Probabilistic Object Recognition using Multidimensional Receptive Field Histograms

Bernt Schiele and James L. Crowley
IMAG-GRAVIR, I.N.P.G., Grenoble, France
email: Bernt.Schiele@imag.fr

Abstract

This paper describes a probabilistic object recognition technique which does not require correspondence matching of images. This technique is an extension of our earlier work on object recognition using matching of multi–dimensional receptive field histograms. In [4] we have shown that multi-dimensional receptive field histograms can be matched to provide object recognition which is robust in the face of changes in viewing position and independent of image plane rotation and scale. In this paper we extend this method to compute the probability of the presence of an object in an image.

The paper begins with a review of the method and previously presented experimental results. We then extend the method for histogram matching to obtain a genuine probability of the presence of an object. We present experimental results on a database of 100 objects showing that the approach is capable recognizing all objects correctly by using only a small portion of the image. Our results show that receptive field histograms provide a technique for object recognition which is robust, has low computational cost and a computational complexity which is linear with the number of pixels.

1. Introduction

In [4] we presented a technique to determine the identity of an object in a scene using multidimensional histograms of a vector of local neighborhood operators. We showed that this technique can be used to determine the most probable object, independent of the object's position, image-plane orientation and scale. We have found that this technique presents a fast and robust method to determine if a specified object is present in an image of a scene. The method is not restricted to a particular family of filters but can be used with a large scale of different filters.

In [5] we evaluated the robustness of the approach to view point changes. The experimental results demonstrate that the approach is robust to such changes. We concluded from these results that the approach can be made more robust by increasing the number of dimensions of the histograms and that the performance of the approach can also be improved by an increase in the number of bins per histogram axis.

In both papers we *matched* a multidimensional receptive field histogram of an object in the scene with histograms of the object database. The present paper introduces a technique to determine the probability that an object is in a scene based on arbitrarily chosen points in the image. As a result, we can state that the technique does not rely on image correspondence matching. Using this technique the probability of each object can be calculated only based on the multidimensional receptive field histograms of each object (see section 4).

The next section introduces the multidimensional receptive field histograms as a generalization of the color histogram approach of Swain and Ballard [6] and describes the local receptive fields which we will use throughout the paper. Section 4 develops a technique to determine the probability of each object in the scene in the context of multidimensional receptive field histograms. Experimental results on a database of 100 objects are presented in section 5.

2. Multidimensional Receptive Field Histograms

Swain and Ballard [6] 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. Even changes in the shape of an object do not necessarily degrade the performance of their method.

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 *histograms*. Thus it is natural to exploit the power of histograms techniques 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.

One can identify the following parameters for the *multi-dimensional receptive field histogram* approach:

- The choice of local property measurements (section 3),
- The technique for comparing histograms for recognition. In section 4 we propose to calculate a probability for each object only based on receptive field histograms.
- Design parameters of the histograms: number of dimensions of the histogram and the resolution of each axis.

The local properties should be chosen so that they are either invariant or equivariant to scale and 2D–rotation . 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. [2, 1, 3]) (we use also some rotationally invariant filters which should be equivariant to scale-changes). The following section 3 describes receptive field functions based on the Gaussian derivatives.

In [4] we introduced different measurements for the comparison of histograms. We identified the χ^2–test as the most suitable for histogram matching. In section 4 we develop a technique to determine the probability of an object in a scene only based on multidimensional receptive field histograms.

The design parameters of the histograms determine the separability between different objects. In [5] 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 and image–plane rotation, but also diminishes the discrimination between objects. We concluded also that discrimination can be recovered by improving the number of histograms dimensions provided by independent local properties. In this article we will use different filter combinations at multiple scales to increase the ability to discriminate of the approach.

3. Local Characteristics

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. We have found energy normalization (section 3.2) to be the most robust in the presence of additive Gaussian noise [4].

As stated in above, it is desirable that local properties be equivariant to scale and 2D–rotation. In this paper we are only using filters which are based on equivariant Gaussian derivatives. We should mention that the approach can be used with any receptive field function (see [4] for further receptive field functions as i.e. Gabor filter).

