurements. The first column of table 1 shows the filter–pairs. The first row shows the histogram comparison measurement as introduced in section 4: the two χ^2 measurements χ^2_T and χ^2_{TH}, sum of squared differences (SSD) and the intersection measurement. The table shows a recognition rate of 100%, when we choose the filter pair magnitude and direction of gradient, and the comparison measurement χ^2_T.

Fig. 2. Top: The 31 objects of the scale experiment. Bottom: The 6 different scales

Filter	χ^2_T	χ^2_{TH}	SSD	intersection
MD	100.0	98.9	89.8	91.4
DxDy	97.8	97.8	90.9	62.9
LD	97.3	97.3	88.7	86.0
LM	94.1	94.6	82.8	26.9
G3	86.6	86.0	64.0	43.5
G5	93.5	91.4	81.7	57.5
G7	97.8	97.8	92.4	34.4

Table 1. Recognition results with 31 Objects at 6 different scales

Following the results of table 1 we can analyze the different histogram comparison measurements: χ^2_T almost always gives the best results. χ^2_{TH} works nearly as well as χ^2_T. The SSD also gave quite good results nearly all of the time. The intersection (originally used by Swain and Ballard) give good results in some particular cases. Nevertheless the average performance over all of the filter pairs is not satisfactory. To summarize the table we can conclude that the χ^2 are the best, followed by SSD and intersection. In other experiments (e.g. section 7.3, 7.2 and [8]), we did make similar observation in relation to the histogram comparison measurements. Therefore we state that the χ^2 measurements are the best to compare *multidimensional receptive field histograms*.

7.2 2D Rotation experiment

This section presents results of an experiment where we considered the effects of 2D (image plane) rotation of objects at different changes in scale.

In this particular experiment we had 10 objects at 8 different orientations. The difference between the orientations was roughly 45°. Furthermore we took images of each object at 5 different scales, where the difference between each scale was approximately 10%. Therefore the whole image–set contains $10 \times 8 \times 5 = 400$ images.

For the experiment we divided the image set into database and the test–set. The database consists of three different scales, respectively the first, the third and the fifth scale. Therefore $3 \times 8 \times 10 = 240$ histograms are in the database. The remaining 2 scales are then tested against the database (test–set is therefore $2 \times 8 \times 10 = 160$ histograms of images).

We can report here the effects of 2D rotation and scale changes on recognition rates (see table 2) of three filter–pairs (for the description of the abbreviations see section 7.1).

Filter	χ^2_T	χ^2_{TH}	SSD	prod	intersection
DxDy	99.4	99.4	81.3	10.0	66.9
G3	86.9	85.0	53.1	10.6	56.9
G5	88.8	87.5	54.4	18.8	19.4

Table 2. Recognition results with 10 Objects at 5 different scales and 8 different orientations

As we already concluded from the scale experiment, the χ^2 measurements give the best results (χ^2_T slightly better than χ^2_{TH}). SSD gives good results for DxDy and intersection does not give satisfactory results for any of the reported filters.

7.3 Experiment: generalizing scales from one single view

Up to now we always took images of the same object at different scales. Since this is not always practical we want to take only one image of an object and to generalize to a range of scales. This is demonstrated in a second scale experiment.

This second scale experiment uses only one image of each object at one particular scale (of the first series). Starting from this single image we calculate 5 histograms, each corresponding to a different scale of the object. Therefore we have to use "steerable" filters as Gabor filters or Gaussian derivatives. In this particular experiment we used first order Gaussian derivatives (in x and in y direction = dxdy) and the magnitude and direction of the first Gaussian derivative (= magdir) with $\sigma = 0.8, 0.9, 1.0, 1.1$ and 1.2. This was done with all 31 objects of the first experiment (see section 7.1) so that the histogram database contains $5 \times 31 = 155$ histograms. As a test–set we used the images of the 31 objects of the second series at 5 different scales. For each of those images we calculated the histogram with $\sigma = 1.0$. These histograms are then compared to the histogram database.

Filter	χ^2_T	χ^2_{TH}	SSD	intersection
magdir	99.4	100	19.4	99.4
dxdy	98.7	98.1	9.0	91.6

Table 3. Recognition results of the second scale experiment: 31 Objects at 5 different scales, where we generalized the scales from one single view of each object

Table 3 shows the results of the experiment. Once again the χ^2 give the best

results. The intersection measurement gives quite good results too. This time SSD doesn't give good results at all.

This experiment shows that we can "steer" the scale, so that it is possible to calculate all considered scales from one single image of an object.

8 Conclusion and Perspective

In this paper 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 very local filters for gradient and Laplacian, as well as with more noise resistant filters such as Gabor filters and Gaussian derivatives. We have demonstrated that the method is most reliable and robust when the inner product of the receptive field at each neighborhood is normalized by the energy of the neighborhood. Our experiments have also demonstrated that the χ^2 test provides the most reliable form of histogram comparison for this method.

Relatively high recognition rates have been demonstrated with vectors composed of only two receptive field. We showed in [8] that these rates can be made even higher by increasing the number of filters included in the vector. The increase in memory required can be off-set by decreasing the quantization of the histograms.

References

1. J. G. Daugman. High confidence visual recognition of persons by test of statistical independance. *IEEE PAMI*, 15(11):1148–1161, November 1993.
2. F. Ennesser and G. Medioni. Finding waldo, or focus of attention using local color information. *IEEE PAMI*, 17(8):805–809, 1995.
3. 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*, 1991.
4. W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE PAMI*, 13(9):891–906, 1991.
5. B. V. Funt and G. D. Finlayson. Color constant color indexing. *IEEE PAMI*, 17(5):522–529, 1995.
6. G. Healey and D. Slater. Using illumination invariant color histogram descriptors for recognition. In *CVPR*, pages 355–360, 1994.
7. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 2nd edition, 1992.
8. B. Schiele and J. L. Crowley. The robustness of object recognition to rotation using multidimensional receptive field histograms. submitted to International Workshop on Object Representation for Computer Vision, April 1996.
9. B. Schiele and J. L. Crowley. Probabilistic object recognition using multidimensional receptive field histograms. submitted to ICPR'96, August 1996.
10. M.J. Swain and D.H. Ballard. Color indexing. *IJCV*, 7(1):11–32, 1991.
11. C.-J. Westelius. *Preattentive Gaze Control for Robot Vision*. PhD thesis, Department of Electrical Engineering, Linköping University, 1992.

This article was processed using the LaTeX macro package with ECCV'96 style