Gabor Wavelet Networks for Object Representation

Volker Krueger and Gerald Sommer

Computer Science Institute, Christian-Albrechts University Kiel
Preußerstr. 1-9, 24105 Kiel, Germany
Tel: ++49-431-560496, FAX: ++49-431-560481
email: vok@ks.informatik.uni-kiel.de

Abstract. In this article we want to introduce the Gabor wavelet network as a model based approach for an effective and efficient object representation. The Gabor wavelet network has several advantages: invariance to some degree with respect to translation, rotation and dilation, the use of Gabor filters ensured that geometrical and textural object features are encoded, the representation precision ranges from photorealistic to coarse and can be adapted as needed for a specific tast. The feasibility of the Gabor filters as a model for local object features ensures a considerable data reduction while at the same time allowing *any* desired precision of the object representation ranging from a sparse to a photo-realistic representation. The feasibility of the object representation is verified by a pose estimation experiment.

1 Introduction

Recently, model-based approaches for the recognition and the interpretation of images of variable objects, like the bunch graph approach, PCA, eigenfaces and active appearance models, have received considerable interest [12; 8; 2]. These approaches achieve good results because solutions are constrained to be valid instances of a model. In these approaches, the term "model-based" is understood in the sense that a set of training objects is given in form of gray value pixel images while the model "learns" the variances of the gray values (PCA, eigenfaces) or, respectively, the Gabor filter responses (bunch graph).

In this work we want to introduce a novel approach for object representation that is based on Gabor Wavelet Networks. Gabor Wavelet Networks (GWN) are combining the advantages of RBF networks with the advantages of Gabor wavelets: GWNs represent an object as a linear combination of Gabor wavelets where the parameters of each of the Gabor functions (such as orientation and position and scale) are optimized to reflect the particular local image structure. Gabor wavelet networks have several advantages:

1. By their very nature, Gabor wavelet networks are invariant to some degree to affine deformations and homogenous illumination changes,
2. Gabor filters are good feature detectors [7] and the optimized parameters of each of the Gabor wavelets are directly related to the underlying image structure,

3. the weights of each of the Gabor wavelet are directly related to their filter responses and with that they are also directly related to the underlying local image structure,
4. the precision of the representation can be varied to *any* desired degree ranging from a coarse representation to an almost photo-realistic one by simply varying the number of used wavelets.

We will discuss each single point in section 2.

A further point should be mentioned: The use of Gabor filters implies a model for the actual representation of the object information: GWN represents object information as a set of local image primitive, which leads to a higher level of abstraction and to a considerable data reduction. Both, textural and geometrical information is encoded at the same time, but can be split to some degree. Most other approaches, especially those based on PCA and eigenimages, do not apply any model for the actual knowledge representation. Instead, the representation is done on a pixel level. The Gabor-model based representation leads to a considerabel data-reduction. Furthermore, a GWN can be seen as a task oriented optimal filter bank: given number of filters, a GWN defines *that* set of filters that extracts the maximal possible image information. This is an important aspect for several reasons: E.g. for real-time applications one wants to keep the number of filtrations low to save computational resources and it makes sense in this context to relate the number of filtrations to the amount of image information needed for a specific task.

In the following section we will give a short introduction to GWNs. Also, we will discuss each single point mentioned above, including the invariance properties, the abstraction properties and specificity of the wavelet parameters for the object representation and a task oriented image filtration. In section 3 we will present results on a pose estimation experiment where we will exploit the optimality of the filter bank to speed up the response time of the system and to optimize the training of the neural network. In the last section we will conclude with some final remarks.

1.1 Related Work

There are other models for image interpretation and object representation. Most of them are based on PCA, such as the eigenface approach [11]. The eigenface approach has shown its advantages expecially in the context of face recognition. Its major drawbacks are its sensitivity to perspective deformations and to illumination changes. PCA encodes textural information only, while geometrical information is discarded. Furthermore, the alignment of face images into a common coordinate system is still a problem.

