Analysis and Segmentation of Facial Features in Laser Range Data*

R. Koch, M. H. Groß
Institute for Information Systems
Computer Science Department
Swiss Federal Institute of Technology
ETH–Zentrum, CH–8092 Zürich
E–mail: koch@inf.ethz.ch

(*) An extended version of this paper is available under [11].

Abstract

This paper describes a new method for analysis and segmentation of facial laser range data employing complex–valued non–orthogonal Gabor wavelets, principal component analysis and a topological mapping network. The initial data set consisting of shape and texture information is encoded in amplitude and phase of a complex 2D image function. A set of oriented Gabor filters performs a decomposition of the data and allows for retrieving local shape and texture features. In order to cluster the obtained feature vector further subspace mapping methods have to be figured out. For this purpose, a segmentation pipeline is proposed consisting of principal component analysis, normalization and a topological mapping network. This process ultimately renders a R,G,B subspace representation of the multidimensional feature vector.

1. Introduction

One of the major challenges of vision research is the development of methods for automatic modeling of complex geometric objects or scenes. In spite of several efforts over recent decades there is not yet a generic solution to this problem. Recent advancements in the development of laser scanners afford the modeling of shape and color information of the object in question. Hence, sophisticated methods to encode and analyze range image maps have become essential requirements.

Human faces in particular are attractive objects for laser range scanners [20], since on one hand complex geometric features and intrinsic symmetries [16] are scanned, while on the other hand they are very well conditioned for human visual analysis. Accurate geometric data from human faces can be used to enhance simple photographs and allow for development of higher quality techniques for facial surgery, face recognition, face reconstruction, simulation of aging, etc.

Moreover, when considering range data, there needs to be a way to also encode and analyze texture. Apart from that, wavelets and multiresolution analysis, as proposed by [6] or [19] have become very attractive for many applications. In Computer Graphics mostly orthonormal wavelets and their separable 2D and 3D extensions have been used for hierarchical data decomposition and approximation. Especially in image processing, the power of wavelets has been investigated for feature extraction and analysis [3], [21]. While orthonormal wavelets are stressed in many applications, [7] employs nonorthogonal and nonseparable 2D–Gabor wavelets for image analysis. Other interesting research is reported by [4], who combines graph matching with Gabor functions for face recognition.

This paper addresses two fundamental areas of facial feature analysis: First, it describes how to extract multidimensional features from laser range scans using a complex–valued decomposition scheme with Gabor wavelets. Secondly, it provides a generic scheme to visualize these features in orthogonal subspaces. For this purpose we consider shape and texture as amplitude and phase of a complex–valued 2D image respectively. We then perform a hierarchical decomposition with a carefully designed set of Gabor wavelets. Once the decomposition of the data is computed, the result is a multidimensional feature vector at each facial point. To inspect it visually subspace projections have been applied in terms of a principal component analysis of the wavelet features and a modified topological mapping neural network [15].

The organization of the paper is as follows: First of all we introduce in the different steps of our feature analysis pipeline and our encoding scheme for laser range data sets is elaborated. Light is shed on the mathematical basics of the Gabor function and its relationship to non–orthogonal wavelet decomposition. Furthermore it is illustrated how to get local data features from this transform. Principal component analysis and a topological mapping network then briefly adressed. Finally, we present results from the performance of our method applied to range data of human faces. R,G,B representations of multidimensional face features are depicted as derived from the Gabor wavelets

and it is outlined how this method generalizes onto different data sets.

2. The Face Analysis Pipeline

2.1. Overview

We propose a decomposition and segmentation pipeline as fig. 1. This concept comprises an embedded framework using different methods proposed by one of the authors in [12] and [14]. Once the feature vectors are extracted from the decomposition, they are fed into a subsequent processing pipeline of PCA–analysis, normalization and clustering into R,G,B color space with a Kohonen map.