3.1. Local Characteristics based on 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 $G(x,y)$:

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

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

$$\begin{aligned} G_x(x,y) &= -\frac{x}{\sigma^2}G(x,y) \\ G_y(x,y) &= -\frac{y}{\sigma^2}G(x,y) \end{aligned}$$

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

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

This property of the Gaussian derivative is known as "steerability" [2]. 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 $\alpha - 90°$ as *Dy*.

In this paper we are using also rationally invariant filters as the magnitude of the first Gaussian derivative and the Laplace operator:

$$\begin{aligned} Mag(x,y) &= \sqrt{G_x^2 + G_y^2} \\ G_{xx}(x,y) &= (\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2})G(x,y) \\ G_{yy}(x,y) &= (\frac{y^2}{\sigma^4} - \frac{1}{\sigma^2})G(x,y) \\ Lap(x,y) &= G_{xx}(x,y) + G_{yy}(x,y) \end{aligned}$$

3.2. Normalization of the local characteristics by Energy

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 [4]). 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 *Energy* and Normalization by *Variance* [4]. We found that the *Energy* normalization (formula (2)) gives the most robust results in the presence of additive Gaussian noise. In the experiments described below we achieve quite satisfactory results with this normalization.

$$I_{ene}(x,y) = \frac{\sum_{i,j} I(x+i,y+j)M(i,j)}{\sqrt{\sum_{i,j} I(x+i,y+j)^2}\sqrt{\sum_{i,j} M(i,j)^2}}$$

with $I(x,y)$ the image and $M(i,j)$ the convolution mask of a filter.

4. Probabilistic Object Recognition

In the context of probabilistic object recognition we are interested in the calculation of the probability of the object O_n given a certain local measurement M_k. This probability $p(O_n|M_k)$ can be calculated by the Bayes rule:

$$p(O_n|M_k) = \frac{p(M_k|O_n)p(O_n)}{p(M_k)} = \frac{p(M_k|O_n)p(O_n)}{\sum_n p(M_k|O_n)p(O_n)}$$

with

- $p(O_n)$ the a priori probability of the object O_n,
- $p(M_k)$ the a priori probability of the filter output combination M_k and
- $p(M_k|O_n)$ is the probability density function of object O_n, which differs from the histogram of an object O_n only by a normalization factor.

Having two local measurements M_k and M_l from the same object O_n we can rewrite this formula:

$$p(O_n|M_k \wedge M_l) = \frac{p(M_k \wedge M_l|O_n)p(O_n)}{\sum_n p(M_k \wedge M_l|O_n)p(O_n)}$$

Under the assumption of independence of M_k and M_l we obtain:

$$p(O_n|M_k \wedge M_l) = \frac{p(M_k|O_n)p(M_l|O_n)p(O_n)}{\sum_n p(M_k|O_n)p(M_l|O_n)p(O_n)}$$

Having K independent local measurements $M_1, \ldots, M_K$ we can calculate the probability of each object O_n by:

$$p(O_n|\bigwedge_k M_k) = \frac{p(\bigwedge_k M_k|O_n)p(O_n)}{\sum_n p(\bigwedge_k M_k|O_n)p(O_n)} = \frac{\prod_k p(M_k|O_n)p(O_n)}{\sum_n \prod_k p(M_k|O_n)p(O_n)} \quad (2)$$

In our context the local measurement M_k corresponds to a single multidimensional receptive field vector. Therefore K local measurements M_k correspond to K receptive field vectors which are typically from the same region of the image. To guarantee the independence of the different local measurements we choose the minimal distance $d(M_k, M_l)$ between two measurements M_k and M_l sufficiently large (in the experiments described below we choose the minimal distance $d(M_k, M_l) \geq 2\sigma$. This distance is sufficient to guarantee independence of the receptive fields from a signal processing point of view).

For the experiments we assume that all objects have the same probability $p(O_n) = \frac{1}{N}$, where N is the number of objects. Therefore formula (2) simplifies to:

$$p(O_n|\bigwedge_k M_k) = \frac{\prod_k p(M_k|O_n)}{\sum_n \prod_k p(M_k|O_n)} \quad (3)$$

The probabilities $p(M_k|O_n)$ are directly given by the multidimensional receptive field histograms. Therefore formula (3) shows a calculation of the probability for each object O_n only based on the multidimensional receptive field histograms of the N objects.

