Global Gabor features for rotation invariant object classification

Ioan Buciu
Electronics Department
Faculty of Electrical Engineering and Information Technology
University of Oradea 410087, Universitatii 1
Romania
ibuciu@uoradea.ro

Ioan Nafornita
Electronics and Communications Faculty
"Politehnica" University of Timisoara
Bd. Vasile Parvan, no.2
300223 Timisoara
Romania
ioan.nafornita@etc.upt.ro

Ioannis Pitas
Department of Informatics
Aristotle University of Thessaloniki
GR - 541, BOX 451
Thessaloniki
Greece
pitas@aiia.csd.auth.gr

Abstract

The human visual system can rapidly and accurately recognize a large number of various objects in cluttered scenes under widely varying and difficult viewing conditions, such as illuminations changing, occlusion, scaling or rotation. One of the state-of-the-art feature extraction techniques used in image recognition and processing is based on the Gabor wavelet model. This paper deals with the application of the aforementioned model for object classification task with respect to the rotation issue. Three training sample sizes were applied to assess the method's performance. Experiments ran on the COIL - 100 database show the robustness of the Gabor approach when globally applied to extract relevant discriminative features. The method outperforms other state-of-the-art techniques compared in the paper such as, principal component analysis (PCA) or linear discriminant analysis (LDA).

1. Introduction

Despite the ease with which the biological visual system performs object recognition, this is a very complex computational task, requiring a quantitative trade-off between invariance to certain object transformations on the one hand, and specificity for individual objects on the other. For instance, object recognition ought to be invariant across huge variations in the appearance of objects such as building, for instance, due to viewpoint, illumination, or occlusions, but, at the same time, the system should maintain specificity, i.e., the ability to discriminate between different buildings. These complex constraints are taken into account when complex artificial object recognition systems are built. To date, there is no universal accepted computer based object recognition system which satisfactory copes with all the above mentioned requirements.

As the object recognition has been at the heart of computer vision research for the last decades, plenty of works in the field has been reported in the literature. Although it is impossible to exhaustively mention all methods and approaches proposed to deal with this task, some of them are worth mentioning. Lighting-invariant approaches have been developed by Lowe [1], and Mikolajczyk and Schmid [2]. In the former, the author proposed a local image descriptor based on scale invariant feature transform which utilized an orientation histogram using a Gaussian weighted window. Mikolajczyk and Schmid introduce affine invariant points where a Harris detector is used. Viewpoint-invariant image representation has been developed by Huttenlocher et al. [3] who made use of Hausdorff distance for finding image similarity. Another viewpoint invariance recognition method has been more recently proposed by Rothganger et al. [4], where multi-view constraints are associated with groups of patches for 3D object recognition. A shape model which both segments and recognize objects was proposed by Leibe et al. [5]. However, their approach relies on a probabilistic principle which can be crucially affected by

its precision. Caputo et al [6] presents an algorithm that combines support vector machines with local features via a new class of Mercer kernels for performing scalar products on feature vectors consisting of local descriptors, computed around interest points (like corners). A same strategy was employed by Caputo and Dorkó [7] which combine color and shape information, retrieved by kernel features, for a more robust object recognition accuracy (applied to the COIL-100 database). Several local descriptors (including Gaussian and Laplacian filters) are analyzed for the same COIL-100 database in [8]. However, no object recognition performance or comparison were reported for any of these local descriptors.

A particular approach for extracting relevant features is given by the so-called "elastic graph matching" [9], and its relative, named'elastic bunch graph matching" [10], which have been successfully applied for face recognition. "Elastic bunch graph matching" is based on applying a set of Gabor filters to special representative landmarks on the face (corners of the eyes and mouth, the contour of the face). These filters named Gabor jets represent the multiscale nature of receptive fields, as each component has a unique combination of orientation, frequency tuning and scale. The face is represented by a list of values that comprise the amount of contrast energy that is present at spatial frequencies, orientations and scales included in the jet. For the object recognition task, similar Gabor filters have been used by Würtz [11] to extract local features that are robust to translations, deformations, and background changes.

