

Face Representation and Tracking Using Gabor Wavelet Networks

Michal Spanel, Supervisor: Doc. Dr. Ing. Pavel Zemcik
Brno University of Technology, Faculty of Information Technology
Computer Graphics and Multimedia Department
E-mail: xspane01@stud.fit.vutbr.cz

Abstract

Many approaches to human gesture recognition exist. This work presents one of the approaches. Especially the problems of human face representation and tracking in video sequences are discussed. A method, where a discrete face template is represented by linear combination of continuous 2D odd-Gabor wavelet functions (Gabor Wavelet Network-GWN) is proposed. The weights and 2D parameters (position, scale and orientation) of each wavelet are determined optimally. Using this representation, an effective face tracking method is achieved that is robust to illumination changes and deformations of the face image such as smile. Possibilities of cooperation between a GWN and others methods for robust facial expressions recognition are also discussed, for example: PCA (Principal Component Analysis) and parameterized models of optical flow.

Keywords: Image, Object representation, Feature-based, Template-based, Gabor Wavelet Networks, Facial features, Tracking, Facial expressions, Parameterized models of optical flow, Principal Component Analysis

1 Introduction

The automatic face and facial features tracking in video sequence is a fundamental problem of gesture recognition. A new method for face tracking based on the approach of Krüger [1] is introduced; the potential of a Gabor Wavelet Networks (GWNs) for face template representation is also shown. This tracking method is quite robust and it is insensitive to homogeneous illumination changes and affine deformations of the face image.

The face template is represented by continuous 2D odd-Gabor wavelets. A continuous model of a discrete face image is obtained by linear combination of wavelets. The weights and 2D parameters (position, orientation and dilatation) of each wavelet are determined optimally. With respect to earlier defined approaches to object representation, it could be said that GWN combines template-based and feature-based approaches. Each wavelet represents a model of some image feature. On the other side, the overall face geometry is presented in this model as well.

The number of wavelets may be chosen by the user. With increasing number of wavelets it is possible to obtain a more accurate face template (face representation becomes more specific). As the number decreases, the representation becomes more general; thus, it is possible to suppress effect of different individuals' faces. The results produced by GWN tracking may be used for facial expressions recognition. These possibilities are discussed further in the paper.

Face localization in the first image of a video sequence is a problem that is not easily solvable. Many methods are, however, available (skin color segmentation, facial features detection and grouping, etc.). At the moment, the focus is not on these "starting methods". Several ideas are, however, presented in Future Work section.

2 Related work

Many approaches have been proposed for 2D object representation. Only those that are usable for face tracking and facial expressions recognition are described.

One of the most successful approaches to template-based object representation is based on Principal Component Analysis (PCA). PCA approximates texture only, geometrical information about object is not evaluated. The major drawback of PCA is its sensitivity to deformations and illumination changes. The eigentracking approach [3] builds on (and extends) eigenspace (like PCA) representations, robust estimation techniques, and parameterized optical flow estimation. An affine transformation between the eigenspace and the image is robustly estimated.

The paper [7] represents other template-based approaches, where object representations are found implicitly through application of artificial neural networks (ANNs). The inputs to the neural network are subsampled gray-value images of the object class.

Feature-based approaches form other group. The features are detected in an image through spatial filters, and filter responses are grouped according to geometric constraints. Probabilistic frameworks are used to reinforce probabilities and to evaluate the likelihood that the candidate is certain object.

Faces and facial features tracking is fundamental task. Recently, color-based systems have been widely used to accomplish it. Many works present statistical

skin-color model, invariant to different people, which is used to track face blob along image sequence. In this way, a working preprocessor can be used for initial face localization in our GWN approach.

The papers [4][5] describe the representation and recognition of human motion using parameterized models of optical flow. A person's limbs, face and facial features are represented as patches whose motion in an image sequence can be modeled by low-order polynomials. A robust optical flow estimation technique is used to recover the motion of these patches. Recovered motion parameters provide description which can be used to recognize human gestures and facial expressions.

