Wavelet Networks for Face Processing

V. Krueger

Center for Automation Research

University of Maryland

4456 A.V. Williams Building

College Park, MD 20742

vok@cfar.umd.edu

http://www.cfar.umd.edu/ vok

G. Sommer

Inst. für Informatik und Prakt. Mathematik

University of Kiel

Preusser Str 1-9

24105 Kiel, Germany

gs@ks.informatik.uni-kiel.de

Wavelet networks (WNs) were introduced in 1992 as a combination of artificial neural radial basis function (RBF) networks and wavelet decomposition. Since then, however, WNs have received only little attention. We believe, that the potential of WNs has been generally underestimated. WNs have the advantage, that the wavelet coefficients are directly related to the image data through the wavelet transform. In addition, the parameters of the wavelets in the WNs are subject to optimization, which results in a direct relation between the represented function and the optimized wavelets, leading to considerable data reduction (thus making subsequent algorithms much more efficient) as well as to wavelets that can be used as an optimized filter bank. In this paper, we analyze some of their properties and hightlight their advantages for object representation purposes. We then present a series of experimental results where we have used WNs for face tracking in which we exploit the efficiency due to data reduction, for face recognition and face-pose estimation where we exploit the optimized filter bank principle of the WNs.

OCIS codes: 100.7410,100.5010

1. Introduction

Wavelets networks were first mentioned by Zhang and Benveniste[1] in the context of non-parametric regression of functions in $\mathbb{L}^2(\mathbb{R}^2)$. In wavelet networks, the radial basis functions of RBF-networks are replaced by wavelets. During the training phase, the network weights as well as the degrees of freedom (position, scale, orientation) of the wavelet functions are optimized. Zhang and Benveniste realized that wavelet networks inherit the properties of wavelet decomposition and mention espe-

cially their universal approximation property, the availability of convergence rates and the explicit link between the network coefficients and the wavelet transform.

However, since their introduction in 1992, wavelet networks (WN) have received little attention in recent publications. Szu et. al.[2,3] have used WNs for signal representation and classification. They have explained how a WN template, a *superwavelet*, can be generated and showed how they can be used for pattern matching. In addition, they mention the large data compression achieved by such a WN representation. Zhang[4] showed that WNs are able to handle nonlinear regression of moderately large input dimension with sparse training data. Holmes and Mallick[5] analyzed WNs in the context of a Bayesian framework. Reyneri[6]has recently analyzed the relations between artificial neural networks (ANNs), fuzzy systems and WNs have been discussed.

It appears, that in the cited works, WNs have only been applied to certain problems but that their properties have not been investigated. Starting from a wavelet representation as described by Zhang[4] we have analyzed the properties such a representation has. Zhang and Benveniste[1] have mentioned, e.g., that there is an explicit link between the weights (wavelet coefficients) and some appropriate transform. This link is established through wavelet theory. We have further investigated the following properties of wavelet networks:

- that the explicit link mentioned above can be exploited to find optimized filter banks.
- that there exists an additional explicit link between the parameters of the optimized wavelet network functions and the represented function; and that the chosen mother wavelet introduces model information for image features that the optimized wavelets in a WN will represent.
- that the optimized wavelets are linearly independent, when the optimization scheme presented here, which is similar to the one proposed by Zhang and Benveniste,[1] is followed.

- that the wavelets of the network form a low-dimensional subspace in the $\mathbb{L}^2(\mathbb{R}^2)$ space and that its dual is a vector space over $\mathbb{R}$, the *wavelet subspace* of the vectors of wavelet coefficients.

We will exploit the above properties for object representation. In particular we will show that tracking and recognition is facilitated by the above properties: Both can be carried out efficiently in the low-dimensional wavelet subspace while the mapping of an input image into the wavelet subspace can be established with a small number of local image projections. We have carried out a small set of experiments, affine face tracking and face recognition, in order to support our claims. Some of the presented ideas such as the use of WN template (superwavelets) enhance the ideas mentioned in Szu et. al..[2,3] In addition as mentioned above, WNs can be used to optimize image filtering. We have used the optimized wavelets as filters in a face-pose estimation experiment. Having reached an estimation error of 0.65° using non-optimized filters, the error decreased to 0.21° using the optimized wavelets.

2. Introduction to Wavelet Networks

To define a WN, we begin by taking a family of N wavelet functions $\mathbf{\Psi} = \{\psi_{\mathbf{n}_1}, \ldots, \psi_{\mathbf{n}_N}\}$ with parameter vectors $\mathbf{n} = (c_x, c_y, \theta, s_x, s_y)^T$ of some mother wavelet ψ:

$$\psi_{\mathbf{n}}(\mathbf{x}) \quad = \quad \psi(\mathbf{SR}(\mathbf{x} - \mathbf{t})) \; . \tag{1}$$

The c_x, c_y defines the translation $\mathbf{t}$ of the wavelet, s_x, s_y defines the dilation $\mathbf{S}$ and θ defines the orientation $\mathbf{R}$. The parameters vector $\mathbf{n}$ (translation, orientation and dilation) of the wavelets may be chosen arbitrarily at this point. According to wavelet theory, any function $f \in \mathbb{L}^2(\mathbb{R}^2)$ can be losslessly represented by their continuous wavelet transform and thus, with arbitrary precision, by a wavelet network. We therefore interpret the image f to be a function of the space $\mathbb{L}^2(\mathbb{R}^2)$ and assume further, without loss of generality that f is DC-free. In order to find the WN for image f

we minimize the energy functional

$$E \quad = \quad \min_{\mathbf{n}_i, w_i \text{ for all } i} \| f - \sum_{i=1}^{N} w_i \psi_{\mathbf{n}_i} \|_2^2 \tag{2}$$

with respect to the weights w_i and the wavelet parameter vectors $\mathbf{n}_i$. Equation (2) says that the w_i and $\mathbf{n}_i$ are optimized, i.e. translation, dilation and orientation of each wavelet are chosen such that the image f is optimally approximated by the weighted sum of wavelets $\psi_{\mathbf{n}_i}$. We therefore define a wavelet network as follows:

Definition: Let $\psi_{\mathbf{n}_i}$, $i = 1, \ldots, N$ be a set of wavelets, f a DC-free image and w_i and $\mathbf{n}_i$ chosen according to the energy functional (2). The two vectors

$$\mathbf{\Psi} \quad = \quad (\psi_{\mathbf{n}_1}, \ldots, \psi_{\mathbf{n}_N})^T \text{ and } \mathbf{w} = (w_1, \ldots, w_N)^T$$

define then the *wavelet network* $(\mathbf{\Psi}, \mathbf{w})$ for image f.

