amcs

# DETECTION OF MOVING OBJECTS IN IMAGE SEQUENCES USING 3D VELOCITY FILTERS

SAM SCHAULAND, JOERG VELTEN, ANTON KUMMERT

Faculty of Electrical, Information and Media Engineering
University of Wuppertal, 42097 Wuppertal, Germany
e-mail: {schauland, velten, kummert}@uni-wuppertal.de

A movement analysis of objects contained in visual scenes can be performed by means of linear multidimensional filters, which have already been analyzed in the past. While the soundness of the results was convincing, interest in those systems declined due to the limited computational power of contemporary computers. Recent advances in design and implementation of integrated circuits and hardware architectures allow realizing velocity filters if the $n$-D system is carefully adapted to the analyzed problem. In this paper, the fundamental principles of visual scene analysis by linear multidimensional filters are examined with respect to possible sources of degradation. The extraction of movement information and its practical use are demonstrated using a wave digital filter (WDF) implementation.

**Keywords:** Multidimensional signal processing, velocity filters, motion filters, object detection, wave digital filters.

## 1. Introduction

In the field of visual scene analysis, object detection and segmentation are among the most challenging research topics. As the range of possible applications is wide (from robotics to operational safety to automotive applications), numerous application-specific challenges occur, such as

- cluttered background textures,

- abrupt illumination changes,

- high intra-class variability of the target object class,

- unknown object size and shape.

They complicate the design of algorithms capable of accurate detection and classification. Especially, this is the case if target objects are complex, like pedestrians or vehicles.

A popular and effective way to overcome these difficulties is using various different kinds of active sensors. (Gandhi and Trivedi (2006) give an overview on the state of the art in pedestrian collision warning systems.) Although many systems presented in the past using active sensors produce promising results, they often reveal drawbacks including low data acquisition speed, low resolution, and high costs. Therefore, optimized exploitation of the extensive environmental information offered by

vision-based systems will be a long term goal in research on collision warning systems.

A typical vision-based object detection and tracking system consists of three major steps: region of interest (ROI) detection, object detection and object tracking. In the first step, the location and size of object candidates are estimated. Afterwards, features are extracted and a classifier is used to determine if a candidate is a member of the target object class. The final step tracks the recognized objects and refines the recognition results.

The aim of the work presented in this paper is not to fully replace the existing approaches. In fact, we propose a method to improve the available methods and/or make them more efficient. Instead of using commonly applied image processing tools and features, we interpret the image sequence recorded by a camera as a three-dimensional signal and use methods known from signal processing applications to detect moving objects in a scene. The advanced theoretical knowledge on system description and handling (using state-space methods like the Givone-Roesser model (Lu and Antoniou, 1992)) available for these filters is a big advantage compared with the most recently presented approaches.

In detail, the image is filtered by 3D linear shift invariant (LSI, the multidimensional extension of linear time invariant (LTI) filters known from the one-dimensional

case) filters targeting the separation of signals due to different velocity or directional components (Runze, 2005). These velocity filters are successfully applied and well known in the domain of geology, where they are used to understand and forecast seismic wave migration, cf. (Dudgeon and Mersereau, 1984). Linear $n$-D filters of low order have also been successfully used to extract velocity information from visual scenes (Bruton and Bartley, 1985), and a framework for the treatment and development of fan filters for image sequence analysis by means of a 3-D Radon transform was presented in (Marzetta, 1994). Until that time, a crucial point for the realization of these filters had been the missing available hardware capacity (Marzetta, 1994): computing and analyzing visual scenes of reasonable resolutions in temporal and spatial directions require high-end hardware to process $n$-D filters in real time. Advances in the design and implementation of integrated circuits and hardware architectures will allow realizing of $n$-D linear filters of higher order in the near future.

The remainder of this paper is structured as follows: In the following section, theoretical deliberations on motion in image sequences as well as an introduction to velocity filters are presented. Section 3 is focused on system design, i.e. the design of transfer functions and the relevant implementation. Afterwards, we provide experimental results followed by a summary of the proposed approach and related research directions in the future.

## 2. Theoretical background

This section introduces theoretical background on the representation of moving objects in the spatio-temporal and spectral domains as well as the basics on the discrimination of moving objects by speed and/or direction.

### 2.1. Two-dimensional objects moving in a three-dimensional space.
Choosing an adequate signal representation is strongly connected to the developed application: while in the case of geophysical, SONAR or RADAR applications, for example, measured signals are often interpreted as planar waves, moving objects in an image sequence can be represented by signals constant on straight lines in the spatio-temporal domain. Bartley and Bruton (1986; 1987) refer to the latter class as "linear trajectory signals" ( for a more detailed comparison of the two representations, see (Bolle, 1992)).

A descriptive introduction to the Fourier description of multidimensional visual scene representations is subsequently given by the investigation of a temporally moving 1-D object. Let $f(x)$ be a spatially limited 1-D signal describing the object, i.e. $f(x) \neq 0$ for $a < x < b$, with $a, b, x \in \mathbb{R}$, and let $F(jk_x)$ be its respective Fourier transform. A linear movement of this object in the positive direction of $x$ with constant speed $v$ is characterized by

the 2-D signal

$$s_c(t,x) = f(x - vt). \tag{1}$$

The Fourier transform of $s_c(x,t)$ with respect to the spatial variable $x$ can be evaluated by the Fourier shift theorem as

$$\mathcal{F}_x\{s_c(t,x)\} = F(jk_x) \cdot e^{-jvtk_x}, \tag{2}$$

where $\mathcal{F}_x$ denotes the Fourier transform with respect to the variable $x$.

The Fourier transform with respect to the temporal variable $t$ is evaluated by using the modulation theorem, which leads to the 2-D Fourier transform of $s_c(t,x)$ :

$$\begin{aligned} S_c(j\omega_t, jk_x) &= \mathcal{F}_t\{F(jk_x) \cdot e^{-jvk_xt}\} \\ &= F(jk_x) \cdot 2\pi \cdot \delta(\omega_t + vk_x). \end{aligned} \tag{3}$$

The defined movement can thus be decomposed into complex exponential waves for which the relation between $k_x$ and $\omega_t$ is fixed, namely, $k_x = -\omega_t/v$, i.e. the respective Fourier components are located on a plane including the origin. The latter is clarified by Fig. 1. On the left-hand side it displays the locus of $f(\boldsymbol{r})$ and on the right-hand side the respective non-zero points of $S_c(j\omega_t, jk_x)$.

