in: Proc. Workshop on Virtual Intelligence - Dynamic Neural Networks, VI-DYNN'98, pp. 476-485

3D Neural Fields and Steerable Filters for Contour-based Person Localization

Andrea Corradini, Ulf-Dietrich Braumann, Hans-Joachim Boehme and Horst-Michael Gross

Department of Neuroinformatics, Technical University of Ilmenau, D-98684 Ilmenau, Germany

ABSTRACT

This paper introduces a way to locate persons in visual images of cluttered scenes using a *shape-of-contour approach.* The contour which we refer to is that of the upper body of frontally aligned persons.

As we know from the Gestalt psychologists,[1] perception utilizes mechanisms to combine discrete sensations in order to assess their spatial and/or temporal relationships (grouping mechanisms). From an abstract viewpoint, Gestalt effects appear if some quality criteria based on local features are fulfilled. For those features we restrict to oriented edge elements extracted by means of oriented filters modeling oriented receptive fields.[2–4] What is important is the correct integration of pieces of edges in order to obtain separate objects.

The most important Gestalt laws related with this approach are *good continuation* and *symmetry* both describing effects which necessitate grouping mechanisms.[5,1] Since a shape or a silhouette might be partially concealed or occluded, the detection is just a variant of contour completion.

After deriving an approximation of it using a set of example images we take a *spatial arrangement of steerable filters* to determine the pointwise orientation along the contour.

However, the application of the filter arrangement typically yields a coarse distributed outcome. To select the most promising location we apply a dynamic pattern formation within a *three-dimensional dynamic neural field* to get the location even considering the distance of a person. It turned out that by means of simple homogeneous internal interaction rules the dynamic neural field can find robust localization solutions. The activity of the field-neurons can be considered as internal state enabling a permanent localization helpful for tracking the person.

Keywords: Computer Vision, 3D Dynamic Neural Field, Winner-Take-All, Steerable Filter, Person Localization

1. INTRODUCTION

The present work deals with the visual detection and localization of persons in the context of gesture-based human-robot interaction. One should consider that *detection* does not mean to *recognize* just a certain pattern of color, brightness or intensity, since persons obviously can be worn in huge variety. Further we think, on those scales interesting for the localization task the contour shape represents a really high invariant, especially against the background of real-world scenes. Therefore, we refer to a person based on a typical shape of contour. Our simple contour shape prototype model consists of an arrangement of oriented filters (e. g. "Gabor-shaped"[6] or steerable ones[7]) doing a piecewise approximation of the upper shape (head, shoulder) of a frontally aligned person. The arrangement itself was formed based on a set of training images. Applying such filter arrangement in a multi-resolution manner,[8] this leads to a robust localization of frontally aligned persons even in depth.

The central problem of selecting the most promising (salient) image region is treated by means of a *three-dimensional dynamic neural field* performing a winner-take-all (WTA) process (blob-like pattern formation[9]). Because of the three-dimensional nature of the input (several pyramidal levels) a three-dimensional dynamic neural field is required. To our knowledge, this type of field is novel. The main advantage of dynamic neural fields is the use of simple homogeneous interaction rules leading to an implicit solution which occurs as a local blob of active neurons as equilibrium state of the field. From the mathematical point of view the field can be compared to a recurrent nonlinear dynamic system. Therein, the activity of the neurons can be considered as internal states providing a *permanent localization* helpful for tracking.

Send correspondence to `{andreac,ulf}@informatik.tu-ilmenau.de` or fax to ++49 3677 69 1665, or call ++49 3677 69 1302/1305. More information can be obtained at `http://cortex.informatik.tu-ilmenau.de`

2. ARRANGEMENTS OF STEERABLE FILTERS

2.1. Motivation and Related Work

Our idea refers just to a description of the outer shape of head and shoulders, whereas the interesting and independently developed approach of OREN and PAPAGEORGIOU[10] considers the complete body of persons (pedestrians) using *Haar wavelet templates.* The common aspect between the two approaches is a set of locally distributed oriented filters used to determine the strength of certain orientations of visual "structure" for a small region. In our application it cannot be ensured that one can capture the whole body of a person so that our model refers to just the head-shoulder-region.

Further, using steerable filters (see subsection 2.3) we can be more precise in determining local orientations while fortunately keeping the computational load quite low. Since steerability means that one needs just a fixed basis set of filters for convolutions followed by an analytical maximization to determine the angle of the most dominant orientation, the application of steerable filters can be considered really elegant for our purposes.

The idea of taking a set of oriented filters for orientation determination is based on some physiological knowledge on the mammalian visual system. The light stimuli entering the eye and striking the receptor cells in the retina are first converted into electrical signals and then, after some processing stages, sent to the visual cortex. Here, at the back of the brain, the incoming information is processed by cells having orientation selectivity doing local decompositions of the visual information with respect to the frequency space. These so-called *simple cells* are orientation sensitive showing a maximum response to a pattern of a given orientation and a less one to patterns of some other orientation. YOUNG[4] developed a mathematical model of simple cell's receptive fields based on Gaussian Derivative (GD) functions. JONES and PALMER[11] proposed another model to fit the spatial shapes of cortical simple cells visual receptive fields based on Gabor functions. Although we can find in the literature other approaches[2] there are some reasons for preferring GDs rather than others. These non-GD models have some drawbacks compared to the GD ones. They require more parameters, their basis functions are not orthogonal and not separable in space and frequency. In addition, the Gabor functions get closer and closer to the GD functions of order n as n approaches infinity.