It should be mentioned that it was proposed before[7–9] to use an energy functional (2) in order to find the optimal set of weights w_i for a *fixed* set of non-orthogonal wavelets $\psi_{\mathbf{n}_i}$. The WN concept enhances these approaches by finding also the optimal parameters $\mathbf{n}_i$ for each (not-necessarily orthonormal) wavelet $\psi_{\mathbf{n}_i}$. WNs also appear to enhance the RBF neural network approach considerably. This was pointed out recently,[6] even though a considerably simplified version of WNs with radial wavelets, which considerably limits the potentials of the WNs, was investigated.

The parameters $\mathbf{n}_i$ are chosen from *continuous* phase space and the wavelets are positioned with sub-pixel accuracy. This is precisely the main advantage over the discrete approach.[7,9] While in the case of a discrete phase space local image structure has to be approximated by a combination of wavelets, only a *single* wavelet needs to be chosen in the continuous case to precise reflect the local image structure. This assures that a maximum of the image information can be encoded with only a small number of wavelets.

In order to find a WN ($\mathbf{\Psi}$, $\mathbf{w}$) for a function f, we use the Levenberg-Marquardt method.[10] As initialization, we distribute the wavelets homogeneously over the region of interest. The orientations are initialized randomly, the scales are initialized to a constant value that is related to the density with which the wavelets are distributed. We constrain the wavelet parameters to prevent degenerated wavelet shapes. For the two wavelet types in this paper (odd Gabor, difference-of-Gaussian) we have used constrains[11] to prevent the wavelets parameter from diverging. In several experiments we have found that this rough initialization is sufficient. Also, we apply a coarse-to-fine strategy by first optimizing a set of wavelets initialized to coarse scale, followed by the optimization of a set of wavelets, initialized to a finer scale. Within each set of wavelets, we optimized one wavelet after another. The order in which the wavelets are optimized introduces a bias. We have done experiments to optimize all wavelets at once; but even though the convergence is much slower simultanious optimization does not lead to a considerable decrease of the minimal energy (2). In general, the effect of the bias seems to be very little.

Intuitively, a coarse-to-fine strategy for optimization makes sense because this minimizes the energy functional (2) more efficiently. To optimize a WN with 16 wavelets it takes about 30s on a 750 MHz Pentium processor.

Using the optimal wavelets $\mathbf{\Psi}$ and weights $\mathbf{w}$ of the wavelet network of an image f, f can be (closely) reconstructed by a linear combination of the weighted wavelets:

$$\hat{f} \;=\; \sum_{i=1}^{N} w_i \psi_{\mathbf{n}_i} = \mathbf{\Psi}^T \mathbf{w} \; . \tag{3}$$

The quality of image representation and reconstruction depends on the number N of wavelets used (see Fig. 1). The quality may be varied from a coarse representation to an almost photo-realistic one.

Fig. 1 goes here

Examples of reconstructions can be seen in Fig. 1: A WN with $N = 216$ wavelets was optimized on the inner-face region of image I (rightmost image) (we will generally use the notation $f, g, \ldots$ to refer to images. We will use the notation $I, J, \ldots$ when we want to refer explicitly to gray value images used in experiments). The odd Gabor function has been used as mother wavelet. The reconstructions were computed using Eq. (3) for various N, and we have chosen in each case the N wavelets according to the order they were optimized. This means that a WN, optimized with $N = 16$ wavelets, results in the WN shown in the leftmost image. The quality can be further increased by choosing the wavelets with the N largest coefficients.

The images in Fig. 2 show the relation between the optimized positions of the wavelets (right), and their reconstructions (left) for two different mother wavelets: the odd Gabor function (top) and an anisotropic difference-of-Gaussian (DOG) (bottom).

Fig. 2 goes here.

A. Direct Calculation of Weights

Wavelet functions are not necessarily orthogonal. For a given family $\mathbf{\Psi}$ of wavelets it is therefore not generally possible to calculate a wavelet coefficient w_i directly by a simple projection of the wavelet $\psi_{\mathbf{n}_i}$ onto the considered function. It was therefore proposed[7,8] to use Eq. (2) to find the optimal coefficients w_i for each fixed wavelet. Because optimization is a slow process, we suggest a direct calculation for the case of a finite wavelet family. The correct coefficients w_i are computed by projecting the *dual* wavelets $\tilde{\psi}_{\mathbf{n}_i}$. The wavelet $\tilde{\psi}_{\mathbf{n}_i}$ is the dual wavelet to the wavelet $\psi_{\mathbf{n}_i}$ if

$$\langle \psi_{\mathbf{n}_i}, \tilde{\psi}_{\mathbf{n}_j} \rangle \quad = \quad \delta_{i,j} \; . \tag{4}$$