This description is easily generalized to the $n$-D case by replacing scalars by vectors. The object is now an $n$-D object $f(\mathbf{r}), \mathbf{r} \in \mathbb{R}^n$ with its $n$ - D Fourier transform $F(j\mathbf{k})$, where $j\mathbf{k} = (jk_x, jk_y, jk_z)'$ in the 3-D case. Here $\mathbf{v} = (v_x, v_y, v_z)'$ contains the velocity information corresponding to each component of $\mathbf{r}$. The Fourier transform of the $(n+1)$-D signal

$$s_c(t, \boldsymbol{r}) = f(\mathbf{r} - \mathbf{v}t) \tag{4}$$

is given by

$$\mathcal{F}\{f(\mathbf{r} - \mathbf{v}t)\} = F(j\mathbf{k}) \cdot 2\pi \cdot \delta(\omega_t + \mathbf{v}'\mathbf{k}). \tag{5}$$

If the object defined by (1) is transferred into a higher dimensional space, e.g.

$$s_c(t, \mathbf{r}) = f(x - vt), \tag{6}$$

the resulting wave is a combination of planar waves with propagation speed $v$ and an identical propagation direction. The shape of the wave in this direction is defined by Fourier components along the respective plane, i.e. the respective composition of plane waves. While in (6) this wave propagates in the direction of the first space coordinate, combinations of plane waves propagating in the direction of the normed vector $\mathbf{r}_0$ can be described by

$$s_c(t, \mathbf{r}) = f(\mathbf{r_0'}\mathbf{r} - vt), \tag{7}$$

where $\mathbf{v} = \mathbf{r}_0 \cdot |\mathbf{v}| = \mathbf{r}_0 \cdot v$. In the $(n+1)$-D Fourier description, such waves are thus defined by spectral components that are located on a plane including the origin.

The angle between the $\omega_t$-axis and the normal vector of this plane determines the propagation speed of the wave.

Plane waves that move at constant speed $|\mathbf{v}|$ into any direction in the $n$-D space can be detected by an analysis of its $(n+1)$-D spatio-temporal Fourier spectrum. For each temporal frequency $\omega_t$, Fourier components of this type of waves are located on the surface of an $n$-D hypersphere with radius $\omega_t/|\mathbf{v}|$. For the 3-D case of image sequence processing (two spatial and one time coordinate), the 3-D Fourier components of such waves are located on the surface of a right circular double cone, centred at the origin. Directional movements of constant speed are defined by their speed vector configuration, i.e. the particular value for each component.

*Example 1.* The following example shows a typical application for a motion-based object detection system.
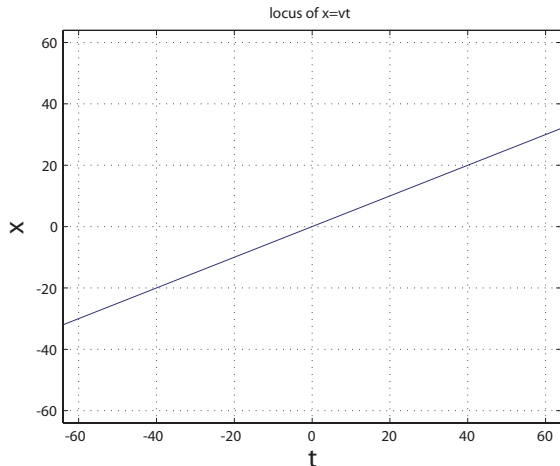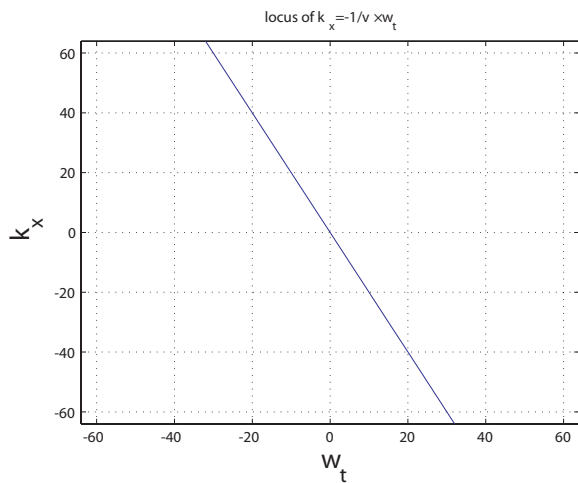


Fig. 1. Locus of $f(\boldsymbol{r})$ on the left- and right-hand sides of the respective non-zero points of $S_c(j\omega_t, jk_x)$.
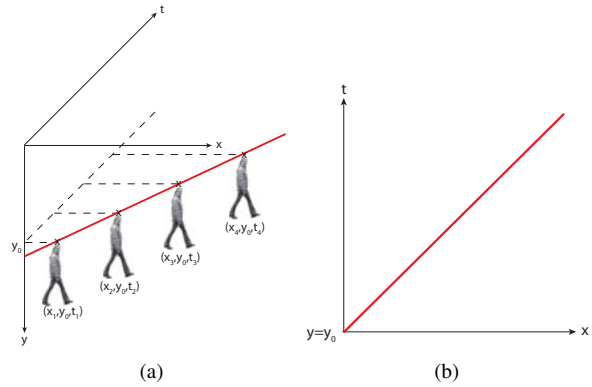


Fig. 2. Example of an object linearly moving from the left to right with constant $y = y_0$. The position of the pedestrian's head point in 2(b) over time describes a straight line in the $xt$-plane. Fig. 2(b) depicts the "top view" of the corresponding motion path.

Figure 2(a) shows a pedestrian crossing the street in front of a camera at four different points in time. As there is no vertical movement (i.e. $y = y_0$), the pedestrian, represented by his//her head point, moves on a straight line in the $x-t$ - plane. Figure 2(b) shows a top-view of this path. Note that the targeted object moves on a straight line, i.e. there is no acceleration with respect to the image coordinate system. Note that in the given example the third dimension of the world coordinate system (i.e. the distance between the camera host and the pedestrian) is also constant ($z = z_0$).

According to (3), under these constraints the respective Fourier components are located on a plane perpendicular to the direction of the motion path.

**2.2. Speed information.** Commonly, a visual scene is described by an $(n+1)$-D spatio-temporal signal (3-D or 4-D), and an object in this scene is described by an $n$-D spatial signal (2-D or 3-D, respectively). If the scene contains only one object that moves at constant speed into a fixed direction, then non-zero Fourier components of the spatio-temporal $(n+1)$-D signal are generally located on an $n$-D hyperplane in the $(n+1)$-D space that includes the origin (see (5)). This hyperplane in the Fourier domain, can be decomposed into lines that also include the origin, where each of these lines represents a combination of planar waves of identical propagation speed and direction. The latter underlines the fact that $(n+1)$-D spatio-temporal signals can be decomposed into planar waves.

