ECCV, Dublin, Ireland, 2000

Object Recognition Using Coloured Receptive Fields

Daniela Hall*, Vincent Colin de Verdière** and James L. Crowley***

Projet PRIMA — Lab. GRAVIR–IMAG
INRIA Rhnes–Alpes
655, avenue de l'Europe
38330 – Montbonnot Saint Martin, France

Abstract. This paper describes an extension of a technique for the recognition and tracking of every day objects in cluttered scenes. The goal is to build a system in which ordinary desktop objects serve as physical icons in a vision based system for man-machine interaction. In such a system, the manipulation of objects replaces user commands.
A view-variant recognition technique, developed by the second author, has been adapted by the first author for a problem of recognising and tracking objects on a cluttered background in the presence of occlusions. This method is based on sampling a local appearance function at discrete viewpoints by projecting it onto a vector of receptive fields which have been normalised to local scale and orientation. This paper reports on the experimental validation of the approach, and of its extension to the use of receptive fields based on colour. The experimental results indicate that the second author's technique does indeed provide a method for building a fast and robust recognition technique. Furthermore, the extension to coloured receptive fields provides a greater degree of local discrimination and an enhanced robustness to variable background conditions.
The approach is suitable for the recognition of general objects as physical icons in an augmented reality.

Keywords: Object Recognition, Texture & Colour, Appearance-Based Vision, Phicons

1 Introduction

This article addresses the problem of the recognition of objects with a wide variety of features under changing background conditions. The proposed system is to be used in the context of an augmented reality system. In this system,

* email: Daniela.Hall@inrialpes.fr
** email: Vincent.Colin-de-Verdiere@inrialpes.fr
*** email: Jim.Crowley@inrialpes.fr

physical icons (phicons) are used to enhance the man-machine interface. Physical icons are physical objects to which a virtual entity can be attached. Such a virtual entity can represent system commands and their parameters [13,5]. A classical example for the use of phicons are editing operations. An eraser can stand for deleting, scissors can stand for cutting, and tape can stand for pasting. An appropriate selection of phicons allow users to quickly adapt to the graspable interface. Our problem is to build such a system to investigate the improvement in usability provided by phicons.

In an augmented reality system, one or more cameras observe a region of interest in which interaction can take place. Such a region can be a desk or more general a three dimensional space within a room. In such an environment the background and the lighting is variable. Translation of objects invoke differences in the view point of the camera and object pose. These problems require a system that is robust to such differences and make the recognition and pose estimation of phicons in an augmented reality an interesting challenge for computer vision.

An important constraint in a phicon based interface is that the user may select the object which serve as his personal interface. This imposes the constraint that the computer vision system can not be engineered for specific classes of objects. The system must be completely general. In addition, the computer vision system must not interfere with natural interaction. Thus the vision system must have a very low latency (on the order of 50 milliseconds in the case of tracking), and a very low failure rate.

The acceptance of objects with a wider variety of features increases the difficulty of recognition and pose estimation. Although there already exist many different approaches, most established methods work well for restricted classes of objects.

In this article an approach is proposed that allows the view-variant recognition of objects in a desk-top scene observed with a steerable camera. A possible solution could be provided by colour histograms [12,10]. However, this approach is not suitable for pose estimation. The extension to pose estimation in 2D and 3D is an important factor for the design of the approach. For this reason receptive fields are preferred to colour histograms.

The second author [3] has recently demonstrated a technique for the recognition of objects over changes in view-point and illumination which is robust to occlusions. In this approach, local scale and orientation are estimated at each point in an image. A vector of receptive fields is then normalised to this scale and orientation. The local neighborhood is projected onto this vector. This provides a representation which can be used by a prediction-verification algorithm for fast recognition and tracking, independent of scale an image orientation. View invariant recognition is obtained by sampling this representation at regular intervals over the view sphere. Because the method uses local receptive fields, it is intrinsically robust to occlusions.

In this article we adapt this technique to the problem of recognising and tracking physical icons. The technique extended by employing coloured receptive fields. The proposed approach allows the recognition of a wide variety of common objects, including objects with features that make recognition difficult, such as specularity and transparency. Evaluation of the experiments show that good results are obtained, even in an environment with variable background.

The next section reviews the description of the local appearance function by projection onto normalised receptive fields vectors. We then describe how this approach can be extended to coloured receptive fields. We then provide experimental results which validate the second author's approach using grey scale images, and then demonstrate the contribution of colour.