With $\tilde{\mathbf{\Psi}} = (\tilde{\psi}_{\mathbf{n}_1}, \ldots, \tilde{\psi}_{\mathbf{n}_N})^T$, we can write

$$\left[\left\langle \mathbf{\Psi}, \tilde{\mathbf{\Psi}} \right\rangle\right] \;=\; \mathbb{I} \tag{5}$$

and we find $\tilde{\psi}_{\mathbf{n}_i}$ to be

$$\tilde{\psi}_{\mathbf{n}_i} \;=\; \sum_j (\Psi)^{-1}_{i,j}\, \psi_{\mathbf{n}_j}\ , \tag{6}$$

where $(\mathbf{\Psi})_{i,j} = \langle \psi_i, \psi_j \rangle$. Given a family $\mathbf{\Psi}$ of optimized wavelets of a WN for the function f, we can compute the orthogonal projection of a function g into the subspace $< \mathbf{\Psi} > \subseteq \mathbb{L}^2(\mathbb{R}^2)$ (see (3)), i.e.

$$\hat{g} \;=\; \sum_{i=1}^{N} w_i \psi_{\mathbf{n}_i} \text{ with } \mathbf{w} = \tilde{\mathbf{\Psi}} g\ . \tag{7}$$

The method to compute the orthogonal projection of a function g into the subspace $< \mathbf{\Psi} >$ is mathematically equivalent to using the pseudo-inverse of $\mathbf{\Psi}$ directly. However, using the dual wavelets of Eq. (6) will prove to be computationally more efficient: For our tracking experiment we will have to deform the entire WN affinely, which means that the pseudo-inverse has to be recomputed. The matrix $(\mathbf{\Psi})_{i,j}$, on the other hand, is invariant, except for a factor, to affine deformations of the WN, and only the projections $\langle g, \psi_{\mathbf{n}_i} \rangle$ need to be recomputed.

B. Wavelet basis and Wavelet subspace

Considering the optimized family of wavelets $\mathbf{\Psi}$, its closed linear span constitutes a subspace $< \mathbf{\Psi} >$ in the $\mathbb{L}^2(\mathbb{R}^2)$ space. With Eq. (7) any function can be orthogonally projected into that subspace. It is interesting to ask whether $\mathbf{\Psi}$ constitutes a basis, because then the projection is unique. That this is indeed so can be shown with induction over the number of wavelets: Consider n wavelets $(\psi_{\mathbf{n}_1}, \ldots, \psi_{\mathbf{n}_n})$, that minimize the energy functional (2) and that form a basis. Let us choose a new wavelet that approximates best the residual between the function f and its approximation with

the first n wavelets. After optimization of the nth $+ 1$ wavelet, the energy functional (2) is smaller than before (for the n wavelets):

$$\|f - \sum_{i=1}^{n} w_i \psi_{\mathbf{n}_i}\| > \|f - \sum_{i=1}^{n} w_i \psi_{\mathbf{n}_i} - w_{n+1} \psi_{\mathbf{n}_{n+1}}\| \; .$$

Assuming now, that

$$< \psi_{\mathbf{n}_1}, \ldots, \psi_{\mathbf{n}_n} > = < \psi_{\mathbf{n}_1}, \ldots, \psi_{\mathbf{n}_{n+1}} >$$

we have in particular

$$< \psi_{\mathbf{n}_1}, \ldots, \psi_{\mathbf{n}_n} >^{\perp} = < \psi_{\mathbf{n}_1}, \ldots, \psi_{\mathbf{n}_{n+1}} >^{\perp} \; .$$

This again means that

$$f - \sum_{i=1}^{n} w_i \psi_{\mathbf{n}_i} \in (< \psi_{\mathbf{n}_1}, \ldots, \psi_{\mathbf{n}_{n+1}} >)^{\perp} \; ,$$

which implies

$$\langle f - \sum_{i=1}^{n} w_i \psi_{\mathbf{n}_i}, \psi_{\mathbf{n}_{n+1}} \rangle = 0 \; .$$

This, however, contradicts the choice of $\psi_{\mathbf{n}_n}$ in the optimization step, where $\psi_{\mathbf{n}_n}$ was selected such that

$$\langle f - \sum_{i=1}^{n} w_i \psi_{\mathbf{n}_i}, \psi_{\mathbf{n}_{n+1}} \rangle \neq 0 \; .$$

Let us call the closed linear span of $< \mathbf{\Psi} >$ the $\mathbb{L}^2(\mathbb{R}^2)$ *(image) subspace*. The dual wavelets $\tilde{\psi}_{\mathbf{n}_i}$ are linearly independent, and the projection

$$\mathbf{w} = \tilde{\mathbf{\Psi}} g$$

establishes an isomorphism from $\mathbb{L}^2(\mathbb{R}^2)$ (or the image space, respectively) into $\mathbb{R}^n$ which is the space of the n-vectors containing the wavelet coefficients. This space is dual to the image subspace and we call it the *wavelet subspace*.

Fig. 3 goes here.

C. Euclidean Distance in Wavelet Sub-space

An interesting question is, how to compute the distance between two vectors of wavelet coefficients. Let us consider two vectors $\mathbf{v}$, $\mathbf{w}$ of some wavelet subspace, define w.r.t. WN $\mathbf{\Psi}$. Computing the Euclidean distance between the two vectors, $\|\mathbf{v}-\mathbf{w}\|_2$, fails to reflect the different influences (e.g. due to different scales) of the wavelets in the sum (3). Instead, we suggest to compute the Euclidean distance between the WNs of $\mathbf{v}$ and $\mathbf{w}$ as follows: Starting out from the Euclidean distance in the (image) subspace $< \mathbf{\Psi} >$

$$\left\| \sum_{i=1}^{N} v_i \psi_{\mathbf{n}_i} - \sum_{i=1}^{N} w_i \psi_{\mathbf{n}_i} \right\|_2 \quad , \tag{8}$$

algebraic transformations lead to

$$\begin{aligned} \|\mathbf{v}-\mathbf{w}\|_{\mathbf{\Psi}} &:= \left[\sum_{i,j} (v_i - w_i)(v_j - w_j) \langle \psi_{\mathbf{n}_i}, \psi_{\mathbf{n}_j} \rangle \right]^{\frac{1}{2}} \\ &= (\mathbf{v}-\mathbf{w})^t \, (\mathbf{\Psi})_{i,j} \, (\mathbf{v}-\mathbf{w}) \, . \end{aligned} \tag{9}$$

$\|\cdot\|_{\mathbf{\Psi}}$ computes the Euclidean distance between the two appropriate points in $< \mathbf{\Psi} >$ and thus considers the different parameters of the wavelets. For orthogonal wavelets, the matrix $(\mathbf{\Psi})_{i,j} = \langle \psi_{\mathbf{n}_i}, \psi_{\mathbf{n}_j} \rangle$ is the unity matrix and no weighting is needed.