An $n$-D object that moves according to (7) has planar wave components, for which the propagation speed is bounded by $|\mathbf{v}| = |\mathbf{r}_0 v| = v$. This maximal propagation speed exists for planar waves that propagate into the movement direction $\mathbf{r}_0$ of the object. For all other planar waves composing this movement, the propagation speed

is lower than this maximum. Figure 3 shows a plane and a right double cone in the 3-D Fourier domain as mentioned in Section 2.1. The plane denotes non-zero points
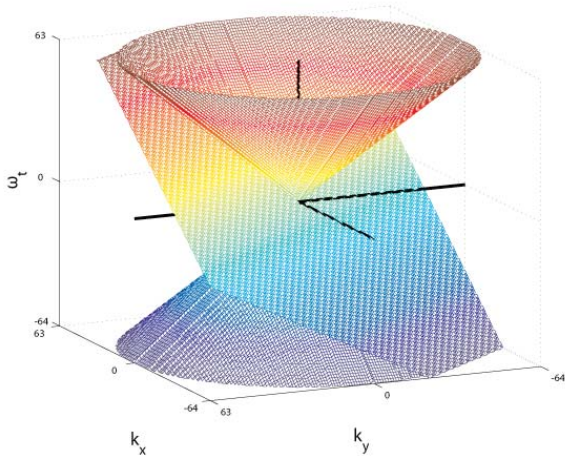


Fig. 3. Plane and a right double cone in the 3-D Fourier domain. The plane denotes non-zero points of a linear movement at constant speed of a 2-D object. The double cone denotes all planar waves with the same propagation speed as the moving object, but in an undefined direction.

of a linear movement of constant speed of a 2-D object. The double cone denotes all planar waves with the same propagation speed as the moving object, but in an undefined direction. Their intersection denotes the planar waves that propagate into the same direction as the moving object. Fourier components on this intersection line describe Fourier components of the object with respect to the propagation direction (a slice through the object spectrum). Perpendicularly to this direction, the object is described by all other planar waves contained in the plane.

This shows that a proper discrimination by speed, independently of its movement direction, is not possible for the case of general objects (i.e. objects with an infinitely extended Fourier spectrum). Only if the object representation has a narrow spectrum perpendicular to the propagation direction $\mathbf{r_0}$, the discrimination of objects by speed is possible. In the latter case, all bands of the Fourier representation that may contain object information can be selected. Components of objects of higher speed may also be contained. For the spatio-temporal 3-D case mentioned in Section 2.1, the surface of the right circular double cone would have to be enlarged by a certain surrounding, where the width of the surrounding depends on the bandwidth of the moving object. This area in the Fourier domain contains all objects of maximum bandwidth $\Delta\omega$ moving at speed $v$, but also components of objects of higher speed.

**2.3. Directional information.** Directional filters that allow discriminating movement directions independently

of speed need passbands that include all frequency components between the hyperplanes for the lowest and highest possible object speeds. Unfortunately, two objects moving at constant speeds only have Fourier descriptions by nearly disjoint sets of coefficients if the objects move into opposed directions. The latter is clarified in the following. Figure 4 shows a 3-D plane that denotes non-zero points of the Fourier description of an object that moves at constant speed into the positive $x$-direction.
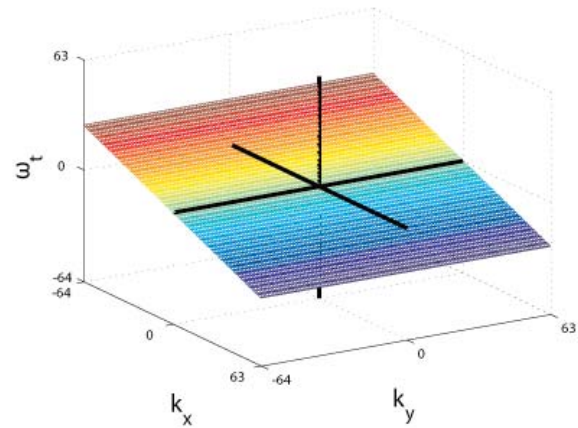


Fig. 4. 3-D plane that denotes non-zero points of the Fourier description of an object that moves at constant speed in the positive $x$-direction.
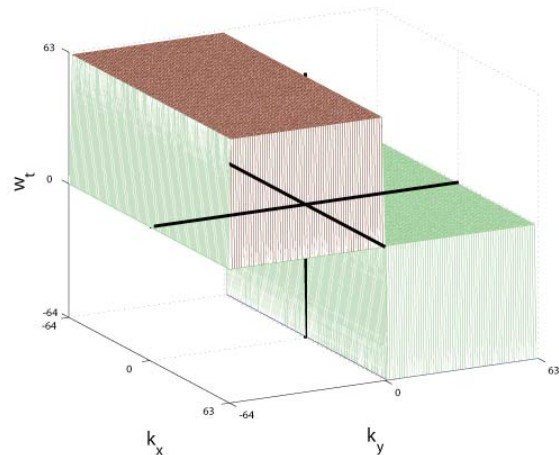


Fig. 5. Ideal 3-D transfer function for retaining all movement information in the positive $y$-direction. All points of $H(j\omega_t, jk_x, jk_y) = 1$ are marked.

Figure 5 shows an ideal transfer function for retaining all object movement information into the positive $y$-direction. Planar waves propagating in the $y$-direction can pass a filter with this transfer function completely and independently of propagation speed. Figure 6 shows the resulting non-zero points after filtering. It shows that only
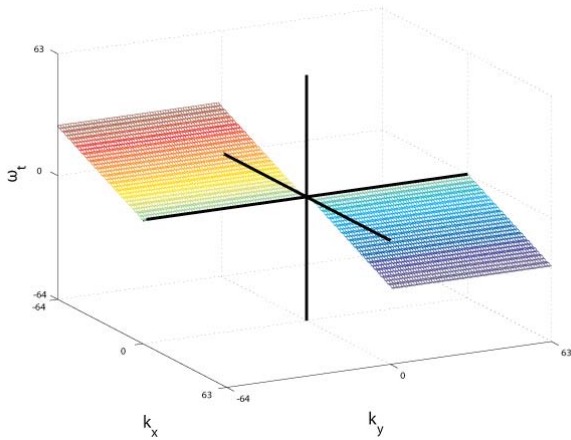
Fig. 6. Resulting non-zero points after filtering. Observe that only the planar waves of the negative $y$-direction components are filtered out.

the planar waves of the negative $y$-direction components are filtered out. Movements of similar directions have similar regions of support of Fourier components, i.e. the discrimination of objects of similar movement directions is a challenging task due to overlapping bands. The overlapping effect is stronger for lower speeds. If directional information of a movement has to be extracted, it may be beneficial to give a lower bound for the speed considered.