2 Describing local appearance

In 1991 Adelson and Bergen [2] reported a function that derives the basic visual elements from structural visual information in the world. This function is called the plenoptic function (from "plenus", full or complete, and "opticus", to see). The plenoptic function is the function of everything that can be seen. In machine vision the world is projected onto an image, which is a sample of the plenoptic function:

$$P(x, y, t, \lambda, V_x, V_y, V_z) \tag{1}$$

where (x, y) are the image coordinates, t, the time instant, λ the response wavelength, and (V_x, V_y, V_z) the view point. If the plenoptic function for an object is known it would be possible to reconstruct every possible image of the object; that is from every possible view, at every moment, for every image pixel, at every wavelength.

Adelson and Bergen propose to analyze samples of the plenoptic function using low order derivatives as feature detectors. Koenderink [8] expands the image signal by the first terms of its Taylor decomposition, that is in terms of the derivatives of increasing order. The vector of this set is called "Local Jet". The Local Jet is known to be useful for describing and recognising local features [11]. The signal derivatives are obtained by convolution of the signal by a set of basis functions.

2.1 Gaussian derivatives

Gaussian derivatives provide a basis for a Taylor series expansion of a local signal. This means that a local image neighborhood can be reconstructed by a linear combination of weighted Gaussian derivative filters. This reconstruction becomes an approximation which increases in error as the number of filters is reduced. The formula for the n^{th} 1D Gaussian derivative with respect to the dimension, x, is:

$$\delta_{x^n} g(x, \sigma) = \frac{d^n g(x, \sigma)}{dx^n} = \left(\frac{-1}{\sigma}\right)^n He_n\left(\frac{x}{\sigma}\right) g(x, \sigma), \tag{2}$$

$$\text{with } g(x, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

where He_n stands for the n^{th} Hermite "type e"polynomials [1].

Gaussian derivatives have an explicit scale parameter, σ, and can though be generated at any scale. With steerable filters proposed by Freeman [6] Gaussian derivatives can be oriented in any arbitrary direction. With automatic scale selection [9] the local scale of a feature can be determined. The object in an image can be normalised by scale which allows recognition under scale changes. The determination of the dominant orientation of a neighborhood allows to normalise by orientation. These two properties are used by all techniques presented in this article.

3 Sampling local appearance

In the technique proposed by Colin de Verdière [3] a training set consists of all overlapping image neighborhoods, referred to as imagettes, of all model images. An imagette is projected onto a single point in the descriptor space R. Each model image can be represented as a grid of overlapping imagettes. The projections of these imagettes form a surface, a local appearance grid, which models the local appearance of the image in R (see figure 1).

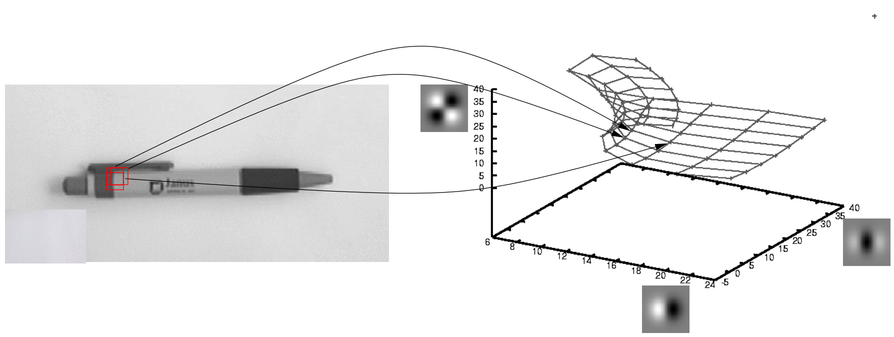


Fig. 1. An image as a surface in a subspace of R

Each object is represented by a set of images from different view points. As every image results in a local appearance grid, each object is modeled by the set of surfaces in R. The recognition process equals the search of the corresponding surface for the projection of a newly observed imagette. The basis of all surfaces in R are stored in a structural way, so that the searched surface can be obtained by table lookup. The resulting surface contains information about the object identity, the view point of the camera and information about the relative location of the imagette to the object position. The information from several points allow to estimate the pose of the object.

