14 Motion

14.1 Introduction

Motion analysis long used to be a specialized research area that had not much to do with general image processing. This separation had two reasons. First, the techniques used to analyze motion in image sequences were quite different. Second, the large amount of storage space and computing power required to process image sequences made image sequence analysis available only to a few specialized institutions that could afford to buy the expensive specialized equipment. Both reasons are no longer true. Because of the general progress in image processing, the more advanced methods used in motion analysis no longer differ from those used for other image processing tasks. The rapid progress in computer hardware and algorithms makes the analysis of image sequences now feasible even on standard personal computers and workstations.

Therefore we treat motion in this chapter as just another feature that can be used to identify, characterize, and distinguish objects and to understand scenes. Motion is indeed a powerful feature. We may compare the integration of motion analysis into mainstream image processing with the transition from still photography to motion pictures. Only image sequence analysis allows us to recognize and analyze dynamic processes. Thus far-reaching capabilities become available for scientific and engineering applications including the study of flow; transport; biological growth processes from the molecular to the ecosystem level; diurnal, annual, and interannual variations; industrial processes; traffic; autonomous vehicles and robots — to name just a few application areas. In short, everything that causes temporal changes or makes them visible in our world is a potential subject for image sequence analysis.

The analysis of motion is still a challenging task and requires some special knowledge. Therefore we discuss the basic problems and principles of motion analysis in Section 14.2. Then we turn to the various techniques for motion determination. As in many other areas of image processing, the literature is swamped with a multitude of approaches. This book should not add to the confusion. We emphasize instead the basic principles and we try to present the various concepts in a unified way as filter operations on the space-time images. In this way, the interrelations between the different concepts are made transparent.

B. Jähne, Digital Image Processing
ISBN 3-540-24035-7

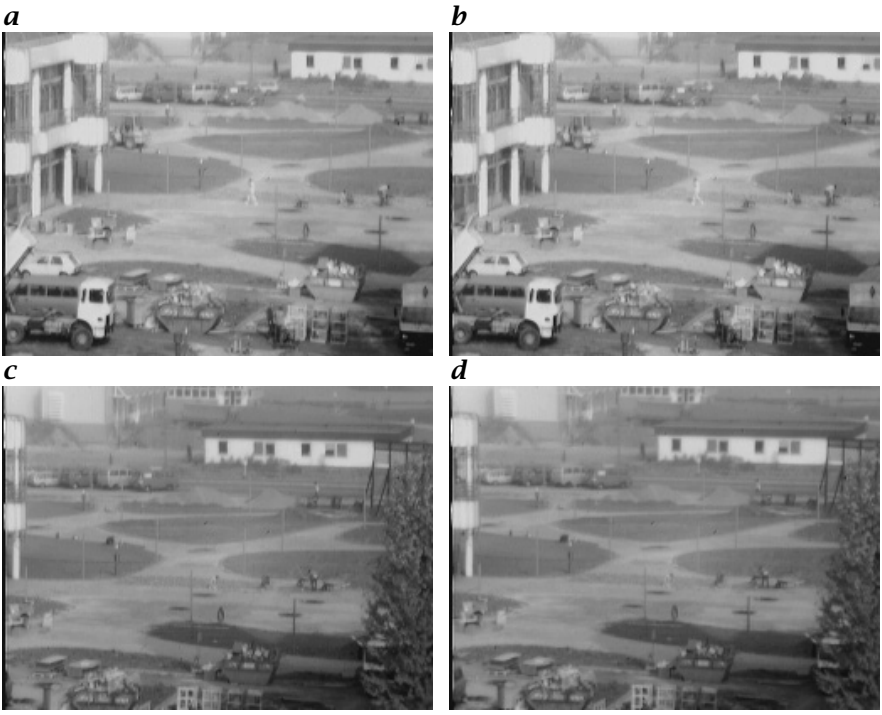

Figure 14.1: a–d *Two pairs of images from the construction area for the new head clinic at Heidelberg University. What has changed from the left to the right images?*

In this sense, we will discuss differential (Section 14.3), tensor (Section 14.4), correlation (Section 14.5), and phase (Section 14.6) techniques as elementary motion estimators.