This work presents GWNs as an object representation approach which is sparse and efficient. It is shown that GWNs can be used for robust affine face tracking insensitive to illumination changes and deformations of the face image.

3 Function Approximation

The wavelet representation for a face template is obtained by an approximation of image function. The problem can be defined as searching for an unknown continuous function $f': R^2 \rightarrow R$. A grayscale image that is described by a intensity function $f: R^2 \rightarrow R$ is used. In this case, the value of each pixel of the image is determined as the value of the function sample (x_i, y_i) where x_i is pixel position and y_i is pixel intensity.

3.1 Neural Networks

Neural networks have been intensively studied for function approximation. It has been shown that a feedforward neural network with only one hidden layer is sufficient to approximate any continuous function which is defined on a hypercube with edges $(0,1)$. In other words, given any continuous function f defined on $[0,1]^n$ and any $\varepsilon > 0$, there is a sum $f'(x)$ of the form:

$$f'(\vec{x}) = \sum_{i=1}^M \omega_i \sigma(\vec{a}_i^T \vec{x} + b_i),$$

for which $|f'(x) - f(x)| < \varepsilon$ for all $x \in [0,1]^n$. Function σ is non-linear continuous, limited and monotonically increasing. Parameters $\omega_i, b_i \in R$ and $a_i \in R^n$.

3.2 Wavelet Networks

Wavelet networks were introduced as an alternative to feedforward neural networks for function approximation. This concept was inspired by the wavelet decomposition and neural networks. It is well known that wavelet decomposition allows decomposition of any function $f(x) \in L^2(R^n)$ using family of functions obtained by dilating and translating a single mother wavelet function $\psi: R^2 \rightarrow R$. Function f may be expressed as a linear combination of wavelets,

where wavelet weights are estimated by the decomposition process. The number of the wavelets is elective and the parameters are optimized by a learning process. The more wavelets are used, the more precise approximation is achieved.

The typical architecture can be seen in Figure 1. Mathematically, it can be expressed in the following way:

$$f'(x) = \sum_{i=1}^M w_i \psi_{n_i}(x) + \bar{f}$$

where $w_i \in R$, ψ_{n_i} is a wavelet function and n_i is parameters vector (position, dilatation, orientation).

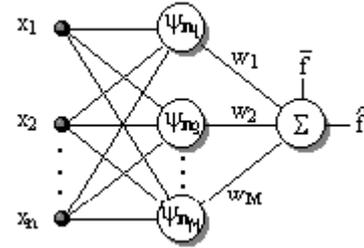


Fig. 1: Wavelet network

3.3 Gabor Wavelet Networks

In this case, the face template is represented by a wavelet network where the mother wavelet is an 2D odd-Gabor function. Figure 2 illustrates used function.

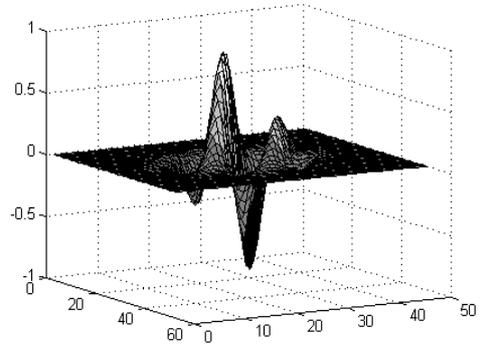


Fig. 2: 2D odd-Gabor wavelet function

First, a family of M 2D odd-Gabor functions $\psi = \{\psi_{n1}, \psi_{n2}, \dots, \psi_{nM}\}$ is defined. Then parameters vector $n_i = \{c_x, c_y, \theta, s_x, s_y\}$ modify the shape of the mother wavelet, c_x, c_y are the translation (position), s_x, s_y denote the dilatation, and θ denotes orientation of wavelet.

$$\psi_{n_i}(x, y) = \exp\left(-\frac{1}{2\pi^2} [s_x((x-c_x)\cos\theta + (y-c_y)\sin\theta)]^2 + [s_y(-(x-c_x)\sin\theta + (y-c_y)\cos\theta)]^2\right) * \sin(s_x((x-c_x)\cos\theta - (y-c_y)\sin\theta))$$

The wavelet function above represents used mother 2D odd-Gabor function as well.