It is important to note that the locations of the measurements can be chosen arbitrarily. Therefore the technique is fast (only a certain number of local receptive field vectors have to be calculated) and robust to occlusion (the approach is strictly local). Furthermore the technique works without correspondence between the object database and the test image.

5. Experimental Results

This section describes two experiments using a database of 100 objects (figure 1). In the first experiment we examine the performance of the approach in the presence of scale changes. In a second experiment we allow image-plane rotation as well. In these experiments we are particular interested in the use of multi-scale histograms. As earlier experimental results indicate, recognition rates can be increased by adding dimensions to the histograms. This can be done either by adding independent filter or by using the same filter at multiple scales.

Figure 1. 30 of the 100 objects

Throughout this section we use the following filter combinations: *Mag-Lap* for the magnitude of the first Gaussian derivative and the Laplace operator and *Dx-Dy-Lap* for the first Gaussian derivatives in x– and y–direction and the Laplace operator. These filter combinations are used in this paper at to different scales which are f.e. $\sigma = 2.0$ and 4.0. In the following the abbreviation *Dx-Dy-Lap-2+4* corresponds to a 6-dimensional receptive field histogram of the filter combination *Dx-Dy-Lap* calculated at two scales $\sigma = 2.0$ and 4.0.

5.1. Scale experiment

In a first experiment we show that the approach can accommodate changes in scale up to a scale change of ± 40%. As mentioned earlier we use the "steerability" of the Gaussian derivatives in order to handle scale. This implies that we can calculate receptive field histograms which correspond to different scales of an object only based on one image of the object. In the experiment described here we are using f.e. the histograms *Dx-Dy-Lap-1.67+3.3*, *Dx-Dy-Lap-2+4* and *Dx-Dy-Lap-2.4+4.8*, which correspond to ± 20% scale change of the object. In the context of multidimensional histograms we can either use each of these scaled histograms independently or we can add them to one single histogram per object. The results described here are based one histogram per object since this method gives similar results and reduces the number of histograms significantly.

In this first experiment we calculated for each of the 100 objects one histogram which corresponds to ± 40% scale change. As test images we used 130 images of the same objects where 70 objects were taken under approximately the same scale. Of the remaining 30 objects we took two different images with an arbitrary scale change in the range of ± 40% (of the camera). Figure 2 shows the results of the approach on this test images. The y-axis corresponds to the recognition rate whereas the x-axis shows the number of points which we used to calculate the probability of each object (cf. formula 3). These points are equi-distributed in the center of each image (with a distance of 2σ between each point). We could have chosen any other image location. But since the objects are usually centered in the image this choice ensures that the receptive field vectors contains information from the object itself.

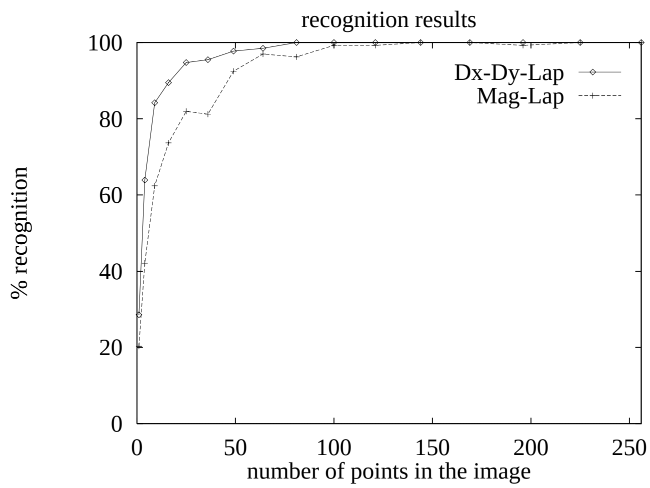

Figure 2. Recognition results for the scale-experiment

As the graph in figure 2 shows, we obtain better recognition results by using the histogram *Dx-Dy-Lap* than by using the histogram *Mag-Lap*. This observation can be explained by the higher dimensionality of the first histogram. For the 6-dimensional *Dx-Dy-Lap* histogram we need only 81 points to recognize all 133 test images correctly, which corresponds to 14.4% of the whole image. In the case of the 4-dimensional *Mag-Lap* histogram one needs 144 points (22.7% of the image) to recognize all objects correctly (by using 100 points (17% of the image) one obtains one misclassification).