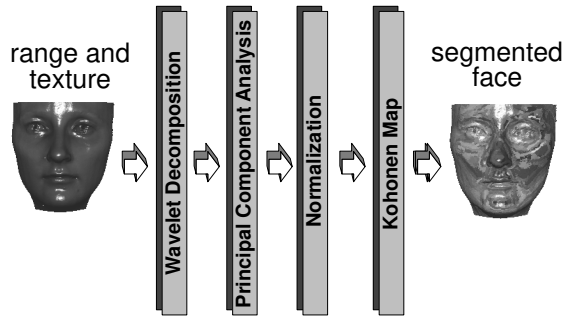


Figure 1: Proposed analysis pipeline

First of all the Gabor wavelets provide non–orthonormal expansions of the data. Hence the resulting feature vector is either more or less highly correlated, depending on the choice of the decomposition parameter. The decorrelation can be accomplished by further expansion into its principal components.

The coefficients of the feature vector transformed by the PCA are decorrelated and have to be normalized. A self–organizing neural network clusters the data and maps them automatically onto the R,G,B color space. Similar data features are represented by similar colors. The method has the advantage of reducing features to a limited set of clusters rather than to visualize a whole subspace. Therefore, the results obtained are much more expressive than an immediate mapping of the first three eigenvectors into R,G,B.

In the following subsections, we describe the processing steps of this pipeline.

2.2. Data Preprocessing

We employ laser range data from a Cyberware laser scanner, which provides highly accurate shape and color values of scanned objects. The data sets can be interpreted as R,G,B,range images defined on a cylindrical coordinate system. The grid size is 512 x 480 at a resolution in the range of about 0.3 mm.

Due to the cylindrical coordinate system of the scanner, the data have to be projected into Cartesian coordinates representing the 2D image function in (x,y). The projection can be thought to as take the normalized z_S–coordinate of the scanner system as the y–coordinate of the image and the angular coordinate γ_S as it's x–coordinate. The result is an operation which enrolls the cylinder onto a plane.

$$z_S \mapsto y \ , \quad \frac{\gamma_S}{2\pi} \mapsto x \tag{1}$$

where (z_S, r_S, γ_S) denote the cylinder coordinates of the scanning system.

Now, the question arises how to encode both color and shape in order to decompose them efficiently with the Gabor pyramid. A straightforward way would be to treat R,G,B and range separately as four different decompositions. This is however very expensive and provides highly correlated and high–dimensional features. The Gabor function consists of both a symmetric and an antisymmetric oscillation term, accounting for the real and the imaginary components of the data, respectively. Hence, in our approach, range and color are interpreted as amplitude and phase of a complex–valued image function that is convoluted with a complex–valued wavelet.

Let $r(x,y)$, $g(x,y)$, $b(x,y)$, $s(x,y)$ be the color and range information at a particular point (x,y). In order to compute the amplitude $a(x,y)$ we extract luminance according to the rules of colorimetry:

$$\begin{aligned} a(x,y) \ = \ & 0.31 \ r(x,y) \\ & + \ 0.59 \ g(x,y) + \ 0.10 \ b(x,y) \end{aligned} \tag{2}$$

The range function is taken immediately as the phase term of the image function and the description of $i(x,y)$ is obtained as:

$$i(x,y) \ = \ a(x,y) \ e^{2\pi j s(x,y)} \tag{3}$$

Where $s(x,y)$ is normalized between 0 and 1.

Once the data is encoded as in eq. 3 it can be fed into the wavelet decomposition step of our pipeline.

2.3. Wavelet Decomposition

2.3.1. The Gabor Function

Gabor functions of different frequency range and orientation can provide a non–orthogonal function basis for any finite energy function $f \in \boldsymbol{L^2R^n}$. One way to rapidly obtain data features is to compute a Gabor transform as an expansion of f. The 2D version of the complex–valued Gabor function can be expressed [8] as:

$$\psi^G(x,y) \ = \ e^{-\pi\left(\frac{(x-x_0)^2}{\sigma_x{}^2}+\frac{(y-y_0)^2}{\sigma_y{}^2}\right)} \ e^{-2\pi j\left(u_0(x-x_0)+v_0(y-y_0)\right)} \tag{4}$$

(x_0, y_0) stand for the translation of it's origin and (u_0, v_0) specify the modulation coordinates in the frequency domain. The effective width and length are given by (σ_x, σ_y) that specify the elliptic envelope.