**2.4. Data sampling.** Another source of degradation is data sampling, which is necessary for digital signal processing. Similarly to common temporal 1-D signal processing problems, where sampling frequency usually complies with Shannon's theorem, spatio-temporal sampling of visual scenes ought to be compliant with Shannon's theorem generalized to the multidimensional case. However, this restriction is typically violated, i.e. scenes recorded by a camera are not constrained to a motion compliant with Shannon's theorem. The latter will be illustrated subsequently.

The temporal variation of the value for one point in a grey scale 3-D space representation can be defined by a temporal 1-D function. A moving object that passes this point causes a transition of this function from the value that represents the background to the one that represents the object. The duration of this transition may be very short, even for slowly moving objects, since objects appear instantaneously and not gradually. This instantaneous transition from one value to another yields high frequency components in the respective Fourier representation and thus requires a very high sampling frequency for an alias-free description in discrete time and space.

The effect is visualized for a spatio-temporal 2-D signal in Figs. 7 to 9 for different speeds. Figures 7 and 8 show a 2-D representation of a rectangular object that

is moved by one sample per time step and its frequency spectrum, respectively. The 2-D signal is shown as an image, where the signal value is expressed in pixel brightness. The black colour represents the value one and white represents zero. In Figs. 9 and 10 the same is given for an object speed of two pixels per time step. Figure 8
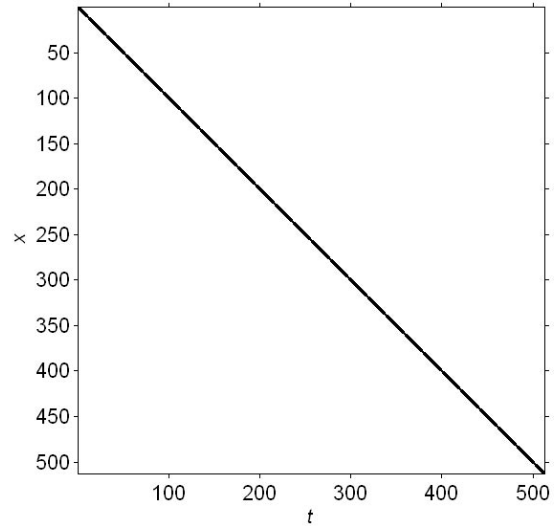


Fig. 7. Spatio-temporal discrete 2-D signal representing a moving rectangular (1-D) object. The speed is one sample per time step.
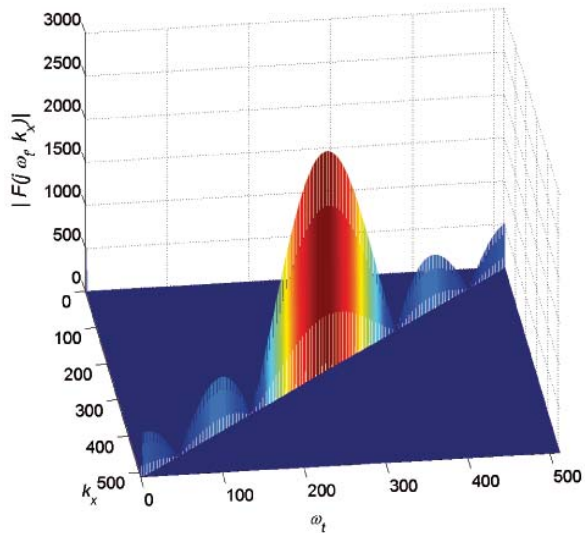


Fig. 8. Absolute value of the resulting discrete Fourier transform. It shows the expected si-function spread over one diagonal line.

shows the expected result, which seems to coincide with the continuous result. The Fourier transform of the rectangular object function, the si-function in the $k_x$-direction, is spread along the $\omega_t$ direction by the $\delta$-function according to (3).

Accompanying problems are disclosed by Figs. 9 and 10, which differ only in the underlying object speed.
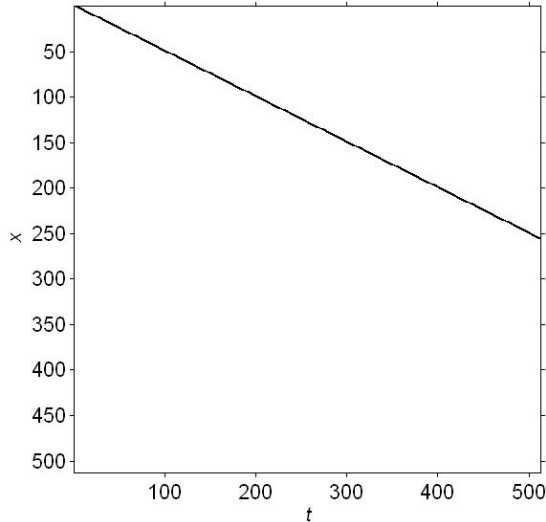


Fig. 9. Object moving at the speed of two samples per time step.
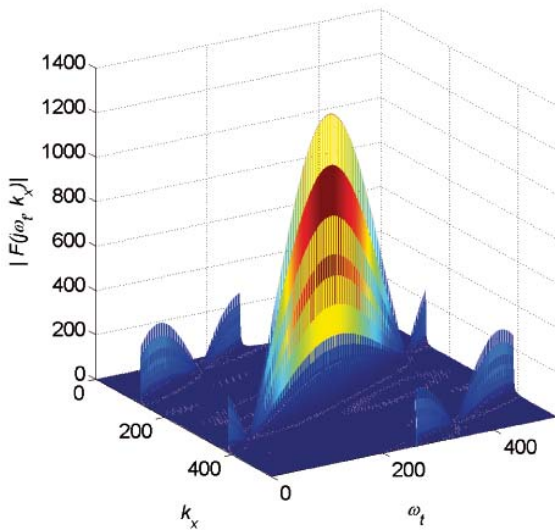


Fig. 10. Absolute value of the resulting discrete Fourier transform of the image shown in Fig. 9. It shows additional alias components caused by the infinitely extended si-function.

The Fourier transform now shows additional spectral values at lines parallel to the original one. These additional parts are aliasing components stemming from the periodical replication of the infinitely extended si-function. These components also exist and disturb the signal represented in Fig. 7, but in the given configuration they only appear on the original line and can thus not be recognized in the given plot of Fig. 8.