Same techniques can be used to derive additional distance or similarity measures, such as, e.g., the *normalized cross correlation*.

D. Relation Between the Filter Responses and the Optimized Parameters

The results of the optimization of a WN on an image depends largely on the choice of the mother-wavelet. An example can be seen in Fig. 4. A WN is optimized by increasing the number of wavelets until either a maximal wavelet number N or an energy threshold is reached. Each new wavelet is thus optimized based on the residual between the original function f and the already optimized

wavelets:

$$R = f - \sum w_i \psi_{\mathbf{n}_i} \ .$$

The optimization procedure parameterizes each new wavelet such that

$$\sum_{\mathbf{x}} (R(\mathbf{x}) - w_{\text{new}} \psi_{\mathbf{n}\text{new}}(\mathbf{x}))^2 = \min$$

is minimized which is true, where the correlation between the wavelet and the residual R is maximal:

$$\sum_{\mathbf{x}} R(\mathbf{x}) \psi_{\mathbf{n}\text{new}}(\mathbf{x}) = \max \ .$$

In case of the odd Gabor function, which is an excellent edge detector, the optimized wavelets will end at edges (see figs. 2 and 4). The chosen mother wavelets seems to introduce a model for local image features. While, e.g., the odd Gabor function seems to models edge segments, the anisotropic DOF seems to favor homogeneous image regions (Fig. 4).

Fig.4 goes here.

3. Experiments on Wavelet Networks

In this section we will describe the small experiments we have carried out to illustrate the properties that were discussed above. In detail, the experiments include face tracking, face recognition and face-pose estimation.

A. Face Tracking

In this experiment we have verified, whether tracking can be carried out in the wavelet subspace.[12] This subspace method is an enhancement of the approach by Krüger et. al.[13] where tracking was based directly on the gray-value differences considered in the energy functional in Eq. (2). First, we

have optimized a WN $(\mathbf{\Psi}, \mathbf{v})$ on the face image I that we want to track, where $\mathbf{\Psi} = (\psi_{\mathbf{n}_1}, \ldots, \psi_{\mathbf{n}_N})^T$. This WN will serve as our face template. Using Eq. (3) and the notation of Eq. (1), we can affinely deform the template WN:

$$\hat{I}(\mathbf{SR}(\mathbf{x} - \mathbf{t})) \quad = \quad \sum_{i=1}^{N} v_i \psi_{\mathbf{n}_i}(\mathbf{SR}(\mathbf{x} - \mathbf{t})) \tag{10}$$

where $\mathbf{S}$, $\mathbf{R}$ and $\mathbf{t}$ define, as in (1), dilation, rotation and translation. We call $\hat{I}$ a *superwavelet*,[2,3] referring to the fact that a linear combination of wavelets is again a wavelet.

Tracking is established by finding at each time step the appropriate deformation parameters of the superwavelet such that the sum-of-squared difference between the image at time t and the deformed template is minimized. To do so, we project at each time step t the image J_t into the wavelet subspace. This was done by first setting $\mathbf{S}$, $\mathbf{R}$ and $\mathbf{t}$ of the superwavelet $\hat{I}$ in Eq. (10) to roughly appropriate values (e.g. by using the computed deformation values from the previous time step) and by then using the deformed dual wavelets to compute the corresponding wavelet coefficients $\mathbf{w}$. The difference

$$\|\mathbf{v} - \mathbf{w}\|_{\mathbf{\Psi}} \tag{11}$$

measures how well the deformation parameters were chosen. Based on this difference, we can minimize the energy functional

$$E \quad = \quad \min_{\mathbf{S},\mathbf{R},\mathbf{t}} \|\mathbf{v} - \mathbf{w}\|_{\mathbf{\Psi}} \tag{12}$$

to compute optimal deformation values, where $\mathbf{v}$ and $\mathbf{w}$ are given by projections of the wavelets, parameterized by $\mathbf{S}, \mathbf{R}$ and $\mathbf{t}$.

Estimating the optimal deformation values can be done efficiently: Since the linear combination of the wavelets $\psi_{\mathbf{n}_i}$ is a wavelet, $\hat{I}$ from above is again a wavelet and the optimization scheme of Section 2. can be applied. The employed Levenberg-Marquardt algorithm needs a number of cycles

in which the deformation parameters are refined until a certain optimum is reached. In each cycle $\mathbf{w}$ in Eq. (12) has to be recomputed. For a WN with 16 wavelets, this, however, needs just 16 projections of the filters onto the image. The matrix $(\mathbf{\Psi})_{i,j}$ in Eq. (6) is invariant (except for some factor) to the affine changes due to $\mathbf{S}$, $\mathbf{R}$ and $\mathbf{t}$. This means it can be computed in advance, which increases efficiency of the tracking process. Also, this means that the WN $\mathbf{\Psi}$ and its deformed version remain comparable, i.e., the distance measure (12) can still be computed. This will be of importance in the next section. With a WN with $N = 16$ wavelets we have reached 30 fps on a 700 MHz Linux-Pentium. Because of the high frame rate, the differences between successive images were small and the Levenberg-Marquardt algorithm seldom exceeded 7 cycles. An example can be seen in Fig. 5. The white box indicates the tracked inner-face region, on which our template WN was optimized. We have also experimented with different number N of wavelets and noticed a linear decrease in speed, but an increase in precision for larger N ($N < 116$).