Most of the aforementioned papers, including Gabor methods, rely on manually or automatically selected local features which are extensively used for categorizing objects in current systems. While such approaches are reasonably successful in cases when the set of categories is rather limited, it is clear that in more complex situations some structuring of the features is unavoidable. Moreover, one drawback of these feature extraction techniques is the manual annotation of landmarks when the Gabor filters are applied to specific localized points. For large databases this annotation may become prohibited. Even when automatic annotation is carried out, the general performance is very sensitive to the registration and annotation precision. To avoid the shortcomings of the local object descriptors we propose here a global image representation based on the Gabor wavelets. The Gabor wavelets are applied to the whole image instead of choosing image patches. The experiments were run with respect to the rotation issue and revealed an excellent behavior of the global Gabor features for classifying objects compared to other two state-of-the-art methods named PCA [12] and LDA [13] (chosen as baseline), on the COIL-100 database. Three different sizes were chosen for the training data to investigate its robustness with respect to the rotation issue.

Figure 1. 100 objects from the COIL-100 database, each taken at frontal view, i.e. zero pose angle.

The outline of the paper is as follows. A short description of the database involved in experiments is presented in Section 2. Next, Section 3 shortly describes the Gabor functions, along with their convolution process and the new vector formation methodology. Experimental results are reported in Section 4 and conclusions are drawn in Section 5.

2. Database description

The Columbia Object Image Library (COIL-100) database [14] comprising color images of 100 objects has been chosen for the experimental part. The object images were taken at pose intervals of 5 degrees, resulting 72 poses per object, summing 7200 samples. Figure 1 shows the images of the 100 objects at frontal view. Each color image is converted to a gray-scale image and sampled down to 32×32 pixels to reduce the computational load. The training set was formed by selecting each fourth sample (corresponding to each object), starting with the first one (zero angle pose). More precisely, images corresponding to view angle of $0^0, 20^0, 40^0, \ldots, 320^0$ and 340^0, are picked up, yielding thus, 18 samples for each object. The training set, denoted by $\mathbf{X}^{training}$ comprises then 1800 samples. The remaining 5400 images form the test set, denoted by $\mathbf{X}^{test}$. For the sake of brevity we shall call this the *Set 1*. We further construct two more sets by selecting each eight sample, and each twelfth sample, respectively, (for each object) from the database to form the training set, with the remaining samples included in the test set. Therefore, the second training set, named *Set 2* contains 8 samples for each object (800 training samples), whilst the third set, *Set 3*, comprises only 6 samples/object (with a total of 600 training samples). As less samples are included in the training set we can have an insight on how robust are the approaches against the rotation degree for all methods investigated in the paper.

3. Gabor wavelets and feature vector formation

Gabor wavelets (filters or functions) are based on physiological studies of simple cells in the human visual cortex. The cells are selectively tuned to orientation as well as spatial frequency, and their response can be accurately enough approximated by 2D Gabor filters [15]. Thus, the increased popularity of this approach is biologically well justified. A 2D Gabor wavelet transform is defined as the convolution of the image $\mathcal{I}(\mathbf{z})$:

$$\mathcal{J}_{\mathbf{k}}(\mathbf{z}) = \int\int \mathcal{I}(\mathbf{z}')\psi_{\mathbf{k}}(\mathbf{z}-\mathbf{z}')d\mathbf{z}' \tag{1}$$

with a family of Gabor filters [10]:

$$\psi_{\mathbf{k}}(\mathbf{z}) = \frac{\mathbf{k}^T\mathbf{k}}{\sigma^2}\exp(-\frac{\mathbf{k}^T\mathbf{k}}{2\sigma^2}\mathbf{z}^T\mathbf{z})\left(\exp(i\mathbf{k}^T\mathbf{z})-\exp(-\frac{\sigma^2}{2})\right), \tag{2}$$

where $\mathbf{z} = (x, y)$ and $\mathbf{k}$ is the characteristic wave vector:

$$\mathbf{k} = (k_\nu \cos\varphi_\mu \qquad k_\nu \sin\varphi_\mu)^T \tag{3}$$

with

$$k_\nu = 2^{-\frac{\nu+2}{2}}\pi, \;\; \varphi_\mu = \mu\frac{\pi}{8}, \tag{4}$$
$$\nu = 0,1,2,3,4, \;\; \mu = 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}.$$

The parameters ν and μ define the frequency and orientation of the filter. Four orientations 0, $\frac{\pi}{4}$, $\frac{\pi}{2}$, $\frac{3\pi}{4}$ are used in our experiments. Two frequency ranges, i.e., high frequencies (hfr) with ν = 0,1,2 and low frequencies (lfr) with ν = 2,3,4 are also considered. Figure 2 depicts the 8th object from Figure 1 representing a car sampled down to 32×32 resolution. The magnitude for the Gabor high frequency range convolution result corresponding to the car image is shown in Figure 3, while Figure 4 illustrates the low frequency case.

