CSAIL

massachusetts institute of technology — computer science and artificial intelligence laboratory

Rotation Invariant Object Recognition from One Training Example

Jerry Jun Yokono and Tomaso Poggio

AI Memo 2004-010
CBCL Memo 238

April 2004

Abstract.

Local descriptors are increasingly used for the task of object recognition because of their perceived robustness with respect to occlusions and to global geometrical deformations. Such a descriptor--based on a set of oriented Gaussian derivative filters--is used in our recognition system. We report here an evaluation of several techniques for orientation estimation to achieve rotation invariance of the descriptor. We also describe feature selection based on a single training image. Virtual images are generated by rotating and rescaling the image and robust features are selected. The results confirm robust performance in cluttered scenes, in the presence of partial occlusions, and when the object is embedded in different backgrounds.

1 Introduction

Object recognition in a real world environment is one of the challenging tasks in computer vision research. Recent reports describe good performances under various conditions such as different illuminations and clutter backgrounds. Within recognition, we refer to categorization as the task to classify generalized categories such as faces and cars and to identification as the task to recognize a specific object. Interestingly, the task of object identification but not categorization is addressed in the framework of one-shot object learning.

The key to object categorization is how to build a generalized object model in feature space to classify the objects into a category using statistical learning. Schneiderman [23] used a histogram-based technique and Osuna [27] used the Support Vector Machine (SVM) classifiers to separate face and non-face. Both approach showed good categorization ability. Recently, Heisele [21][22] presented a novel SVM-based method using facial components and their geometrical configuration. Components are selected automatically using a statistical error bound; a SVM classifier is then generated for each component, followed by a higher-level combination classifier. Viola and Jones [28] described a real-time face detection system using features similar to Haar-like wavelets introduced in object recognition by Poggio [29]. They use a cascade processing to implement their AdaBoost learning algorithm. This learning procedure not only tunes the thresholds for the wavelet-like features but also effectively performs feature selection.

On the other hand, for an object identification system is crucial to find invariant features for various kinds of transformations such as rotation, scale, illumination, and deformations. Schiele and Crowley [17] proposed a multi-dimensional histogram approach which can identify 3D objects. In their system, Gaussian derivatives are used to create a multi-dimensional feature space. Mel [20] developed a histogram-based object recognition system that is inspired by properties of the human visual cortex model and uses multiple low-level attributes such as color, local shape and texture. Although these systems are robust to changes in rotation, translation, and deformation, they cannot handle recognition in a cluttered scene when similar patterns appear in the background. In general, it should be clear that significant invariance to rotation and deformation can only be achieved at the expense of specificity. Recently, Lowe [9] developed an object recognition system that works well in cluttered scenes and achieves rotational and scale invariance by using a unique local descriptor inspired by Edelman and Poggio [30]. The descriptor is constructed from several collections of orientation histograms; it is normalized with the canonical orientation of that local region. Matching is performed by efficient nearest neighbor search and by generalized Hough transform followed by solving the affine parameters. The system can learn an object model from a single image.

In most of the categorization approaches reported in the literature, a large number of training images is needed off-line to solve problems such as face detection. Our main goal in this report is to describe an on-line learning scheme than can be effective after just one example but can also improve incrementally when more examples become available.

As a first step, we have evaluated the performance of a local descriptor based on sets of oriented Gaussian derivatives [5]. Gaussian derivative filters are known for their selectivity to specific orientation and frequency. Comparisons of several local descriptors have been done using various combinations of Gaussian derivatives with different orientations and scales with the simplest local descriptor: gray value patches. We have performed the comparison in terms of criterion of selectivity and invariance. In [5], we have shown that our Gaussian descriptor is robust against affine changes such as rotation and scale changes. We also implemented a simple recognition system that can learn object model from only single image and achieved good recognition ability even if objects are partially occluded and are in different illumination environments. Since the local descriptor only responds to particular pattern, the system is not able to detect rotated objects. In this paper, we achieve a system that can recognize rotated objects as well as robust feature selection.