Evidently, Gabor functions are located in both space and frequency and they moreover uniquely

achieve the theoretical lower bound of uncertainty as it is dictated by the Heisenberg relation:

$$\Delta_{xy}\,\Delta_{uv} \geqq \frac{1}{4\pi^2} \tag{5}$$

Fortunately, the Fourier transform $G(u,v)$ of g has exactly the same functional form with inverted or interchanged parameters. For functions located at the origin it reduces to a simple Gaussian in the frequency plane. Introducing the spatial frequencies (u, v) we obtain the following expression:

$$G(u,v) = e^{-\pi\left(\frac{(u-u_0)^2}{\sigma_u^2}+\frac{(v-v_0)^2}{\sigma_v^2}\right)}\; e^{-2\pi j\left(x_0(u-u_0)+y_0(v-v_0)\right)} \tag{6}$$

with $\sigma_u = 1/\sigma_x$ and $\sigma_v = 1/\sigma_y$

Evidently, implementations of Gabor filters can be accomplished easily through employing eq. 6.

The constant size of the envelope leads to an uniform partitioning of the frequency plane when expanding f. Thus the localization of the Gabor transform is equal throughout the entire spectrum. However, sophisticated data analysis should provide a high spatial resolution for high frequencies and a low resolution for lower frequencies. Therefore, we have to adapt the size of the Gaussian window to its frequency position. Due to the scaling theorem of the Fourier transform, shifting and increasing the width and length of the Gaussian in frequency space go in hand with different modulation and a decrease of width and length of the wavelet's envelope in the spatial domain. It is clear, a whole family of self–similar functions of this type can be generated, simply by shifting and scaling a Gaussian in the frequency plane.

2.3.2. Gabor Wavelets

The idea of building an expansion basis from self–similar functions of different sizes and orientations immediately leads to the generic wavelet transform [19], [6]. Basically, it can be considered as the inner product of any finite energy function f with a set of self–similar basis functions $\psi_{a,b}$ that are derived from each other by scaling and shifting of one prototype using the parameters $a, b \in \mathbb{R}$ respectively. It's 2D representation can be addressed as:

$$CWT = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi_{a,b}(x,y)\, f^*(x,y)\; dxdy = \langle \psi_{a,b}\;,\; f \rangle \tag{7}$$

with

$$\psi_{a,b}(x,y) = \frac{1}{\sqrt{|a_x\; a_y|}}\psi\left(\frac{x-b_x}{a_x}, \frac{y-b_y}{a_y}\right) \tag{8}$$

The brackets <,> denote the inner product operation. Its Fourier transform is given by

$$\Psi_{a,b}(u,v) = \sqrt{|a_x\; a_y|}\; e^{-2\pi j(b_x\, +\, b_y)}.\Psi(a_x u, a_y v) \tag{9}$$

Note that the normalization forces the bases to hold Parseval's energy equivalent.

2.3.3. Frequency Decomposition with Gabor Wavelets

Depending on the job they are tailored for, additional constraints of the bases are often demanded, such as compact support, smooth shape, fast decay, closed–form descriptions, etc. It has turned out that the construction of bases satisfying these criteria is a non–trivial optimization process and usually a compromise has to be found, semi–orthonormality for example [5]. However, relaxing the orthonormality and bandpass constraint, we can construct a wavelet basis using the Gabor functions of different scale, shift and orientation [7]. In this case a complete set of self–similar functions is defined by

$$\psi^G_{mpq\Theta}(x,y) = a^m\; \psi^G(x',y') \tag{10}$$

$$\begin{aligned} x' &= a^m\; ((x\,\cos\Theta + y\,\sin\Theta) - p) \\ y' &= a^m\; ((-x\,\sin\Theta + y\,\cos\Theta) - q) \end{aligned} \tag{11}$$

Here, we assume a circular scaling with $a_x = a_y = a$. The parameters represent transformations of the mother wavelet in scale (a,m), translation (p,q) and orientation (Θ). Note, that due to the definition of the Gabor function each wavelet has an intrinsic DC–fraction in it's real (cosine) component.