In the case of images of a real world scenery that are provided by a video camera, this instantaneous transition is partly compensated for by the finite aperture of the video camera sensor. Nevertheless, usual camera configurations (768 × 576 pixels, 50 Hz) cannot produce alias-free scene descriptions. It is thus necessary to take these unavoidable components into account during filter design and analysis.

## 3. System design

In this section, we discuss issues related to the implementation of velocity filters. We propose two kinds of exemplary transfer functions differing in computational complexity and filter accuracy.

### 3.1. Short introduction to wave digital filters.
The filters presented in the following subsections have been implemented on the basis of two-dimensional wave digital filters (WDF). In this subsection, we briefly introduce the WDF concept.

WDFs were introduced by Fettweis (1986) in the 1970s. As they are directly derived from classic analog circuits (reference circuits), every delay element in a WDF can be interpreted physically as holding the current state of an inductor or a capacitor, which distinguishes them from most other digital filter types.

Instead of the voltages and currents of the reference circuit, the so-called wave variables are used as signals. The introduction of these wave variables enables us to meet one major realizability constraint for a digital filter signal-flow diagram: it must not contain delay-free directed loops. In this way a class of digital filters is obtained whose passivity and therefore stability constraints can be easily assured—even using a finite arithmetic (Fettweis and Meerkoetter, 1975).

Multidimensional wave digital filters ($m$-D WDF) result from the generalization of the one-dimensional concept based on multiple independent variables. $M$-D WDFs are derived from $m$-D reference circuits depending on $m$ complex frequency variables $\psi_1, \ldots, \psi_m$. The correspondence between the frequency domain of the WDF and the reference filter is defined by the bilinear transform

$$\psi_i = \frac{z_i - 1}{z_i + 1}, \quad i = 1, \ldots, m. \qquad (8)$$

A detailed treatment of WDFs can be found in (Fettweis, 1986).

### 3.2. Detection of a vertically moving 2D-object.
Consider a simple object such as a two-dimensional Gaussian pulse given by the equation

$$s_c(x, y) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left(\left(\frac{x}{\sigma}\right)^2 + \left(\frac{y}{\sigma}\right)^2\right)\right) \qquad (9)$$
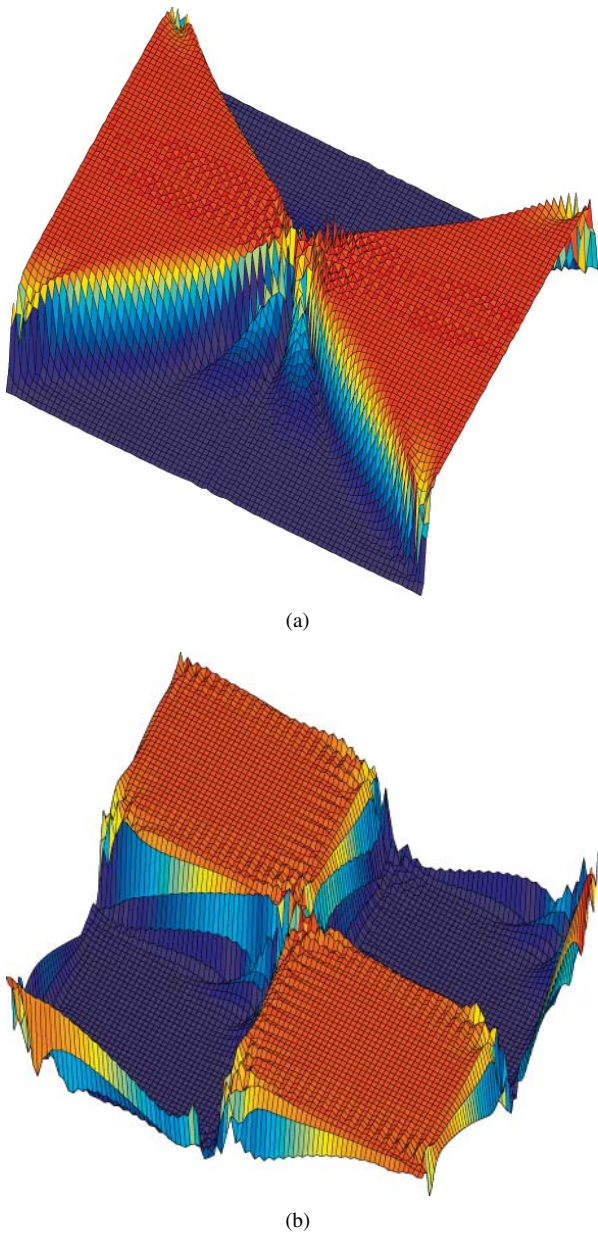
(a)



(b)

Fig. 11. Transfer functions of WDFs used in cascade: (a) and (b) show a plane in the $\omega_x\omega_y$ and $\omega_t\omega_y$ directions, respectively.

(see Fig. 12(a)) moving from the left to right at a velocity of one spatial unit per temporal unit (in the case of a digital video sequence: pixel per frame). A plot of the corresponding 3D frequency spectrum is given in Fig. 12(a), verifying the deliberations given in Section 2.1: the non-zero points of the spectrum are located on a plane with a normal vector pointing into the direction of the object movement in the spatio-temporal domain.

Thus, this plane has to be included in the passband of the transfer function of a filter applied for the detection of objects moving at the mentioned velocity and direction.

In the following, we present two examples of diffe-

rent ways to implement such a filter with both implementations offering different advantages in terms of accuracy and computational complexity.
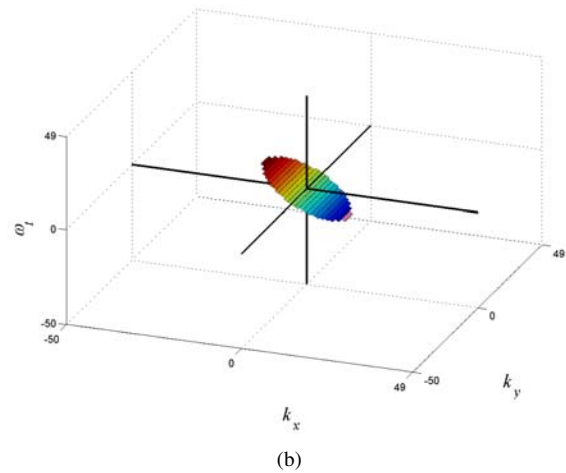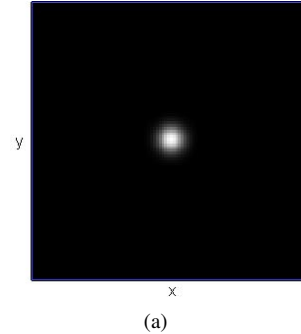


(a)



(b)

Fig. 12. Two-dimensional Gaussian pulse (a) and the magnitude function of the resulting Fourier spectrum in the case of a moving Gaussian pulse.

