TR－A－0134

Applicability of Oriented Filters to Edge Detection and Motion Analysis

Christine Lacombe

1992. 2. 24
（1992. 1.28受付）

ATR視聴覚機構研究所
〒619-02 京都府相楽郡精華町光台2-2　☎07749-5-1411

ATR Auditory and Visual Perception Research Laboratories
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan
Telephone: +81-7749-5-1411
Facsimile: +81-7749-5-1408

Contents

Abstract

This paper deals with spatio-temporal oriented filters applied to motion analysis in image sequences. As velocity corresponds to orientation in the three-dimensional space, oriented filters are of the highest importance. Adelson and Freeman[1] have shown that the orientation which gives the greatest output for these filters gives the slope of the contour, so the velocity in the 3D case. Our approach will consist of two steps; first the theory and the use of these oriented filters will be presented in the two-dimensional case applied to traditional edges detectors. We will compare the different approaches and show the importance of edge's orientation detection. As a second step, this formalism of oriented filters will be generalized to the 3D case and a discussion about their use in motion analysis will be presented.

This paper was prepared and researched at ATR as part of a six month internship from INT, Paris, France. The research was performed under the direction of Ed Gamble with the support of K. Arimura and K. Ueno.

1 Facilities Overview

Before beginning the discussion of oriented filters we present a very brief overview of ATR and the Connection Machine.

1.1 The ATR Research Center

ATR, Advanced Telecommunications Research Institute International, was established on March 22, 1986 with support from various sections of industry, academia

and government. It was created just after the privatization of NTT and a strong collaboration between NTT and ATR exists. In the spring of 1989 the research laboratory moved from Osaka to the new Kansai Science City in Kyoto Prefecture. ATR is made up of the following four corporations.

ATR Auditory and Visual Perception Research Laboratories Psychologists, physiologists and engineers are working on the basic mechanisms of perception and cognition in the human senses of sight and hearing. This laboratory is composed of three departments which perform research in the following areas:

Visual Perception Department: Basic mechanisms of visual perception, character and pattern recognition, and scene analysis and understanding.

Cognitive Processes Department: cognitive processes for visual information, parallel computing principles, and learning and motor theories of perception.

Hearing and Speech Perception Department: Auditory models, speech perception and recognition.

ATR Interpreting Telephony Research Laboratories Researchers in this laboratory are all working on an automatic interpreting telephone which would make it possible for speakers of different nationalities to converse using different languages. Research topics include: speech recognition, speech synthesis, research on the interface between speech and language, and machine translation.

ATR Communication Systems Research Laboratories Research is based on a human-oriented intelligent communication system: communications with realistic sensations, nonverbal interfaces, automatic generation of communication software and security.

ATR Optical and Radio Communications Research Laboratories The members of this laboratory performs research on optical and electronic devices related to optical intersatellite communications and "Key Technologies" for future mobile communications.

Japan's economic growth has mainly been based on the development of applied technology. It is one of their goals to promote basic technological research and the exchange of researchers and information between research institutions in Japan and abroad.

Personally, I worked for 6 months as a researcher in the ATR Auditory and Visual Perception Research Laboratories, in the Visual Perception Department. My main interest is computer vision and my research subject was on the "Applicability of Oriented Filters to Edge Detection and Motion Analysis" as reported in this paper. In order to study real-time processing and to allow for efficient computation, we used a Connection Machine.

1.2 The Connection Machine

The Connection Machine is a data parallel computing system. The main idea is to get rid of Von Neumann architecture (two part design in a computer: the memory and the processing) and to build a computer in which memory and processing are combined. The Connection Machine is made up of a large number of tiny processor/memory cells connected by a programmable communication network. Local computations are performed in the processing cell associated with a cell of storage where the data are stored. All the cells are connected in a programmable pattern. Because thousands or millions of processing cells work simultaneously, the computation proceeds much more rapidly than on a conventional machine.

Nowadays different versions of Connection Machines exist: from the CM-1 to the newest CM-5. These machines differ by the number and type of processors, by the interconnectivity of the processors, and thereby by the processing speed. The one we used is the CM-2 which has a maximum of 65536 processors/memory cells. It has a peak instruction rate of about 10 Gigaflops. In the configuration at ATR, the Connection Machine had 16k processors each with 256k bits of memory per processor.

2 Introduction

In computer vision the study of **visual motion** is of the highest importance. Motion provides meaningful information about the perception of space and the external

environment. Changes in an image due to motion of objects can be related to the structure of the objects in space. For example we can determine the positions of points in the space by their observations in a sequence of images.

The applications of dynamic scene analysis cover a broad range of fields, for example:

- in medicine the automatic analysis of image sequences of the human heart is used to establish diagnosis after heart surgery,
- in meteorology the study of satellite images provides information about the movements of atmospheric disturbances,
- in a military field, analysis of image sequences are used to recognize and track targets,
- in robotics three dimensional vision and in particular motion in 3D are of significant importance; in this area it is necessary to control the visual space in order to locate, track and avoid mobile objects,
- survaillence, guidance, traffic monitoring, autonomous navigation are others applications for which *3D motion information is of the greatest importance.*

Many researchers have been interested in the psychophysical study of visual motion. Some of them outlined the presence of motion-sensitive cells in the primary visual cortex which are directionally selective and tuned to spatiotemporal frequencies (Andersen and Siegel[2], Adelson and Bergen[3], Poggio and Reichardt[4]). In order to detect this spatiotemporal orientation, Adelson and Bergen[3] have proposed **spatiotemporal oriented filters** named motion energy filters. The use of this kind of filters is motivated by the fact that image motion is characterized by orientation in space time domain.

As the **velocity** of an object corresponds to **orientation** in space time domain, 3D oriented filters, two-dimensions for space plus one for time, are efficient in velocity determination and in motion analysis.

Our main task is to study the applicability of **3D oriented filters to motion analysis.** In order to reach our goal, we will as a first step towards a more comprehensive analysis present the theory, design and use of these oriented filters in

two-dimensions; we will apply these **2D oriented filters to edge detection** and we will compare them to other known edge detectors. After this experience with 2D oriented filters, we will study them in the 3D case and we consider their applicability to analysis of motion in temporal image sequences.

It is very interesting to apply the oriented filters to the two-dimensional case because edge detection is one important task in image segmentation. In this field the main task is to separate the components of an image in order to analyze and recognize them. The components correspond to the physical objects in the scene. The gray-level values have two basic properties: discontinuity and similarity. The discontinuities occur at the boundaries of the objects where the intensity is assumed to change abruptly. The objects are supposed to have homogeneous surfaces that means regions of constant or smoothly varying intensity. But these assumptions are not always valid so many errors occur in boundary detection. Problems arise when physical boundaries have similar surfaces, when surface texture exists, and naturally when noise is present. Most often images contain a variety of edges; these edges have different length and contrast. But what is more important is that they occur at any orientation and consequently an efficient edge-detection procedure has to be able to distinguish intensity changes at different angles. This motivates the use of oriented filters for edge detection.

In this paper we present the applicability of oriented filters to the two dimensional case as edge detectors and to the three dimensional case in motion analysis. In the first part we review the traditional edge detectors (Canny, Laplacian) then we present the theory, the design and the use of oriented filters applied to the edge detectors. In the last part we discuss the applicability of these steerable filters to motion detection (via the concept of 3D oriented filters) and the limitations we encountered.

3 Traditional Edges Detectors

3.1 Introduction

One common assumption in image analysis is that **abrupt intensity changes occur at objects boundaries.** It is for this reason that edge detection is of fundamental

importance in image analysis. For our analysis we idealize an edge as a step function which has a magnitude and a direction. (Other types of edges exist e.g. 'roof edges', we consider only step edges in this discussion).

We will use the following properties of a step function in order to detect the edges: *an abrupt intensity change gives rise to a peak when we consider its first derivative and to a zero crossing when we consider its second derivative.* These properties are illustrated Figure 1.

A suitable filter is the two dimensional is the Gaussian distribution:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \tag{1}$$

There are 3 reasons for using the Gaussian:

- it smoothes the image so it decreases the influence of noise (due to the sampling system, transmission channel, texture variations, etc). The space constant of the Gaussian σ is used to modify the smoothing (the more smoothed we want the image to be, the larger should be the value of σ). The Gaussian distribution has the property to be smooth and localized in both spatial and frequency domains.

- the first and second derivative of the Gaussian are easy to compute analytically which allow us to detect peaks in the first derivative and zero crossings in the second derivative.

- the Gaussian is separable and due to the Central Limit Theorem it can be composed of many 'smaller' functions; the Gaussian is efficient to compute.