In section 2, we review our local descriptor based on the Gaussian derivatives which is implemented using "steerable filters" [6]. In section 3, we show how our descriptor achieves rotation invariance by steering the filter response using local gradient orientation. Robust feature selection is discussed in section 4. Section 5 shows experiments and section 6 concludes the paper.

2 Gaussian derivative based local descriptor

2.1 Steerable filters

Gaussian derivatives are widely used filters which have spatial orientation selectivity as well as frequency selectivity. These filters, which show robustness to illumination changes and viewpoint changes, have many interesting properties. The first and second derivatives correspond to the edge feature and the bar feature, respectively. The steerable filter response for the *n*th order Gaussian derivative $G_n(\theta)$ to an arbitrary orientation θ is, given the Gaussian distribution *G:*

$$
\begin{aligned}
&G = e^{-\frac{x^2+y^2}{2\sigma^2}} \\
&G_1(\theta) = \cos(\theta)G_1(0^o) + \sin(\theta)G_1(90^o) \\
&G_2(\theta) = k_{21}(\theta)G_2(0^o) + k_{22}(\theta)G_2(60^o) + k_{23}(\theta)G_2(120^o) \\
&k_{2i}(\theta) = \frac{1}{3}\{1 + 2\cos(2(\theta - \theta_i))\} \\
&G_3(\theta) = k_{31}(\theta)G_3(0^o) + k_{32}(\theta)G_3(45^o) + k_{33}(\theta)G_3(90^o) + k_{34}(\theta)G_3(135^o) \\
&k_{3i}(\theta) = \frac{1}{4}\{2\cos(\theta - \theta_i) + 2\cos(3(\theta - \theta_i))\}
\end{aligned}
\tag{1}
$$

where $k_{in}(\theta)$ is the coefficient of the basis.

We use Gaussian widths of 1, 2, and 4. The filters cover enough orientations and frequencies by using three scales of Gaussian derivatives. At each pixel location, the local information captured by the filter depends on the width of the Gaussian function. When the Gaussian is wider, the filter captures information from larger neighborhoods and it responds to lower frequency.

2.2 Local descriptor

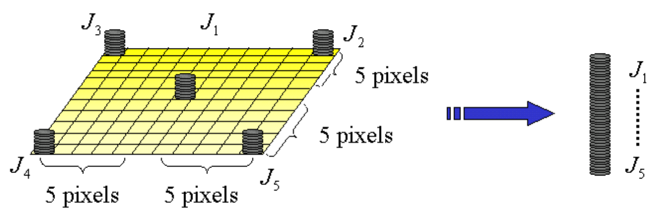

Figure1. Gaussian derivatives up to the third order with four orientations and three scales are computed and expanded using four neighboring pixel locations $J = \{J_1, J_{2,} \cdots, J_i\}$, where $J_i = \{G(\lambda, \theta, \sigma)\}$, $\lambda : order, \theta : orientation, \sigma : GaussianWidth$

Local descriptors should have selectivity to specific local patterns as well as invariance to affine changes such as rotation, scale and translation. If selectivity is poor, many incorrect matches will appear, making recognition difficult. We have evaluated the selectivity and the invariance of the descriptor with the filter as a function of the width of the Gaussian, the number of widths, and the orders of the derivatives [5]. Normally, using high order derivatives makes the descriptor more selective to the specific local pattern. Of course, there is a trade-off between selectivity and invariance. A descriptor might be very specific but show poor invariance to affine changes. We have showed that the third order descriptor has good performance under both criteria of selectivity and invariance. We use Gaussian derivatives up to the third order, with four orientations and three widths. The vector length of the jet descriptor associated with a location in the image is 3x3x4=36. Although the descriptor has good discriminant power only from one pixel location, it is not sufficient when patterns might be simple. The descriptor can be expanded, combining the neighboring four jets, which are five pixels away from the center pixel and the length of the local descriptor is 36x5=180. The local descriptor used in the experiments is shown in Figure 1.