As we know from the Gestalt psychologists,[1] perception utilizes mechanisms to combine discrete sensations in order to assess their spatial and/or temporal relationships. Such effects are referred to as grouping mechanisms. The perception of form and Gestalt is an elementary facility of the visual system. The underlying mechanisms are assumed to work at rather early stages of visual processing using specialized visual pathways. From an abstract viewpoint, Gestalt effects appear if some quality criteria based on local features are fulfilled. For those features, we restrict to oriented edge elements extracted by means of oriented filters modeling oriented receptive fields. What is important is the correct integration of pieces of edges in order to obtain separate objects from them.

For our application we conceptualized the approach of an *arrangement* of spatially distributed oriented filters to describe persons as visual objects by means of the outer shape of head and shoulders. The most important Gestalt laws related with this approach are *good continuation* and *symmetry* which both obviously describe effects which necessitate grouping mechanisms.[5] Since this shape might be partially concealed or occluded for some reasons, the detection task is just a variant of contour completion.

The problem of detecting and locating persons is just a binary problem and no selection between alternatives. Therefore we do not need e. g. some deformable contour models to adapt to different contour shapes in order to find the most certain but can restrict to one static contour. However, the matching process with the filter arrangement is controlled by means of a *dynamic selection process* based on mutually interacting models of dynamic neuron. The process can find *that* position where the most heaped responses occur and considers this to be *the most promising position of a complete contour.*

However, each section of the contour should be approximated by a special oriented filter. Thus, searching a person would require possibly as much *differently oriented* filters as filters belong to the arrangement, which is computationally very costly.

2.2. Determining the Course of Contour

In our previous work[12] we use an arbitrary model of the contour based on a manually design restricting to just four filter orientations. Obviously, one would have a more precise model using more than four orientations, i. e. the contour model should be closely related to *real data.*

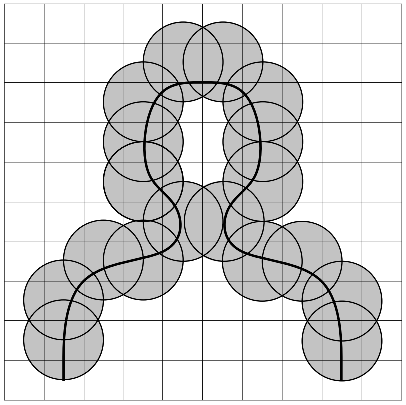

Figure 1. A Schematic sketch of the arrangement of steerable filters.

Steerable filters have the nice property that an initially limited number of convolutions is sufficient to derive any orientation information within an image. Thus, their use provides an extended set of orientations, avoids lots of additional filters and enables a more accurate computation of the course of contour.

Our complete data set consists of images showing ten persons in front of a homogeneous background under three different viewing angles ($0°, +10°$ and $-10°$, where $0°$ correspond to an exactly frontal aligned face). Additionally, in order to have a symmetrical response the whole data set was vertically mirrored extending the data set to 60 images. Following, the 256×256-images (grayscale) were lowpass-filtered and downscaled resulting in the size 16×16. Then, we applied a Sobel operator* to the images enhancing the edges of each image. This operator (similarly to others,[13] e. g. Robinson-, Kirsch-, Prewitt-operators) approximates the first derivative. In the case of a 3×3 convolution mask, the gradient is estimated in eight possible directions, and the convolution result of highest magnitude indicates the gradient direction. Next, all of those edge-marked intermediate images are averaged, since the contour to determine should *on average* match the real outer contour. To find that edge representing the typical contour shape we thresholded. High threshold values give gaps within the contour, whereas low ones yield too broad contours†.

Now, we have the course of the contour of interest resulting in a 16×16 binary matrix where the elements along the contour are set to 1, the others remain 0. We refer to this contour matrix as $\mathbf{\Lambda}$. What is further of interest is the local orientation of each contour element. It is achieved by means of the steerable filters (see below) applied to the binary contour shape so that for each element of $\mathbf{\Lambda}$ an angle of orientation can be determined.

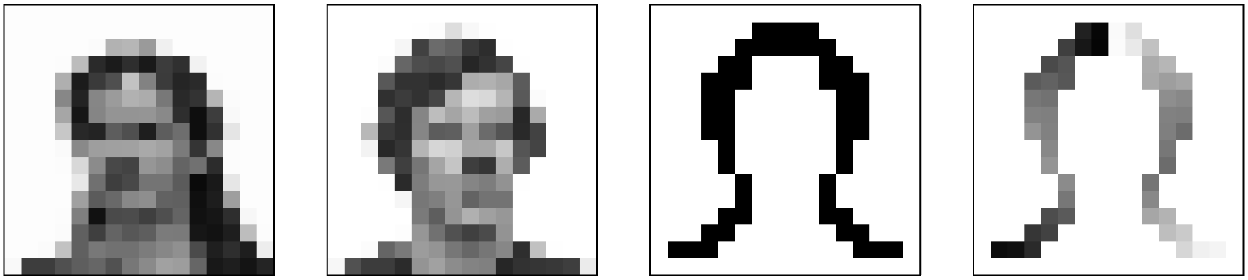

Figure 2. From left (a) to right (d): Two example images (16×16) out of the data set; determined binary shape of contour supp $(\mathbf{\Lambda})$; orientation angles coded with gray values $\mathbf{\Lambda}$ (0°: black; 90°: medium gray; 180°: white). Note that around the parting transitions 180° to 0° occur. Since the original data set was mirrored the contour model is symmetric.