4 Using GWNs

The Gabor wavelet networks have been already defined. In this section, others aspects of using GWNs are discussed. An important part of this method is parameters and weights optimization. For image f an energy (error) function is specified which is minimized by the means of learning process that respects the desired wavelet network parameters.

$$E = \min_{\tilde{n}_i, w_i \forall i} \left\| f - \left(\sum_i w_i \psi_{\tilde{n}_i} + \tilde{f} \right) \right\|_2^2$$

The energy minimization problem is useful to solve by the Levenberg-Marquardt (LM) gradient descent method (for details see [8]). The method might get stuck in local minima; therefore, a careful selection of initial parameters is important.

It can be said that the two optimized vectors $\psi = \{\psi_{n1}, \psi_{n2}, \dots, \psi_{nM}\}$ and $w = \{w_1, w_2, \dots, w_M\}$ constitute an optimized Gabor Wavelet Network (GWN) for the specific face image f . The reconstruction of the original image is done using the following expression:

$$f' = \sum_{i=1}^M w_i \psi_{\tilde{n}_i} + \tilde{f}.$$

The quality of the reconstruction largely depends on the number M of used wavelets. The obtained continuous template has several advantages. The degree of generalization depends on the number of wavelets. The model is quite insensitive to affine deformations of the face region (eye blinking, smile). Since the GWN is free of mean image value \tilde{f} , the model is also insensitive to homogeneous illumination changes.

4.1 Optimization of GWN

The procedure of GWN optimization uses pyramidal scheme and is distributed into several layers. First, the 4x4 coarse wavelets are equidistantly positioned in the inner face region. These wavelets define the first pyramid layer. They are roughly initialized and then optimized with respect of energy function. The result is a GWN_{16} representing the image I'_{16} .

In a second step the difference between the original image and its reconstruction $D = I - I'_{16}$, which is approximated by 6x6 finer wavelets, is calculated. These wavelets form the second pyramid layer. Both layers merged together define GWN_{52} and the reconstructed image I'_{52} . So the face template is described by the distribution of 52 wavelets. This way, it is possible to efficiently minimize energy of reconstructed image. It is always necessary to proceed from coarse wavelets to finer wavelets.

The initial orientations are random and the initial dilatations are constant in each layer, and their values are chosen with respect to the distances to the neighbouring wavelets. The pyramid layer is not minimized globally because faster is process the wavelets one by one.

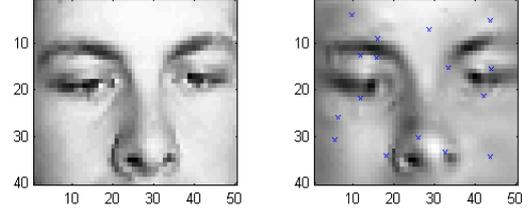


Fig. 3: Optimized GWN_{116} (original on the left side)