The edge detection procedure requires a two stage process: First, we have to smooth the input image (with a Gaussian, for the reasons we exposed previously) in order to eliminate the displacement errors but while also preserving the discontinuities and second, we have to detect the edges. The second stage is simply Canny's non-maximum suppression and hysteresis[5] applied to maxima in the gradient magnitude or zero crossings detection applied to the Laplacian.

The mathematical property of the convolution allows us to use the different paths for computation of the Gaussian derivatives. These paths are illustrated below with the following notion: convolving the image, I, with a Gaussian, G, is $G * I$, taking

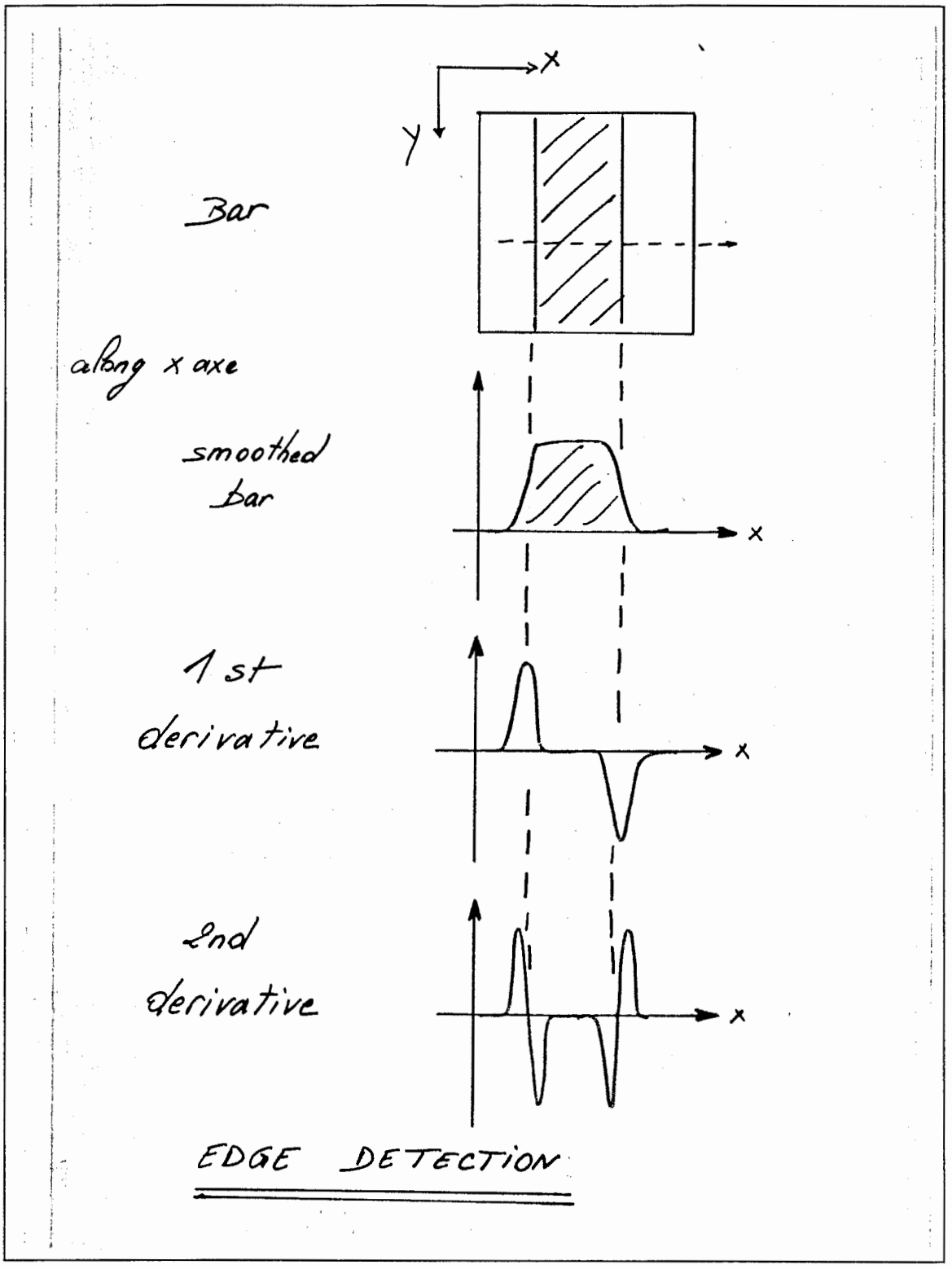

Figure 1: From the top of the figure to the bottom: an example of an object represented as a light bar, a smoothed cross section of it along the x axis, its first derivative (edges occur at the locations of the maxima) and its second derivative (edges occur at the locations of zero-crossings).

the first derivative is $\bigtriangledown(G * I)$, and taking the second derivative is $\bigtriangledown^2(G * I)$. This notion:

$$I \xrightarrow{(*G)} (I * G) \xrightarrow{\bigtriangledown or \bigtriangledown^2} \bigtriangledown (I * G) \; or \; \bigtriangledown^2 (I * G)$$

$$I \xrightarrow{\bigtriangledown or \bigtriangledown^2} (\bigtriangledown I) \; or \; (\bigtriangledown^2 I) \xrightarrow{(*G)} (\bigtriangledown I * G) \; or \; (\bigtriangledown^2 I * G) \tag{2}$$

$$I \xrightarrow{(*\bigtriangledown G) or (*\bigtriangledown^2 G)} (I * \bigtriangledown G) \; or \; (I * \bigtriangledown^2 G)$$

The choice of a computational path is based upon computation efficiency and conceptual simplicity.

3.2 With the First Derivative

We now consider the use of the first derivative for the computation of intensity edges in images. First we describe techniques for smoothing the image; that is followed by a description for computation of the gradient and extraction of the edges.

3.2.1 Smoothing the Input Image

They are many different ways to smooth an image with a Gaussian kernel:

- using the *binomial convolution*; iterated application of the mask: [1 2 1] along the x axis on the input image to simulate the smoothing with the Gaussian along this axis. Smoothing with a 2D Gaussian is just iteratively applying the mask along the two axes x and y.
- using the *Fourier Transform*; according to the property of the convolution, smoothing an image with a Gaussian can be computed by taking the Fourier Transform of the input image and of the Gaussian, multiplying both and taking the inverse Fourier Transform of the result, $(I * G) \Longleftrightarrow \mathcal{F}^{-1}(\mathcal{F}(I).\mathcal{F}(G))$ where $\mathcal{F}$ denotes the Fourier transform and $\mathcal{F}^{-1}$ is the inverse Fourier transform.

The binomial convolution is an approximation method that works well if the Gaussian σ is large; the Fourier transform is of course exact.

3.2.2 Gradient of the smoothed image

The gradient of the smoothed image $S(x,y) = (I * G)(x,y)$ is defined as a two dimensional vector:

$$\nabla [S(x,y)] = \begin{bmatrix} \frac{\partial S}{\partial x} \\ \frac{\partial S}{\partial y} \end{bmatrix}.$$

This vector points in the direction of maximum rate of change in S(x,y), it is a vector perpendicular to the contour. The gradient can be computed in different ways:

1. using Sobel operators: ∇y is computed by applying the following 3x3 mask

$$differentiation \begin{cases} \nearrow \\ \searrow \end{cases} \overbrace{\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}}^{averaging}.$$

   A similar mask for ∇x can be easily found by rotating the above mask by 90 degrees. These masks compute both a differentiation and a local averaging.

2. using simple derivative operators:

$$\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

3.2.3 Detection of the Maxima of the Gradient

As we have shown previously in Figure 1, the edges occur at the locations of peaks in the gradient of the smoothed image. **In order to find the edges, we have to detect maxima in the magnitude of the gradient of the smoothed image.** To detect maxima we use the direction of the gradient which is given by: $\arctan(\nabla y / \nabla x)$. **A point is a local maximum if it is a maximum in the direction of the gradient.** We compare its value to those of its two neighbors located in the direction of the gradient.

Edge detection is improved if weak edges are eliminated. Simple thresholding of some measure of edge strength is insufficient because if the threshold is too high

few edges exist and if the threshold is too low too many edges exist. Canny has proposed **hysteresis thresholding** to improve the extraction of edges. The main idea of hysteresis thresholding is to use **two thresholds**. At a certain location in the image, if the value of the magnitude of the gradient is *above the highest threshold*, we are sure that it is a point of the edge; if it is *below the lowest threshold* we are sure that it is not a point of the edge. If the value of the magnitude of the gradient is *between the two thresholds*, we look at the neighborhood of the considered point. If one of its neighbors is still a point of the edge, the point itself becomes part of the edge and we continue linking all the edges with this procedure. This allows us to get filled, linked edges and to eliminate the false edges. See Figure 2 for a illustration and further description.