5.2. Scale and 2D-rotation experiment

In a second experiment we wanted to examine the capability of the approach to handle image-plane rotation. Therefore we used the same database of the 100 objects and calculated the 4-dimensional *Mag-Lap* histogram (which are rotationally invariant). In the above mentioned test-images we replaced for 22 objects there test image with two arbitrarily rotated test images. Therefore we have 152 test images (with 60 scaled images and 44 rotated images).

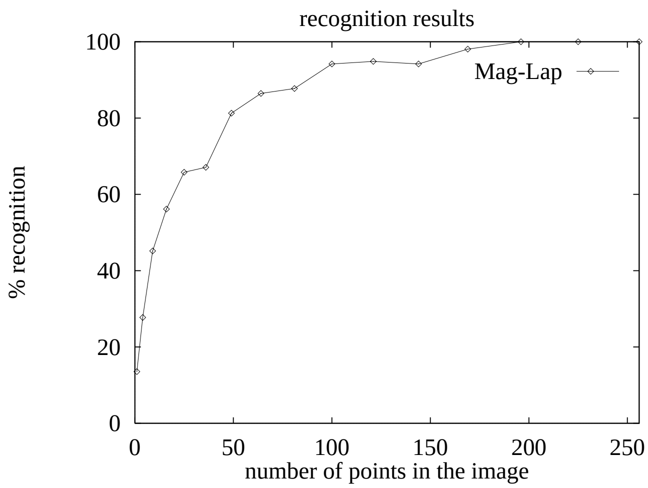

Figure 3. Recognition results for the scale-rotation-experiment

Figure 3 shows the result of the scale-rotation experiment. For the 4-dimensional receptive field histogram *Mag-Lap* we need 196 points to recognize all 152 test images correctly, which corresponds to 29% of the image. With 100 points (17% of the image we obtain 7 misclassification). This results can certainly be improved by using higher dimensional receptive field histograms, which we are currently investigating.

Overall we can conclude, that the developed probabilistic object recognition technique is well suited for multidimensional receptive field histograms. The number of points needed to recognize all 100 objects correctly is relatively low and therefore computationally inexpensive. The image support for these points is between 15% and 23% of the image in the scale experiment and slightly higher for the rotation-scale experiment. These results show the robustness of the approach to partial occlusion. The performance of the approach can be increased either by adding new independent Filters or by using the same filters at multiple scales.

6. Conclusion

In this paper we have extended the matching of multidimensional receptive field histogram to probabilistic object recognition without correspondence. We have developed a technique to determine the probability that an object may be found in an image based only on multidimensional receptive field histograms. Promising results have been shown on a database of 100 objects in the presence of scale changes and image-plane rotation.

References

[1] L. Florack, B. Romeny, J. Koenderink, and M. Viergever. General intensity tranformations and second order invariants. In *SCIA'91*, pages 338–345, 1991.

[2] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *PAMI*, 13(9):891–906, 1991.

[3] D. G. Jones and J. Malik. A computational framework for determining stereo correspondence from a set of linear spatial filters. In *ECCV'92*, pages 395–410, 1992.

[4] B. Schiele and J. L. Crowley. Object recognition using multidimensional receptive field histograms. In *ECCV'96, Vol I*, pages 610–619, 14–16 April 1996.

[5] B. Schiele and J. L. Crowley. The robustness of object recognition to view point changes using multidimensional receptive field histograms. ECIS-VAP, Object Recognition Day. Available via WWW[1], March 1996.

[6] M. Swain and D. Ballard. Color indexing. *IJCV*, 7(1):11–32, 1991.

[1] http://pandora.imag.fr/Prima/schiele/