3 Rotation Invariance

Rotational invariance of the descriptor is achieved by computing a main orientation of the local region, and "steer" the neighboring pixel locations and their filter responses.

We consider three possibilities to compute the main orientation based on the:

(i) gradient orientation at a center pixel location.
(ii) a peak in the orientation histogram of the local region.
(iii) orientation of the eigenvector of the second moment matrix of the local region.

Although above three techniques are all well-known, we perform experiments to decide which one is suitable for our system. We perform experiments regarding robustness to rotation changes, scale changes, and noise to evaluate the robustness of (i) and (ii). (iii) is a good way if we use eigenvalues to detect interest points since interest point detection and orientation estimation can be done at the same time. However, computing eigenvalues at all pixel locations is expensive. Since we use the Harris [8] corner detector, we can use the relative ratio of the eigenvalues instead of computing the eigenvalues.

3.1 Robustness to rotation

Using the COIL-100 image database, a main orientation at each interest point is computed. Due to the discreteness of the digital image, a gradient orientation of a pixel changes when the pattern is rotated. Using the Gaussian kernel, gradient orientation robustness can be improved. The gradient orientation can be computed from the first order derivative response of our descriptor.

$$\alpha = atan2(Gx, Gy; \sigma) \qquad (2)$$

where Gx and Gy are the first order derivatives of the Gaussian and σ is the width of the Gaussian, respectively.
We compare (i) and (ii) regarding robustness of orientation estimation. (i) has the advantage that we can compute it directly from the filter response. The orientation histogram is computed using an 11x11 neighborhood.
We conducted the following experiments to evaluate the two techniques:

1. Run Harris corner detector to find interest corner-like points.
2. Compute main orientations α_1 using both (i) and (ii).
3. Rotate the image by φ degrees.
4. Compute the main orientations α_2 at the corresponding pixel locations.
5. $\beta = \alpha_2 - \alpha_1$ is the estimated rotation angle.
6. Count the number of feature points which are within tolerance ε.
   $\varphi - \varepsilon < \beta < \varphi + \varepsilon$.

9679 points from all COIL 100 images are used. We use $\varepsilon = 15$ for the tolerance at which the descriptor still shows high correlation for the slight rotation and this value is found out reasonable. Result is shown in Figure 2. If we use wider width of Gaussian at center pixel, the orientation estimation shows the best performance. Note that when the rotation degree is $-\pi, -\pi/2, 0, \pi/2, \pi$, there is no intensity interpolation during the rotation transformation that the intensity is exactly the same as original image and rotation estimation is perfect. If we use the Gaussian width of two or four, the performance is better than using orientation histogram. Since interest points are corner-like points, there are sometimes two peaks in the histogram in which case the incorrect direction might be chosen as the main direction.

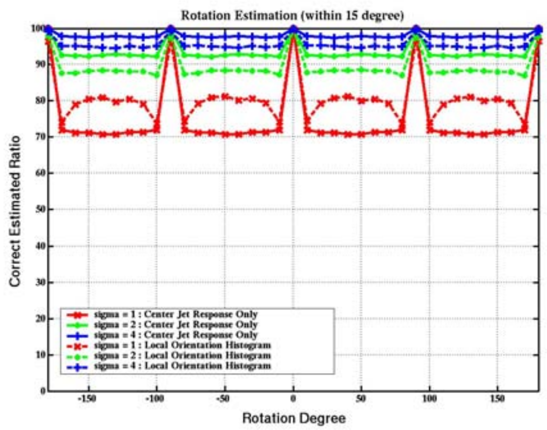

Figure 2. Orientation estimation is performed using COIL images. If we use larger width of the Gaussian, the gradient orientation of the first order derivatives shows robust estimation.

3.2 Robustness to scale changes

We also perform an orientation estimation for various scale changes. Images are resized linearly and a main orientation is computed at each point. The same 9679 points as before are used in the experiment. Here, (i) and (ii) show almost the same results for the Gaussians with large width.