The thresholds are computed using a **histogram** of the values of the magnitude of the gradient. Canny has proposed that the 'true' edges in a noisy image (white Gaussian noise) represent 20% of the points and that 20 other percent are possible edges pixels. The highest threshold is given by the value of the magnitude of the gradient when 80% of the points are 'eliminated' and the lowest threshold when 60% of the points are 'eliminated'. These assumptions are not valid if the input image is not noisy or if it is highly textured; that is the noise is not spatially white, Gaussian noise.

3.3 With the Second Derivative

Edge detection with the second derivative will be considered next. Smoothing of the image is identical to that described in the previous section. We begin here with a description of the computation of the second derivative.

3.3.1 Laplacian of the smoothed image

The Laplacian of the point S(x,y) of the smoothed image is given by the following expression: $\nabla^2 [S(x,y)] = \frac{\partial^2 S}{\partial x^2} + \frac{\partial^2 S}{\partial y^2}$. The Laplacian can be easily computed by using one of the following 3x3 masks:

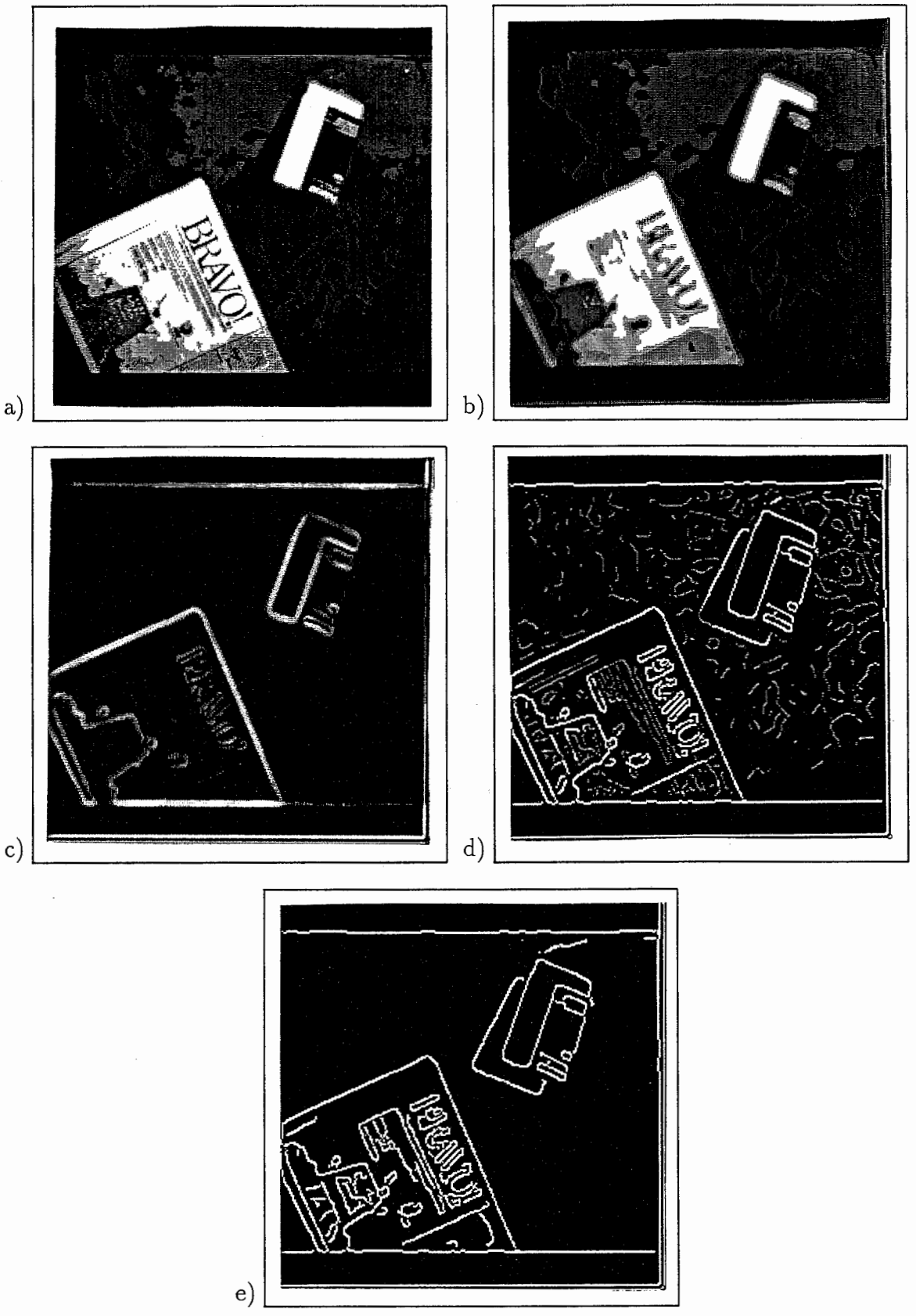

Figure 2: a) Input image, b) Smoothed image (with a Gaussian of $\sigma = 1.5$), c) Gradient of the smoothed image, d) First stage of edge detection, e) Edge detection after Canny's thresholding hysteresis

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad , \quad \begin{bmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{bmatrix}$$

These masks are derived based upon the discrete approximation to a Laplacian on a rectangular lattice of points.

3.3.2 Zero Crossings Detection

In the case of the **second derivative** the locations of the edges occur at the positions of the **zero crossings** (that means where the value of the function passes from positive to negative or from negative to positive depending on the sign of the slope (Figure 1). In order to improve the zero crossings detection we compute *the gradient of the Laplacian* (only the zero crossings which have a high value for their gradient correspond to edges). A point is considered part of an edge if one of its neighbors has an opposite sign and if the value of its gradient is above a certain threshold. This procedure allows the elimination of false edges.

The filters we applied in this section, first and second derivatives of Gaussians, are both fixed orientation filters. However, edges in an image have an orientation which has to be determined in order to enhance the detection and to give some more information for motion analysis. The notion of oriented filters is fundamental.

4 Theory of Oriented Filters

Our main goal is to accurately compute edges and their orientation. This could be accomplished by applying a filter optimized for each and every orientation but this approach would be computationally prohibitive. Adelson and Freeman[6] have shown that **a filter of any orientation can be synthesized as a linear combination of basis filters which are rotated versions of itself.** We now will review the mains ideas of their work.

A function of orientation θ can be written as a linear combination of rotated versions of itself: $f^{\theta}(x,y) = \sum_{j=1}^{M} k_j(\theta) f^{\theta_j}(x,y)$ where $k_j(\theta)$ are the interpolation