Fig. 5 goes here.

B. Face Recognition independent of Gesture

In this section we will present the results of a face recognition experiment, using two small face databases, the Yale face database with 15 individuals and eight images per person, and the Manchester database with 30 individuals and 10 images per person. In both databases, the individuals show different facial expressions on all of their images (happy, sad, surprised, etc.). The goal was to recognize each subject independent of the expression. We proceeded as follows. First, a WN $(\mathbf{\Psi}_i, \mathbf{v}_i)$ was optimized for each gallery face I_i which resulted in a set of template WNs (see Section A.) We chose the faces with the "normal" expressions to be our gallery faces

The recognition of a probe face was then carried out by first finding optimal deformation values

for the template WNs and by then computing the optimal wavelet coefficient vectors. This resulted optimal coefficient vectors $\mathbf{w}_i$ for each of the template WNs $(\mathbf{\Psi}_i, \mathbf{v}_i)$ in the gallery. The technique of the previous Section A. was employed to accomplish this. Fig. 6 illustrates what happens when for the same individual the optimal coefficient vectors are computed with a correct (left) and with a wrong template WN (right). Eq. (3) was used to compute the two reconstructions shown, using the optimal weight vectors.

Fig. 6 goes here.

Having computed an optimal coefficients vectors $\mathbf{w}_i$ with each of the template WNs $(\mathbf{\Psi}_i, \mathbf{v}_i)$ in the gallery, they are compared each with the vector $\mathbf{v}_i$ of the template WNs, using $\|\mathbf{v}_i - \mathbf{w}_i\|_{\mathbf{\Psi}_i}$. The top match identifies the probe face. As mentioned above, the distance measure $\|\mathbf{v}_i - \mathbf{w}_i\|_{\mathbf{\Psi}_i}$ remains invariant (except for a scaling factor) under the affine deformation of the wavelet network $\mathbf{\Psi}_i$ so that even when the weight vector $\mathbf{w}_i$ is computed by an affinely deformed WN (while $\mathbf{v}_i$ is the weight vector of the original WN), the above distance can still be computed.

Examples can be seen in figs. 7 and 8. Fig. 7 shows reconstructions of optimal coefficients vectors of subject 01 in the Yale database, showing different expressions, but computed with the template WN optimized for that subject, whereas Fig. 8 shows the reconstructions of optimal coefficients vectors of subjects in the Yale database other than subject 01, but computed with the same WN as was used in Fig. 7.

Fig. 7 goes here.

Fig. 8 goes here.

Fig. 9 goes here.

The visual impression of figs. 7 and 8 were reflected when we computed the distance between the vectors $\mathbf{w}_i$ and $\mathbf{v}_i$, $\|\mathbf{v}_i - \mathbf{w}_i\|_{\mathbf{\Psi}_i}$. Table 9 show a clear difference between the probe images that show different gestures of the original subject and the probe images that show different subjects.

All gallery WNs used $N = 52$ wavelets. As mother wavelet, we chose the odd Gabor function. In case of the Yale Face Database 96% of the top matches were the correct matches, while in case for the Manchester Database 93.3% of the top matches were correct. For all subjects in the Yale database, the "surprised" expression was the expression with the lowest similarity (see table 9). Without this expression, 97.8 % of the top matches were correct.

It should be mentioned that a direct comparison with other face recognition approaches is difficult, as the employed face databases are too small.[14]

C. Pose Estimation

In this section we present the results of two face-pose estimation experiments that were carried out to verify the "optimal filter bank" principle of the WNs.[15]

For both experiments we connected a doll's head to a robot arm and let the robot move the doll's head in front of a fixed camera. With this the correct pose was always known. In the first experiment we tracked the doll's head with a color blob tracker and distributed 4×4 sets of 4 complex Gabor filters with the different orientations of 0, $\frac{\pi}{4}$, $\frac{\pi}{2}$ and $\frac{3}{4}\pi$ over the tracked inner face region. The resulting 128 complex projections of these filters were then fed into an artificial neural

LLM network (ANN).[16,17] This was done so for training as well for testing. A precise description of this experiment can be found in.[18] The mean pan/tilt error that we reached was $\approx 0.58^\circ$, computed as $\sqrt{(\delta p)^2 + (\delta t)^2}$.

It is reasonable to assume that the choice of better Gabor filters would result an even lower mean pan/tilt error. In our second experiment, we therefore optimized a template WN for the doll's face with $N = 52$ wavelets (see Fig. 10). As mother wavelet, the odd Gabor function was used, Fig. 10 shows the reconstruction based on the optimized WN.

Fig. 10 goes here.

As in the first experiment, the doll's head was connected to a robot arm, so that the pan/tilt ground truth was known. During the training of the ANN and testing, the doll's head was first tracked using our face tracking method of Section A. and then the optimal wavelet coefficient vectors $\mathbf{w}$ were computed. Fig. 11 shows example images of the tracked doll's head. Fig. 11 goes here.

The optimal coefficients vectors $\mathbf{w}$ were then fed into the ANN. The employed ANN was of the same type in both experiments. In this second experiment the dimensionality of the feature vectors was smaller: instead of the 128 complex values coefficients of the first experiment, we used in the second experiment only 52 real valued coefficients. We used 400 training images in both experiments. With this, we reached a mean pan/tilt error of $\approx 0.23^\circ$ with a processing speed of ≈ 10 fps on a 450 MHz Linux Pentium. The experiments have been repeated several times, and the variations of the estimated mean pan/tilt error over several experiments were small (0.02° for

the WN experiment).

4. Conclusions

In this paper we have discussed some properties of wavelet networks used for object representation. Wavelet networks are a combination of RBF networks and wavelet decomposition, where radial basis functions are replaced by wavelets. We have shown, that

- the optimized parameters as well as the wavelet coefficients are directly related to the underlying image structure,
- the coefficients can be computed from the projections of the wavelets onto the considered function and
- the optimized wavelets are linearly independent.