*For simplicity reasons we did not use the true definition of the non-linear Sobel operator but a faster linearized approximation made up as linear superimposition.

†In our work we chose a threshold of 17%.

2.3. Applying Steerable Filters

The previous task provides a binary image representing an averaged head-shoulder-portrait but give no information about the local orientation at a given contour point. After determining the contour, we measure the local orientation by means of the application of a set of filters which are oriented in every direction. This again could be done, e. g. using the conventional *Gabor-type* filters but it requires the choice of certain directional (pair of) filters each of them differing from the others by a certain small rotation. In this case each filter pair corresponds to that angle the filter is tuned to. It also means that the orientation at a point of the contour is provided by the filter pair which has maximal response in this point. Unfortunately, by such an approach there is a trade-off between the required exactness of the value of the angle and the number of filters. The more exactly the measure of the orientation has to be, the more filters (e.g. certain orientation) we have to choose. In this paper we consider a different approach using the *steerable filters*[7] for orientation estimation. This approach provides an efficient filtering output by applying a few *basis filters* corresponding to a few angles and then interpolating the basis filter responses in the desired direction. Steerable filters are computationally efficiency and do not suffer from the orientation selection problem.

The mathematical *directional derivative* operator of a two-dimensional function embodies the principle of steerability:

$$\begin{aligned} \frac{\partial}{\partial \vec{n}} : \boldsymbol{\mathcal{C}}^1(\boldsymbol{\Omega}, \mathbb{R}) &\longrightarrow \boldsymbol{\mathcal{C}}^0(\boldsymbol{\Omega}, \mathbb{R}) \\ f &\longrightarrow \nabla f \cdot \vec{n} \end{aligned} \tag{1}$$

Here $\boldsymbol{\mathcal{C}}^k(\boldsymbol{\Omega}, \mathbb{R})$ indicates the space of the k-fold differentiable functions defined from some interval $\boldsymbol{\Omega} \subseteq \mathbb{R}^2$ into $\mathbb{R}$ and $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$ is the gradient operator. This directional operator is always with respect to a unit vector $\vec{n} = (n_x, n_y)$ which unequivocally determines a direction. By means of simple trigonometrical considerations the unit vectors $\vec{n}$ can be written as function of a *unique* angle ϑ in the range $[0, 2\pi)$ as $\vec{n} = (\cos\vartheta, \sin\vartheta)$. The definition of the directional derivative operator in accordance with the above considerations yields:

$$\begin{aligned} \frac{\partial}{\partial \vec{n}} &= n_x \frac{\partial}{\partial x} + n_y \frac{\partial}{\partial y} \\ &= \cos\vartheta \frac{\partial}{\partial x} + \sin\vartheta \frac{\partial}{\partial y} \end{aligned} \tag{2}$$

The equation states that the directional derivative operator can be synthesized at any arbitrary orientation ϑ from a linear combination of the operator tuned to 0° and 90°[‡]. The operators $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ are called *basis set* while the functions $\sin\vartheta$ and $\cos\vartheta$ are referred to as the *interpolation functions*.

In general, a function $f(\cdot)$ is considered steerable if it satisfies the following:

- the basis set is made up of their M rotated copies $f^{\alpha_1}(\cdot) \ldots f^{\alpha_M}(\cdot)$ on any certain angles $\alpha_1 \ldots \alpha_M$
- a rotated copy $f^{\vartheta}(\cdot)$ of it on some angle ϑ can be obtained by superimposition of its basis set times interpolation functions $I^{(\alpha_j)}(\vartheta)$ by

$$f^{\vartheta}(\cdot) = \sum_{j=1}^{M} I^{(\alpha_j)}(\vartheta) f^{\alpha_j}(\cdot) \tag{3}$$

In our work we take a quadrature pair[§] by using the *second order* derivative of a Gaussian and an approximation of its Hilbert transform by a third-order polynomial times a Gaussian. We refer to the second derivative of a Gaussian and its Hilbert transform approximation respectively as G and H. To measure the orientation along the contour we use the phase independent squared sum of the output of the quadrature pair. This squared response as a function of the filter orientation at a point (x, y) represents a local *oriented energy*[14] and is computed as

$$E^{(x,y)}(\vartheta) = \left(G^{(x,y)}(\vartheta)\right)^2 + \left(H^{(x,y)}(\vartheta)\right)^2 \tag{4}$$

[‡] If $\vec{n} = (1, 0)$ or $\vec{n} = (0, 1)$ the directional derivative is exactly the partial derivative with respect to x or y, respectively.