14.2 Basics

14.2.1 Motion and Gray Value Changes

Intuitively we associate motion with changes. Thus we start our discussion on motion analysis by observing the differences between two images of a sequence. Figure 14.1a and b shows an image pair of a construction area at Heidelberg University. There are differences between the left and right images which are not evident from direct comparison. However, if we subtract one image from the other, the differences immediately become visible (Fig. 14.3a). In the lower left of the image a truck has moved, while the car just behind it is obviously parked. In the center of the image we discover the outline of a pedestrian which is barely visible

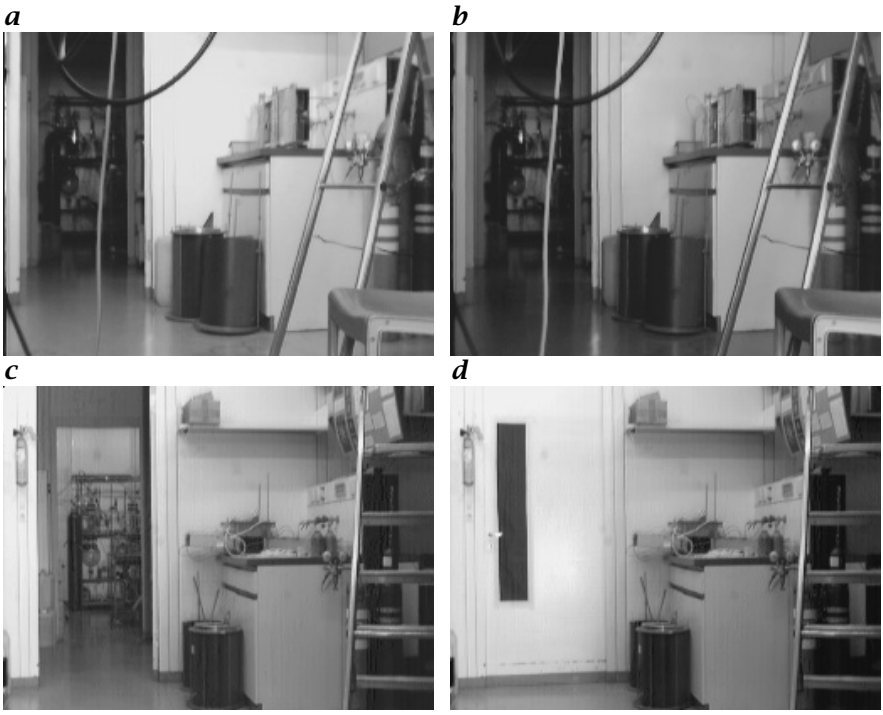

***Figure 14.2: a** to **d** Two pairs of images from an indoor lab scene. What changes can be seen between the left and right images?*

in the original images. The bright spots in a row at the top of the image turn out to be bikers moving along a cycle lane. From the displacement of the double contours we can estimate that they move faster than the pedestrian. Even from this qualitative description, it is obvious that motion analysis helps us considerably in understanding such a scene. It would be much harder to detect the cycle lane without observing the moving bikers.

Figure 14.1c and d show the same scene. Now we might even recognize the change in the original images. If we observe the image edges, we notice that the images have shifted slightly in a horizontal direction. What has happened? Obviously, the camera has been panned. In the difference image Fig. 14.3b all the edges of the objects appear as bright lines. However, the image is dark where the spatial gray value changes are small. Consequently, we can detect motion only in the parts of an image that show gray value changes. This simple observation points out the central role of spatial gray value changes for motion determination.

So far we can sum up our experience with the statement that motion *might* result in temporal gray value changes. Unfortunately, the reverse

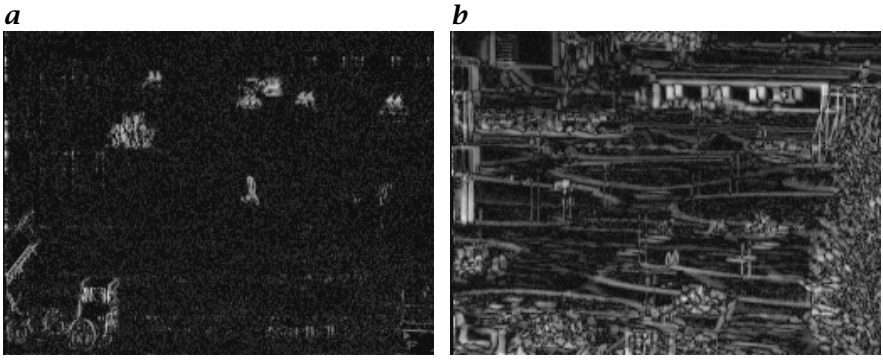

Figure 14.3: *Magnitude of the difference between* ***a*** *images* ***a*** *and* ***b*** *in Fig. 14.1;* ***b*** *images* ***c*** *and* ***d*** *in Fig. 14.1.*