4.2 Direct Calculation of Weights

It is not necessary to calculate the wavelet weights just as others parameters (using L-M gradient descent method). V. Krüger have presented algorithm for the direct calculation of weights [1]. His approach is based on principles of bi-orthogonality and dual wavelets. Gabor wavelets are non-orthogonal – that means it is not possible to compute weights by simple projection wavelets onto the image. This problem can be solved by considering the bi-orthogonal family of wavelets ψ' . We say that two families of wavelets $\psi = \{\psi_{ij}\}$ and $\psi' = \{\psi'_{ij}\}$ are bi-orthogonal if for all i, j they satisfy the condition:

$$\langle \psi_i, \psi'_{i'} \rangle = \delta_{i, i'},$$

where $\delta_{i, j}$ is “Dirac” function

$$\delta_{i, j} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

and $\langle f, g \rangle$ is scalar product defined as

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)g(x)dx$$

The wavelet ψ' is called dual wavelet of ψ . How the dual wavelets are found? Just by the following substitution:

$$\psi'_{\tilde{n}_i} = \sum_{j=1}^N (\Psi_{i, j})^{-1} \psi_{\tilde{n}_j}$$

where Ψ is the matrix of pairwise scalar products:

$$\Psi_{i, j} = \langle \psi_{\tilde{n}_i}, \psi_{\tilde{n}_j} \rangle.$$

Knowing dual wavelets and orthogonality principle it is possible to write the simple equation, which is used for direct calculation of weights:

$$w_i = \langle \psi'_{\tilde{n}_i}, f \rangle$$

4.3 Repositioning of a GWN

It has been shown how the continuous wavelet representation for a face template is obtained – the background and basic ideas of Gabor Wavelet Network. Now it will be shown how this representation can be affinely repositioned in a new face image so that its wavelets are distributed on the same facial features as in the original image. This process is called GWN repositioning.

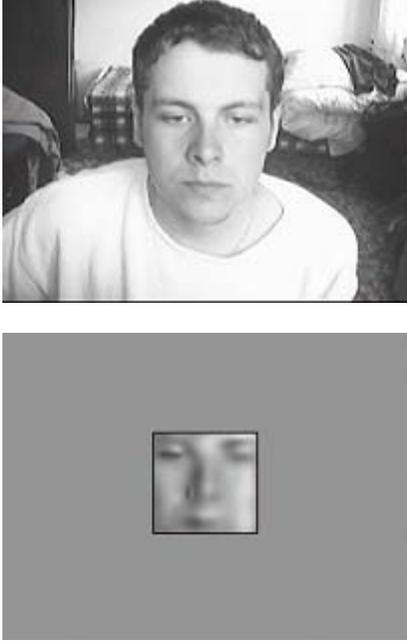


Fig. 4: Face template-GWN₅₂

For example, let us consider the face template shown in Figure 4. Now assume face motion so the face position changes in the new image. In the repositioning process, all the wavelets are positioned correctly on the same facial features in the warped (new) image. It is important to note that GWN repositioning may cover any affine transformation (translation, dilatation, rotation and shearing) applied to the original face region.

The GWN repositioning consists in determination of correct parameters of this transformation. First it is necessary to define the so-called Gabor superwavelet. Let us assume we have a GWN described by two vectors $\boldsymbol{\psi} = \{\psi_{n1}, \psi_{n2}, \dots, \psi_{nM}\}$ and $\boldsymbol{w} = \{w_1, w_2, \dots, w_M\}$. Gabor superwavelet Ψ_n (GSW) is defined as a linear combination of the wavelets ψ_{ni} such that:

$$\Psi_n(x) = \sum_i w_i \psi_{ni}(SR(x - C) + C + T)$$

where the vector \boldsymbol{n} of parameters of the superwavelet Ψ_n determines the dilatation matrix \boldsymbol{S} , the rotation matrix \boldsymbol{R} , the translation vector \boldsymbol{T} , and the vector \boldsymbol{C} that contains coordinates of the face region centre

$$\boldsymbol{S} = \begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix},$$

$$\boldsymbol{R} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix},$$

$$\boldsymbol{T} = (t_x, t_y),$$

$$\boldsymbol{C} = (c_x, c_y).$$

For a given new image g , it is possible to arbitrarily deform the superwavelet by optimizing its parameters with respect to the energy function:

$$E = \min_{\bar{n}} \|g - \Psi_{\bar{n}}\|_2^2.$$

Note, please, that the superwavelet parameters include only translation, dilatation, and rotation. Even so, shearing can be included and thus any affine deformation of GSW is allowed; for this purpose, the dilatation matrix \boldsymbol{S} with the new parameter s_{xy} can be rewritten as shown below:

$$\boldsymbol{S} = \begin{pmatrix} s_x & s_{xy} \\ 0 & s_y \end{pmatrix}.$$

In order to find the optimal parameters, the energy function must be minimized using the previously declared Levenberg-Marquardt method. Usually, the initialization for gradient descent method are satisfactory within the range $\pm 10\%$ in position $\pm 20\%$ in dilatation, and/or $\pm 10\%$ in rotation.

The essential property of wavelet representation, the ability to generalize face template using small number of wavelets, remains valid and the wavelet representation works quite well for different individuals.

4.4 Face Tracking Using GWN

In the previous section, the GWN repositioning has been described. Its principle may be also applied to a video sequence providing the way to solve the fundamental problem of gesture recognition and face tracking. In this approach, the face is considered a planar object that is viewed under an orthographic projection.

For each frame J_t (acquired in the time t), the Gabor superwavelet is optimized according to the energy function:

$$E = \min_{\bar{n}_t} \|J_t - \Psi_{\bar{n}_t}\|_2^2.$$

GSW parameters in J_{t-1} are used as initial values for optimization in frame J_t . As the frame to frame image changes are mostly small, the optimization process's convergence is faster.

After the initialization at time t_0 (face localization in the first frame), the wavelet representation for face region is obtained using a GWN. This template is then affinely repositioned in the next frame as described above. Face tracking is then performed applying the found transformations to the selected points.

This tracking method includes the overall geometry of face; therefore, it is robust and insensitive to facial feature deformations such as eye blinking.

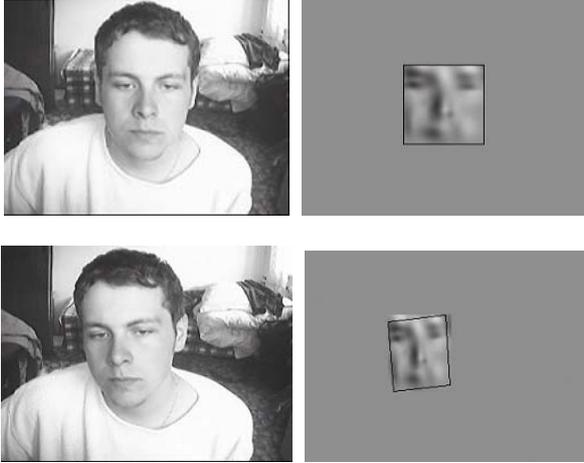


Fig. 5: GWN repositioning

5 Potential of GWNs

In this section, the ability of Gabor wavelet networks in relation to facial expressions recognition is discussed. Tracking of the facial features (eyes, brows, limbs, and others) is crucial problem of facial features recognition. Recovering the position of these features can be done using our tracking algorithm. Affinely transformed positions of the the wavelets around the selected features determine features position in a new frame of the used video sequence. Recognizing the changes of those facial feature regions, we can realize a facial expressions recognizer.

5.1 Models of optical flow

First, an approach which uses parameterized models of optical flow (see [5]). A person's limbs, eyes and brows are represented as patches whose motion in an image sequence can be modeled by low-order polynomials. A robust optical flow estimation technique is used to recover the motion of these patches and recovered motion parameters provide a rich description, which can be used to recognize human gestures and facial expressions.

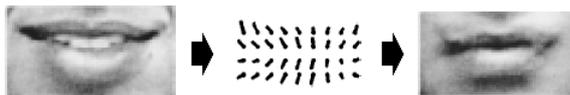


Fig. 6: Parameterized model of optical flow

The image motion of a planar patch of the face can be described by the parameters illustrated in Figure 7.