[§] A pair of filters is in quadrature if they have the same frequency response but differ in phase by $\frac{\pi}{2}$.

Because of the symmetry of the functions $G^{(x,y)}(\vartheta)$ and $H^{(x,y)}(\vartheta)$, the energy at every pixel is periodic of period π. Moreover, from the steering theorems[7] it follows that a linear combination of $M_G = 3$ and $M_H = 4$ basis functions is sufficient to synthesize every version rotated to any angle ϑ of respectively G and H.

Because the property of steerability is not dependent on the selection of the basis functions being oriented at some certain directions we select the three α_j with $j = 1 \dots 3$ evenly spaced orientations 0°, 60° and 120° for G and the four α_j with $j = 4 \dots 7$ likewise evenly spaced orientations 0°, 45°, 90° and 135° for H. On account of the choice of the above directions, the Gaussian derivative part $G^{(x,y)}(\vartheta)$ of the filter pair tuned on some angle ϑ can be written as

$$\begin{aligned} G^{(x,y)}(\vartheta) &= I_G^{(0^\circ)}(\vartheta)G^{(x,y)(0^\circ)} + I_G^{(60^\circ)}(\vartheta)G^{(x,y)(60^\circ)} + \\ & \quad I_G^{(120^\circ)}(\vartheta)G^{(x,y)(120^\circ)} \end{aligned} \tag{5}$$

with interpolation functions[7]

$$I_G^{(\alpha_j)}(\vartheta) = \frac{1}{4}\left[2\cos(\vartheta - \alpha_j) + 2\cos\left(3(\vartheta - \alpha_j)\right)\right] \quad j = 1 \dots 3 \tag{6}$$

For the same reason its approximated Hilbert transform $H^{(x,y)}(\vartheta)$ can be written in the form

$$\begin{aligned} H^{(x,y)}(\vartheta) &= I_H^{(0^\circ)}(\vartheta)H^{(x,y)(0^\circ)} + I_H^{(45^\circ)}(\vartheta)H^{(x,y)(45^\circ)} + \\ & \quad I_H^{(90^\circ)}(\vartheta)H^{(x,y)(90^\circ)} + I_H^{(135^\circ)}(\vartheta)H^{(x,y)(135^\circ)} \end{aligned} \tag{7}$$

with corresponding interpolation functions[7]

$$I_H^{(\alpha_j)}(\vartheta) = \frac{1}{5}\left[1 + 2\cos\left(2(\vartheta - \alpha_j)\right) + 2\cos\left(4(\vartheta - \alpha_j)\right)\right] \quad j = 4 \dots 7 \tag{8}$$

The whole quadrature pair can also be expressed with $M = M_G + M_H = 7$ basis filters.

From equations 5 and 7 and by means of trigonometric identities the equation 4 can be written as a Fourier series containing only the even frequencies:

$$E^{(x,y)}(\vartheta) = a_0 + \sum_{k=1}^{3} b_k \sin(2k\vartheta) + \sum_{k=1}^{3} c_k \cos(2k\vartheta) \tag{9}$$

We then use this expression for the oriented energy to accurately estimate the *dominant* local orientation by pointwise maximizing the oriented energy. The simplest way to solve this maximization problem consists of calculating the energy for each angle within the set $[0, \pi)$ (with a certain angular step) and then taking

$$\vartheta_{MAX}^{(x,y)} = \arg\max\{E^{(x,y)}(\vartheta) \mid \vartheta \in [0, \pi)\} \tag{10}$$

The smaller the step is chosen, the finer is the solution. However, to find this maximum value we do not search degreewise because of the computational cost indeed we look for some analytical solution.

For complexity reasons we restrict to $E^{(x,y)}$ including only the first order term so that $E^{(x,y)}$ is simplified to:

$$E^{(x,y)}(\vartheta) \approx a_0 + b_1 \sin(2\vartheta) + c_1 \cos(2\vartheta) \tag{11}$$

Now we search the maximum of $E^{(x,y)}(\vartheta)$ among the zeroes of $\frac{\partial E^{(x,y)}(\vartheta)}{\partial \vartheta}$.
Defining

$$A = \frac{c_1}{\sqrt{b_1^2 + c_1^2}} \tag{12}$$

the analytical expression for $\vartheta_{MAX}^{(x,y)}$ is:

$$\vartheta_{MAX}^{(x,y)} = \arg\max\left\{E(\vartheta), \quad \forall\vartheta \quad | \quad \vartheta \in \left\{\frac{\pm\arccos(\pm A)}{2}\right\}\right\} \tag{13}$$