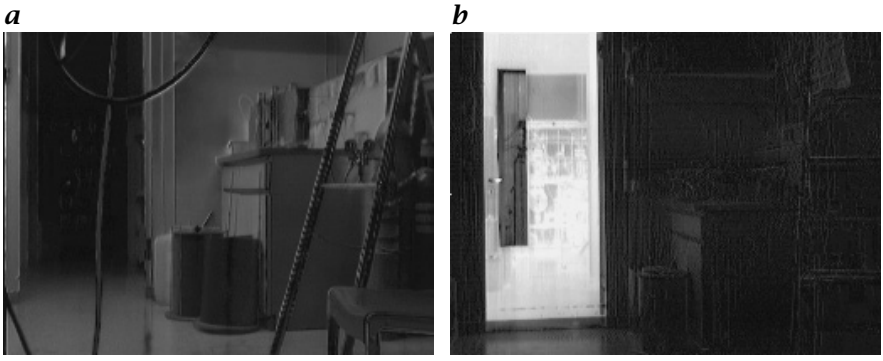

Figure 14.4: *Difference between* ***a*** *images* ***a*** *and* ***b*** *in Fig. 14.2;* ***b*** *images* ***c*** *and* ***d*** *in Fig. 14.2.*

conclusion that all temporal gray value changes are due to motion is not correct. At first glance, the pair of images in Fig. 14.2a and b look identical. Yet, the difference image Fig. 14.4a reveals that some parts in the upper image are brighter than the lower. Obviously the illumination has changed. Actually, a lamp outside the image sector shown was switched off before the image in Fig. 14.2b was taken. Can we infer where this lamp is located? In the difference image we notice that not all surfaces are equally bright. Surfaces which are oriented towards the camera show about the same brightness in both images, while surfaces facing the left hand side are considerably brighter. Therefore we can conclude that the lamp is located to the left outside of the image sector.

Another pair of images (Fig. 14.2c and d) shows a much more complex scene, although we did not change the illumination. We just closed the door of the lab. Of course, we see strong gray value differences where the door is located. The gray value changes, however, extend to the

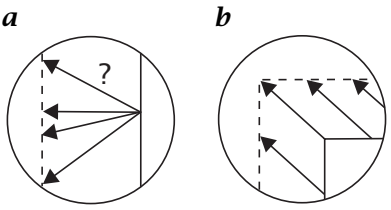

Figure 14.5: *Illustration of the aperture problem in motion analysis:* ***a*** *ambiguity of displacement vectors at an edge;* ***b*** *unambiguity of the displacement vector at a corner.*

floor close to the door and to the objects located to the left of the door (Fig. 14.4b). As we close the door, we also change the illumination in the proximity of the door, especially below the door because less light is reflected into this area.

14.2.2 The Aperture Problem

So far we have learned that estimating motion is closely related to spatial and temporal gray value changes. Both quantities can easily be derived with local operators that compute the spatial and temporal derivatives. Such an operator only "sees" a small sector — equal to the size of its mask — of the observed object. We may illustrate this effect by putting a mask or aperture onto the image.

Figure 14.5a shows an edge that moved from the position of the solid line in the first image to the position of the dotted line in the second image. The motion from image one to two can be described by a *displacement vector*, or briefly, *DV*. In this case, we cannot determine the displacement unambiguously. The displacement vector might connect one point of the edge in the first image with any other point of the edge in the second image (Fig. 14.5a). We can only determine the component of the DV normal to the edge, while the component parallel to the edge remains unknown. This ambiguity is known as the *aperture problem*.

An unambiguous determination of the DV is only possible if a corner of an object is within the mask of our operator (Fig. 14.5b). This emphasizes that we can only gain sparse information on motion from local operators.

14.2.3 The Correspondence Problem

The aperture problem is caused by the fact that we cannot find the corresponding point at an edge in the following image of a sequence, because we have no means of distinguishing the different points at an edge. In this sense, we can comprehend the aperture problem only as a special case of a more general problem, the *correspondence problem*. Generally