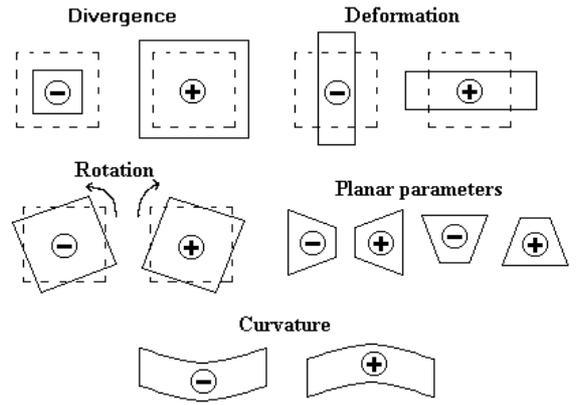


Fig. 7: Various parameters used to represent motion

For a faster estimation of model parameters, we can employ recovered GSW parameters as the initial values. Robust regression method based on the brightness constancy assumption (details in [4]) then converges quickly. Curves of parameters obtained through image sequence may be used as inputs for gesture recognizer. It can be based on Hidden Markov Models (HMMs) or Dynamic Time Warping (DTW).

5.2 Principal Component Analysis

One of the best-known approaches for object, in our context facial expressions, recognition is Principal Component Analysis (PCA).

Formally, PCA is defined as follows: Let $\{x_1, \dots, x_N\}$ be a set of n-dimensional gallery images. We want to define a linear, orthogonal mapping W from the n-dimensional space into an m-dimensional feature space, with $m < n$. Using W , a new feature vector y_k can be calculated for each image x_k :

$$y_k = W^T x_k.$$

The covariance matrix C is defined as

$$C = \sum_{k=1}^N (x_k - \mu)(x_k - \mu)^T,$$

where N is the number of gallery images and μ is their mean image. In PCA, the linear mapping W_{PCA} is chosen so that the covariance matrix of the feature vector is diagonal matrix and that the determinant of $W^T C W$ is maximized:

$$W_{PCA} = \arg \max |W^T C W| = [w_1, \dots, w_m].$$

The set $\{w_1, \dots, w_m\}$ is the set of n-dimensional eigenvectors of covariance matrix C that correspond to the m largest eigenvalues. The eigenvectors are often referred to as eigenpictures or eigenfaces.

It was said that PCA approximates texture only. The geometry of the object is not evaluated. The major drawbacks of PCA are its sensitivity to deformations and illumination changes. The eigentracking approach [3] builds on and extends the eigenspace representations, robust estimation techniques and

parameterized optical flow estimation. Here, an affine transformation between the eigenspace and the image is robustly estimated.

The PCA principle applied to the gallery images of separated facial features may be used for a facial expressions recognition. Also, an advantage can be taken from the known GSW transformation that can produce better initial values.

5.3 Recognition Using GWNs

In paper [1], the author proposed various aspects of automatic image recognition with Gabor wavelet networks. The idea behind the matching using GWNs is that a $GWN(\psi, w)$ which is optimized for a particular facial expression appears to be specific for that particular expression. Any other facial expression is not well represented by the same GWN.

In the first step, the gallery images of separated facial features are encoded. Each feature and expression is thus represented by a specific GWN. In order to recognize facial expression in a test image, the image needs to be encoded using each GWN in the gallery. If a GWN exists, which allows a good representation of the probe image, that GWN identifies the expression. In order to represent the test image using the GWN, we need to reposition GWN in the image and then directly calculate new optimal weights. The new and original weights determine the difference (e.g. Euclidean distance can be used) between the expression in the test image and the original facial expression. The nearest gallery GWN is then identified as the facial expression in the test image.

6 Experimental results

The presented face tracking algorithm was tested for different video sequences. A common USB web camera with resolution 320x240 pixels was used.