Another PCA based approach is the active appearance model (AAM)[2]. This approach enhances the eigenface approach considerably by including geometrical information. This allows an alignment of image data into a common coordinate system while the formulation of the alignment technique can be elegantly done with techniques of the AAM framework. Also, recognition and tracking applications are presented within this framework. An advantage of this approach was

demonstrated in [2]: they showed the ability of the AAM to model, in a photo-realistic way, almost any face gesture and gender. However, this is undoubly an expensive task and one might ask in which situation such a precision is really needed. In fact, a variation to different precision levels in order to spare computational resources and to restrict considerations to the data actually needed for a certain application seems not easily possible.

The bunch graph approach [12], on the other hand, is based on the discrete wavelet transform. A set of Gabor wavelets are applied at a set of hand selected prominent object points, so that each point is represented by a set of filter responses, called *jet*. An object is then represented by a set of jets, that encode each a single local texture patch of the object. The jet topology, the so-called *image graph*, encodes geometrical object information. A precise positioning of the image graph onto the test image is important for good matching results and the positioning is quite a slow process. The feature detection capabilities of the Gabor filters are not exploited since their parameters are fixed and a variation to different precision levels has not been considered so far.

2 Introduction to Gabor Wavelet Networks

The basic idea of the wavelet networks is first stated by [14], and the use of Gabor functions is inspired by the fact that they are recognized to be good feature detectors [7]. To define a GWN, we start out, generally speaking, by taking a family of N odd Gabor wavelet functions $\mathbf{\Psi} = \{\psi_{\mathbf{n}_1}, \ldots, \psi_{\mathbf{n}_N}\}$ of the form $\psi_{\mathbf{n}}(x,y) = \exp\left(-\frac{1}{2}\Big[s_x\left((x-c_x)\cos\theta - (y-c_y)\sin\theta\right)\Big]^2 + \Big[s_y\left((x-c_x)\sin\theta + (y-c_y)\cos\theta\right)\Big]^2\right)\times \sin\Big(s_x\left((x-c_x)\cos\theta - (y-c_y)\sin\theta\right)\Big)$, with $\mathbf{n} = (c_x, c_y, \theta, s_x, s_y)^T$. Here, c_x, c_y denote the translation of the Gabor wavelet, s_x, s_y denote the dilation and θ denotes the orientation. The choice of N is arbitrary and is related to the maximal representation precision of the network. In order to find the GWN for image I, the energy functional $E = \min_{\mathbf{n}_i, w_i \text{ for all } i} \|I - \sum_i w_i \psi_{\mathbf{n}_i}\|_2^2$ is minimized with respect to the weights w_i and the wavelet parameter vector $\mathbf{n}_i$. A Gabor wavelet network is defined as follows:

Definition: Let $\psi_{\mathbf{n}_i}$, $i = 1, \ldots, N$ be a set of Gabor wavelets, I a DC-free image and w_i and $\mathbf{n}_i$ chosen according to the energy functional. The two vectors $\mathbf{\Psi} = (\psi_{\mathbf{n}_1}, \ldots, \psi_{\mathbf{n}_N})^T$ and $\mathbf{w} = (w_1, \ldots, w_N)^T$ define then the *Gabor wavelet network* $(\mathbf{\Psi}, \mathbf{w})$ for image f.

The optimization of each wavelet with respect to the underlying image is precisely the main advantage over the discrete approach used in[12]. While in case of a discrete phase space local image structure has to be approximated by a combination of wavelets, a *single* wavelet can be chosen selectively in the continuous case to reflect *precisely* the local image structure. This assures that a maximum of the image information is encoded.

Using the optimal wavelets $\mathbf{\Psi}$ and weights $\mathbf{w}$ of the Gabor wavelet network of an image f, I can be (closely) reconstructed by a linear combination of the weighted wavelets: $\hat{I} = \sum_{i=1}^{N} w_i \psi_{\mathbf{n}_i} = \mathbf{\Psi}^T \mathbf{w}$. Of course, the quality of the image

Fig. 1. The very right image shows the original face image I, the other images show the image I, represented with 16, 52, 116 and 216 Gabor wavelets (left to right). In the very left image, the positions of the first 16 wavelets are indicated.

Fig. 2. The figure shows images of a wooden toy block on which a GWN was trained. The black line segments sketch the positions, sizes and orientations of all the wavelets of the GWN (left), and of some automatically selected wavelets (right).

representation and of the reconstruction depends on the number N of wavelets used and can be varied to reach almost any desired precision (see fig. 1).