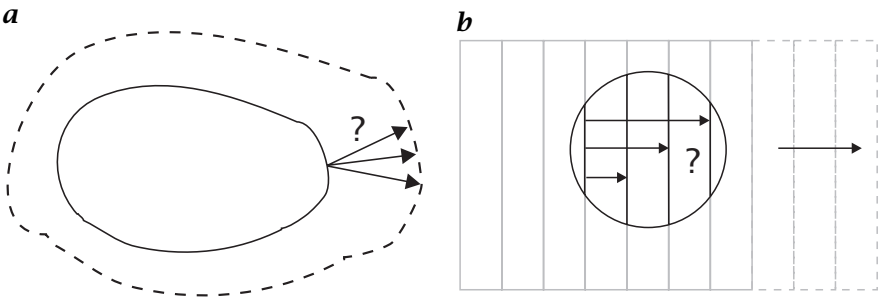

Figure 14.6: *Illustration of the correspondence problem:* ***a*** *deformable two-dimensional object;* ***b*** *regular grid.*

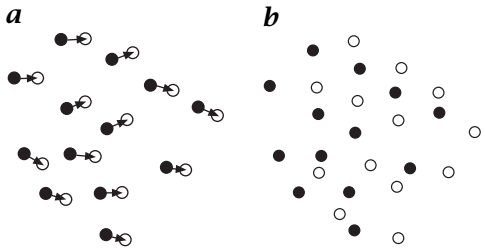

Figure 14.7: *Correspondence problem with indistinguishable particles:* ***a*** *mean particle distance is larger than the mean displacement vector;* ***b*** *the reverse case. Filled and hollow circles: particles in the first and second image.*

speaking, this is that we are unable to find unambiguously corresponding points in two consecutive images of a sequence. In this section we discuss further examples of the correspondence problem.

Figure 14.6a shows a two-dimensional deformable object — like a blob of paint — which spreads gradually. It is immediately obvious that we cannot obtain any unambiguous determination of displacement vectors, even at the edge of the blob. In the inner part of the blob, we cannot make any estimate of the displacements because there are no features visible which we could track.

At first we might assume that the correspondence problem will not occur with rigid objects that show a lot of gray value variations. The grid as an example of a periodic texture, shown in Fig. 14.6b, demonstrates that this is not the case. As long as we observe the displacement of the grid with a local operator, we cannot differentate displacements that differ by multiples of the grid constant. Only when we observe the whole grid does the displacement become unambiguous.

Another variation of the correspondence problem occurs if the image includes many objects of the same shape. One typical case is when small particles are put into a flow field in order to measure the velocity field

(Fig. 14.7). In such a case the particles are indistinguishable and we generally cannot tell which particles correspond to each other. We can find a solution to this problem if we take the consecutive images at such short time intervals that the mean displacement vector is significantly smaller than the mean particle distance. With this additional knowledge, we can search for the nearest neighbor of a particle in the next image. Such an approach, however, will never be free of errors, because the particle distance is statistically distributed.

These simple examples clearly demonstrate the basic problems of motion analysis. On a higher level of abstraction, we can state that the *physical correspondence*, i. e., the real correspondence of the real objects, may not be identical to the *visual correspondence* in the image. The problem has two faces. First, we can find a visual correspondence without the existence of a physical correspondence, as in case of objects or periodic object textures that are indistinguishable. Second, a physical correspondence does not generally imply a visual correspondence. This is the case if the objects show no distinctive marks or if we cannot recognize the visual correspondence because of illumination changes.

14.2.4 Motion as Orientation in Space-Time Images

The discussion in Sections 14.2.1–14.2.3 revealed that the analysis of motion from only two consecutive images is plagued by serious problems. The question arises, whether these problems, or at least some of them, can be overcome if we extend the analysis to more than two consecutive images. With two images, we get just a "snapshot" of the motion field. We do not know how the motion continues in time. We cannot measure accelerations and cannot observe how parts of objects appear or disappear as another object moves in front of them.

In this section, we consider the basics of image sequence analysis in a multidimensional space spanned by one time and one to three space coordinates. Consequently, we speak of a *space-time image*, a *spatiotemporal image*, or simple the $\boldsymbol{xt}$ space.

We can think of a three-dimensional space-time image as a stack of consecutive images which may be represented as an *image cube* as shown in Fig. 14.9. At each visible face of the cube we map a cross section in the corresponding direction. Thus an xt slice is shown on the top face and a yt slice on the right face of the cube. The slices were taken at depths marked by the white lines on the front face, which shows the last image of the sequence.

In a space-time image a pixel extends to a *voxel*, i. e., it represents a gray value in a small volume element with the extensions Δx, Δy, and Δt. Here we confront the limits of our visual imagination when we try to grasp truly 3-D data (compare the discussion in Section 8.1.1).