**3.3. Transfer function design.** Bruton and Bartley (1985) present simple filter structures exactly matching the given requirements described in Section 3.2, as they possess passbands located on planes in the 3D Fourier space. However, in practice these filter structures do not offer much flexibility as the passbands are very narrow and therefore the filter used has to exactly match the velocity of the target object. In order to overcome this problem, we design wedge filters possessing transfer functions which do not only comprise the demanded but also a certain range of "neighbouring" planes (see Fig. 13). The transfer function depicted in Fig. 13(b) shows a WDF implementation of the ideal wedge filter shown in Fig. 13(a), created using the second reactance transformation described in (Bolle, 1992) on a quadrantal filter in the $(k, \omega_t)$-plane. The resulting WDF includes 32 multipliers and 44 delay elements. A less complex (but less accurate) solution of the given filter problem will be presented in Section 3.4.
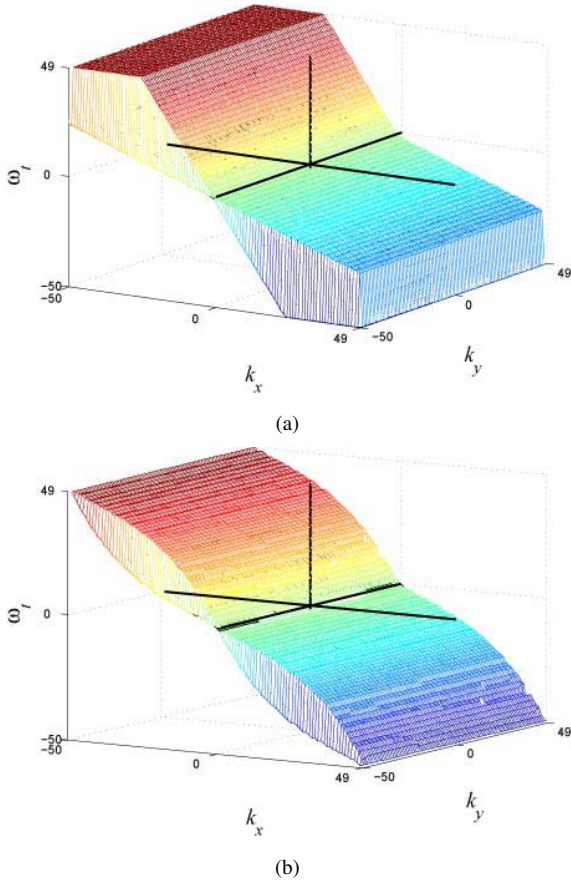
(a)



(b)

Fig. 13. Ideal transfer function of a 3D wedge filter (ideal (a) and realized using a wave digital filter (b)).

### 3.4. Design of computationally effective transfer functions.

Obviously, the more complicated the transfer function, the more complex the filter structure and the more arithmetic operations are necessary. That is why the complexity of transfer functions has to be as high as necessary but, at the same time, as low as possible.

Figure 14 shows transfer functions capable of filtering objects moving in a certain $x$-direction (right to left or left to right) without velocity restrictions. However, in the case of Figs. 14(a) and 14(b), where stop and pass regions take up four full octants (resulting in simple and easy to implement filter structures using 2D fan filters (Bolle, 1992)), objects moving in diagonal and even almost vertical directions are not suppressed by the filter. Depending on the scene, this can be a strong drawback, as noise and unwanted detections can be a consequence. Thus, the pass regions are trimmed to match a V-shape in order to limit the filter result to primarily horizontal motion. Figures 14(c) and (d) show these optimized pass regions with a "V-angle" of 90 degrees, which is sufficient for the application examples covered in this work (see Section 4).

We define the transfer functions in Figs. 14(a) and (b)

as $H_1(j\boldsymbol{\omega})$ and $H_2(j\boldsymbol{\omega})$, those shown in Figs. 14(c) and (d) as $H_3(j\boldsymbol{\omega})$ and $H_4(j\boldsymbol{\omega})$, with

$$\boldsymbol{\omega} = \begin{pmatrix} \omega_t \\ \boldsymbol{k} \end{pmatrix}. \qquad (10)$$

Then each of these transfer functions can be described using the unit step function (Heaviside function):

$$u(\omega) = \begin{cases} 1 & \text{if } \omega > 0, \\ \frac{1}{2} & \text{if } \omega = 0, \\ 0 & \text{otherwise}, \end{cases} \qquad (11)$$

$$H_1(j\boldsymbol{\omega}) = u(\omega_x)u(\omega_t) + u(-\omega_x)u(-\omega_t), \qquad (12)$$

$$H_2(j\boldsymbol{\omega}) = u(-\omega_x)u(\omega_t) + u(\omega_x)u(-\omega_t), \qquad (13)$$

$$H_3(j\boldsymbol{\omega}) = u(\omega_x - |\omega_y|)u(\omega_t)$$
$$+ u(-\omega_x - |\omega_y|)u(-\omega_t), \qquad (14)$$

$$H_4(j\boldsymbol{\omega}) = u(\omega_x - |\omega_y|)u(-\omega_t)$$
$$+ u(-\omega_x - |\omega_y|)u(\omega_t). \qquad (15)$$
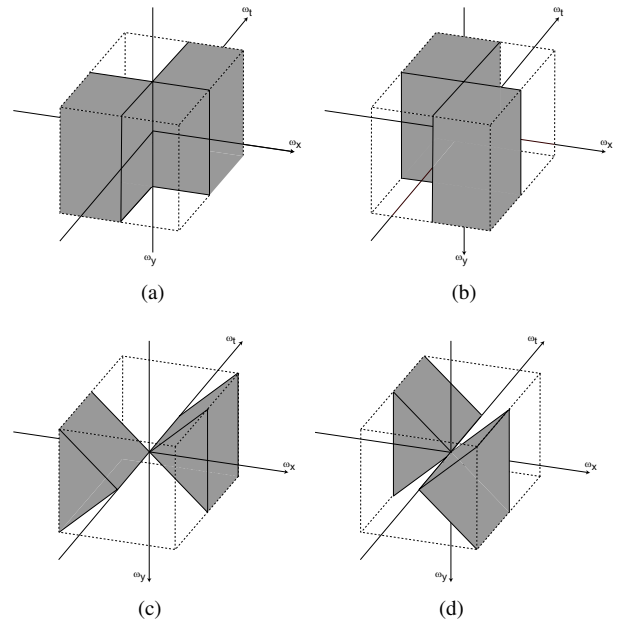


(a)

(b)

(c)

(d)

Fig. 14. Here (a) and (b) show transfer functions for filtering all objects moving from the right to left ($H_1(j\boldsymbol{\omega})$) and left to right ($H_2(j\boldsymbol{\omega})$), respectively, including diagonal (and almost vertical) directions that only incorporate small parts of the particular horizontal direction. Furthermore, (c) and (d) ($H_3(j\boldsymbol{\omega})$ and $H_4(j\boldsymbol{\omega})$) are optimized for the detection of the horizontal movement.