It was mentioned above that the Gabor wavelets are recognized to be good feature [7] detectors, that are directly related to the local image features by the energy functional. This means that an optimized wavelet has e.g. ideally the exact position and orientation of a local image feature. An example can be seen in fig. 2. The figure shows the image of a little wooden toy block, on which a Gabor wavelet network was trained. The left image shows the positions, scales and orientations of the wavelets as little black line segments. By thresholding the weights, the more "important" wavelets may be selected, which leads to the right image. Ideally, each Gabor wavelet should be positioned *exactly* on the image line after optimization. Furthermore, since large weights indicate that the corresponding wavelets represents an edge segment (see sec. 2.1), these wavelets encode local geometrical object information.

The use of Gabor filters as a model for local object primitives leads to a higher level of abstraction where object knowledge is represented by a set of local image primitives. The Gabor wavelets in a network that represent edge segments can be easily identified. How to identify wavelets, however, that encode specific textures is not really clear, yet, and subject to future investigation.

2.1 Direct Calculation of Weights and Distances

As mentioned earlier, the weights w_i of a GWN are directly related to the filter responses of the Gabor filters $\psi_{\mathbf{n}_i}$ on the training image.

Gabor wavelet functions are not orthogonal. For a given family $\mathbf{\Psi}$ of Gabor wavelets it is therefore not possible to calculate a weight w_i directly by a simple projection of the Gabor wavelet $\psi_{\mathbf{n}_i}$ onto the image. Instead one has to consider the family of dual wavelets $\tilde{\mathbf{\Psi}} = \{\tilde{\psi}_{\mathbf{n}_1} \dots \tilde{\psi}_{\mathbf{n}_N}\}$. The wavelet $\tilde{\psi}_{\mathbf{n}_j}$ is the dual wavelet to the wavelet $\psi_{\mathbf{n}_i}$ iff $\langle \psi_{\mathbf{n}_i}, \tilde{\psi}_{\mathbf{n}_j} \rangle = \delta_{i,j}$. With $\tilde{\mathbf{\Psi}} = (\tilde{\psi}_{\mathbf{n}_1}, \dots, \tilde{\psi}_{\mathbf{n}_N})^T$, we can write $\left[\langle \mathbf{\Psi}, \tilde{\mathbf{\Psi}} \rangle\right] = \mathbb{1}$. In other words: $w_i = \langle I, \tilde{\psi}_{\mathbf{n}_i} \rangle$. We find $\tilde{\psi}_{\mathbf{n}_i}$ to be $\tilde{\psi}_{\mathbf{n}_i} = \sum_j \left(\mathit{\Psi}^{-1}\right)_{i,j} \psi_{\mathbf{n}_j}$, where $\mathit{\Psi}_{i,j} = \langle \psi_{\mathbf{n}_i}, \psi_{\mathbf{n}_j} \rangle$.

The equation $w_i = \langle I, \tilde{\psi}_{\mathbf{n}_i} \rangle$ allows us to define the operator $\mathcal{T}_{\mathbf{\Psi}} : \mathbb{L}^2(\mathbb{R}^2) \longmapsto < (\psi_{\mathbf{n}_1}, \dots, \psi_{\mathbf{n}_N}) >$ as follows: Given a set $\mathbf{\Psi}$ of optimal wavelets of a GWN, the

operator $\mathcal{T}_{\mathbf{\Psi}}$ realizes an orthogonal projection of a function J onto the vector subspace $< \mathbf{\Psi} >$ (see fig. 4), i.e. $\hat{J} = \mathcal{T}_{\mathbf{\Psi}}(J) = J\tilde{\mathbf{\Psi}}\mathbf{\Psi} = \sum_{i=1}^{N} w_i \psi_{\mathbf{n}_i}$ with $\mathbf{w} = J\tilde{\mathbf{\Psi}}$. The direct calculation of the distance between two families of Gabor wavelets, $\mathbf{\Psi}$ and $\mathbf{\Phi}$, can also be established by applying the above to each of the wavelets $\phi_i \in \mathbf{\Phi}$: $\mathcal{T}_{\mathbf{\Psi}}(\phi_j) = \sum_i \left[\langle \phi_j, \tilde{\psi}_i \rangle\right] \psi_i$, which can be interpreted as the representation of each wavelet ϕ_j as a superposition of the wavelets ψ_i. With this, the distance between $\mathbf{\Psi}$ and $\mathbf{\Phi}$ can be given directly by $\sqrt{\left[\sum_j \frac{\|\phi_i - T_{\mathbf{\Psi}}(\phi_i)\|}{\|\phi_i\|}\right]^2 + \left[\sum_j \frac{\|\psi_i - T_{\mathbf{\Phi}}(\psi_i)\|}{\|\psi_i\|}\right]^2}$, where $\|\cdot\|$ is the euclidian norm. With this distance measurement, the distance between two object representations can be calculated very efficiently.