The obtained results confirmed the robustness of the method in respect of the affine deformations and homogenous illumination changes. The number of required computational resources (speed) highly depends on the number of the used wavelets where lower number increases generality of the face template. Figure 8 illustrates two frames of a test sequence during tracking, while the black bounding box represents the repositioned Gabor wavelet network.

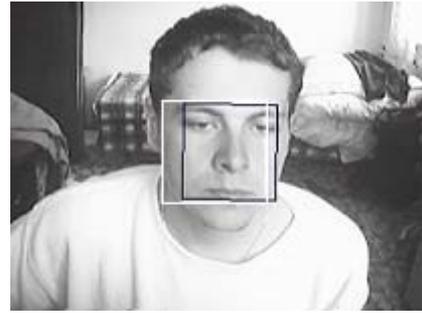
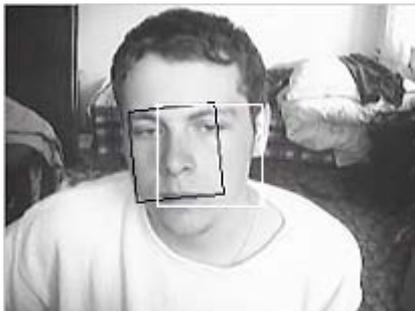


Fig. 8: Example of face tracking

The results were compared with other tracking method, the parameterized models of optical flow described in the previous sections. The results show that the GWN approach is more computationally expensive, but much more precise.

7 Conclusions and Future Work

The objective of this paper is presentation of the new method for face representation and tracking with use of the Gabor wavelet networks. The method is based on continuous wavelet representation of a discrete face template. This representation may be face specific or generic, depending on the number of used wavelets. The face specific representation can be used e.g. for face recognition while the generic representation can be used e.g. for facial expression detection.

The face tracking algorithm was implemented by repositioning the Gabor superwavelet for each frame. This approach appears to be quite perspective, insensitive to certain deformations and iconic changes in face region (smile, eye blinking), insensitive to homogenous illumination changes, and also generally very robust.

It is worth mentioning that the GWN technique has recently been used to not only face tracking, but also face recognition and face position estimation.

The previously proposed gesture recognition approaches are not yet fully implemented and the implementation work is in progress. Thus, in the future work, we intend to combine the GWN approach together with others approaches (such as PCA, parameterized models of optical flow, and/or skin-color models) to achieve robust face tracking and facial expressions recognizer.

Acknowledgements

I am grateful to my supervisor Doc. Dr. Ing. Pavel Zemcik for revision of this paper. The work is sponsored by „Multi Modal Meeting Manager“, EU-HLT, IST-2001-34485 grant project.

References

- [1] V. Krüger: *Gabor Wavelet Networks for Object Representation*. Dissertation, Christian-Albrechts-Universität, Kiel, 2001.
- [2] R. S. Feris, R. M. Cesar Junior: *Tracking Facial Features Using Gabor Wavelet Networks*. Brasil, 2001.
- [3] M. J. Black, A. D. Jepson: *EigenTracking: Robust Matching and tracking of Articulated Objects Using a View-Based Representation*. European Conference of Computer Vision, ECCV'96, Cambridge, England, April 1996.
- [4] M. J. Black, P. Anandan: *The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields*. Academic Press, California, 1996.
- [5] M. J. Black, et al.: *Recognizing human motion using parameterized models of optical flow*. USA, 1996.
- [6] K. Toyama, G. Hager: *Incremental Focus of Attention for Robust Vision-Based Tracking*. Yale University, 1998.
- [7] H. Rowley, S. Baluja, T. Kanade: *Rotation invariant neural network-based face detector*. In Proc. IEEE Conf. On Computer Vision and Pattern Recognition, Santa Barbara, 1998.
- [8] L. Finshi: *An Implementation of the Levenberg-Marquardt Algorithm*. Eidgenössische Technische Hochschule, Zürich, April 1996.