The approach based on Gaussian derivatives proposed in [3] serves as benchmark for the evaluation of the results. This approach is fast due to efficient storage and recursive filters [14], rotation invariant due to steerable filters [6], invariant to scale due to automatic scale selection [9], and robust to occlusions due to receptive fields. It produces good results for compact textured objects (see section 5.1). The approach fails completely for objects with sparse texture or objects of small sizes or with holes. The reason is that the Gaussian derivatives are computed only from the luminance image. In the luminance image the structure is very well preserved but the chromatic information is lost, and thereby the ability to distinguish objects by their colour. Small or non compact objects can not be recognised because the imagette contains part of the variable background. If the portion of the background is important the imagette is projected on a different point within the descriptor space. The detection of a surface belonging to another object or no surface at all is possible.

The approach described in this section serves as a starting point for the development of an improved approach. For the discrimination of poorly structured objects, chromatic information is indispensable. In the case of other objects, chrominance improves discrimination. A system that employs structural and chromatic information describes an additional dimension of the plenoptic function. Because this dimension includes more information, it can be expected to produce superior recognition results, at the cost of increased computation. Most of the additional cost may be avoided by keeping the number of receptive fields constant. We compensate the addition of receptive fields for chrominance with a reduction in the number of receptive fields for higher order derivatives. Our experiments show that chrominance is more effective than third order derivatives in discrimination of local neighborhoods.

4 Coloured receptive fields

A new descriptor space is needed that is based on Gaussian derivatives and capable of processing colour images. A direct approach would be to filter each colour channel separately. The advantage would be that no information is lost and no new technique needs to be developed. The disadvantage is that the normalisation process would need to be duplicated independently for each colour channel.

An alternative is to maintain the use of the luminance channel, and to complement this with two channels based on chrominance. The chrominance channels are described using colour-opponent receptive fields. Luminance is known to describe object geometric structure while chrominance is primarily useful for discrimination. Thus a receptive field vector is used in which chrominance receptive fields are normalised with the scale and orientation parameters computed from the luminance channel.

4.1 Selection of an appropriate colour space

This section addresses the problem of designing the colour opponent receptive fields for chrominance.

The RGB coordinate system is transformed according to following transformation

$$\begin{pmatrix} l \\ c_1 \\ c_2 \end{pmatrix} = T \begin{pmatrix} r \\ g \\ b \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & -1 \end{pmatrix} \begin{pmatrix} r \\ g \\ b \end{pmatrix} \tag{3}$$

This transformation, illustrated in figure 2, moves the origin to the center of the colour cube. One axis corresponds to the luminance axis, which will be used for structure analysis. The other two axis are orthogonal to the luminance axis and are used for colour analysis. We note that the two axis coding colour information are sensitive to red green differences and blue yellow differences, inspired by models of the human visual system [7].

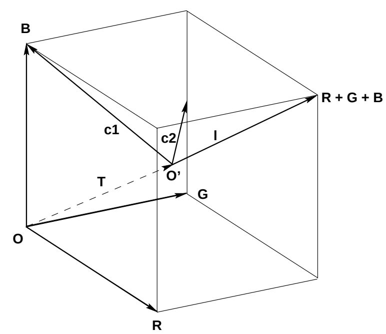


Fig. 2. Transformation of the RGB coordinate system.

Projection of the image neighborhood onto the luminance axis provides a description of geometric structure. Projection onto the colour difference channel improves discrimination.

5 Experimental Results

The experiment is based on 8 ordinary objects form an office desktop, that are appropriate to serve as physical icons (shown in figure 3). This set of objects is used to demonstrate the capability of the approaches to cope with general objects, among them objects with difficult features. The set contains textured and uniform objects, compact objects and objects with holes, specular and transparent objects. Some of the objects can be discriminated easily by their structure (eraser, sweets box), or by their colour (pen, scissors). Other objects exhibit specularities and transparencies which would render most object recognition techniques unreliable (tape, pencil sharpener, protractor). Recognition of such objects is difficult, because small changes of illumination or background conditions invoke important changes in the appearance of these objects.