2.2 Reparameterization of Gabor Wavelet Networks

The "reverse" task of finding the position, the scale and the orientation of a GWN in a new image is most important because otherwise the filter responses are without any meaning. For example, consider an image J that shows the person of fig. 1, left, possibly distorted affinely. Given a corresponding GWN we are interested in finding the correct position, orientation and scaling of the GWN so that the wavelets are positioned on the same facial features as in the original image, or, in other words, how should the GWN be deformed (warped) so that it is aligned with the coordinate system of the new object. An example for a successful warping can be seen in fig. 1, where in the very right image the wavelet positions of the *original* wavelet network are marked and in fig. 3, where in new images the wavelet positions of the *reparameterized* Gabor wavelet network are marked. Parameterization of a GWN is established by using a

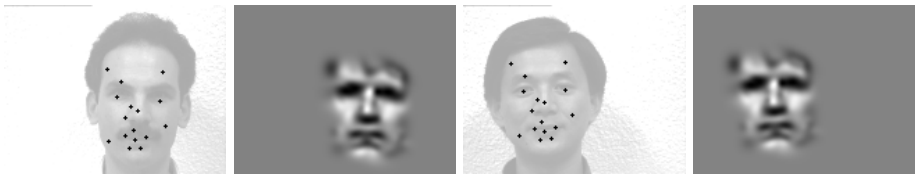

Fig. 3. The images show the positions of each of the 16 wavelets after reparameterizing the wavelet net and the corresponding reconstruction. The reconstructed faces show the same orientation, position and size as the ones they were reparameterized on.

superwavelet [10]:

Definition: Let $(\mathbf{\Psi}, \mathbf{w})$ be a Gabor wavelet network with $\mathbf{\Psi} = (\psi_{\mathbf{n}_1}, \ldots, \psi_{\mathbf{n}_N})^T$, $\mathbf{w} = (w_1, \ldots, w_N)^T$. A *superwavelet* $\mathbf{\Psi}_{\mathbf{n}}$ is defined to be a linear combination of the wavelets $\psi_{\mathbf{n}_i}$ such that $\mathbf{\Psi}_{\mathbf{n}}(\mathbf{x}) = \sum_i w_i \psi_{\mathbf{n}_i}(\mathbf{SR}(\mathbf{x} - \mathbf{c}))$, where the parameters of vector $\mathbf{n}$ of superwavelet $\mathbf{\Psi}$ define the dilation matrix $\mathbf{S} = \mathrm{diag}(s_x, s_y)$, the rotation matrix $\mathbf{R}$, and the translation vector $\mathbf{c} = (c_x, c_y)^T$.

A superwavelet $\mathbf{\Psi}_{\mathbf{n}}$ is again a wavelet (because of the linearity of the sum) and in particular a continuous function that has the wavelet parameters dilation, translation and rotation. Therefore, we can handle it in the same way as we handled

each single wavelet in the previous section. For a new image J we may arbitrarily deform the superwavelet by optimizing its parameters $\mathbf{n}$ with respect to the superwavelet energy functional E: $E = \min_{\mathbf{n}} \|J - \mathbf{\Psi}_{\mathbf{n}}\|_2^2$. Equation defines the operator $\mathcal{P}_{\mathbf{\Psi}} : \mathbb{L}^2(\mathbb{R}^2) \mapsto \mathbb{R}^5$, $g \longrightarrow \mathbf{n} = (c_x, c_y, \theta, s_x, s_y)$, where $\mathbf{n}$ minimizes the superwavelet energy functional E; $\mathbf{\Psi}$ is defined to be a superwavelet. For optimization of the superwavelet parameters, the same optimization procedure as for the energy functional may be used.