The first point is in our opinion very important: While, e.g., the radial basis functions in RBF networks do not reflect any properties of the represented functions, the wavelets of an optimized WN on the other hand do closely resemble properties of the represented function. Of course, as we have shown, it depends on the used mother-wavelet, what properties are reflected. In addition to the above properties, we have shown, that, apart from certain constrains taken from Daugman,[8] it is fairly straight forward, to optimize WNs (definition of a region of interest is needed, of course) and that the optimization time with under a minute for a mid-sized network is acceptable.

The above properties have been used in three small experiments, in which we have made extensive use of the wavelet subspace and the fact, that the wavelets form a basis. The wavelet subspace is isomorphic to the subspace of the $\mathbb{L}^2(\mathbb{R}^2)$, spanned by the optimized wavelets. It is low dimensional and invariant to affine deformations of a template WN which makes computations in our tracking experiments more efficient. The pose estimation experiment showed that by carefully

selecting the filters (e.g. by using a WN) both the error and the filtering effort can be minimized. All experiments would not have been successful, if the mapping from the $\mathbb{L}^2(\mathbb{R}^2)$ into the subspace hadn't been unique. Our experiments have mainly dealt with faces, but we think that the properties of WNs are general enough to be applied to general objects.

Recently, the relations between ANNs, WNs and fuzzy systems have been discussed,[6] but WNs were considered only in a very simplified fashion: Only radial wavelets were considered, which limits the potential of wavelet networks considerably. We would like to argue that, because of the close relation between the data and the basis functions, WNs offer new potential that goes beyond the potential of RBF Networks. At least for 2-D functions and the shapes of human faces, this has been partially shown here. We think that this can be generalized to other N-D functions.

Acknowledgment

The images used are derived from the Yale Face Database. This work as partially supported by the DFG-Grant So 322/1-2 and 1-3, and by the Human ID Grant program under the ONR Grant N00014-00-1-0908. We would like to thank Rama Chellappa for his valuable comments. The images used in this publication are derived from the "Yale Face Database".[19]

References

1. Q. Zhang and A. Benveniste. Wavelet networks. *IEEE Trans. Neural Networks*, 3:889–898, 1992.

2. H. Szu, B. Telfer, and S. Kadambe. Neural network adaptive wavelets for signal representation and classification. *Optical Engineering*, 31:1907–1961, 1992.

3. H. Szu, B. Telfer, and J. Garcia. Wavelet transforms and neural networks for compression and recognition. *Neural Networks*, 9:695–708, 1996.

4. Q. Zhang. Using wavelet network in nonparametric estimation. *IEEE Trans. Neural Networks*, 8(2):227–236, 1997.

5. C. C. Holmes and B. K. Mallick. Bayesian wavelet networks for nonparametric regression. *IEEE Trans. Neural Networks*, 11:27–35, 2000.

6. L. Reyneri. Unification of neural and wavelet networks and fuzzy fystems. *IEEE Trans. Neural Networks*, 10:801–814, 1999.

7. I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Information Theory*, 36:961–1005, 1990.

8. J. Daugman. Complete discrete 2D Gabor transform by neural networks for image analysis and compression. *IEEE Trans. Acoustics, Speech, and Signal Processing*, 36:1169–1179, 1988.

9. T. S. Lee. Image representation using 2D Gabor wavelets. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18:959–971, 1996.

10. W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes, The Art of Scientific Computing*. Cambridge University Press, Cambridge, UK, 1986.

11. J. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized two-dimensional visual cortical filters. *J. Opt. Soc. Am.*, 2:1160–1168, 1985.

12. R. Feris, V. Krüger, and R. Cesar Jr. Efficient real-time face tracking in wavelet subspace. In *Proceedings of the Int. Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems*, pages 113–118, Vancouver, BC, Canada, 2001, in conjunction with the ICCV'01, 2001. IEEE Computer Society.

13. V. Krüger and G. Sommer. Affine real-time face tracking using gabor wavelet networks. In *Proc. Int. Conf. on Pattern Recognition*, Barcelona, Spain, Sept. 3-8, 2000. IEEE Computer Society.

14. A. Pentland. Looking at people: sensing for ubiquitous and sensable computing. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22:107–119, 2000.

15. V. Krüger, S. Bruns, and G. Sommer. Efficient head pose estimation with gabor wavelet networks. In *Proc. British Machine Vision Conference*, Bristol, UK, Sept. 12-14, 2000. BMVC.

16. J. Bruske and G. Sommer. Dynamic cell structure learns perfectly topology preserving map. *Neural Computation*, 7:845–865, 1995.

17. H. Ritter, T. Martinez, and K. Schulten. *Neuronale Netze*. Addison-Wesley, 1991.

18. V. Krüger. Gabor wavelet networks for object representation. Technical Report CS-TR-4245, University of Maryland, CFAR, May 2001.

19. P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19:711–720, 1997.

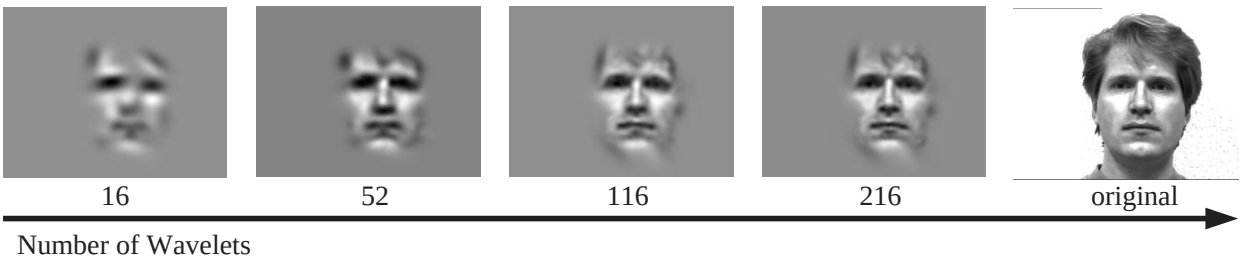

Fig. 1. The images indicate the variability in precision with a varying number of wavelets.