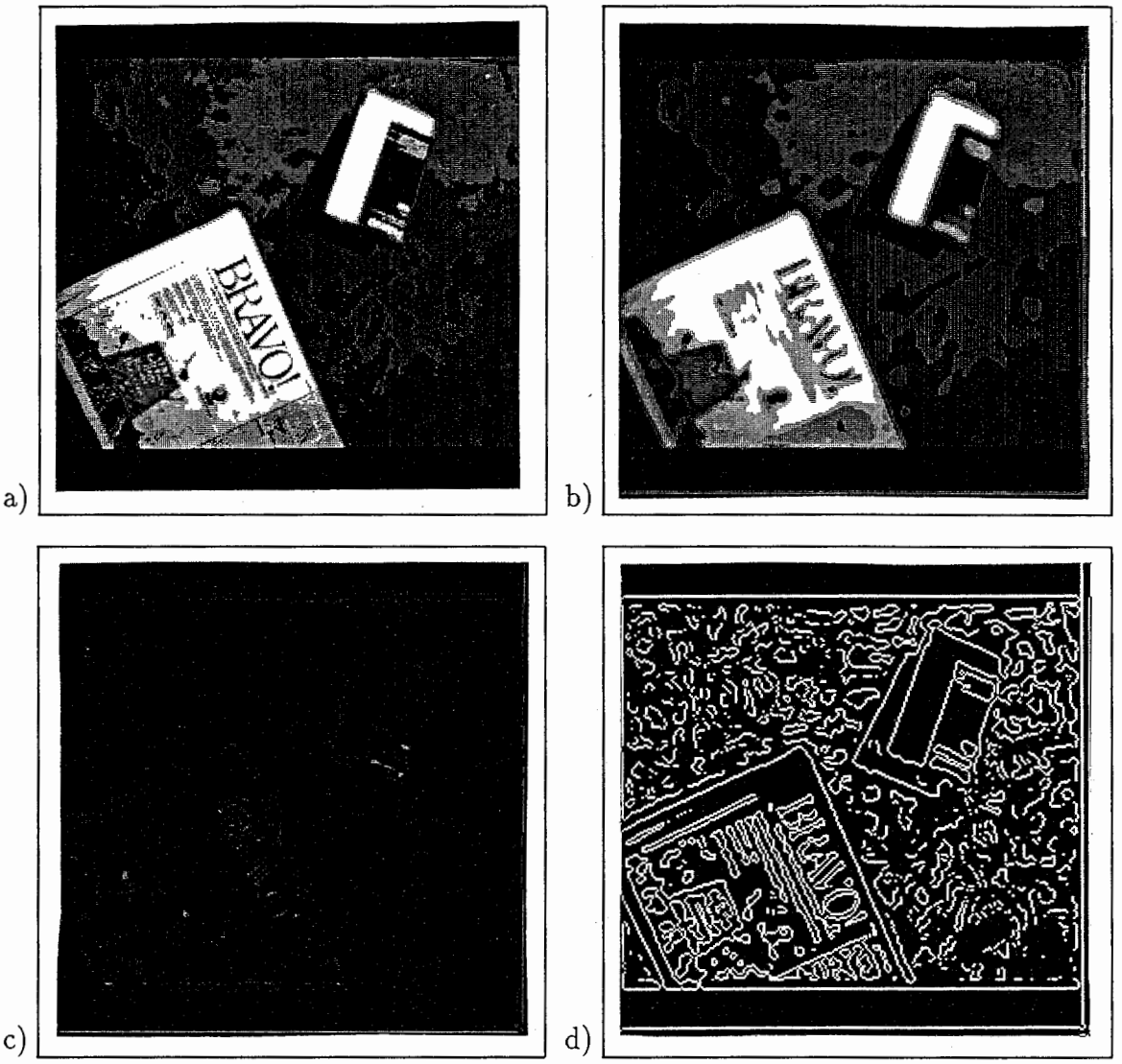

Figure 3: a) Input image, b) Smoothed image (with a Gaussian of $\sigma = 1.5$), c) Laplacian of the smoothed image, d) Zero-crossing detection

functions and $f^{\theta_j}(x,y)$ are basis filters. Convolving the image with this oriented filter $f^{\theta}(x,y)$ is just convolving it with the basis filters, multiplying the results by the correct interpolation functions (with a value dependent on the angle θ) and adding the results.

Two important questions must be resolved: how many basis filters are necessary to steer the function in any orientation and what is the expression of the interpolation functions? Freeman and Adelson have presented two theorems which will allow us to answer these questions.

- **Theorem 1.** If the function f can be written as a product of a windowing function W and a Nth order polynomial P_N, (2N + 1) basis filters are necessary to synthesize f rotated to any angle. $f(x,y) = W(x,y)P_N(x,y)$ if P_N contains only odd or even terms in x and y, (N + 1) basis filters can synthesize f rotated to any angle: $f^{\theta}(x,y)$

- **Theorem 2.** If f is any function which can be expanded in a Fourier series in polar coordinates: $f(r,\phi) = \sum_{n=-N}^{N} a_n(r)\, e^{in\phi}$ where $r = \sqrt{x^2+y^2}$ and $\tan\phi = y/x$, then the interpolations functions $k_j(\theta)$ from $f^{\theta}(x,y) = \sum_{j=1}^{M} k_j(\theta)\, f^{\theta_j}(x,y)$ are solutions of the following system:

$$\begin{pmatrix} 1 \\ e^{i\theta} \\ \vdots \\ e^{iN\theta} \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ e^{i\theta_1} & e^{i\theta_2} & \dots & e^{i\theta_M} \\ \vdots & \vdots & \dots & \vdots \\ e^{iN\theta_1} & e^{iN\theta_2} & \dots & e^{iN\theta_M} \end{pmatrix} \begin{pmatrix} K_1(\theta) \\ K_2(\theta) \\ \vdots \\ K_M(\theta) \end{pmatrix}$$

This gives us a general formalism; we will apply it to a few examples for the first and second derivatives.

5 Application of oriented filters to edge detection

As we did in the first part of this report, we will consider the first and second derivatives of the Gaussian distribution and we will now present their expression tuned to any direction. As it is well known, derivatives of the Gaussian are steerable in orientation.

5.1 Using the first derivative of a Gaussian distribution

Our main goal in this section is to find **the orientation of the contour**; using the **first derivative of the Gaussian**, we have to find the orientation θ_M for which the filter gives the greatest output. The different steps of the procedure are as followed:

Step 1. convolution of the image with the first derivative of the Gaussian tuned to any direction; according to Adelson and Freeman's theorems that means

convolution with only the basis filters and multiplication by the interpolation functions.

Step 2. determination of the orientation θ_M which gives the maximum output for the filter, f^{θ_M}: $f^{\theta_M}(x,y) = \sum_{j=1}^{M} k_j(\theta_M) f^{\theta_j}(x,y)$ for $f()$ being the first derivative of the Gaussian.

θ_M is given by the derivative of the filter output along θ: $\theta / \frac{\partial f^{\theta}(x,y)}{\partial \theta} = 0$

Step 3. Computation of the filter response for the maximum orientation.

Step 4. Detection of the local maxima in the perpendicular direction of the maximum orientation. Followed by thresholding with hysteresis and edge linking.

We now consider these steps in detail.

Step 1. The preliminary step consists in finding the number of basis filters and the expression of the interpolation functions which synthesize the first derivative of a Gaussian in any direction. The well known expression of a 2D Gaussian distribution is:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \tag{3}$$

Considering its first derivative along the x axis, we get:

$$\frac{\partial G(x,y)}{\partial x} = \frac{-x}{2\pi\sigma^4} e^{-(x^2+y^2)/2\sigma^2} = P_1(x,y)W(x,y). \tag{4}$$

It is written as a 1st order polynomial $P_1(x,y) = -x/2\pi\sigma^4$ times a windowing function: $W(x,y) = e^{-(x^2+y^2)/2\sigma^2}$. Theorem 1 can be applied and by consequence two basis filters can synthesize the first derivative of the Gaussian distribution in any direction for this two dimensional case.

$$G_1^{\theta^\circ}(x,y) = K_1(\theta) G_1^{\theta_1}(x,y) + K_2(\theta) G_1^{\theta_2}(x,y). \tag{5}$$

An obvious choice for θ_1 and θ_2 is respectively 0° and 90° so that:

$$G_1^{\theta^\circ}(x,y) = K_1(\theta) G_1^{0^\circ}(x,y) + K_2(\theta) G_1^{90^\circ}(x,y). \tag{6}$$

The two basis function can be computed in a traditional way, for example with Sobel operators and are defined by $G_1^{0°}(x,y) = \frac{\partial G(x,y)}{\partial x}$ and by $G_1^{90°}(x,y) = \frac{\partial G(x,y)}{\partial y}$.

The expressions of the interpolation functions are given by the second theorem. In polar coordinates we get:

$$G_1^{0°}(r,\phi) = \frac{-r\cos\phi}{2\pi\sigma^4} e^{-r^2/2\sigma^2} = \frac{-re^{-r^2/2\sigma^2}}{2\pi\sigma^4}\left(\frac{e^{i\phi}+e^{-i\phi}}{2}\right) \tag{7}$$

This is the expression of the first derivative of a Gaussian expanded in its Fourier series:

$$f(r,\phi) = \sum_{n=-N}^{N} a_n(r)e^{in\phi} \tag{8}$$

with $N = 1$.

The 2 interpolation functions are solutions of the following system :

$$e^{i\theta} = \begin{pmatrix} e^{i\theta_1} & e^{i\theta_2} \end{pmatrix} \begin{pmatrix} K_1(\theta) \\ K_2(\theta) \end{pmatrix} \tag{9}$$

with $\theta_1 = 0°$ and $\theta_2 = 90°$. Resolving the system gives us: $K_1 = \cos\theta, K_2 = \sin\theta$ for θ increasing clockwise and, from Equation 6,

$$G_1^{\theta°} = \cos\theta G_1^{0°} + \sin\theta G_1^{90°} \tag{10}$$

Thus, the first derivative of the Gaussian is now steerable in any direction. Smoothing the input image with this oriented filter is simply given by

$$(I * G_1^{\theta°}) = \cos\theta(G_1^{0°} * I) + \sin\theta(G_1^{90°} * I) \tag{11}$$

Due to the property of the convolution, convolving the input image with the filter tuned to any direction is the same as convolving it with the basis filters and multiplying the results by the correct weights and adding.