The reparameterization (warping) works quite robust: Using the superwavelet of fig. 1 we have found in several experiments on the various subjects with $\approx$ 60 pixels in width that the initialization of $\mathbf{n}_0$ may vary from the correct parameters by approx. ± 10 px. in x and y direction, by approx. 20% in scale and by approx. $\pm 10^\circ$ in rotation. Compared to the AAM, these findings indicate a much better robustness [2]. Furthermore, we found that the warping algorithm converged in 100% of the cases to the correct values when applied on the *same* individual, independently of pose and gesture. The tests were done on the images of the Yale face database and on our own images. The poses were varied within the range of $\approx \pm 20^\circ$ in pan and tilt where all face features were still visible. The various gestures included *normal, happy, sad, surprised, sleepy, glasses, wink*. The warping on other faces depended certainly on the similarity between the training person and the test person and on the number of used wavelets. We found that the warping algorithm always converged correctly on $\approx$ 80% of the test persons (including the training person) of the Yale face database. The warping algorithm has also been successfully applied for an wavelet based affine real-time face tracking application [6].

3 Experiments: Pose Estimation

In this section we will present results of our experiments for estimating the pose of a face. There exist many different approaches for pose estimation, including pose estimation with color blobs [3], pose estimation applying a geometrical approach [4], stereo information [13] or neural networks [1], to cite just a few. Color blob approaches give only approximate orientation information. The precision of the geometrical approach [4] was extensively tested and verified in [9]. The minimal mean pan/tilt error that was reached was $> 1.6^\circ$. In comparison to this, the neural network approach in [1] reached a minimal pan/tilt error of $> 0.58^\circ$. The good result in [1] was reached by first detecting the head using a color tracking approach. Within the detected color blob region, 4×4 sets of 4 complex Gabor filters with the different orientations of 0, $\frac{\pi}{4}$, $\frac{\pi}{2}$ and $\frac{3}{4}\pi$ were evenly distributed. The 128 complex projections of these filters were then fed into a neural RBF network. At this point, it is reasonable to assume that a precise positioning of the Gabor filters would result into an even lower mean pan/tilt error. In our experiments we therefore trained a GWN on an image I showing a doll's head. For the training of the GWN we used again the optimization scheme introduced in section 2 with $N = 52$ Gabor wavelets.

In order to be comparable we used in our experiments *exactly* the same neural network and the same number of training examples as described in [1]. The doll's head was connected to a robot arm, so that the pan/tilt ground truth

was known. During the training and testing, the doll's head was first tracked using our wavelet based face tracker [6]. For each frame we proceeded in two steps:

1. optimal repositioning of the GWN by using the positioning operator $\mathcal{P}$
2. calculating the optimal weights for the optimally repositioned GWN by using the projection operator $\mathcal{T}$.

See fig. 4 for example images. The weight vector that was calculated with the

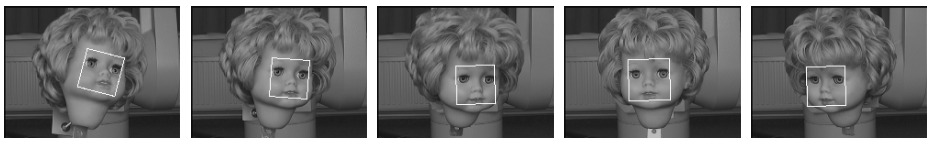

Fig. 4. The images show different orientations of the doll's head. The head is connected to a robot arm so that the ground truth is known. The white square indicates the detected position, scale and orientation of the GWN.

operator $\mathcal{T}$ was then fed into the same neural RBF network that was used in [1]. The training was done exactly as it was described in [1]: We used 400 training images, evenly distributed within the range of $\pm 20°$ in pan and tilt direction (this is the range where all face features appeared to be visible). With this, we reached a minimal mean pan/tilt error of $0.19°$ for a GWN with 52 wavelets and a minimal mean pan/tilt error of $0.29°$ for a GWN with 16 wavelets. The theoretical speed of the system on a 450 MHz Linux Pentium should reach $>\approx 5$ fps for the 52 wavelet network and $>\approx 10$ fps for the 16 wavelet network. The experiments were carried out on an experimental setup, that has not yet been integrated into a complete, single system.