The wavelets were chosen in the order they were optimized.

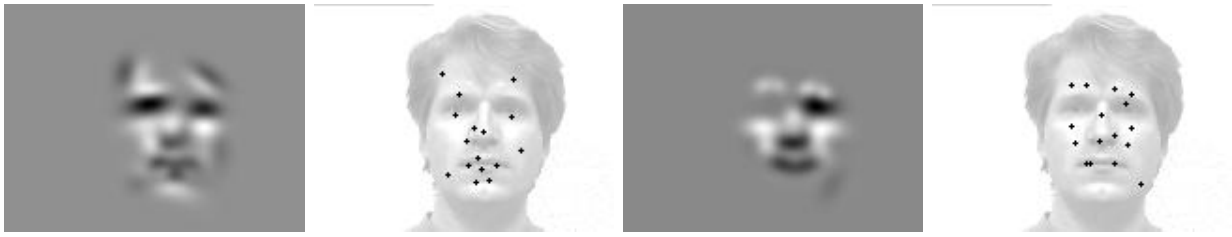

Fig. 2. These images show examples for two different mother wavelets, the odd gabor function (two left images), and the non-isotropic difference-of-Gaussian (two right images). The images show the reconstruction with 16 wavelets, and their superimposed optimized wavelet positions.

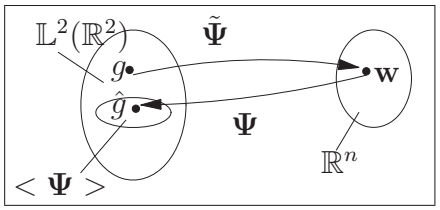

Fig. 3. *A function* $g \in \mathbb{L}^2(\mathbb{R}^2)$ *is mapped by the linear mapping* $\tilde{\mathbf{\Psi}}$ *onto the vector* $\mathbf{w} \in \mathbb{R}^N$ *in the wavelet subspace. The mapping of* $\mathbf{w}$ *into* $\mathbb{L}^2(\mathbb{R}^2)$ *is achieved with the linear mapping* $\mathbf{\Psi}$*. Both mappings constitute an orthogonal projection of a function* $g \in \mathbb{L}^2(\mathbb{R}^2)$ *into the (image) subspace* $< \mathbf{\Psi} > \subset \mathbb{L}^2(\mathbb{R}^2)$*.*

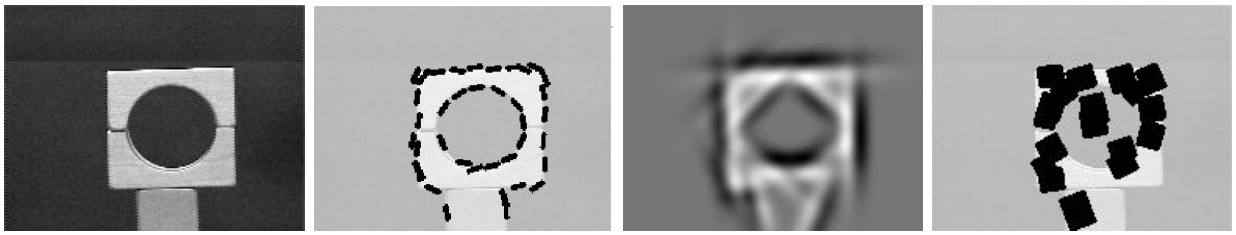

Fig. 4. This figure shows images of a wooden toy block on which a WN was trained. The black line segments sketch the positions, sizes and orientations of all the wavelets of the WN. The third image (from left) shows the residual image R between the original image and the approximation by the wavelets. The right image sketches the parameters of the largest optimized anisotropic DOG wavelets.

Fig. 5. Sample frames of our wavelet subspace tracking experiment. Note that the tracking method is robust to facial expressions variations as well as affine deformations of the face image.

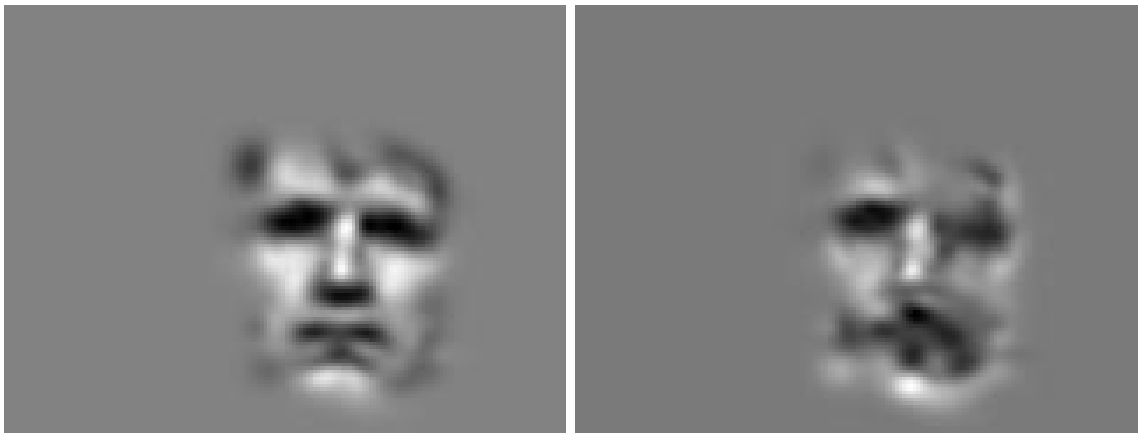

Fig. 6. These images show what happens when for the same individual the optimal coefficients vector is computed with a correct (left) and with a wrong (right) template WN.