Step 2. The orientation which gives the greatest filter's output is given by the first derivative along θ of the expression of $(G_1^{\theta°} * I)$. The solution of this derivative set to

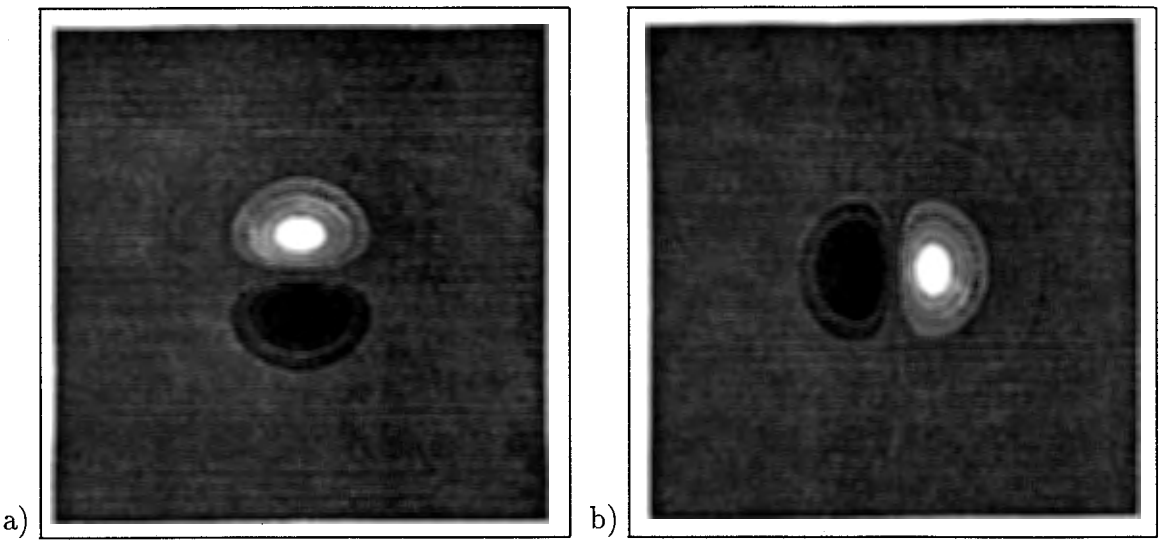

Figure 4: Basis filters for the first derivative of the Gaussian: a) $G_1^{0°}$, b) $G_1^{90°}$

zero is the value of the maximum orientation θ_M:

$$\theta_M | \quad \frac{d(G_1^{\theta°} * I)}{d\theta} = (G_1^{0°} * I)\frac{dK_1}{d\theta} + (G_1^{90°} * I)\frac{dK_2}{d\theta} = 0 \tag{12}$$

where only K_1 and K_2 depend on θ as given following Equation 9. For these K_1 and K_2 the solution to Equation 12 is

$$\theta_M = \tan^{-1}\left(\frac{G_1^{90°} * I}{G_1^{0°} * I}\right). \tag{13}$$

Step 3. In order to detect the edges, it is necessary to compute the filter's response for the maximum orientation at each point of the image. This is simply

$$G_1^{\theta_M} = K_1(\theta_M)(G_1^{0°} * I) + K_2(\theta_M)(G_1^{90°} * I) \tag{14}$$

Step 4. The final step of edge detection in this procedure is the same as with Canny's algorithm. The local maxima are detected after comparison with their neighbours situated in the direction perpendicular to the maximum orientation. A hysteresis thresholding is used to improve the detection of maxima and by consequence to improve the edge detection. As previously the final step consists of linking the edges.

In what follows we continue the discussion of oriented filters as applied to edge detection and present the use of these oriented filters with the second derivative of the Gaussian distribution.

5.2 Using the Second Derivative of a Gaussian Distribution

We will now use the **second derivative of the Gaussian tuned to any angle** (as we did previously with the first derivative) to compute the orientation of the contour and we will compare the two methods. The procedure consists of two steps: first convolving the input image with the second derivative of the Gaussian tuned to any direction and second; find the orientation which gives the "best" zero crossing.

Step 1. Convolving the input image with the second derivative of the Gaussian tuned to any angle consists in convolving it with the basis filters which are rotated versions of this function and multiplying the results by the proper weights. Applying Adelson and Freeman's theorems gives the basis filters and the interpolation functions.

The expression for the second derivative of a Gaussian along the x axis is

$$G_2^{0°}(x,y) = \frac{\partial^2 G(x,y)}{\partial x^2} = \frac{1}{2\pi\sigma^4}(\frac{x^2}{\sigma^2} - 1)e^{-(x^2+y^2)/2\sigma^2}. \tag{15}$$

This is the expression of a windowing function $W(x,y) = e^{-(x^2+y^2)/2\sigma^2}$ times a second order polynomial $P_2(x,y) = \frac{1}{2\pi\sigma^4}(\frac{x^2}{\sigma^2} - 1)$. By consequence, Adelson and Freeman's theorem 1 gives the number of basis filters needed to synthesized this function to any direction as three. Consequently, the second derivative of the Gaussian distribution can be synthesized to any direction using 3 basis filters:

$$G_2^{\theta°}(x,y) = K_1(\theta)G_2^{\theta_1} + K_2(\theta)G_2^{\theta_2} + K_3(\theta)G_2^{\theta_3}. \tag{16}$$

An obvious choice for θ_1, θ_2 and θ_3 is respectively $0°, 45°, 90°$. These basis functions can be easily computed; they correspond respectively to the second derivative of the Gaussian along x axis, 45° axis and y axis. For example they can be computed with

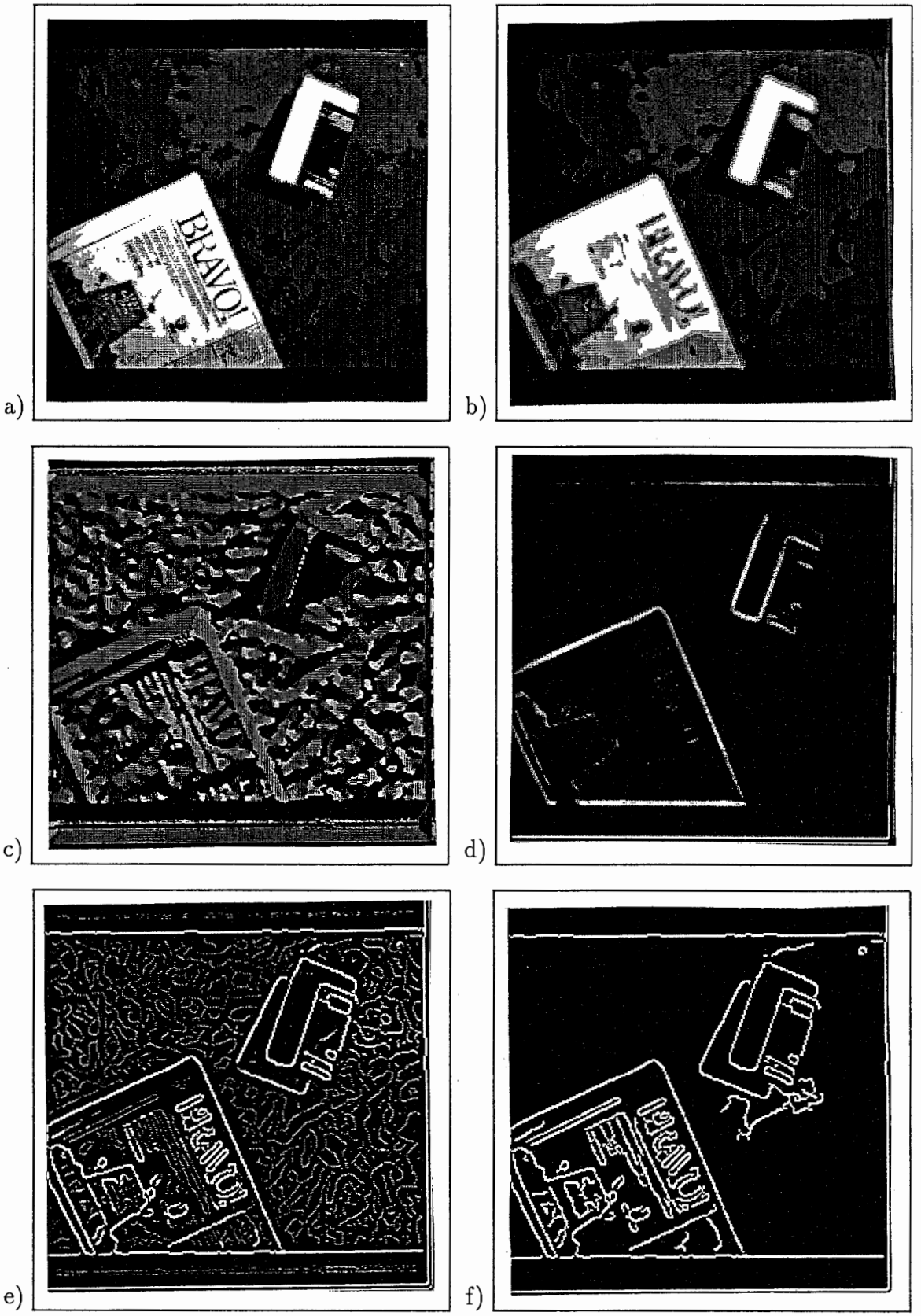

Figure 5: a) Input image, b) smoothed image (with a Gaussian of $\sigma = 1.5$), c) orientation which gives the greatest output, d) value of the first derivative of the Gaussian for these maximums orientations, e) first stage of edge detection, f) edge detection after hysteresis thresholding. 19