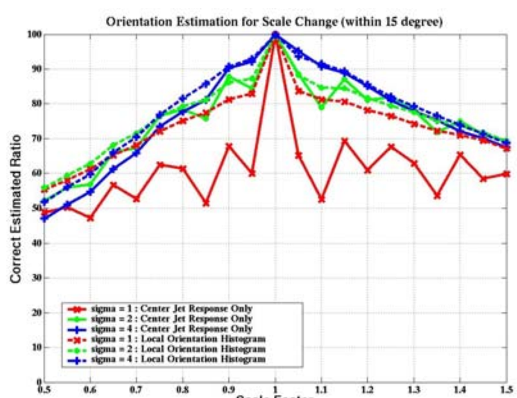

Figure 3. Orientation estimation for scale changes is performed. If we use larger width of the Gaussian, performance is almost the same.

3.3 Robustness to Gaussian noise

We also evaluate the robustness to Gaussian noise. The same 9679 points are used to see the performance. Again, larger width of Gaussian response shows best performance. Even when 10% of Gaussian noise is added, more than 90% of feature points are estimated correctly.

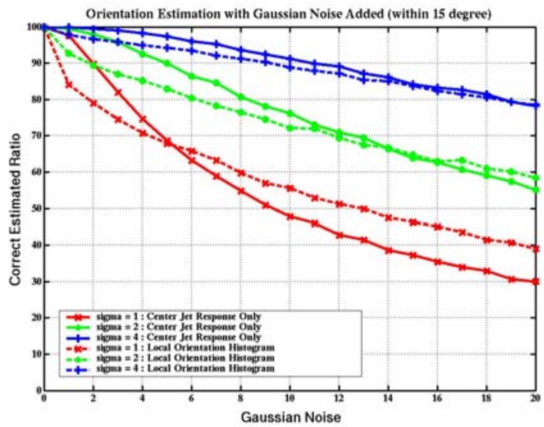

Figure 4. Robustness of the descriptor to additive Gaussian noise is examined. Larger width of the Gaussian makes estimation robust. Local orientation at a center pixel location shows enough robustness for the recognition.

3.4 Steering local descriptor

Experiments described in the previous section suggest the use of a gradient orientation with larger width of Gaussian at a center pixel location as a main orientation. Advantage of this strategy is that we can use it directly from the filter responses of the basis functions of the first order derivatives. Once we compute the main orientation of the center pixel, the filter responses at the center pixel are steered. Four neighboring pixel locations and their filter responses are also steered to create a rotational invariant descriptor as shown in figure 5. The steered *n*th order Gaussian derivative is computed by following:

$$G_n = G_n(\theta_i + \alpha), \quad \theta_1 = 0, \theta_2 = \frac{pi}{2}, \theta_3 = pi, \theta_4 = \frac{3pi}{2} \qquad (3)$$

where α is the main orientation at a center pixel.

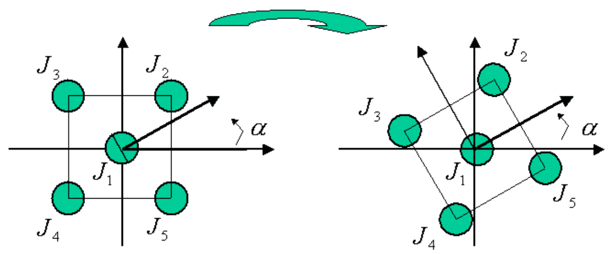

Figure 5. Rotation invariant local descriptor is realized by steering filter response of a center pixel to a main orientation. Neighboring four pixel locations and their filter responses are also steered by the response of the center pixel.