4. Experimental Results

Each 32×32 image has been convolved with 12 Gabor filters corresponding to high frequency range and the four orientations, downsampled further by a factor of two, resulting an image of 16×16 pixels and scanned row by row to form a vector of dimension 256×1 for each Gabor filter. The 12 outputs have been concatenated to form a novel longer feature vector of dimension 3072×1. Hence the final matrix $\mathbf{X}_{Gabor}^{training}$ is $3072 \times n^{training}$, where the training feature vectors have been stored in the columns of $\mathbf{X}_{Gabor}^{training}$. The same procedure has been applied for the low frequency range. We took only the magnitude of Gabor representation, because it varies slowly with the position, while the phase is very sensitive to it. The procedure is applied to both $\mathbf{X}^{training}$ and $\mathbf{X}^{test}$. PCA and LDA is also applied to $\mathbf{X}^{training}$ by computing the eigenimages for various number of principal components (PCs), i.e., PCs $\in \{5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110120, 130, 140\}$. The PCA feature vectors used for classification are formed as $\mathbf{F}_{PCA}^{training} = \mathbf{W}^T\mathbf{X}^{training}$, where $\mathbf{X}^{training}$ is zero mean training samples, and $\mathbf{W}$ denotes the eigenimage matrix. A new test feature vector $\mathbf{f}^{test}$ is then formed as $\mathbf{f}_{PCA}^{test} = \mathbf{W}^T\mathbf{x}^{test}$, where $\mathbf{x}^{test}$ is a zero mean test sample. We should notice that, while the PCA and LDA involve a learning process to extract the relevant statistical (for PCA) and discriminant features (for LDA), the "learning" term for Gabor is somehow inappropriate, as the Gabor filter parameters are tuned by "by hand".

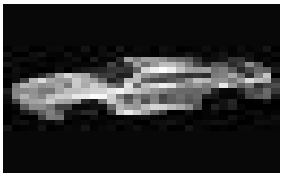

Figure 2. The 8th object from Figure 1, representing a car sampled down to 32×32 resolution.

The 100 objects form the set $\mathcal{L} = \bigcup_{j=1}^{100} \mathcal{L}_j$, where $\mathcal{L}_j$ is a class that corresponds to a particular object. The recognition rate is defined as $RR = \#\{p(\mathbf{c}_{test}) = l(\mathbf{c}_{test})\}$, where $l(\mathbf{c}_{test})$ is the ground truth for $\mathbf{c}_{test}$, and $p(\mathbf{c}_{test}$ is the predicted value of the classifier. Once Gabor features are extracted and the new feature vector is formed for all three procedures, as described above, two classifiers are employed to classify a test object image:

1. *Cosine similarity measure* (CSM). This approach is based on the nearest neighbor rule and uses as similarity the angle between a test vector and a training one. Let $\mathbf{x}_j^{training}$ (in the case of Gabor) be a column vector of $\mathbf{X}_{Gabor}^{training}$ that corresponds to the nearest class $\mathcal{L}_j$. Let also $\mathbf{x}_j^C$ be the nearest $\mathcal{L}_j^C$ class neighbor column vector for a test coefficient

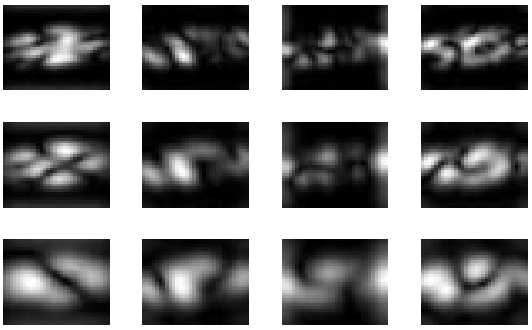

Figure 3. Magnitude of Gabor representation for the car image of Figure 2, convolved with 12 Gabor filters for ν = 2,3,4 (high frequency range - hfr) and μ = 0, $\frac{\pi}{4}$, $\frac{\pi}{2}$, $\frac{3\pi}{4}$.