Sobel operators (applied twice times to get the second derivative):

$$G_2^{0^\circ} \rightsquigarrow \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}, \quad G_2^{45^\circ} \rightsquigarrow \begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}.$$

$$G_2^{90^\circ} \rightsquigarrow \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

With these basis angles the expression to synthesis the second derivative of the Gaussian along any orientation is

$$G_2^{\theta^\circ}(x,y) = K_1(\theta)G_2^{0^\circ} + K_2(\theta)G_2^{45^\circ} + K_3(\theta)G_2^{90^\circ}. \tag{17}$$

According to Adelson and Freeman's second theorem, the interpolation functions are solutions of the following system:

$$\begin{pmatrix} 1 \\ e^{i2\theta} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ e^{i2\theta_1} & e^{i2\theta_2} & e^{i2\theta_3} \end{pmatrix} \begin{pmatrix} K_1(\theta) \\ K_2(\theta) \\ K_3(\theta) \end{pmatrix}. \tag{18}$$

Considering the real and imaginary parts this system becomes:

$$\begin{cases} 1 = K_1 + K_2 + K_3 \\ \cos 2\theta = K_1 \cos 2\theta_1 + K_2 \cos 2\theta_2 + K_3 \cos 2\theta_3 \\ \sin 2\theta = K_1 \sin 2\theta_1 + K_2 \sin 2\theta_2 + K_3 \sin 2\theta_3 \end{cases} \tag{19}$$

For the previous choices for θ_1, θ_2 and θ_3 as $0^\circ, 45^\circ$ and 90° the system becomes:

$$\begin{cases} 1 = K_1 + K_2 + K_3 \\ \cos 2\theta = K_1 - K_3 \\ \sin 2\theta = K_2 \end{cases} \tag{20}$$

and resolution of this simple system produces

$$\begin{cases} K_1 = \cos\theta(\cos\theta - sin\theta) \\ K_2 = \sin 2\theta = 2\sin\theta\cos\theta \\ K_3 = \sin\theta(\sin\theta - \cos\theta) \end{cases} \tag{21}$$

so that given these weighting functions and Equation 17 the convolution of the input image with the tuned second derivative is

$$(G_2^{\theta^\circ} * I) = K_1(\theta)(G_2^{0^\circ} * I) + K_2(\theta)(G_2^{45^\circ} * I) + K_3(\theta)(G_2^{90^\circ} * I). \tag{22}$$

This equation shows that it is just necessary to convolve the input image with the basis filters and to multiply the results by the interpolation functions to synthesize the response at any orientation.

Step 2. When smoothing with a second derivative, the edges occur at the locations of the zero crossings. The orientation of the contour is given by the "best" response of the filter tuned to a certain orientation. But for a zero-crossing the response is not an extrema. Consequently the important problem is how to find the "best zero crossing"? Should the gradient about the zero-crossing be used as an estimate of the zero-crossings strength and for subsequent thresholding?

To address the above questions, Adelson and Freeman have proposed the use of the Hilbert Transform. They have defined the notion of **"energy filters"** as being the summation of the squared second derivative of the Gaussian tuned to any angle and the squared Hilbert transform of this second derivative tuned to any angle:

$$E_2(\theta) = [G_2^\theta]^2 + [H_2^\theta]^2. \tag{23}$$

We will now review some definitions and properties of the Hilbert Transform of a function. The Hilbert Transform is a filter with transfer function $-i\ \mathrm{sgn}(f)$. where $\mathrm{sgn}(f)$ is the function

$$\mathrm{sgn}(f) = \begin{cases} +1 & \text{if } f \geq 0 \\ -1 & \text{otherwise} \end{cases}.$$

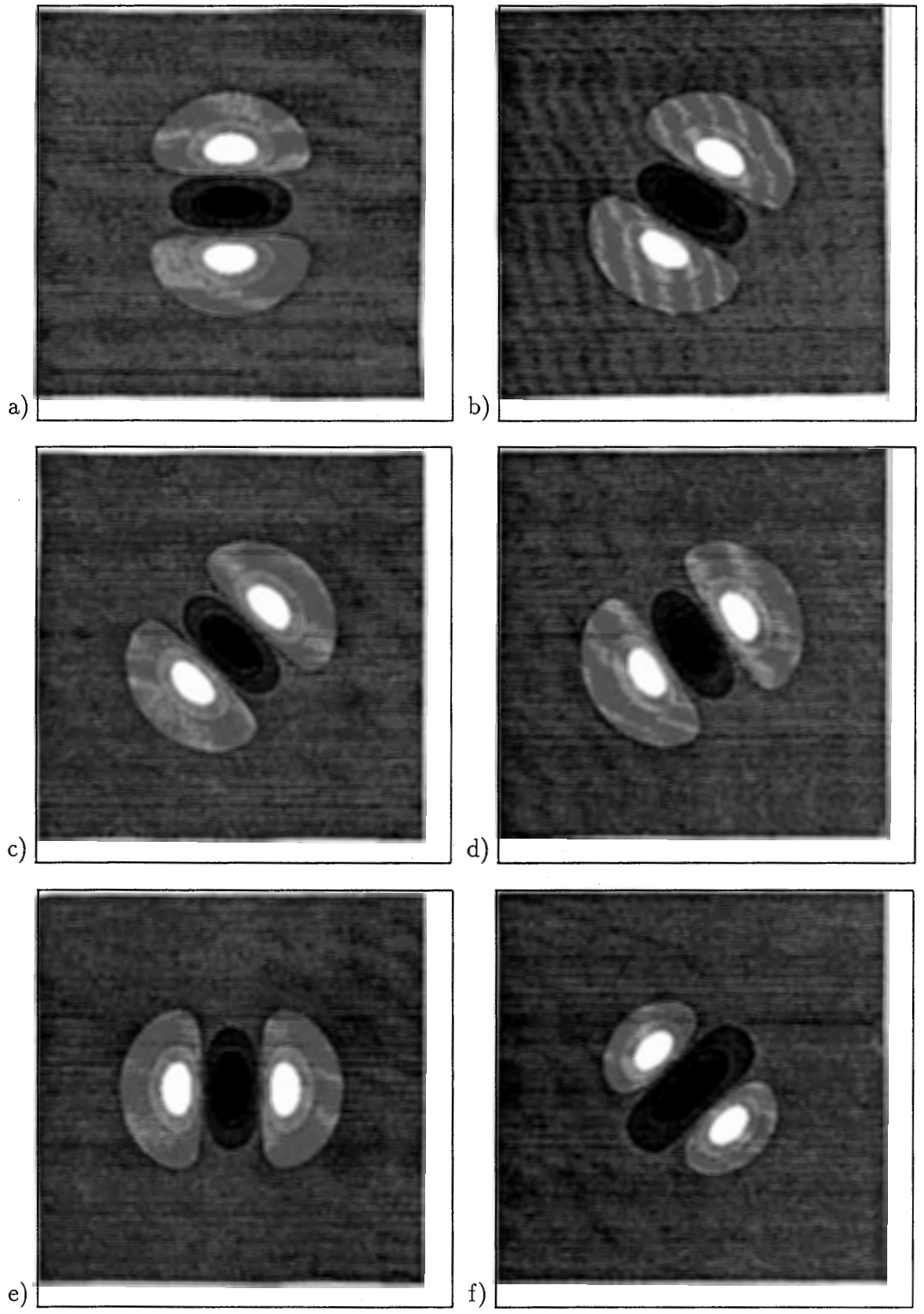

Figure 6: Second derivative of a Gaussian tuned to any direction: a) 0°, b) 30°, c) 45°, d) 60°, e) 90°, f) 135°.