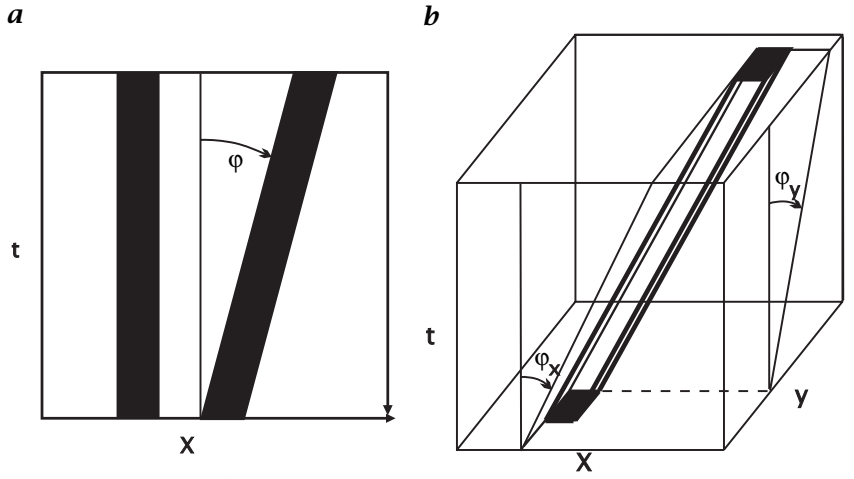

Figure 14.8: *Space-time images:* ***a*** *two-dimensional space-time image with one space and one time coordinate;* ***b*** *three-dimensional space-time image.*

Therefore, we need appropriate representations of such data to make essential features of interest visible.

To analyze motion in space-time images, we first consider a simple example with one space and one time coordinate (Fig. 14.8a). A non-moving 1-D object shows vertically oriented gray value structures. If an object is moving, it is shifted from image to image and thus shows up as an inclined gray value structure. The velocity is directly linked to the orientation in space-time images. In the simple case of a 2-D space-time image, it is given by

$$u = -\tan\varphi, \tag{14.1}$$

where φ is the angle between the t axis and the direction in which the gray values are constant. The minus sign in Eq. (14.1) is because angles are positive counterclockwise. The extension to two spatial dimensions is straightforward and illustrated in Fig. 14.8b:

$$\boldsymbol{u} = -\begin{bmatrix} \tan\varphi_x \\ \tan\varphi_y \end{bmatrix}. \tag{14.2}$$

The angles φ_x and φ_y are defined analogously to the angle between the x and y components of a vector in the direction of the constant gray values and the t axis.

A practical example for this type of analysis is shown in Fig. 14.9. The motion is roughly in the vertical direction, so that the yt cross section can be regarded as a 2-D space-time image. The motion is immediately apparent. When the cars stop at the traffic light, the lines are horizontally

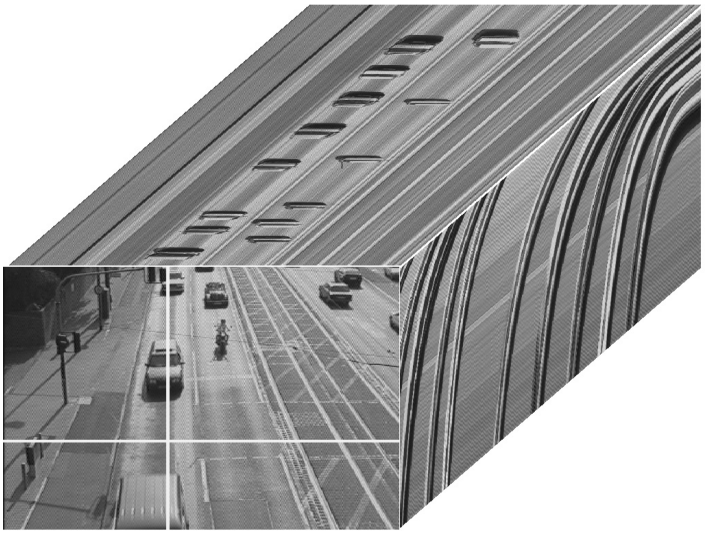

Figure 14.9: *A 3-D image sequence demonstrated with a traffic scene in the Hanauer Landstraße, Frankfurt/Main represented as an image cuboid. The time axis runs into the depth, pointing towards the viewer. On the right side of the cube a yt slice marked by the vertical white line in the xy image is shown, while the top face shows an xt slice marked by the horizontal line (from Jähne [88]).*

oriented, and phases with accelerated and constant speed can easily be recognized.