4 Selection of robust features for object learning

Harris corner detector finds many interest points in the object image, normally couple of hundreds if we use 240x320 image. There are feature vectors which might not take a important role in the recognition. Those feature vectors are normally not robust to rotation changes or scale changes. If we can remove those vectors, it might be useful to discard outliers and also can reduce computational cost. Figure 5 and figure 6 show robustness to rotations changes and scale changes, respectively. The x-axis is the feature index. In this case, there are 106 feature vectors in the object model. The left y-axis is the histogram (blue bar graph) showing how many times the descriptor estimated consistent orientation and right y-axis is the usage ratio (red plot graph) of actually used in the recognition. By rotating the image with ten degrees interval, if the rotation degree is estimated correctly, vote the feature index incrementally. The number of bins is 360/10=36. The red plot indicates the feature vectors that are used in the actual recognition experiment. We resize the image from $1/\sqrt{2}$ to $\sqrt{2}$ linearly. It is obvious that either rotationally or scale unstable features are less used during the recognition. Those feature vectors are not stored in the object model. Normally from five to ten percent of those features are eliminated. If many images are available for training, it would be useful to remove feature vectors that are not used for the recognition. For the one-shot learning system, feature selection is done by generating virtual images and removing unstable features.

5 Object recognition system

To confirm the viability of the descriptor, we implemented an object recognition system that is capable of learning from one example of an object.

5.1 Learning an object model

To realize scale invariant recognition, a multi-scale representation of an object model

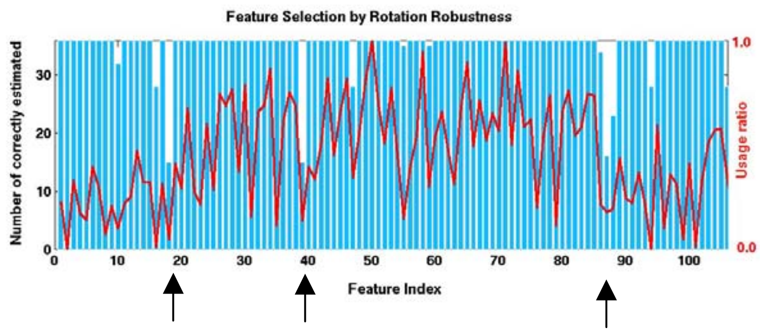

Figure 6. Rotation unstable features are not used in the object model. These features can be chosen by computing the histogram of correct rotation estimations. Blue histogram shows how many counts the feature estimates the correct orientation. Red plot is the usage ratio of the features actually used in the recognition. Arrow indicates the feature indexes that are not robust to rotations.

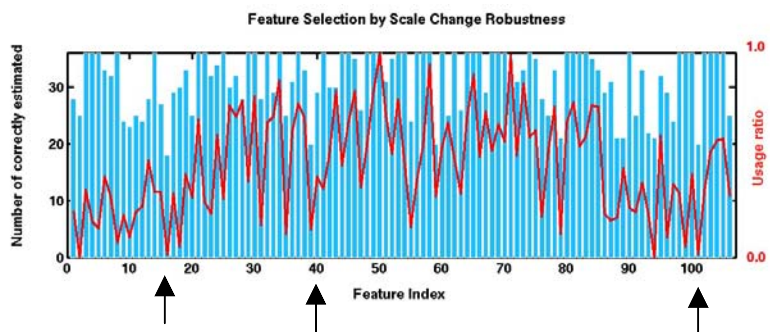

Figure 7. Features that are not robust to scale changes are also not stored in the object model. From five to ten percent of the feature vectors are eliminated. Arrow indicates the feature indexes that are not robust to scale changes.

is constructed using a Gaussian kernel to generate the scale-space. The width of the Gaussian function increases by $\sqrt{2}$ and the image is subsampled by the factor of $\sqrt{2}$ as the level of the scale goes up. The Harris corner detector runs at each scale to find interest points. Among the interest point detectors, the Harris corner detector is found to be the best under criteria of repeatability and information content [12]. The criterion for corner detection is based on the relative ratio of eigenvalues in local region avoiding directly computing of the eigenvalues. At each interest point, its rotation invariant local information described in the previous sections is computed and if it satisfies the criterion of rotation robustness and scale robustness, the steered feature vectors are stored with their local main orientation. Rotation robustness and scale robustness features are more than half of the histogram shown in figure 6 and 7.