Fig. 7. Various images of "subject01" (top) and their projections into the image subspace. The applied WN was optimized on the "normal" expression of "subject01".

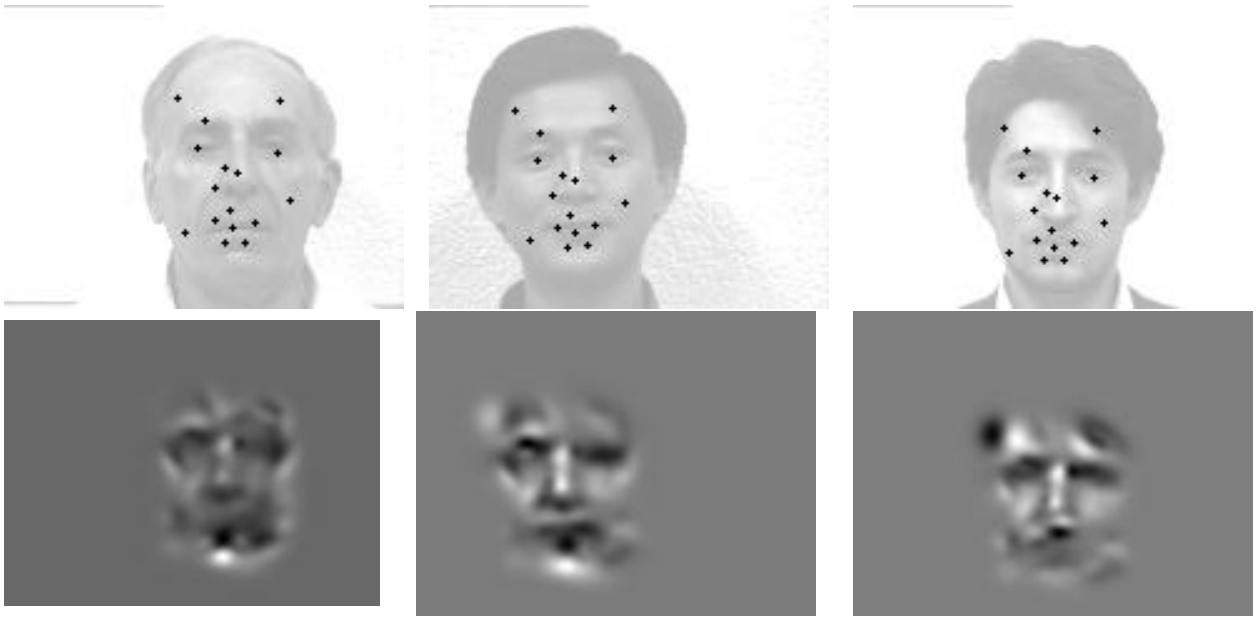

Fig. 8. Various images of subjects other than "subject01" (top) and their projections into the image subspace. The applied WN was optimized on the "normal" expression of "subject 01".

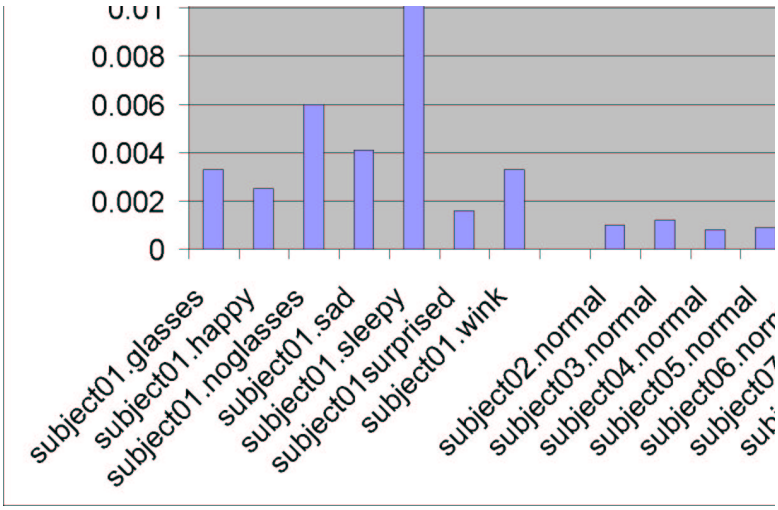

Fig. 9. The table shows the distance measurements $1/\|\cdot\|_{\mathbf{\Psi}}$ of the images of the various subjects in the face database to the gallery WN $(\mathbf{\Psi}_{01}, \mathbf{v}_{01})$ of subject 01. Higher values indicate a smaller difference between the two compared wavelet coefficient vectors. One sees that the values in the left part of the tables (subject 01) indicate a much smaller difference than the values in the right part of the table (different subjects).

Fig. 10. The left image shows the original doll face image I, the right image shows its reconstruction $\hat{I}_{52}$ using formula (3) with an optimized WN $\mathbf{\Psi}$ of just $N = 52$ odd Gabor wavelets, distributed over the inner face region.

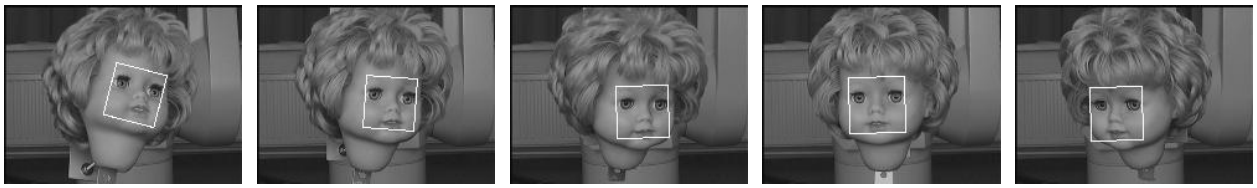

Fig. 11. The images show different orientations of the doll's head. The head is connected to a robot arm so that the ground truth is known. The white square indicates the detected position, scale and orientation of the WN.

List of Figures