In our implementation, we define the decomposition using a lower frequency bound ω_0 and an angular resolution $\Delta\Theta$. Figure 2 shows how the frequency plane is decomposed by superimposing Gaussian filters. The mathematical details of the construction rules we propose here are described in details in the extended version of this paper [11].

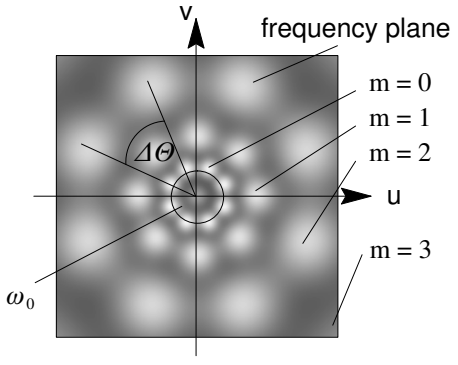


Figure 2: Decomposition of the frequency plane with Gabor wavelets.

Let $i(x,y)$ be our complex facial image and $I(u,v)$ be it's Fourier transform, we obtain the feature vector $g_l(x,y)$ by multiplying $G_{m,i}(u,v)$ with $I(u,v)$ and subsequent inverse Fourier transform:

$$g_l = g_{m,i}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G_{m,i}(u,v)I(u,v)e^{2\pi j(ux+vy)}dudv \quad (12)$$

This scheme provides a more effective way of building up the wavelet pyramid – a way which is preferable to computing all inner products of $<i(x,y),\psi_{mpq\Theta}>$ explicitly.

2.4. Principal Component Analysis

Generally, dimensionality reduction is strongly connected with finding subspaces that satisfy respective optimization criteria. In statistical data analysis, we can find several techniques for this task. One of the most famous methods, taking the optimization as an eigenvalue problem, is the principal component analysis or Karhunen–Loève expansion. This method aims to find a subset of principal directions in a data set of any statistical distribution. The corresponding basis vectors satisfy the constraint of orthonormality and the subspace defined by these vectors diagonalizes the covariance matrix of the data.

The eigenvectors $\boldsymbol{o}_k$ $(k = 1,..,K)$ ordered according to their eigenvalues of the covariance matrix define a subspace that minimizes the average error of information lost by the reduction of the number of dimensions.

The dimensionality reduction is obtained by projecting the feature vectors $\boldsymbol{g}_l$ into the eigenspace spanned by the $\boldsymbol{o}_k$

$$\phi_{lk} = \boldsymbol{o}_k^T \boldsymbol{g}_l \quad (13)$$

and leaving out the coefficients of higher, i.e. less significant, dimensions.

2.5. Normalization

The principles of PCA force most of the data energy to be concentrated in a few significant eigenvectors. Therefore the PCA based feature vectors $\boldsymbol{\Phi}_l$ have to be normalized for further processing. Since the a priori probability of these vectors is unknown, we assume uniform distribution and normalize their components according to:

$$\phi_k^{min} = \min_l(\phi_{lk}) \qquad \phi_k^{max} = \max_l(\phi_{lk}) \quad (14)$$

$$\tilde{\phi}_{lk} = \frac{\phi_{lk} - \phi_k^{min}}{\phi_k^{max} - \phi_k^{min}} \quad (15)$$

One straightforward way to visualize the obtained feature space is to encode its 3 most significant basis vectors into R,G,B. Although this presentation reveals a first sketch of similarities in the multidimensional feature space in terms of similar colors, it does not yet provide a sophisticated feature extraction and the visual interpretation of the resulting images remains difficult. Hence, a additional clustering is required to group similarities in the data. This can be reached by a topological mapping neural network as explained below.