5.2 Recognizing objects in cluttered scene

To recognize objects in cluttered scenes, the Harris corner detector is applied to find interest points and jets are computed at each point as in the learning part of the

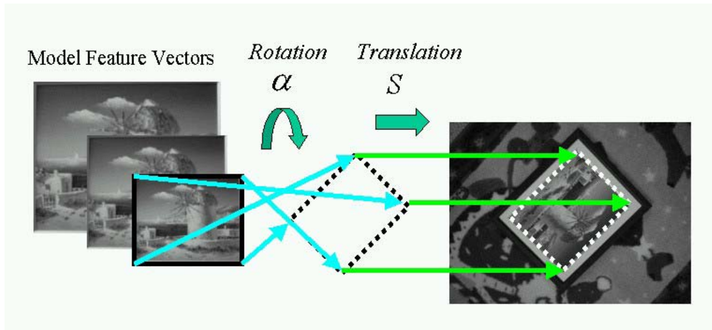

Figure 8. Corresponding pairs are computed by matching rotation invariant feature vectors. Rotation degrees α from the object model is estimated by the angle distribution of the matched pairs. Each matched pairs are then rotated based on the peak angle to create translation space S and coordinates of matched pairs are subtracted.

procedure. For each scale level and for each jet, a highest correlating jet is searched in the object model. We use the normalized correlation as a distance metric.

$$Correlation(v_1, v_2) = \frac{(v_1 - \overline{v_1})(v_2 - \overline{v_2})}{\sqrt{(v_1 - \overline{v_1})^2 (v_2 - \overline{v_2})^2}} \qquad (4)$$

At this point, we usually have many outliers on the background. At each level of the scale space, corresponding pairs are first filtered out by the orientation difference which can be computed by subtracting the main orientation of the model jet from the orientation of the detected jet (figure 9: bottom-left). If there are peaks in the distribution, those corresponding pairs are fed into a location translation filter (figure 9: bottom-right). The coordinates of the correspondence pairs are subtracted to create a voting space. Here, model coordinates are rotated to the peak orientation shown in figure 8. If any peaks are found above the threshold in the translation space, the objects candidates are detected. The threshold for voting differs at each scale level. The final decision is performed by estimating the object position in the scene. Using matching pairs of the translation space peak, homography is estimated by RANSAC [4].

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \qquad (5)$$

For the estimation, 6 points from correspondent pairs are first chosen randomly to estimate the planar perspective transformation. Then a score of the homography is computed by measuring how many points of remaining correspondent pairs satisfy this transformation. This cycle repeats 500 iterations and the best homography is used as the object position.

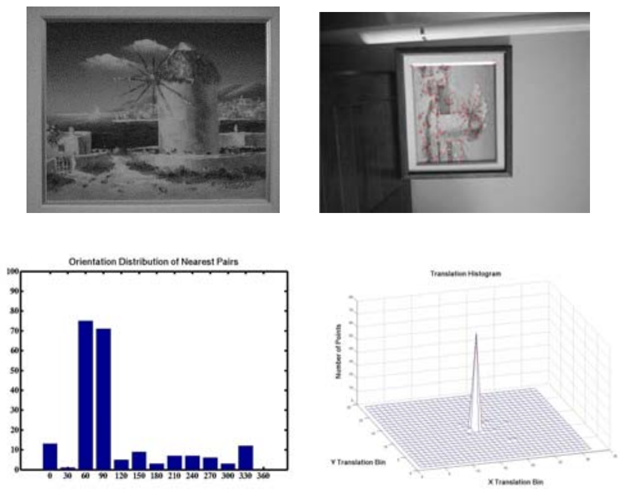

Figure 9. Model image (top-left) and a novel image: 90 degree rotated and also changed in scale (top-right). Rotation difference distribution of the feature vectors shows that many features with 90 degree rotated are in the image (bottom-left) and translation distribution (bottom-right) of the correspondent pairs shows rotated features have similar translation.

6 Performance of the recognition system

Figure 10. Training images used in the experiment. The system needs only one example image. Painting (left), toy phone (middle), and CD cover of Jerry & KENT (right).