4 Conclusions

The contribution of this article is twofold: First, we introduced the concepts of the *Gabor wavelet network* and the *Gabor superwavelet* that allow a data abstraction, a data reduction and a selective filtering:

- The representation of an object with variable degree of precision, from a coarse representation to an almost photo-realistic one,
- the definition of an optimal set of filters for a selective filtering
- the representation of object information on a basis of local image primitives and
- the possibility for affine deformations to cope with perspective deformations.

In the second section we discussed these various properties in detail. In [5; 6], GWNs have already been used successfully for wavelet based affine real time face tracking and pose invariant face recognition. It is future work, to fully exploit the advantages of the data reduction by reducing considerations to the vector space over the set of Gabor wavelet networks. We exploited all these advantages of the GWN for the estimation of the head pose. Second, the experimental results showed quite impressively that it is sensible for an object representation to reflect the specific individual properties of the object rather than being independent of

the individual properties such as general representations are. This can especially be seen when comparing the presented approach with the one in [1]: While having used the same experimental setup and the same type of neural network, the precision of the presented approach is twice as good with only 16 coefficients (vs. 128), and three times as good with only about half the coefficients. Furthermore, the experiment shows, how the precision in pose estimation and the system speed change with an increasing number of filters. A controllable variability of precision and speed has a major advantage: The system is able to decide how precise the estimation should be in order to minimize the probability that the given task is not fulfilled satisfactorily.

Acknowledgment The images used are derived from the Yale Face Database. This work was supported by the DFG grant Ei 322/1-2.

References

1. J. Bruske, E. Abraham-Mumm, J. Pauli, and G. Sommer. Head-pose estimation from facial images with subspace neural networks. In *Proc. of Int. Neural Network and Brain Conference*, pages 528–531, Beijing, China, 1998.
2. T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. In *Proc. Fifth European Conference on Computer Vision*, volume 2, pages 484–498, Freiburg, Germany, June 1-5, 1998.
3. T. Darrell, B. Moghaddam, and A. Pentland. Active face tracking and pose estimation in an interactive room. In *IEEE Conf. Computer Vision and Pattern Recognition, CVPR*, pages 67–72, Seattle, WA, June 21-23, 1996.
4. A. Gee and R. Cipolla. Determining the gaze of faces in images. *Image and Vision Computing*, 12(10):639–647, 1994.
5. V Krüger and G. Sommer. Gabor wavelet networks for object representation. In *Proc. of the Int. Dagstuhl 2000 Workshop*, 2000. to be published.
6. V. Krüger and Gerald Sommer. Affine real-time face tracking using gabor wavelet networks. In *Proc. Int. Conf. on Pattern Recognition*, pages 141–150, Barcelona, Spain, Sept. 3-8, 1999.
7. B.S. Manjunath and R. Chellappa. A unified approach to boundary perception: edges, textures, and illusory contours. *IEEE Trans. Neural Networks*, 4(1):96–107, 1993.
8. B. Moghaddam and A. Pentland. Probabilistic visual learning for object detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(7):696–710, Juli 1997.
9. Eleni Petraki. Analyse der blickrichtung des menschen und er kopforientierung im raum mittels passiver bildanalyse. Master's thesis, Technical University of Hamburg-Harburg, 1996.
10. H. Szu, B. Telfer, and S. Kadambe. Neural network adaptive wavelets for signal representation and classification. *Optical Engineering*, 31(9):1907–1961, 1992.
11. M. Turk and A. Pentland. Eigenfaces for recognition. *Int. Journal of Cognitive Neuroscience*, 3(1):71–89, 1991.
12. L. Wiskott, J. M. Fellous, N. Krüger, and C. v. d. Malsburg. Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):775–779, July 1997.
13. M. Xu and T. Akatsuka. Detecting head pose from stereo image sequences for active face recognition. In *Int. Conf. on Automatic Face- and Gesture-Recognition*, pages 82–87, Nara, Japan, April 14-16, 1998.
14. Q. Zhang and A. Benviste. Wavelet networks. *IEEE Trans. Neural Networks*, 3(6):889–898, Nov. 1992.