2.6. Topological Mapping

The 3D extension of the Kohonen map was introduced by the authors in [10] and employed in different applications. Figure 3 shows the topology of the network.

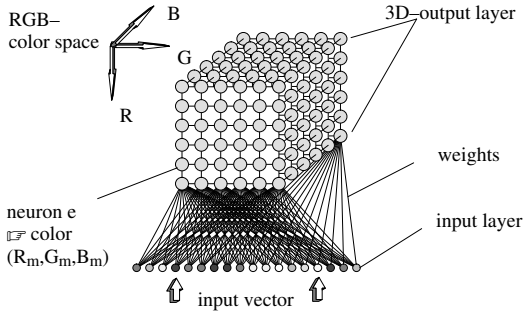


Figure 3: 3D Kohonen feature map

The training of the network is figured out by presenting feature vectors $\tilde{\boldsymbol{\Phi}}_l$ randomly to the input layer of the network whose connection weight vectors $\boldsymbol{m}_h$ of all competitive neurons h are initialized by random values. We choose K input neurons according to the data dimension and define an Euclidean distance d_h between $\tilde{\boldsymbol{\Phi}}_l$ and $\boldsymbol{m}_h$ with

$$d_h = \| \tilde{\boldsymbol{\Phi}}_l - \boldsymbol{m}_h \| = \sqrt{\sum_{k=1}^{K}(\tilde{\phi}_{lk} - m_{hk})^2} \quad (16)$$

The neuron e with the minimum distance is then activated.

The updating of the weights m_{hk} associated with the neurons is only performed within a proximity ($h \in N_e(t)$) of e. This proximity $N_e(t)$ is reduced with increasing training time t. The updating conforms to Eq. 17, where $\eta(t)$ represents a time-dependent learning rate:

$$m_{hk}^{(t+1)} = m_{hk}^{(t)} + \Delta m_{hk}^{(t)} \quad (17)$$

$$\Delta m_{hk}^{(t)} = \begin{cases} \eta(t)(\tilde{\phi}_{lk} - m_{hk}^{(t)}), & if\ h \in N_e(t) \\ 0, & else \end{cases} \quad (18)$$

$$\eta(t) = \eta_0(1 - \frac{t}{T_{max}}),\ t \in [0,..,T_{max}] \quad (19)$$

This rule refers directly to c-means clustering, where the $\boldsymbol{m}_h$ represent the centroids.

Since the neurons in the competitive layer are ordered topologically, neighboring neurons react to similar data vectors in the input space. The mapping can be interpreted as reduction of any K-dimensional input pattern into 3 dimensions, preserving the topology of the data as much as possible. The resolution of this discrete 3D space is given by the number of competitive neurons, i.e., by the clusters.

Referencing the axes of the cube with the primaries R, G, and B, each neuron of the competitive layer represents a discrete entity in R, G, B space, i.e., it corresponds to a particular color triplet. Thus, the similarity of the color provided by the reacting neuron refers to a neighborhood in K-dimensional space.

3. Results

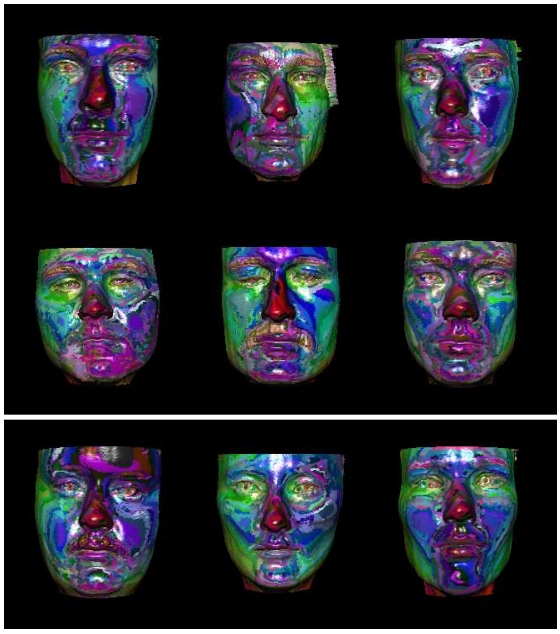
Our data set consists of 10 faces, where 7 have been used for training and 3 for testing generalisation. Figure 4a illustrates the results obtained by the method without PCA. It is clear that the wavelets extract similar facial features that appear as similar colors on each face, such as the tip of the nose in red–orange colors, or cheeks that come up either in blue–green or violet. The mouth, mostly revealed in red–pink, and moustache boundaries are extracted, as are eye brows (light orange). It is interesting to verify, that the left and right hemisphere are clearly distinguished and because of the different colors associated, belong to different clusters. Some of the results reveal a more or less symmetric or antisymmetric color map, depending on the person analyzed.