*Filter design:* For the design of a filter possessing a transfer function like, for example, $H_4(j\boldsymbol{\omega})$ (see Eqn. (15) and Fig. 14(d)), we use a cascade of filters ($H_{4a}(j\boldsymbol{\omega})$ and

$H_{4b}(j\boldsymbol{\omega})$), one operating in the $xy$-plane and one in the $xt$-plane:

$$H_{4a}(j\boldsymbol{\omega}) = \widetilde{H}_{4a}(j\omega_x, j\omega_y), \qquad (16)$$

$$H_{4b}(j\boldsymbol{\omega}) = \widetilde{H}_{4b}(j\omega_x, j\omega_t). \qquad (17)$$

In other words, we combine two two-dimensional filters (see Fig. 11) to obtain one three-dimensional filter:

$$Y(j\boldsymbol{\omega}) = H_{4a}(j\boldsymbol{\omega})H_{4b}(j\boldsymbol{\omega})X(j\boldsymbol{\omega}). \qquad (18)$$

Figure 15 shows a signal flow diagram of the resulting filter cascade.

$$X(j\omega) \Longrightarrow \boxed{H_{4a}(j\omega)} \Longrightarrow \boxed{H_{4b}(j\omega)} \Longrightarrow Y(j\omega)$$
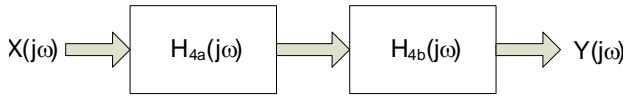
Fig. 15. Cascade of two two-dimensional filters.

## 4. Experimental verification

**4.1. Filtering of moving 2D Gaussian pulses.** For a verification of the results presented in Section 3, an image sequence showing two Gaussian pulses moving from the left to right at different velocities (one and three pixels per frame, respectively) was created (Figs. 16(a) and (b) show two frames of this test sequence). It can be easily observed that in the output sequence shown in Figs. 16(c) and (d) (which correspond to the test sequence frames mentioned above) the faster pulse is almost fully eliminated. The results were obtained using a WDF filter with a transfer function of Fig. 13(b).

**4.2. Pedestrian detection.** The second experiment focuses on a more complex object: a test image sequence showing a pedestrian crossing a street in front of a camera host (Fig. 17) is filtered using the filters presented in Section 3. In detail, we distinguish between two different test scenarios: while the first one is processed using the filter presented in Section 3.3, the second test case uses the optimized version proposed in Section 3.4. Note that the WDF implementation of the respective filter allows us to process the video sequentially in the spatio-temporal domain, i.e. no transformations to the spectral domain are necessary.

Figures 17(a) and (b) show two frames of a video sequence containing a pedestrian crossing the street in front of the camera. Note that in this example the host vehicle does not move—the correction of ego motion using the methods presented in this paper is one important objective for future work but it has not been evaluated yet. The second row (Figs. 17(c) and (d)) shows the two frames after filtering using the filter with the transfer function presented in Section 3.3. Obviously, everything but the
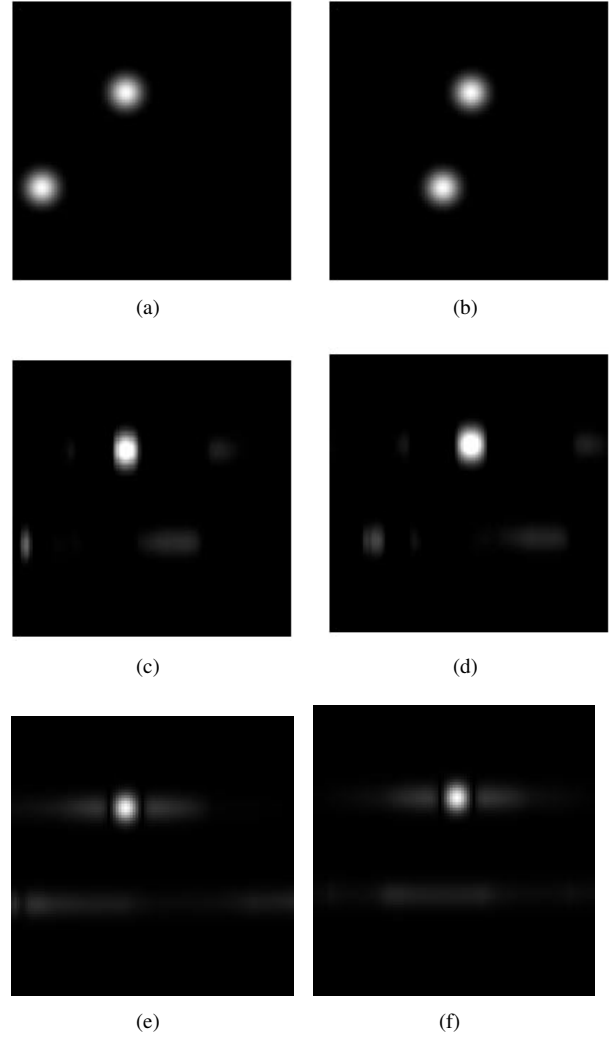


Fig. 16. Panels (a) and (b): two frames of a test sequence showing two Gaussian pulses, the upper one moving at a speed of one pixel per frame, the other at three pixels per frame. Panels (c) and (d): filter results using the transfer function depicted in Fig. 13(b). Panels (e) and (f): filter results using the ideal transfer function depicted in Fig. 13(a).

pedestrian is successfully suppressed and the result could be easily used as an input for further object detection modules.