For imagettes at variable scales, extracted from images of objects in the real world, object background regions tend to be generally lighter or generally darker

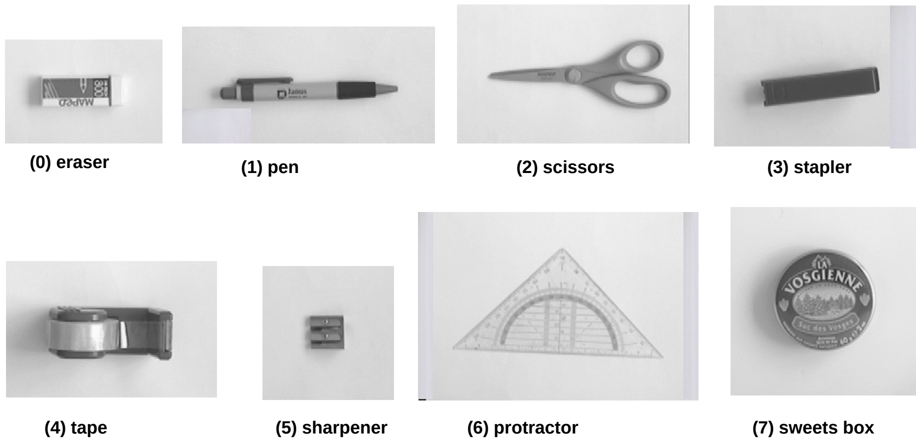


Fig. 3. Object set used in the experiments.

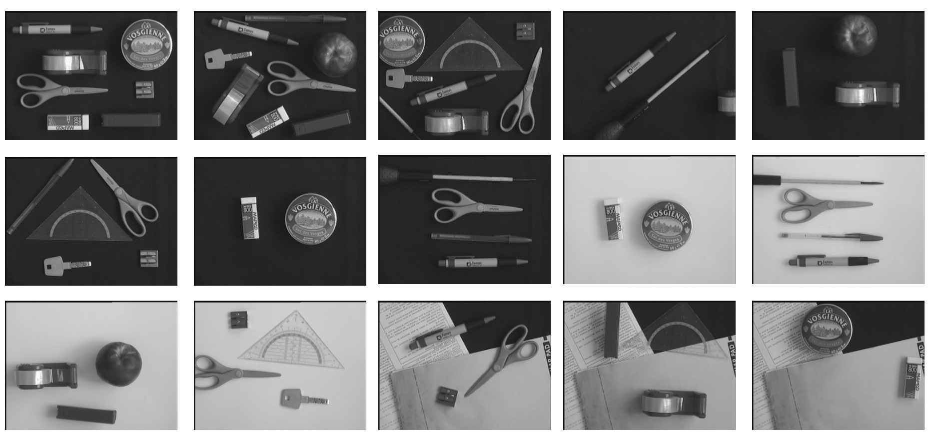


Fig. 4. Test scenes used in the experiments.

than the object. While clutter can introduce mixed backgrounds, such cases tend to be rare. In order to assure recognition over a range of backgrounds, we train models placing our objects on both black background and white background during training. Figure 3 shows the training images on white background. Section 5.1, 5.2, 5.3 applies the technique to images with uniform background, section 5.4 shows results on cluttered background.

The training phase results in a separate data structures for purely luminance based receptive field vectors up to third order, and for a receptive field vector which includes both luminance and chrominance, but are limited to second order. A recognition cycle was run on the test images. A set of 15 test images are used that contain between 2 to 6 different objects of the test set (see Figure 4). The orientation and the position of the objects in the test images is different from the orientation and position in the training images. The distance from the camera is constant and the camera is pointing on the desk. A grid of image neighborhood locations were selected for evaluation using a step size of 5 pixels between neighborhoods. At each neighborhood, the local scale and orientation are determined. The local neighborhood is then projected onto a vector of receptive fields which has been normalised to this scale and orientation. The vector was then used as an index to generate a list of hypotheses for possible objects and image neighborhoods having similar appearance.

For recognition the hypothesis list of the current test point is evaluated. No previous knowledge is used. We point out that the performance of the system can be increased by combining hypotheses with previous knowledge in a prediction-verification algorithm. Comparing the recognition results based on the hypotheses of one single point only gives more precise information about the precision and reliability of the different approaches. The approach can be generalised to recognition under different view points by including images from sample points along the view sphere in the training set.

For each neighborhood, the method produces a sorted list of image neighborhoods from all the trained objects with a similar appearance. Similarity in appearance is determined by the distance between the vector of responses to the receptive fields. A list of neighborhoods within a tolerance distance (epsilon) are returned. This list is sorted by similarity. If the list is too large, then the neighborhood is judged to be non-discriminant and is rejected. Similarly, if no neighborhoods are found within a tolerance, the neighborhood is judged to be unstable, and is rejected. Neighborhoods for which a small number of similar matches are found are labeled as “accepted” in the experiments below.