Figure 10 shows some of the training images. We use only one image example for the experiment. We collected positive and negative images (240x320 pixels) for the test. Object images are taken under various environments. Some are taken with flashlight, some close to a window and some in a dark room. Sometimes the object of interest is occluded. We use 1000 negative cluttered images to examine false alarms. Some of the sample images are shown in Figure 11.

Figure 11. Negative test samples: 1000 cluttered scene images are used and a 100% true negative rate is obtained.

Figure 12. Objects are detected under different environment in different pose. 320x240 images taken under various environments with changes in illumination, scales, rotations and perspectives. Even a CD taken with flash and a painting with occlusion are robustly detected.

The system shows good performance on the test images even with partially occluded objects. Some of the detection results are shown in figure 12. Although the descriptor shows good selectivity and invariance to illumination changes, point-to-point correspondences tend to fail when the ambient illumination is different. It is interesting to mention that under these conditions, the system only uses the local patterns that are not much influenced by illumination changes. Surprisingly, we did not find any false alarm when matching with the these difficult images in the negative test set.

7 Conclusion

This paper presented a rotation invariant local descriptor based on the Gaussian derivatives. We use "steerable filter" to implement the derivative responses. Rotation invariance is achieved by "steering" the descriptor to the main orientation at a center pixel location. An advantage of this strategy is that the main orientation can be computed directly from the first order derivative responses. Although the technique is not novel, comparisons have been made to decide which one is appropriate for our system among other different well-known techniques. Many image processing algorithms in the literature used “steerable filter” to capture local information, in a way which is not discriminant enough for specific object recognition. To make the descriptor more discriminant, we found that neighboring pixel locations which are five pixels away from the center pixel location may be combined. These pixel locations are also geometrically steered to the main orientation together with their filter responses. We also consider feature selection in the case where only a single example is available. Virtual images are generated by rotating and rescaling the image. Rotationally and scale unstable features are computed by estimating the main orientation at the center pixel location and count the number of correctly estimated. Unstable features are removed during the learning step. The resulting object recognition system performs robustly under various illumination changes, viewpoint changes, scale changes, and rotation in the image plane, and under partial occlusion. The system shows extremely low false alarm rate. In our experiments, 100% true negative rate is obtained when 1000 cluttered images are used in the test. We are now working on an object recognition system which should be capable of handling incremental learning when additional example images of the object of interest become available.

Acknowledgements

This research has been done during the visit to CBCL, MIT with the support of Sony Corporation. Author would like to thank to Bernd Heisele and Pascal Paysan for useful suggestions and exciting discussions. Author also would like to thank to Kevin Yao-Jen Chang for useful advice for computing homography.

References

[1] A.Yuille and T.Poggio. “Scaling Theorems for Zero Crossings”. IEEE Trans. Pattern Analysis and Machine Intell., vol.8, pp.15-25, 1986

[2] A.Yuille and T.Poggio. “Scaling and Fingerprint Theorems for Zero-Crossings”. Advances in Computer Vision, pp.47-78, Lawrence Erlbaum, 1988

[3] T.Lindeberg. “Scale-Space Theory in Computer Vision”. The Kluwer International Series in Engineering and Computer Science, Kluwer Academic Publishers, 1994

[4] M. Fischler and R. Bolles. "RANdom SAmpling Consensus: a paradigm for model fitting with application to image analysis and automated cartography", Commun. Assoc. Comp. Mach., 24:381-95, 1981.

[5] Jerry Jun Yokono and Tomaso Poggio. “Oriented filters for Object Recognition: an empirical study”. Proceedings of the IEEE Conference on Face and Gesture Recognition (FG2004). Seoul, Korea. 2004.

[6] W. Freeman and E. Adelson. “The design and use of steerable filters”. PAMI, 13(9):891–906, 1991.