Figures 17(e) and (f) represent results obtained using the less complex filter structures presented in Section 3.4. It can be observed that even if the result is not as good as the one presented in Figs. 17(c) and (d), the pedestrian object is still visible and a thresholded version of the image (as depicted in Figs. 17(g) and (h)) shows only the target object, and not the environment.
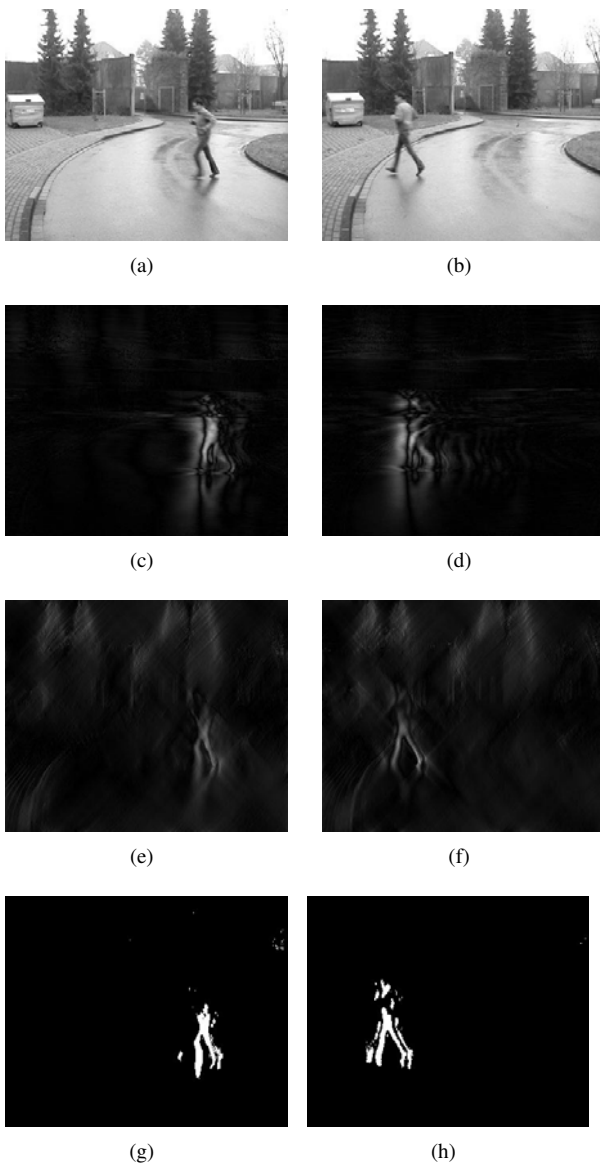
(a)    (b)

(c)    (d)

(e)    (f)

(g)    (h)

Fig. 17. Two frames of a test sequence showing a pedestrian crossing from right to left at different points in time. The top row shows the original image, the second row the filtered image using the filter presented in Section 3.3 and the third one the filtered image using the filter presented in Section 3.4. Panels (g) and (h) display the filter results from (e) and (f) after thresholding.

## 5. Conclusions and future work

Velocity filtering by means of linear $n$-D filters is an adequate approach for the extraction of movement information from visual scenes. Unfortunately, the Fourier representation of these signals is corrupted by several effects. One of these effects is the indispensable violation of Shannon's sampling theorem, by movements of high velocity

and disadvantageous object forms. Nevertheless, the feature of movement direction that can be extracted by linear $n$-D filters can be used to successfully solve object recognition tasks, e.g. pedestrian detection, as presented in Section 4.2.

Experimental results from tests on real world video sequences showing crossing pedestrians demonstrate that our approach is a promising and powerful extension of the existing object detection systems. This work forms a basis for further research on the applicability and potential of linear multidimensional filters in real-time driver assistance applications. Our objective is to benefit from the advantages multidimensional systems theory offers with respect to system description, design, stability, etc. Therefore, extended research on WDF improvability with respect to the addressed applications will be an objective of high priority in the future.

Another focus will be set on the development of more specialized transfer functions capable of filtering objects in more complex scenes and their hardware-effective implementation. Yet another challenge is the solution of one of the most demanding problems in the field of automotive applications using single camera vision: distinguishing between an object movement and a movement solely caused by ego motion (i.e. host vehicle motion) without additional sensors.

By and large, the velocity filters presented in this paper show good performance for motion-based object detection with the potential of leading to new and effective methods in driver assistance system development and other object recognition applications.

## References

Bartley N. and Bruton L. (1987). The design of highly selective adaptive three dimensional recursive cone filters, *IEEE Transactions on Circuits and Systems* **34**(7): 775–781.

Bartley N. R. and Bruton L. T. (1986). The enhancement and tracking of moving objects in digital images using adaptive three dimensional recursive cone filters, *IEEE Transactions on Circuits and Systems* **33**(6): 604–611.

Bolle M. (1992). *Entwurfsverfahren fuer mehrdimensionale diskrete Faecherfilter*, Ph.D. thesis, Ruhr-Universitaet Bochum, Germany.

Bruton L. T. and Bartley N. R. (1985). Three dimensional image processing using the concept of network resonance, *IEEE Transactions on Circuits and Systems* **32**(7): 664–672.

Christie P.A.F, Hugher V.J. and Kennett B.L.N (1983). Velocity filtering of seismic reflection data, *First Break* **1**(3): pp. 9–24.

Dansereau D. and Bruton L. (2007). A 4-D dual fan filter bank for depth filtering in light fields, *IEEE Transactions on Signal Processing.* **55**(2): 542–549.

Dudgeon D. E. and Mersereau R. M. (1984). *Multidimensional Digital Signal Processing.* Prentice-Hall, New Jersey, USA.

Fettweis A. (1986). *Wave Digital Filters: Theory and Practice, European Association for Signal Processing (EURASIP)— EURASIP Short Course*.

Fettweis A. and Meerkoetter K. (1975). Suppression of parasitic oscillations in wave digital filters, *IEEE Transactions on Circuits and Systems* **22**(12): 239–246.

Gandhi T. and Trivedi M. M. (2006). Pedestrian collision avoidance systems: A survey of computer vision based recent studies, *Proceedings of the IEEE Conference on Intelligent Transportation Systems*, Toronto, Canada, pp. 976 – 981.

Lu W.-S. and Antoniou A. (1992). *Two-Dimensional Digital Filters*, Boca Raton, FL, CRC Press.

Marzetta T. L. (1994). Fan filters, the 3-D Radon transform, and image sequence analysis, *IEEE Transactions on Image Processing,* **3**(3): 253–264.

Runze G. (2005). *Entwurf dreidimensionaler digitaler Geschwindigkeitsfilter zur Separation unterschiedlicher Geschwindigkeitskomponenten eines Signals*, Ph.D. thesis, Universitaet Erlangen-Nuernberg, Germany.