The recognition rates must be seen in combination with the acceptance rate. The goal is to obtain high acceptance rates together with high recognition rates. Thus, to evaluate the results of the techniques, three values are presented. First, the percentage of neighborhoods that produced a hypothesis are displayed. The number of such neighborhoods is labeled as the “acceptance rate”. This is the percentage of neighborhoods which are both unambiguous and stable. Secondly, we display the number of neighborhoods for which the most similar recalled neighborhood is from the correct object. These cases are labeled “1st answer correct". A third value presents the number of returned neighborhoods for which the correct object and neighborhood was in the best three returned neighborhoods (correct answer among first 3). Such slightly ambiguous neighborhoods can be employed by a prediction-verification algorithm for recognition. All values are average values over the test scenes with uniform background (section 5.1, 5.2, 5.3) or over the test scenes with cluttered background (section 5.4).

5.1 Local appearance technique based on luminance

This experiment is computed on luminance images according to the technique described in section 3 using recursive filters, automatic scale selection, and steerable filters. This experiment is the benchmark for the following experiments.

object number	0	1	2	3	4	5	6	7
acceptance rate	0.30	0.41	0.65	0.04	0.47	0.54	0.07	0.23
1st answer correct	0.40	0.27	0.62	0.59	0.28	0.12	0.91	0.43
correct answer among first 3	0.77	0.51	0.83	0.82	0.62	0.47	1	0.81

Table 1. Results of technique based on luminance receptive fields. Neighborhoods of objects with discriminant structure are easily recognised. However, luminance provides poor discrimination for uniform and specular objects.

Neighborhoods from objects eraser (0), pen (1), scissors (2), tape (4), sharpener (5) and sweets box (7) produce good acceptance rates. The acceptance rates for neighborhoods from the stapler (3) and protractor (6) are very low which indicates that for most of the observed neighborhoods are unstable or ambiguous. The recognition rates for these objects are thus based on an insufficient number of windows and should not be considered to judge the accuracy of this particular experiment. These two objects are very hard to recognise by a system using only luminance.

Objects eraser (0), scissors (2) and sweets box (7) produce sufficiently high recognition rates and a simple voting algorithm could be used for recognition. A prediction-verification approach would produce a robust recognition for these objects, as reported by Colin de Verdiére [4]. Poor results for recognising neighborhoods are obtained for objects pen (1), tape (4) and sharpener (5).These objects are either uniform or specular, which makes the recognition using only luminance difficult.

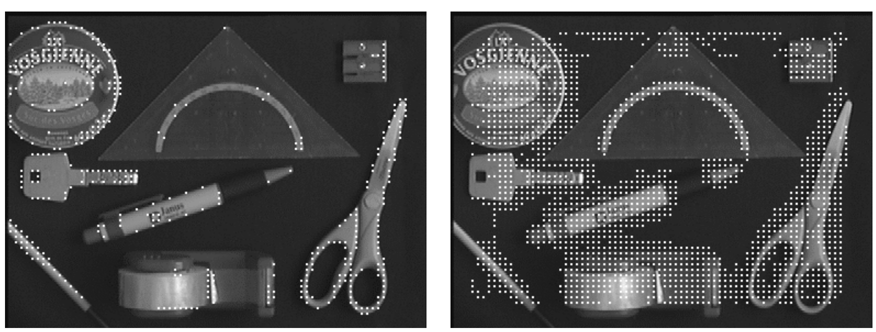
5.2 Coloured receptive field technique using 0th order Gaussian derivative in colour channels

In this experiment two chrominance channels are added to the receptive field vector. These two axis, which are orthogonal to the luminance axis, are encoded with a 0^{th} order Gaussian with size σ as determined by local normalisation. These two channels capture information about the chrominance in the neighborhood of each point. This provides good recognition rates for structural objects in the previous experiment as well as a large improvement in acceptance and recognition for the constant and specular objects.

object number	0	1	2	3	4	5	6	7
acceptance rate	0.82	0.78	0.86	0.97	0.86	0.87	0.19	0.91
1st answer correct	0.87	0.93	0.79	0.96	0.66	0.74	1	0.99
correct answer among first 3	0.94	0.98	0.91	1	0.88	0.96	1	1

Table 2. Results of technique extended to 0^{th} order Gaussian derivative in chrominance channels. High recognition rates are obtained for all objects, although the acceptance rate for transparent objects remains low.