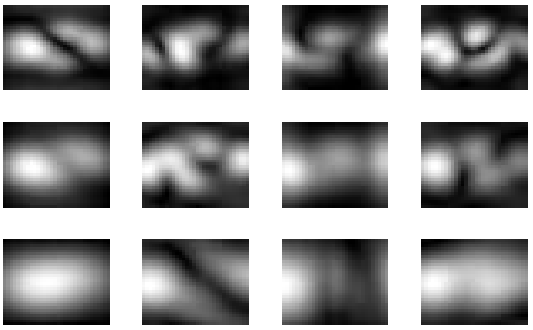

Figure 4. Magnitude of Gabor representation for the car image of Figure 2, convolved with 12 Gabor filters for ν = 0,1,2 (low frequency range - lfr) and μ = 0, $\frac{\pi}{4}$, $\frac{\pi}{2}$, $\frac{3\pi}{4}$.

vector $\mathbf{x}_{Gabor}^{test}$. We compute the quantities:

$$d_j = \frac{\mathbf{x}_{Gabor}^{T(test)} \mathbf{x}_j^{training}}{\|\mathbf{x}_{Gabor}^{test}\| \|\mathbf{x}_j^{training}\|}, \quad (5)$$

$$d_j^C = \frac{\mathbf{x}^{T(test)} \mathbf{x}_j^{C(training)}}{\|\mathbf{x}_{Gabor}^{test}\| \|\mathbf{x}_j^{C(training))}\|},$$

where d_j and d_j^C are the cosines of the angle between a test feature vector and the nearest training one. We assign $\mathbf{x}_{Gabor}^{test}$ to $\mathcal{L}_j$, if $d_j > d_j^C$. Otherwise $\mathbf{x}_{Gabor}^{test} \in \mathcal{L}_j^C$.

2. *Maximum correlation classifier* (MCC). The second classifier is the minimum Euclidean distance classifier. The Euclidean distance from $\mathbf{x}_{Gabor}^{test}$ to $\mathbf{x}_j^{training}$ is expressed as

$$\begin{aligned} \|\mathbf{x}_{Gabor}^{test} - \mathbf{x}_j^{training}\|^2 &= -2[\mathbf{x}_j^{T(training)} \mathbf{x}_{Gabor}^{test} - \\ &\quad \frac{1}{2}\|\mathbf{x}_j^{training}\|^2] + \mathbf{x}^{T(test)} \mathbf{x}_{Gabor}^{test} \\ &= -2h_j(\mathbf{x}_{Gabor}^{test}) + \mathbf{x}_{Gabor}^{T(test)} \mathbf{x}_{Gabor}^{test}, \quad (6) \end{aligned}$$

where $h_j(\mathbf{x}_{Gabor}^{test})$ is a linear discriminant function of $\mathbf{x}_{Gabor}^{test}$. A test image is classified by this classifier by computing two linear discriminant functions $h_j(\mathbf{x}_{Gabor}^{test})$ and $h_j^C(\mathbf{x}_{Gabor}^{test})$ and assigning $\mathbf{x}_{Gabor}^{test}$ to the class corresponding to the maximum discriminant function value.

The results are tabulated in Table 1 which reveals the superiority of the Gabor approach over the others. Despite its discriminative oriented design, LDA leaded to the worst recognition performances. As expected, all three methods performance decreases as the number of samples is reduced. For the Gabor approach, CSM conducted to slightly better results. Regarding the frequency range, a high frequency range favors the recognition performance when the rotation degree is low (expressed by the first data set construction). As the rotation degree gets higher the difference between "hfr" and "lfr" shrinks and the results becomes comparable. As in our case, less training samples means a higher degree of object's rotation expressed by different data sets, the drop in the recognition rate for the methods involved is shown in Table 2. For the Gabor approach, the largest decrease in the recognition rate is 3.79 while the corresponding drop is over 4 and 5 for the PCA and LDA, respectively, when switched from the *Set 1* to *Set 2*. This indicates a higher robustness for the Gabor approach with respect to the rotation degree compared to PCA and LDA methods. The difference is even

bigger when we compare the results for *Set 2* to the ones of *Set 3* as noticed from the Table.

Table 1. Recognition rate expressed in percentage (%) corresponding to all three sets, two classifiers (CSM and MCC) and three feature extraction techniques (Gabor, PCA, and LDA). For the Gabor approach, "hfr" and "lfr" stands for high and low frequency range, respectively. The number of PCs corresponding to the maximum performance is written in parenthesis for PCA and LDA. The highest recognition rate is shown in bold.