The parameters b_1 and c_1 are easy determined by equating equation 4 and 9. One yields

$$\begin{aligned}
b_1 &= \frac{4\sqrt{3}}{9}\left(G^{(60^\circ)}G^{(60^\circ)} + G^{(0^\circ)}G^{(60^\circ)} - G^{(0^\circ)}G^{(120^\circ)} - G^{(120^\circ)}G^{(120^\circ)}\right) \\
&\quad + \frac{\sqrt{2}}{4}\left(H^{(0^\circ)}H^{(45^\circ)} + H^{(90^\circ)}H^{(45^\circ)} + H^{(0^\circ)}H^{(135^\circ)} - H^{(90^\circ)}H^{(135^\circ)}\right) \\
&\quad - \frac{1}{2}\left(H^{(0^\circ)}H^{(90^\circ)}\right) + \frac{3}{4}\left(H^{(45^\circ)}H^{(45^\circ)} - H^{(135^\circ)}H^{(135^\circ)}\right) \qquad (14) \\
c_1 &= \frac{8}{9}\left(G^{(60^\circ)}G^{(120^\circ)} - G^{(0^\circ)}G^{(0^\circ)}\right) - \frac{1}{2}H^{(45^\circ)}H^{(135^\circ)} \\
&\quad + \frac{4}{9}\left(G^{(60^\circ)}G^{(60^\circ)} - G^{(60^\circ)}G^{(0^\circ)} + G^{(120^\circ)}G^{(120^\circ)} - G^{(120^\circ)}G^{(0^\circ)}\right) \\
&\quad + \frac{\sqrt{2}}{4}\left(H^{(45^\circ)}H^{(90^\circ)} - H^{(45^\circ)}H^{(0^\circ)} + H^{(135^\circ)}H^{(90^\circ)} + H^{(135^\circ)}H^{(0^\circ)}\right) \qquad (15)
\end{aligned}$$

The angular value $\vartheta_{MAX}^{(x,y)}$ is a measure related to certain coordinates (x, y). We further refer to the matrix of all these values as $\mathbf{\Theta}$.

Unlike a Gabor-type filter approach the processing scheme by steerable filters requires no additional convolution after the initial pass through the seven basis filters. Moreover, we choose these certain steerable filters because there exists a separable basis set in Cartesian coordinates which considerably lowers the computational costs.

3. DYNAMIC NEURAL FIELDS FOR LOCALIZATION

3.1. Neural Field Input: Responses of the Filter Arrangement

The previous section extensively describes the theory and use of steerable filters. By means of those filters we calculate both the matrix $\mathbf{\Lambda}$ describing a typical course of the head-shoulder-portrait and that matrix $\mathbf{\Theta}$ (corresponding to the image wherein a person is to be found) containing the dominant local orientation values.

At next, we are going to search for the presence of the kernel $\mathbf{\Lambda}$ within the matrix $\mathbf{\Theta}$. To do it, we utilize a matching technique based on a *similarity measure* $m^{(x,y)}$. That measure should fulfill the following conditions:

- $m^{(x,y)}$ is a superimposition as following:

$$m^{(x,y)} = \frac{\displaystyle\sum_{i=0}^{I-1}\sum_{\substack{j=0 \\ \lambda_{i,j}\neq 0}}^{J-1} \widetilde{m}\left(\lambda_{i,j} - \vartheta_{MAX}^{(x+i-\frac{I}{2},y+j-\frac{J}{2})}\right)}{\operatorname{card}\left(\operatorname{supp}\left(\mathbf{\Lambda}\right)\right)} \qquad (16)$$

  where in our work $I = J = 16$ are the dimensions of the matrix $\mathbf{\Lambda}$, $\lambda_{i,j}$ is the element of $\mathbf{\Lambda}$ at position (i, j) and $\vartheta_{MAX}^{(x+i-\frac{I}{2},y+j-\frac{J}{2})}$ is element of $\mathbf{\Theta}$ at position $(x + i - \frac{I}{2}, y + j - \frac{J}{2})$,

- the single function $\widetilde{m}(\gamma)$ describes the similarity between two elements of $\mathbf{\Lambda}$ and $\mathbf{\Theta}$; due to the π-periodicity of the outcome of the steerable filters it should have just the same periodicity,

- $\widetilde{m}(0) = \widetilde{m}(\pi) = \max\limits_{\gamma}\{\widetilde{m}(\gamma)\}$ and $\widetilde{m}\left(\frac{\pi}{2}\right) = \min\limits_{\gamma}\{\widetilde{m}(\gamma)\}$,

- $\widetilde{m}(\gamma)$ decreases for $\gamma \in \left[0, \frac{\pi}{2}\right)$ and increases for $\gamma \in \left[\frac{\pi}{2}, \pi\right)$

Therefore, it seems convenient to use a trigonometrical function as Cosine, so that

$$\widetilde{m}(\gamma) = \frac{\cos\left(2\left|\gamma\right|\right) + 1}{2} \qquad (17)$$

For simplicity reasons one might replace the cosine in equation 17 with a piecewise linear function:

$$\tilde{m}(\gamma) = \left|\frac{2}{\pi}\gamma - 1\right| \tag{18}$$

The normalization to the cardinality of the support[¶] of the matrix $\mathbf{\Lambda}$ in equation 16 ensures $m^{(x,y)} \in [0,1]$ which is needed for the further processing.