The addition of chrominance information raised the acceptance rates from an average of 0.34 in the previous experiment to an average of 0.78. Many fewer neighborhoods are rejected because of ambiguous or unstable structure. Figure 5 illustrates the decrease of the number of ambiguous windows using grey scale and coloured receptive fields. This is an important improvement because even for difficult objects many windows produce a result, which was not the case in the previous experiment. The only object with a low acceptance rate is object protractor (6), which is transparent and particularly difficult to describe.



(a) (b)

Fig. 5. White points mark non-ambiguous (accepted) windows. (a) accepted windows for grey scale receptive fields. (b) accepted windows for coloured receptive fields.

Very good recognition rates are obtained for all objects. The lowest first answer recognition rates are obtained for objects tape (4) and sharpener (5). These objects are highly specular and thus change their appearance with pose and illumination. Even for these objects the recognition rates are sufficiently high that a simply voting scheme could be used for recognition in restricted domains.

5.3 Coloured receptive field technique using 0th and 1st order Gaussian derivatives in colour channels

In this experiment the chrominance information is extended to the first derivatives in order to capture colour gradients that are characteristic for the object. The structure analysis is performed in the 1^{st} and 2^{nd} order derivatives. The 3^{rd} order derivative is abandoned, because its analysis is only interesting when the 2^{nd} order derivative is important [8]. The descriptor space has than 8 dimension which helps to avoid the problems that occur in high dimensional spaces. The comparison of table 1 and table 3 validates that the improvement by using colour is much superior to the loss in structure recognition by abandoning the 3^{rd} order derivative.

object number	0	1	2	3	4	5	6	7
acceptance rate	0.88	0.87	0.91	0.98	0.83	0.98	0.23	0.99
1st answer correct	0.91	0.98	0.86	0.97	0.74	0.77	0.96	1
correct answer among first 3	0.98	0.99	0.94	0.99	0.90	0.97	0.99	1

Table 3. Results of technique extended to 0^{th} and 1^{st} order Gaussian derivatives in chrominance channels. High recognition rates are obtained for all objects. Average results are slightly superior than those in section 5.2.

The acceptance rates are in the average higher than in the previous experiment. The acceptance rate for object protractor (6) is still relatively low. The recognition rates are slightly superior to the recognition rates obtained previous experiment. This shows that colour gradient holds information which improves the discrimination of objects.

5.4 Experiments on cluttered background

The benchmark technique (section 3) produces very low recognition rates (table 4). We obtain a mean of 0.1238 in the first answers, which is even worse than guessing. This means that the background introduces new structures that were not present in the training base. These structures are so important that a correct classification is very difficult.

In the case of a cluttered background, the use of chrominance provides a dramatic improvement in recognition rates. Objects eraser (0), pen (1), tape (4) and sweets box (7) have high recognition rates together with high acceptance rates which allow a reliable classification. There are problems with objects scissors (2),

object number	0	1	2	3	4	5	6	7
acceptance rate	0.04	0.52	0.44	0.33	0.59	0.92	0.54	0.25
1st answer correct	0	0.15	0.39	0	0.03	0	0.17	0.25
correct answer among first 3	1	0.42	0.70	0	0.17	0.15	0.27	0.64

Table 4. Results for objects on cluttered background using technique based on luminance images. Very low recognition rates are observed. Object recognition is difficult.

object number	0	1	2	3	4	5	6	7
acceptance rate	0.65	0.71	0.51	0.80	0.70	0.81	0.53	0.86
1st answer correct	0.91	0.80	0.50	0.34	0.70	0.35	0.29	0.94
correct answer among first 3	0.94	0.90	0.61	0.68	0.87	0.42	0.31	0.98

Table 5. Results for objects on cluttered background obtained by technique extended to 0th order Gaussian derivative in colour channels. Few windows are rejected. Object recognition is possible.

object number	0	1	2	3	4	5	6	7
acceptance rate	0.75	0.60	0.50	0.79	0.60	0.14	0.41	0.96
1st answer correct	0.85	0.82	0.58	0.19	0.77	0.27	0.39	0.99
correct answer among first 3	0.93	0.86	0.64	0.39	0.84	0.55	0.43	0.99

Table 6. Results for objects on cluttered background with technique extended to 0th and 1st order Gaussian derivatives in colour channels. High acceptance rates are observed. Object recognition is possible.

stapler (3), sharpener (5) and protractor (6) either due to low acceptance rates or low recognition rates. For transparent objects such as the object protractor (6) this is expected, because it depends very much on the background conditions. Objects scissors (2), stapler (3) and sharpener (5) are either small, thin or have holes. This means a large amount of neighborhoods contain background information which perturbs the classification.