[7] R. Rao and D. Ballard. “Object indexing using an iconic sparse distributed memory”. ICCV, pp.24-31, 1995

[8] C. Harris and M. Stephens. “A combined corner and edge detector”. In Alvey Vision Conference, pp. 147–151, 1988.
[9] D. G. Lowe. “Object recognition from local scale-invariant features”. In ICCV, pp. 1150–1157, 1999.
[10] K. Mikolajczyk and C. Schmid. “Indexing based on scale invariant interest points”. In ICCV, pp. 525–531, 2001.
[11] C. Schmid and R. Mohr. “Local gray value invariants for image retrieval”. PAMI, 19(5):530–534, 1997.
[12] C. Schmid, R. Mohr, and C. Bauckhage. “Evaluation of interest point detectors”. IJCV, 37(2):151–172, 2000.
[13] J.G.Daugman. “Two-dimensional Spectral Analysis of Cortical Receptive Field Profiles”. Vision Research, vol.20, pages 847-856, 1980
[14] J.G.Daugman. “Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression”. IEEE Transactions on Acoustics, Speech and Signal Processing, vol.36, pages 1169-1179, 1988
[15] M.Lades, J.C.Volbruggen, J.Buhmann, J.Lange, C.Malsburg, R.P.Wurtz and W.Konen. “Distortion Invariant Object Recognition in the Dynamic Link Architecture”. IEEE Transactions on Computers. vol.42, No.3, pages 300-310, 1993
[16] M.Potzsch, N.Kruger and C.Malsburg. “Improving Object Recognition by Transforming Gabor Filter Response”. Computation in Neural Systems, Vol.7, No.2, pages 341-347, 1996
[17] B. Schiele and J. L. Crowley. “Probabilistic object recognition using multidimensional receptive field histogram”. In Proc. Int. Conf. on Patt. Recog., vol. B, pp 50-54, Vienna, August 1996.
[18] Bernt Schiele and James L. Crowley. “Object recognition using multidimensional receptive field histograms”. In ECCV (1), pages 610--619, 1996
[19] B. Schiele and J. L. Crowley. “Recognition without correspondence using multidimensional receptive field histograms”. International Journal of Computer Vision, 36(1):31--50, January 2000. 6
[20] B.W. Mel, "SEEMORE: Combining Color, Shape and Texture Histogramming in a Neurally Inspired Approach to Visual Object Recognition", Neural Computation, 9, pp. 777-804, 1997.
[21] Heisele, B., P. Ho, J. Wu and T. Poggio. “Face Recognition: Component-based versus Global Approaches”. Computer Vision and Image Understanding, Vol. 91, No. 1/2, 6-21, 2003.
[22] Heisele, B., P. Ho and T. Poggio. “Face Recognition with Support Vector Machines: Global Versus Component-based Approach”. International Conference on Computer Vision (ICCV'01), Vancouver, Canada, Vol. 2, 688-694, 2001.
[23] Schneiderman, H. and Kanade, T. “A statistical approach to 3d object detection applied to faces and cars”. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pages 746--751, Hilton Head Island, South Carolina, June 2000.
[24] S. A. Nene, S. K. Nayar and H. Murase, “Columbia Object Image Library (COIL-100)”. Technical Report CUCS-006-96, February 1996.
[25] K. Sung and T. Poggio, "Example-based Learning for View-based Human Face Detection," IEEE TPAMI, Vol. 20(1), pp 39-51, January 1998.
[26] B. Heisele, T. Poggio, and M. Pontil, “Face detection in still gray images”. Technical Report A.I. Memo No. 2001-010, C.B.C.L. Memo No. 197, MIT Center for Biological and Computational Learning (2000)
[27] E. Osuna, R. Freund, and F. Girosi, "Training Support Vector Machines: An Application to Face Detection," IEEE CVPR, 1997.
[28] Paul Viola and Michael J. Jones. “Robust real-time object detection”. Technical Report CRL 2001.
[29] Papageorgiou, C., M. Oren, and T. Poggio. “A General Framework for Object Detection”. In: Proceedings of the Sixth International Conference on Computer Vision (ICCV'98), Bombay, India, 555-562, January 1998.

[30] Edelman, S., Intrator, N., and T. Poggio. "Complex cells and object recognition", 1997