In summary, we come to the important conclusion that motion appears as *orientation* in space-time images. This fundamental fact forms the basis for motion analysis in $\boldsymbol{xt}$ space. The basic conceptual difference to approaches using two consecutive images is that the velocity is estimated *directly as orientation* in continuous space-time images and not as a discrete *displacement*.

These two concepts differ more than it appears at first glance. Algorithms for motion estimation can now be formulated in continuous $\boldsymbol{xt}$ space and studied *analytically* before a suitable discretization is applied. In this way, we can clearly distinguish the principal flaws of an approach from errors induced by the discretization.

Using more than two images, a more robust and accurate determination of motion can be expected. This is a crucial issue for scientific applications, as pointed out in Chapter 1.

This approach to motion analysis has much in common with the problem of reconstruction of 3-D images from projections (Section 8.6). Actually, we can envisage a geometrical determination of the velocity by observing the transparent three-dimensional space-time image from different points of view. At the right observation angle, we look along the

edges of the moving object and obtain the velocity from the angle between the observation direction and the time axis.

If we observe only the edge of an object, we cannot find such an observation angle unambiguously. We can change the component of the angle along the edge arbitrarily and still look along the edge. In this way, the *aperture problem* discussed in Section 14.2.2 shows up from a different point of view.

14.2.5 Motion in Fourier Domain

Introducing the space-time domain, we gain the significant advantage that we can analyze motion also in the corresponding Fourier domain, the $\boldsymbol{k}\nu$ space. As an introduction, we consider the example of an image sequence in which all the objects are moving with constant velocity. Such a sequence $g(\boldsymbol{x}, t)$ can be described by

$$g(\boldsymbol{x}, t) = g(\boldsymbol{x} - \boldsymbol{u}t). \tag{14.3}$$

The Fourier transform of this sequence is

$$\hat{g}(\boldsymbol{k}, \nu) = \int_t \int_{\boldsymbol{x}} g(\boldsymbol{x} - \boldsymbol{u}t) \exp[-2\pi \mathrm{i}(\boldsymbol{k}\boldsymbol{x} - \nu t)] \mathrm{d}^2x \mathrm{d}t. \tag{14.4}$$

Substituting

$$\boldsymbol{x}' = \boldsymbol{x} - \boldsymbol{u}t,$$

we obtain

$$\hat{g}(\boldsymbol{k}, \nu) = \int_t \left[\int_{\boldsymbol{x}'} g(\boldsymbol{x}') \exp(-2\pi \mathrm{i}\boldsymbol{k}\boldsymbol{x}') \right] \exp(-2\pi \mathrm{i}\boldsymbol{k}\boldsymbol{u}t) \exp(2\pi \mathrm{i}\nu t) \mathrm{d}^2x' \mathrm{d}t.$$

The inner integral covers the spatial coordinates and results in the spatial Fourier transform $\hat{g}(\boldsymbol{k})$ of the image $g(\boldsymbol{x}')$. The outer integral over the time coordinate reduces to a δ function:

$$\hat{g}(\boldsymbol{k}, \nu) = \hat{g}(\boldsymbol{k}) \delta(\boldsymbol{k}\boldsymbol{u} - \nu). \tag{14.5}$$

This equation states that an object moving with the velocity $\boldsymbol{u}$ occupies only a two-dimensional subspace in the three-dimensional $\boldsymbol{k}\nu$ space. Thus it is a line and a plane, in two and three dimensions, respectively. The equation for the plane is given directly by the argument of the δ function in Eq. (14.5):

$$\nu = \boldsymbol{k}\boldsymbol{u}. \tag{14.6}$$

This plane intersects the $k_1 k_2$ plane normally to the direction of the velocity because in this direction the inner product $\boldsymbol{k}\boldsymbol{u}$ vanishes. The slope of the plane, a two-component vector, yields the velocity

$$\nabla_k \nu = \nabla_k (\boldsymbol{k}\boldsymbol{u}) = \boldsymbol{u}.$$

The index k in the gradient operator denotes that the partial derivations are computed with respect to the components of $\boldsymbol{k}$.

From these considerations, it is obvious — at least in principle — how we can determine the velocity in an image sequence showing a constant velocity. We compute the Fourier transform of the sequence and then determine the slope of the plane on which the spectrum of the sequence is located. We can do this best if the scene contains small-scale structures, i. e., high wave numbers which are distributed in many directions. We cannot determine the slope of the plane unambiguously if the spectrum lies on a line instead of a plane. This is the case when the gray value structure is spatially oriented. From the line in Fourier space we only obtain the component of the plane slope in the direction of the spatial local orientation. In this way, we encounter the *aperture problem* (Section 14.2.2) in the $\boldsymbol{k}\nu$ space.