Data	Classifier	Feature extraction approach			
		Gabor		PCA	LDA
		lfr	hfr		
Set 1	MCC	97.81	**98.29**	96.93 (70)	95.98 (40)
	CSM	97.93	**98.39**	97.03 (60)	95.89 (40)
Set 2	MCC	94.16	**94.79**	92.65 (40)	90.68 (30)
	CSM	94.14	**95.02**	92.56 (40)	90.19 (50)
Set 3	MCC	**87.14**	87.08	83.25 (20)	75.21 (30)
	CSM	87.61	**88.45**	83.51 (20)	76.15 (140)

5. Conclusions

Gabor filters were successfully applied as feature extraction step to several research image processing or pattern recognition topics, including face recognition, texture recognition, facial expression recognition, etc. Their success is highly motivated due to their ability to extract features that tend to be scale, and rotation invariant. Moreover, the human visual receptive fields which share the same invariant constraints are satisfactory modelled by Gabor filters. In this paper we brought experimental evidence regarding the rotation robustness of the Gabor filters for rotated object recognition, where the experiments were run for two frequencies ranges. Instead of picking up Gabor jets at specific pixel locations, we applied the approach globally, avoiding thus the inconvenient of misalignment for local features. The results indicate more discriminative global feature retrieved by Gabor filters compared to the ones found by PCA and LDA. Future work can include experiments for recognizing similar objects from other databases (not present in the COIL database and training set), or experiments with object images having more complex background. The experiments have been performed by employing simple metric based classifiers. Using advanced classifiers such as multi-class support vector machines may enhance the overall classification performances.

Table 2. Drop in the recognition rate corresponding to Gabor, PCA, and LDA approaches and the two classifiers. The first two rows depicts the difference between the results of *Set 1* and *Set 2*. The last two rows shows the difference between the results of *Set 2* and *Set 3*. The lowest difference is in bold

Classifier	Feature extraction approach			
	Gabor		PCA	LDA
	lfr	hfr		
MCC	3.65	**3.50**	4.28	5.30
CSM	3.79	**3.37**	4.38	5.70
MCC	**7.02**	7.71	9.40	15.47
CSM	**6.53**	6.57	9.05	14.04

References

[1] D. W. Lowe, "Object Recognition from Local Scale-Invariant Features" *Proc. Intl Conf. Computer Vision*, pp. 1150–1157, 1999.

[2] K. Mikolajczyk and C. Schmid, "An Affine Invariant Interest Point Detector," *Proc. European Conf. Computer Vision*, vol. 1, pp. 128–142, 2002.

[3] D. Huttenlocher, G. Klanderman, and W. Rucklidge, "Comparing Images Using the Hausdorff Distance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850–863, Sept. 1993.

[4] F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce, "3D Object Modeling and Recognition Using Affine-Invariant Patches and Multi-View Spatial Constraints," *Proc. Computer Vision and Pattern Recognition*, pp. 272–280, 2003.

[5] B. Leibe and A. Leonardis and B. Schiele, "Combined Object Categorization and Segmentation with an Implicit Shape Model", *ECCV'04 Workshop on Statistical Learning in Computer Vision*, pp. 17–32, 2004.

[6] B. Caputo, C. Wallraven and M-E. Nilsback, “Object Categorization via Local Kernels”, in Proc. *17th International Conference on Pattern Recognition*, vol. 2, pp. 132–135, 2004.

[7] B. Caputo and G. Dorkó, “How to combine color and shape information for 3d object recognition: kernels do the trick”, *Advances in Neural Information Processing Systems,*, pp. 1375–1382, 2002.

[8] J. J. Yokono and T. Poggio, “Evaluation of sets of oriented and non-oriented receptive fields as local descriptors”, *Technical Report*, AIM-2004-007, 2004.

[9] M. Lades, J. C. Vorbrüggen, J. Buhmann, Jörg Lange, C. von der Malsburg, R. P. Würtz, and W. Konen, “Distortion invariant object recognition in the dynamic link architecture,” *IEEE Trans. on Computers*, vol. 42, no. 3, pp. 300–311, 1993.

[10] L. Wiskott, J.-M. Fellous, N. Kruger and C. von der Malsburg, “Face recognition by elastic bunch graph matching,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, 1997.

[11] Rolf P. Würtz, “Object recognition robust under translations, deformations and changes in background,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 769–775, 1997.

[12] I. T. Jolliffe, *Principal Component Analysis*, (2nd ed.), New York: Springer-Verlag, 2002.

[13] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.

[14] http://www.cs.columbia.edu/CAVE.

[15] T. Lee, “Image representation using 2d Gabor wavelets”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 959–971, 1996.