3.2. The 3D Nonlinear Dynamic Field

To achieve a good localization a selection mechanism is needed to make a definite choice. This is not limited to a two-dimensional position. Since we use five fine-to-coarse resolutions we actually can localize persons even in different distances. Therefore, a neural field for selection the most salient region should be three-dimensional. The advantages of using 3D dynamic neural fields are the following:

- the process leads to an implicit solution using *simple homogeneous interaction rules*
- the activity of the neurons considered as internal states provide a *permanent localization* (any-time) helpful for tracking

That field $\mathbf{F}$ can be described as recurrent nonlinear dynamic system. Regarding to the selection task we need a dynamic behavior which leads to *one* local region of active neurons successfully competing against the others, i. e. the formation of one single blob of active neurons as an equilibrium state of the field. The following equations describe the system:

$$\begin{aligned}
\tau \frac{d}{dt} z(\vec{r},t) &= -z(\vec{r},t) - \mathrm{c}_h h(t) + \mathrm{c}_i x(\vec{r},t) \\
&\quad + \mathrm{c}_l \int_{\mathbf{N}} w(\vec{r}-\vec{r}\,')y(\vec{r}\,',t)d^3\vec{r}\,' \quad \text{with} \qquad (19) \\
w(\vec{r}-\vec{r}\,') &= 2\exp(\frac{-3|\vec{r}-\vec{r}\,'|^2}{2\sigma^2}) - \exp(\frac{-|\vec{r}-\vec{r}\,'|^2}{\sigma^2}) \qquad (20) \\
y(\vec{r},t) &= \frac{1}{1+\exp(-z(\vec{r},t))} \qquad (21) \\
h(t) &= \int_{\mathbf{F}} y(\vec{r}\,'',t)d^3\vec{r}\,'' \qquad (22)
\end{aligned}$$

Herein $\vec{r} = (x, y, z)$ denotes the coordinate of a neuron; $z(\vec{r},t)$ is the activation of a neuron $\vec{r}$ at time t; $y(\vec{r},t)$ is the activity of this neuron; $x(\vec{r},t)$ denotes the external input (corresponds to the re-coded similarity measure $m^{\vec{r}}$, cf. equation 16 and see figure 4); $h(t)$ is the activity of a global inhibitory interneuron activated by each neuron over the entire field $\mathbf{F} \subseteq \mathbb{R}^3$; $w(\vec{r} - \vec{r}\,')$ denotes the function of lateral activation of neuron $\vec{r}$ from the surrounding neighbourhood $\mathbf{N} \subseteq \mathbb{R}^3$. Further, τ is the time constant of the dynamical system and σ is the deviance of the Gaussian determining the function of lateral activation. For the computation we used the following values for the constants: $c_h = 0.025$, $c_l = 0.1$, $c_i = 0.1$, $\sigma = 2$ (halved in z-direction), $\tau = 10$ with $\Delta T = 1$ (ΔT sampling rate). The range $\mathbf{I}$ of the function of lateral activation reachs over 5 pixels and 3 pixels in z-direction, respectively (anisotropic neighbourhood).

As illustrated in figure 4 to use a three-dimensional neural field one has to consider that local correspondences between the resolution levels. Therefore, we do a re-coding into a cuboid structure. One side effect is that the coarser a pyramid level is the less one can locate something by means of the similarity measure. However, without particularly treating this effect we just noticed that those levels z of the neural field activated from the rather coarse pyramid levels take little more steps to develop a blob (or a part of a blob, respectively).

[¶] The support of a matrix considers only nonzero elements.