At this point, it may be useful to contrast our method with curvature analysis, or slicing, methods. In particular, when the results on the training data are compared with those obtained on the generalization data set (lower row), this technique evidently outperforms robust results, since color similarities referring to anatomic regions are rather well preserved. In general, within a single face due to the influence of the high–frequency components, there is hardly any large coherently clustered region. The similaritiy of colors in neighbored clusters is, however, preserved in regions of low spatial frequencies.

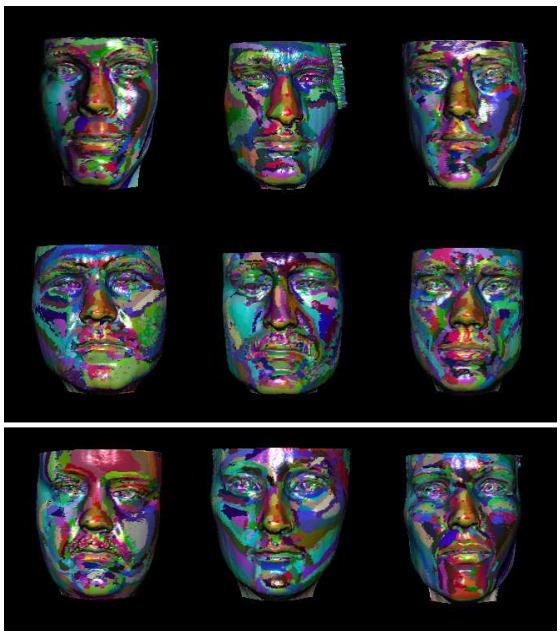
Figure 4b depicts the same investigation, but in this case with a PCA downmapping of the dimensions from an initial 34 to 7, which preserves 99 % of the signal energy encoded in the eigenvectors. The topology preservation of the Kohonen map is not as evident as in the previous figure, since the PCA accounts for the maximum variance in the features. But the regions clustered by the neural network appear in most cases more homogeneous and coherent because some fine grain detail information gets lost with the downscaling procedure. In both computations, the most discontinuous regions appear around the eyes. This is due to the discontinuouities in range, which spread the local spatial frequencies throughout the spectrum. The pupils of most of the test persons, however, are detected in similar colors.

4. Conclusion

We introduced a generic and robust method for feature–based face segmentation by decompositions with Gabor wavelets. Results have been presented that show the capabilities of the method. Although it was primarily applied to encode and analyze laser range data from human faces it is not restricted to and can be extended to any kind of complex data sets.



a)



b)

Figure 4: Face features as similarities of colors mapped as textures to the range data.
a) without PCA (training and generalization data)
b) including PCA (training and generalization data)

[2] for instance, shows that wavelet frame theory, however, defines bounds for the design of dual func-

tion spaces for a perfect reconstruction of Gabor frames. Since Gabor wavelets are powerful feature extractors, further research has to be conducted to find perfect reconstruction frames within these strict limitations. Other alternatives have to be considered, such as nonseparable extensions of biorthonormal wavelets.

Another important aspect is how the method outperforms under noise. Due to the spectral estimation–like method, we expect much more robust results than those, for instance, given by curvature analysis techniques for range data.