As the Inverse Fourier Transform of $-i\ \mathrm{sgn}(f)$ is $\mathrm{vp}(1/\pi x)$, taking the Hilbert Transform of a function g is convolving it with $\mathrm{vp}(1/\pi x)$ in the spatial domain:

$$HT(g(x)) = \hat{g}(x) = g(x) * \mathrm{vp}\left(\frac{1}{\pi x}\right) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{g(\tau)}{x-\tau} d\tau. \tag{24}$$

The notation vp corresponds to the principal value according to Cauchy, that is

$$< \mathrm{vp}(\frac{1}{t}), \phi >= \lim_{\varepsilon \to 0^-} \int_{-\infty}^{\varepsilon} \frac{\phi(t)}{t} dt + \lim_{\varepsilon \to 0^+} \int_{\varepsilon}^{+\infty} \frac{\phi(t)}{t} dt, \tag{25}$$

and the integral of equation 24 must be considered as a Cauchy's integral:

$$\int_{-\infty}^{+\infty} \frac{g(\tau)}{x-\tau} d\tau = \lim_{\varepsilon \to 0^-} \int_{-\infty}^{\varepsilon+x} \frac{g(\tau)}{x-\tau} d\tau + \lim_{\varepsilon \to 0^+} \int_{\varepsilon+x}^{+\infty} \frac{g(\tau)}{x-\tau} d\tau. \tag{26}$$

The Hilbert transform has an interesting property: a filter and its Hilbert transform constitute a *quadrature pair of filters.* The two filters have the same frequency response and they are in quadrature. For example if we consider the second derivative of the Gaussian along the x axis and its Hilbert transform, we show in Figure 7 that their respective Fourier transforms have the same magnitude.

Adelson and Freeman have used this quadrature pair of filters to detect edges; after having computed the dominant orientation θ_d which gives the maximum response for the quadrature pair (G_2^θ, H_2^θ) (where H_2^θ is the Hilbert transform of the second derivative of the Gaussian G_2^θ tuned to any angle θ), they have computed the value of the energy for this dominant orientation:

$$E_2(\theta_d) = [G_2^{\theta_d}]^2 + [H_2^{\theta_d}]^2 \tag{27}$$

They have shown that a point is a point of the edge if the energy $E^2(\theta_d)$ is a local maximum in the direction perpendicular to the local orientation θ_d and they have used Canny's algorithm to threshold the maxima and to enhance the edges.

Also, they have shown that it is possible to make a good approximation of the Hilbert transform of the second derivative of the Gaussian by using a third order polynomial times a Gaussian. They applied their theorems to this approximation of the Hilbert transform and found that four basis filters are necessary to synthesize the

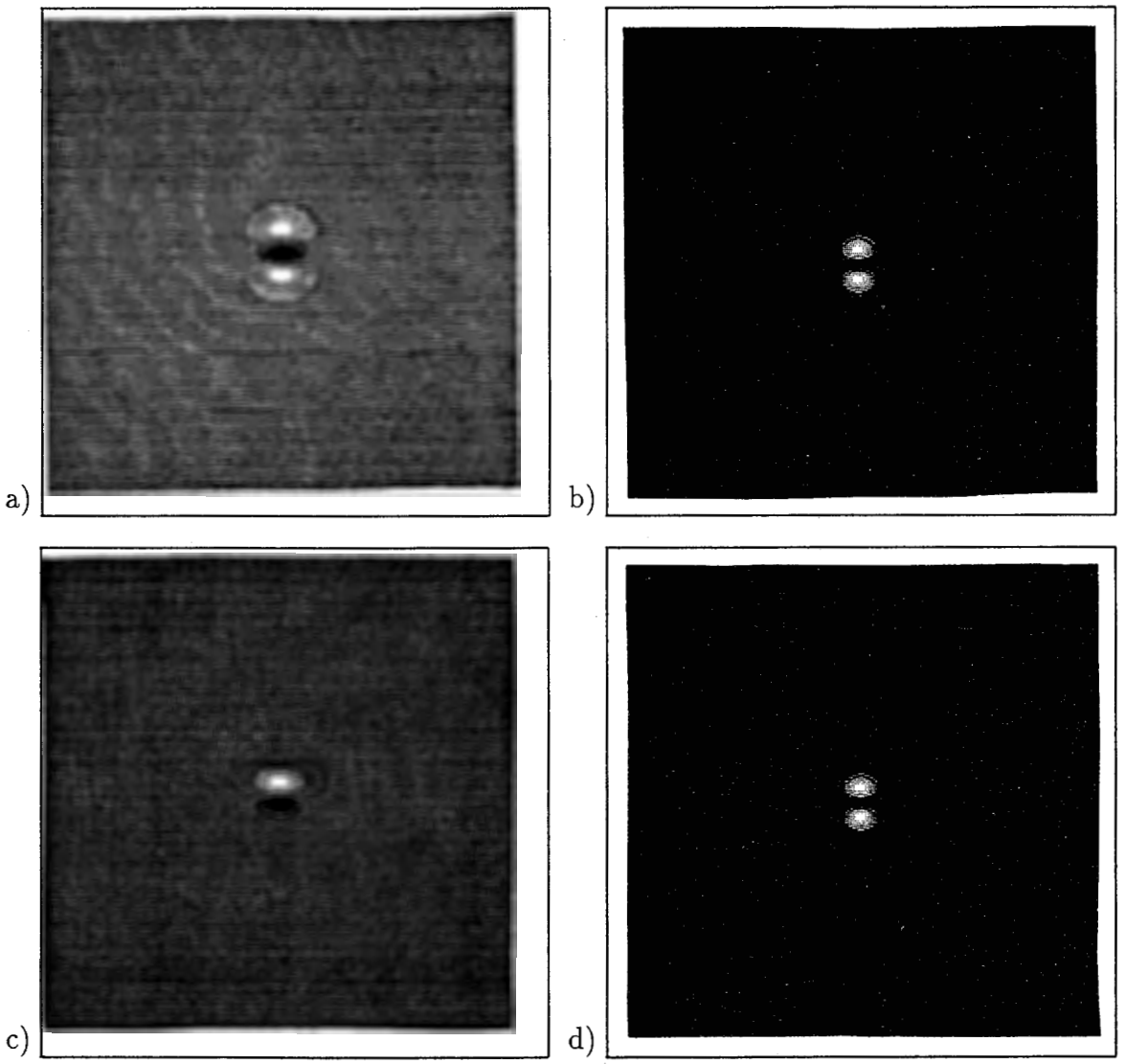

Figure 7: The second derivative of a Gaussian and its Hilbert transform have the same energy spectrum. a) Second derivative of a Gaussian along the x axis, b) its Hilbert transform, c) the magnitude of the Fourier transform of the seconde derivative of the Gaussian (a), and d) the magnitude of the Fourier transform of the Hilbert transform.

Hilbert transform of the second derivative of the Gaussian along any angle. These basis filters are rotated versions of the Hilbert transform. The interpolation functions can be computed with Adelson and Freeman's second theorem.

In using the quadrature pair of filters, we get rid of our previous problem of "best zero crossings" detection; now as Adelson and Freeman have shown, the edges occur at the location of the local maxima of the energy for the maximum orientation; so it

is just an usual problem of maxima detection.

As we have shown previously, the second derivative of the Gaussian can be synthesize as a linear combination of three basis filters given in Equation 17. According to Adelson and Freeman, the Hilbert transform can be synthesize to any angle as a linear combination of four basis filters:

$$H^{\theta^\circ} = J_1(\theta)H^{0^\circ} + J_2(\theta)H^{45^\circ} + J_3(\theta)H^{90^\circ} + J_4(\theta)H^{135^\circ} \tag{28}$$

These basis filters, the Hilbert transform of the second derivative of the Gaussian along $0°, 45°, 90°$ and $135°$ can be computed in finding mathematically the expression of the Hilbert Transform of the second derivative of the Gaussian in polar coordinates (r, ϕ), so in all the space, and to compute it for certain values of ϕ ($\phi = 0°, 45°, 90°, 135°$).

Inspection of Figure 7 shows that the Hilbert transform is very similar to the first derivative of a Gaussian. In the vicinity of an edge the dominant term in the energy measure of Equation 27 will be very similar to the first derivative; the second derivative is small, in fact it is a zero-crossing. Unlike Freeman and Adelson, if the phase information is not used, then the first derivative provides a suitable technique for oriented edge detection; fewer basis filters are required and computation of the Hilbert transform is not required. In this early analysis, the use of Hilbert transforms has not be further employed.

6 Oriented filters in motion analysis

Motion information is of the greatest importance in image segmentation, reconstruction of object shape, scene analysis, moving objects tracking, etc. The determination of this information does not depend on the knowledge of the shape of the object but only on the changes of intensity over time in image sequences. The dynamic scene segmentation consists in partitioning the image into moving regions and stationary ones.