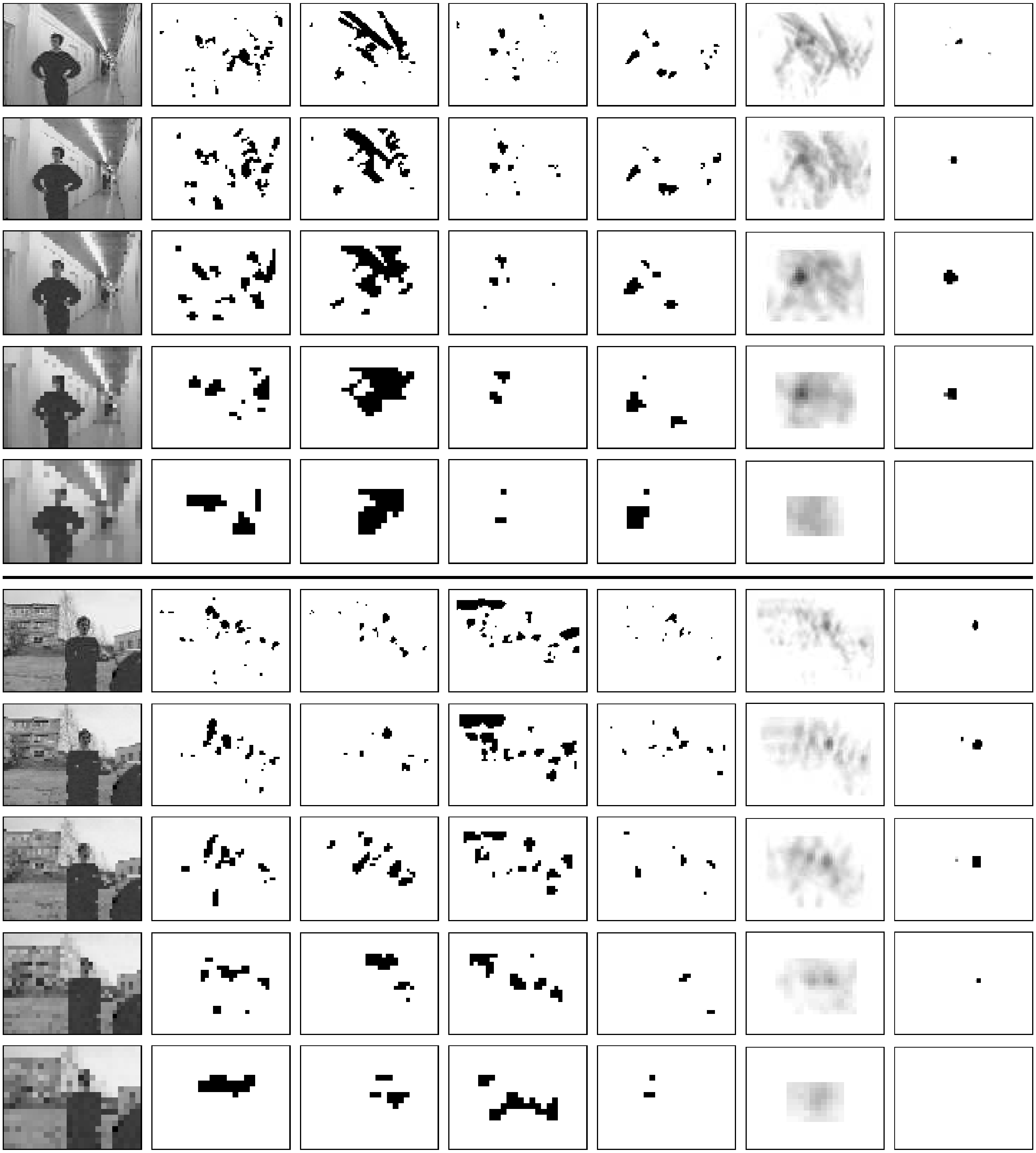

Figure 3. Localization results in an **indoor** (top) and **outdoor environment** (bottom): The localization of a person does not sharply occur at one of the pyramidal planes, the originating spatial blob is most strongly developed on the central of the five planes. Each row contains the results of one of the five computed resolutional levels, from top to bottom 96 × 71, 68 × 50, 48 × 36, 34 × 25 and 23 × 18 pixels. In the seven columns the following results are depicted: input, results of the orientation filtering (0°, 45°, 90° and 135°), the result of the filtering with the filter arrangement and eventually the result of the selection within a three-dimensional field of dynamic neurons.
Note that for clarity reasons of this presentation we have restricted here to just four different orientations.

3.3. Results

The results of the system are qualitatively illustrated in figure 3. Each row contains the results of one of the five (distance $1/\sqrt{2}$) computed resolution steps, from top to bottom 96×71, 68×50, 48×36, 34×25 and 23×18 pixels. The images of the rightmost column show the state of the five layers of the dynamic neural field in a snapshot at that moment when the activity change Δy of the most active neuron became less than 1%. From the second to the fifth column the results of the image filtering task with steerable filter along four certain directions (0°, 45°, 90° and 135°) are shown. The sixth column gives the results of the image filtering task with the arrangement of steerable filters. Finally the leftmost column shows the localization task result.

On average, the system takes less than 11 iteration steps. The range of the blob is not restricted to one plane. To get a more precise specification of the distance of a person one could interpolate the z-coordinate of the blob center.

As far as the system reachs an equilibrium state, another frame is taken and processed the same way. A blob in an equilibrium state does not necessarily indicate the presence of a person in the image, however supposed a person is present it will be easily found. If a localized person moves between two captured frames the correspondent blob will follow over the iteration steps reaching a new equilibrium state.

Our presented results are exemplary, the usage of the shape of contour provides one solution for the person localization problem, even under quite different conditions. The novel approach with a three-dimensional dynamic neural field can be assessed as robust method for the selection process.