14.2.6 Optical Flow

The examples discussed in Section 14.2.1 showed that motion and gray value changes are not equivalent. In this section, we want to quantify this relation. In this respect, two terms are of importance: the *motion field* and the *optical flow*. The motion field in an image is the real motion of the object in the 3-D scene projected onto the image plane. It is the quantity we would like to extract from the image sequence. The optical flow is defined as the "flow" of gray values at the image plane. This is what we observe. Optical flow and motion field are only equal if the objects do not change the irradiance on the image plane while moving in a scene. Although it sounds reasonable at first glance, a more thorough analysis shows that it is strictly true only in very restricted cases. Thus the basic question is how significant the deviations are, so that in practice we can still stick with the equivalence of optical flow and motion field.

Two classical examples where the projected motion field and the optical flow are not equal were given by Horn [81]. The first is a spinning sphere with a uniform surface of any kind. Such a sphere may rotate around any axes through its center of gravity without causing an optical flow field. The counterexample is the same sphere at rest illuminated by a moving light source. Now the motion field is zero, but the changes in the gray values due to the moving light source cause a non-zero optical flow field.

At this point it is helpful to clarify the different notations for motion with respect to image sequences, as there is a lot of confusion in the literature and many different terms are used. *Optical flow* or *image flow* means the apparent motion at the image plane based on visual perception and has the dimension of a velocity. We denote the optical flow with $\boldsymbol{f} = [f_1, f_2]^T$. If the optical flow is determined from two consecutive images, it appears as a *displacement vector* (*DV*) from the features in the

first to those in the second image. A dense representation of displacement vectors is known as a *displacement vector field* (*DVF*) $\boldsymbol{s} = [s_1, s_2]^T$. An approximation of the optical flow can be obtained by dividing the DVF by the time interval between the two images. It is important to note that optical flow is a concept inherent to continuous space, while the displacement vector field is its discrete counterpart. The *motion field* $\boldsymbol{u} = [u_1, u_2]^T = [u, v]^T$ at the image plane is the projection of the 3-D physical motion field by the optics onto the image plane.

The concept of optical flow originates from *fluid dynamics*. In case of images, motion causes gray values, i. e., an optical signal, to "flow" over the image plane, just as volume elements flow in liquids and gases. In fluid dynamics the continuity equation plays an important role. It expresses the fact that mass is conserved in a flow. Can we formulate a similar continuity equation for gray values and under which conditions are they conserved?

In fluid dynamics, the *continuity equation* for the density ϱ of the fluid is given by

$$\frac{\partial \varrho}{\partial t} + \nabla(\boldsymbol{u}\varrho) = \frac{\partial \varrho}{\partial t} + \boldsymbol{u}\nabla\varrho + \varrho\nabla\boldsymbol{u} = 0. \tag{14.7}$$

This equation is valid for two and three-dimensional flows. It states the conservation of mass in a fluid in a differential form. The temporal change in the density is balanced by the divergence of the flux density $\boldsymbol{u}\varrho$. By integrating the continuity equation over an arbitrary volume element, we can write the equation in an integral form:

$$\int_V \left(\frac{\partial \varrho}{\partial t} + \nabla(\boldsymbol{u}\varrho) \right) \mathrm{d}V = \frac{\partial}{\partial t} \int_V \varrho \mathrm{d}V + \oint_A \varrho \boldsymbol{u}\, \mathrm{d}\boldsymbol{a} = 0. \tag{14.8}$$

The volume integral has been converted into a surface integral around the volume using the Gauss integral theorem. $\mathrm{d}\boldsymbol{a}$ is a vector normal to a surface element $\mathrm{d}A$. The integral form of the continuity equation clearly states that the temporal change of the mass is caused by the net flux into the volume integrated over the whole surface of the volume.

How can we devise a similar continuity equation for the optical flux $\boldsymbol{f}$ — known as the *brightness change constraint equation* (*BCCE*) or *optical flow constraint* (*OFC*) — in computer vision? The quantity analogous to the density ϱ is the irradiance E or the gray value g. However, we should be careful and examine the terms in Eq. (14.7) more closely. The left divergence term $\boldsymbol{f}\nabla g$ describes the temporal brightness change due to a moving gray value gradient. The second term with the divergence of the velocity field $g\nabla\boldsymbol{f}$ seems questionable. It would cause a temporal change even in a region with a constant irradiance if the divergence of the flow field is unequal to zero. Such a case occurs, for instance, when an object moves away from the camera. The irradiance at the image plane