Some physiological studies have shown the existence of directionally selective cells in the primary visual cortex. These cells respond to stimuli moving in one particular

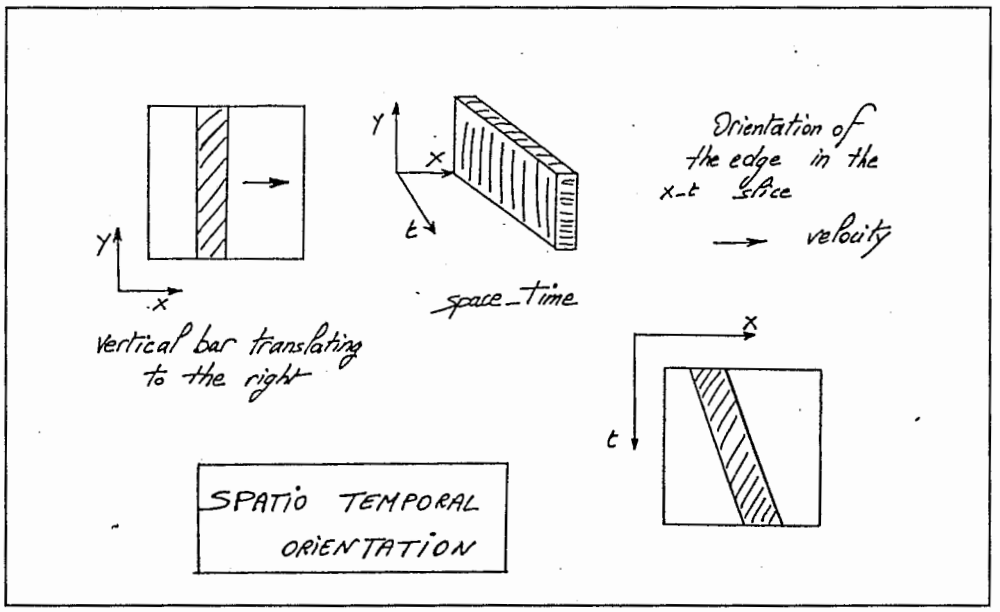

Figure 8: A moving bar in 2 dimensions corresponds to a slanted slab in 3 dimensions and its velocity is the slope of the edge of the object

direction regardless of the sign of the contrast (if it is a directionally asymmetric cell, it's response will depend on the sign of the contrast). Shapley, Reid and Soodak[7] have shown that these cells can be considered as linear spatiotemporal transducers but they have also shown the existence of non-linear spatial interactions.

Adelson and Bergen[3] as well as Watson and Ahumada[8] have been working on spatiotemporal linear systems. The main idea of their work is that **motion is characterized by orientation in space time domain**: x and y for the two spatial dimensions plus t for the temporal dimension. If we consider a bar moving to the right along the x axis as shown in Figure 8, in the space time domain it is similar to a slanted slab; the slope of the edges in the (x,t) domain corresponds to the velocity of the bar along the x axis and the slope of the edges in the (y,t) domain corresponds to the velocity of the bar along y axis (which is zero in our example).

So finding the velocity of the moving object is finding the orientation of the edges in the spatiotemporal domain. We just have to design a spatiotemporal impulse response as we did in the spatial domain. If the spatiotemporal response can be expressed as a temporal impulse response times a spatial impulse response, the spatiotemporal response is separable; it is the case in most of the cells in the cortical area.

Adelson and Bergen[3] used **oriented Gabor filters** because these filters can

detect rightward motion to leftward motion (but any other kind of oriented function can be used so in particular the oriented filters described previously; but as we review their research, we consider Gabor filters. All the following considerations can be applied to the oriented filters described in the previous section).

Adelson and Bergen[3] defined the notion of **spatiotemporal energy**; this energy is the sum of the squared outputs of two Gabor filters in quadrature, one sine phase filter and one cosine filter. The expression of a one-dimensional sine phase Gabor filter is given in the following formula; it is just a sine wave times a Gaussian window (same with the cosine wave instead of the sine wave in the case of the even Gabor filter):

$$g(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{--t^2/2\sigma^2}. \tag{29}$$

In 3 dimension this equation becomes:

$$g(x,y,t) = \frac{1}{(2\pi)^{3/2}\sigma_x\sigma_y\sigma_t} e^{-(\frac{x^2}{2\sigma_x^2}+\frac{y^2}{2\sigma_y^2}+\frac{t^2}{2\sigma_t^2})} \sin(2\pi\omega_{x_0}x + 2\pi\omega_{y_0}y + 2\pi\omega_{t_0}t). \tag{30}$$

The *spatiotemporal oriented filter* responds to one direction of motion and its response oscillates with the contrast. The *spatiotemporal oriented energy* responds to one direction of motion but its response does not oscillate with the contrast and the **opponent energy** which is a combination of leftward and rightward oriented filters response gives an output for both direction regardless of the contrast.

These energy models can not resolve **the aperture problem.** In such a case (see Figure 9), when we visualize only a part of a moving object without having any information about the corners, the extremities of it, the possible values for its velocity constitute a plane; so an infinite number of different orientations in the space time domain.

A very basic approach of the first level of motion detection can consist in applying an "edge detector procedure" in the space-time domain. In an image sequence containing a moving object, along t axis, the intensity of some points in the image change (as it does for a boundary in space). As it is possible to detect edges in an spatial image, it is possible to detect moving part and stationary parts in an sequence of images. The procedure consists in convolving the sequence along the time axis with a low pass causal filter, take the derivative and find the maxima exactly as we did in

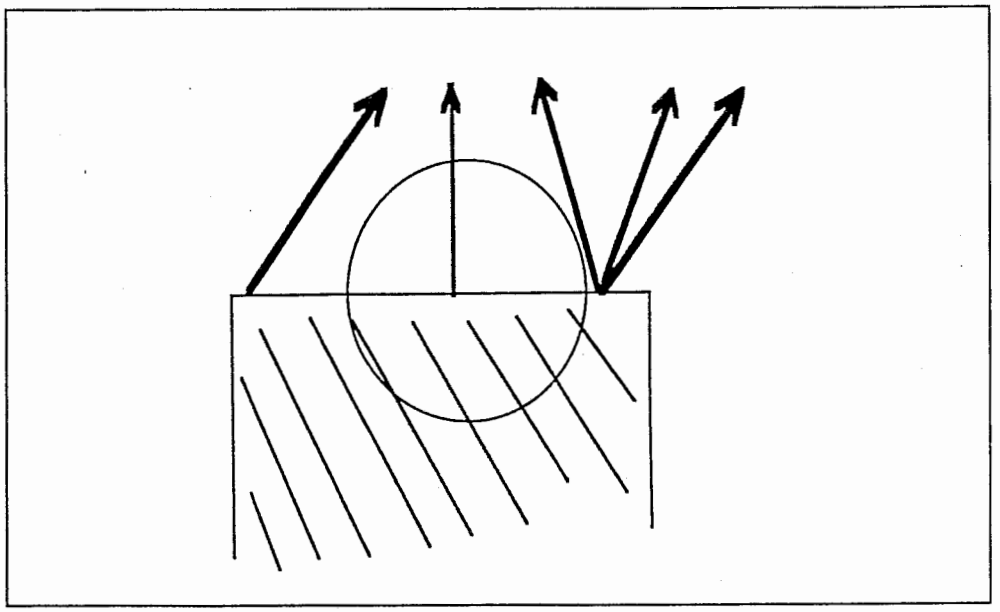

Figure 9: An illustration of the aperture problem. In a small region only the perpendicular component of velocity can be determined for the motion of a straight edge.

the spatial domain. A second step will consist in using the orientation determination and in trying to resolve the ill posed aperture problem.

References

[1] E. H. Adelson and W. T. Freeman. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906, September 1991.

[2] R. A. Anderson and R. M. Siegel. Motion processing in primate cortex. In G. M. Edelman, W. E. Gall, and W. M. Cowan, editors, *Signal and Sense: Local and Global Order in Perceptual Maps*, pages 163–184. New York, 1990.

[3] E.H. Adelson and J.R. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America*, 2(2):284–299, 1985.

[4] Tomaso Poggio and Werner Reichardt. Visual control of orientation behaviour in the fly: Part II: Towards the underlying neural interactions. *Quart. Rev. Biophysics*, 9:377–438, 1976.

[5] John F. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[6] E. H. Adelson and W. T. Freeman. The design and use of steerable filter for image analysis, enhancement and wavelet representation. Technical report, M.I.T. Media Lab Vision Modelling Group and Department of Brain and Cognitive Sciences, M.I.T., Cambridge, Massachusetts 02139, September 1990.

[7] R. Soodak, R. C. Reid, and R. Shapley. Spatiotemporal receptive fields and orientation selectivity.

[8] A.B. Watson and A.J. Ahumada. Model of human visual-motion sensing. *Journal of the Optical Society of America*, 2(7), 1985.