5. Acknowledgement

The authors wish to thank the Computer Graphics Center in Darmstadt for providing the raw laser range data set for this report.

6. References

[1] Akansu, A.; Haddad, R.: Multiresolution Signal Decomposition. Transforms, Subbands, Wavelets. Telecommunications Series, Boston: Academic Press, 1992

[2] Benedetto, Frazier: Wavelets: Mathematics and Applications, CRC–Press, 1994

[3] Bovik, A. C.; Clark, M.; Geisler, W.: Multichannel Texture Analysis Using Localized Spatial Filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, No. 1, pp. 55 – 73, 1990

[4] Buhmann, J.; Lange, J.; van der Malsburg, C.: Distortion Invariant Object Recognition bz Matching Hierarchically labeled Graphs. Proceedings of the IJCNN International Joint Conference on Neural Networks, Washington: IEEE, pp. 155–159, 1989

[5] Chui, C.: An Introduction to Wavelets. Boston: Academic Press, 1992

[6] Daubechies, I.: The Wavelet Transform, Time–Frequency Localization and Signal Analysis. IEEE Transactions on Information Theory, Vol. 36, pp. 961 – 1005, 1990

[7] Daugman, J. G.: High Confidence Visual Recognition of Persons by a Test of Statistical Independance, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 15, No. 11, pp. 1148 – 1161, 1993

[8] Daugman, J, G.: Two–Dimensional Spectral Analysis of Cortical receptive field profiles, Vision Research, Vol. 20, pp. 847 – 856, 1980

[9] Gabor, D.: Theory of Communication. Journal of the Institute of Electrical Engineers, Vol. 93, pp.429–457, 1946

[10] Groß, M. : Visual Computing. Springer–Verlag, 1994

[11] Gross. M., Koch R.: Visualization of Multidimensional Shape and Texture Features in Laser Range Data using Complex–Valued Gabor Wavelets, Yellow Report No. 227, Swiss Federal Institute of Technology, Zürich.

[12] Groß, M.; Koch, R; Lippert, L.; Dreger, A.: Multiscale Image Texture Analysis in Wavelet Space. Proceedings of the IEEE–ICIP'94, IEEE Computer Society Press, Vol. III, pp. 412–416, 1994

[13] Groß, M.; Lippert, L.; Dreger, A.; Koch, R.: A New Method to Approximate the Volume Rendering Equation Using Wavelets and Piecewise Polynomials, Comptuers& Graphics, Vol. 19, No. 1, to appear, 1995

[14] Groß, M.; Luttermann, H.: Combining Principal Component Analysis and Neural Networks for the Recognition of Human Faces - A Case Study for Man-Machine-Communication. Singapore: World Scientific, Proceedings of Pacific Graphics, pp. 176–192, 1993

[15] Groß, M.; Seibert, F.: Visualization of Multidimensional Data Sets using a Neural Network. The Visual Computer, Vol. 10, No. 3, pp. 145–59, 1993

[16] Harmon, L.; Kahn, M.; Lasch, R.; Ramig, P.: Machine Identification of Human Faces. Pattern Recognition, Vol. 13, No. 2, pp. 97–110, 1981

[17] He, Tao; Wang, S.; Kaufman, A.: Wavelet–Based Volume Morphing. ACM Volume Rendering Symposium, 1994

[18] Laine, A.; Fan, J.: Texture Classification by Wavelet Packet Signatures. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, No. 11, pp. 1186 – 1191, 1993

[19] Mallat, S.: A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 7, pp. 674–693, July 1989

[20] Tanaka, T.: Kishimo, F.: Recodering and Visualizing Complex Shape from Range Data. in Kunii (Edt.): Visual Computing, Proceedings of the CGI'92, pp. 331–348, 1992

[21] Unser, M.; Eden, M.: Multiresolution Feature Extraction and Selection for Texture Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 7, pp. 717 – 728, 1989

[22] Vetterli, M.: Wavelets and Filter Banks: Theory and Design. IEEE Transactions on Signal Proc., Vol. 40, pp. 2207 – 2231, 1992