Another interesting observation is that in the case of background clutter, the acceptance rates using only the 0^{th} order Gaussian in the colour channel are slightly higher to the acceptance rates obtained by the technique using the 0^{th} and 1^{st} order derivative in the colour channels. The cluttered background contains a large set of colours, which are not present in the training base. This variety leads to colour gradients at boundaries which have not been observed in training and are thus rejected. Receptive fields based on the 0^{th} order Gaussian are much less sensitive to such background distraction. This is interesting because on uniform background the technique using the colour gradient has been found superior to the technique using only the 0^{th} order Gaussian.

6 Conclusions

The results presented in this article are incremental and primarily experimental. We have experimentally investigated the extension of the technique of [4] to the problem of real time observation of the physical icons for computer human interaction. Certain characteristics of real world objects, such as specularity, transparency or low structure, variable background and changing camera positions make the identification of objects difficult.

The recognition technique evaluated in this article employs local orientation normalisation to provide invariance to image plane rotations. Robustness to scale changes are provided local normalisation using automatic scale selection. The technique can be implemented to operate in real time by recursively computing separable Gaussian filters. Such filters are steered to the local orientation using the steerability property of Gaussian derivatives. Training was performed for the grey scale technique in 237s on a Pentium II 333 MHz. The techniques using colour needed both 278s for 16 training images of average size of 39 212 pixels.

The benchmark technique produces satisfactory recognition results on uniform background. It can clearly be stated that structured objects exhibit higher classification rates. The approach fails for uniform objects, because of a lack of structure. A pure luminance based approach also has problems with difficult objects, such as transparent or specular objects. Recognition rates for cluttered background using only luminance are below chance.

The method is extended by the addition of chrominance information. An chrominance descriptor space is presented that can describe colour images and does not increase the dimensionality greatly in comparison to the starting point technique. Problems with high dimensional spaces are avoided. A system is obtained that preserves the advantages of the pure luminance approach and is capable of classifying a much wider range of objects. It is not significantly more expensive in computation and storage. The experimental section validates that objects with difficult features can be recognised, even on cluttered background. It also indicates that chrominance is more important to recognition than higher order derivatives.

We are currently working to extend the approach to view-variant pose estimation. Recognition under different view points is obtained by including images taken under different view points in the training base. The object pose will be estimated by using the geometrical information of the results from several recognised points. A more robust pose estimation will be obtained by using a prediction-verification algorithm. The result should be a system with high precision, robustness and reliability.

References

1. M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions.* MIT Press, 1965.
2. E.H. Adelson and J.R. Bergen. The plenoptic function and the elements of early vision. *Computational Models of Visual Processing*, pages 3–20, 1991.
3. V. Colin de Verdière. *Représentation et Reconnaissance d'Objets par Champs Réceptifs.* PhD thesis, Institut National Polytechnique de Grenoble, France, 1999.
4. V. Colin de Verdière and J.L. Crowley. A prediction-verification strategy for object recognition using local appearance. Technical report, PRIMA Group, GRAVIR Lab, Grenoble, France, 1999. available at ftp://ftp.inrialpes.fr/pub/prima/.
5. G. Fitzmaurice, H. Ishii, and W. Buxton. Bricks: laying the foundations for graspable user interfaces. In *Proceedings of Computer Human Interaction (CHI '95)*, pages 442–449. ACM Press, 1995.
6. W.T. Freeman and E.H. Adelson. The design and use of steerable filters. *Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906, September 1991.
7. D.H. Hubel. *Eye, Brain, And Vision.* Scientific American Library, New York, USA, 1988.
8. J.J. Koenderink and A.J. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, pages 367–375, 1987.
9. T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.
10. B. Schiele. *Object Recognition using Multidimensional Receptive Field Histograms.* PhD thesis, Institut National Polytechnique de Grenoble, Grenoble, July 1997.
11. C. Schmid and R. Mohr. Local greyvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997.
12. M.J. Swain and D.H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1), 1991.
13. B. Ullmer, H. Ishii, and D. Glas. mediablocks: Physical containers, transports, and controls for online media. In *SIGGRAPH '98*, USA, 1998.
14. I.T. Young and L.J. van Vliet. Recursive implementation of the gaussian filter. *Signal Processing*, pages 139–151, 1995.