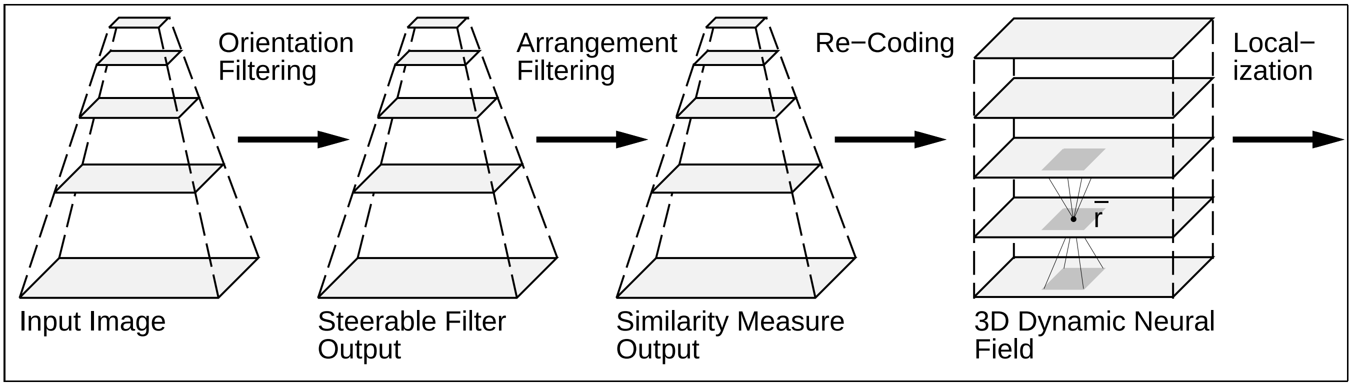

Figure 4. *Processing steps for person localization. Starting from a multi-resolution representation of the image, each level is treated by steerable filters. Applying the filter arrangement we determine a distance measure which is taken as input of a three-dimensional field of dynamic neurons. The resulting blob (locally delimited pattern of active neurons) is used to localize a person.*
The dark marked segments depict the lateral activation region for one (black marked) dynamic neuron covering lateral neurons even from neighboured levels. For the used activation function see equation 20.

4. FUTURE WORK

Once a person is localized, we can precisely analyze the person in detail, e. g. concerning gestures. Gestures are static or flexible postures of hands, arms or the body of a person usually used to convey information from one human being to another. In the context of human-computer interfaces gestures can be used to interact with a computer system. In a framework of an image-based gesture recognition system on board of our mobile robot platform MILVA[‖] the *localization of a user's head* has essential importance, since it prerequisites for any further gesture-related analyses (e. g. distance and angle between hands and/or head).

ACKNOWLEDGMENTS

Andrea Corradini is supported by the EU Fellowship: TMR Marie Curie Research Training Grant # ERB FMBI CT 97 2613. He thanks the Provincia Autonoma di Trento (Italy) for previous partial support, as well as Giuliana Curti for her help.

[‖] **M**ultisensory **I**ntel**l**igent adaptive **V**ehicle in neural **A**rchitecture

All other authors are in part supported by the Thuringian Ministry of Science, Research and Culture (TMWFK): GESTIK-Project.

The authors especially thank Anja Brakensiek and Markus Krabbes for helpful discussions and comments on this work.

REFERENCES

1. Koffka, Kurt, *Principles of the Gestalt Psychology*, Harcourt, Brace & World, New York, 1935.
2. Koenderink, Jan J. and Doorn, Andrea J. van, "Receptive Field Families," *Biological Cybernetics – Communication and Control in Organisms and Automata* **63**, pp. 291–297, 1990.
3. Valois, Russell L. de and Valois, Karen K. de, *Spatial Vision*, no. 14 in Oxford Psychology Series, Oxford University Press, 1988.
4. Young, Richard A., "Oh Say, Can You See? The Physiology Of Vision," Tech. Rep. GMR-7364, General Motors Research Laboratories, Computer Science Department, Warren, MI, May 1991.
5. Bruce, Vicki and Green, Patrick R., *Visual Perception: Physiology, Psychology and Ecology*, Lawrence Erlbaum Associates, London, 2 ed., 1993.
6. Gábor, Dénes, "Theory of Communication," *Journal of the Institute of Electrical Engineers (IEE, London) – Part 3: Radio and Communication Engineering* **93**, pp. 429–457, 1946.
7. Freeman, William T. and Adelson, Edward H., "The Design and Use of Steerable Filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* **13**, pp. 891–906, Sept. 1991.
8. Jolion, Jean-Michel and Rosenfeld, Azriel, *A Pyramid Framework for Early Vision: Multiresolutional Computer Vision*, Kluwer Academic Publishers, 1994.
9. Amari, Shun-ichi, "Dynamics of Pattern Formation in Lateral-Inhibition Type Neural Fields," *Biological Cybernetics – Communication and Control in Organisms and Automata* **27**, pp. 77–87, 1977.
10. Oren, Michael, Papageorgiou, Constantine, Sinha, Pawan, Osuna, Edgar, and Poggio, Tomaso, "Pedestrian Detection Using Wavelet Templates," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97)*, pp. 193–199, IEEE Computer Society, Nov. 1997.
11. Jones, Judson P. and Palmer, Larry A., "An Evaluation of the Two-Dimensional Gabor Filter Model of Simple Receptive Fields in Cat Striate Cortex," *Journal of Neurophysiology* **58**(6), pp. 1233–1258, 1987.
12. Corradini, Andrea, Braumann, Ulf-Dietrich, Brakensiek, Anja, Krabbes, Markus, Boehme, Hans-Joachim, and Gross, Horst-Michael, "Visual Gesture Localization with Dynamic Neural Fields: Towards a Gesture Recognition System," in *Proceedings of the 10-th Italian Workshop on Neural Networks (WIRN'98), Vietri Sul Mare*, May 1998. To appear in November 1998.
13. Sonka, Milan, Hlavac, Vaclav, and Boyle, Roger, *Image Processing, Analysis and Machine Vision*, Chapman & Hall Computing Series, Chapman & Hall, 1993.
14. Knutsson, Hans and Granlund, Gösta H., "Texture Analysis Using Two-Dimensional Quadrature Filters," in *Proceedings of the IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management*, pp. 206–213, IEEE Computer Society, 1983.