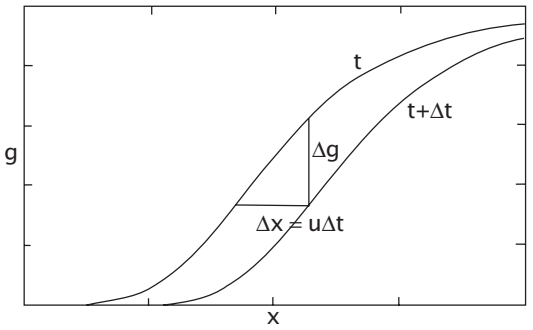

Figure 14.10: *Illustration of the continuity of optical flow in the one-dimensional case.*

remains constant, provided the object irradiance does not change. The collected radiance decreases with the squared distance of the object. However, it is exactly compensated, as also the projected area of the object is decreased by the same factor. Thus we omit the last term in the continuity equation for the optical flux and obtain

$$\boxed{\frac{\partial g}{\partial t} + \boldsymbol{f}\nabla g = 0.} \tag{14.9}$$

In the one-dimensional case, the continuity of the optical flow takes the simple form

$$\frac{\partial g}{\partial t} + f\frac{\partial g}{\partial x} = 0, \tag{14.10}$$

from which we directly get the one-dimensional velocity

$$\boxed{f = -\frac{\partial g}{\partial t} \Big/ \frac{\partial g}{\partial x},} \tag{14.11}$$

provided that the spatial derivative does not vanish. The velocity is thus given as the ratio of the temporal and spatial derivatives.

This basic relation can also be derived geometrically, as illustrated in Fig. 14.10. In the time interval Δt a gray value is shifted by the distance $\Delta x = u\Delta t$ causing the gray value to change by $g(x, t+\Delta t) - g(x, t)$. The gray value change can also be expressed as the slope of the gray value edge,

$$g(x, t+\Delta t) - g(x, t) = -\frac{\partial g(x, t)}{\partial x}\Delta x = -\frac{\partial g(x, t)}{\partial x}u\Delta t, \tag{14.12}$$

from which, in the limit of $\Delta t \to 0$, the continuity equation for optical flow Eq. (14.10) is obtained.

The continuity or BCCE equation for optical flow at the image plane Eq. (14.9) can in general only be a crude approximation. We have already touched this subject in the introductory section about motion and gray value changes (Section 14.2.1). This is because of the complex nature of the reflection from opaque surfaces, which depends on the viewing direction, surface normal, and directions of the incident light. Each object receives radiation not only directly from light sources but also from all other objects in the scene that lie in the direct line of sight of the object. Thus the radiant emittance from the surface of one object depends on the position of all the other objects in a scene.

In computer graphics, problems of this type are studied in detail, in search of *photorealistic* computer generated images. A big step towards this goal was a method called *radiosity* which explicitly solved the interrelation of object emittance described above [52]. A general expression for the object emittance — the now famous *rendering equation* — was derived by Kajiya [100].

In image sequence processing, it is in principle required to invert this equation to infer the surface reflectivity from the measured object emittance. The surface reflectivity is a feature invariant to surface orientation and the position of other objects and thus would be ideal for motion estimation. Such an approach is unrealistic, however, because it requires a reconstruction of the 3-D scene before the inversion of the rendering equation can be tackled at all.

As there is no generally valid continuity equation for optical flow, it is important to compare possible additional terms with the terms in the standard BCCE. All other terms basically depend on the rate of changes of a number of quantities but not on the brightness gradients. If the gray value gradient is large, the influence of the additional terms becomes small. Thus we can conclude that the determination of the velocity is most reliable for steep gray value edges while it may be significantly distorted in regions with only small gray value gradients. This conclusion is in agreement with Verri and Poggio [207, 208] findings where they point out the difference between optical flow and the motion field.

Another observation is important. It is certainly true that the historical approach of determining the displacement vectors from only two consecutive images is not robust. In general we cannot distinguish whether a gray value change comes from a displacement or any other source. However, the optical flow becomes more robust in space-time images. We will demonstrate this with two examples.

First, it is possible to separate gray value changes caused by global illumination changes from those caused by motion. Figure 14.11 shows an image sequence of a static scene taken at a rate of 5 frames per minute. The two spatiotemporal time slices (Fig. 14.11a, c), indicated by the two white horizontal lines in Fig. 14.11b, cover a period of about 3.4 h. The upper line covers the high-rise building and the sky. From the sky it can

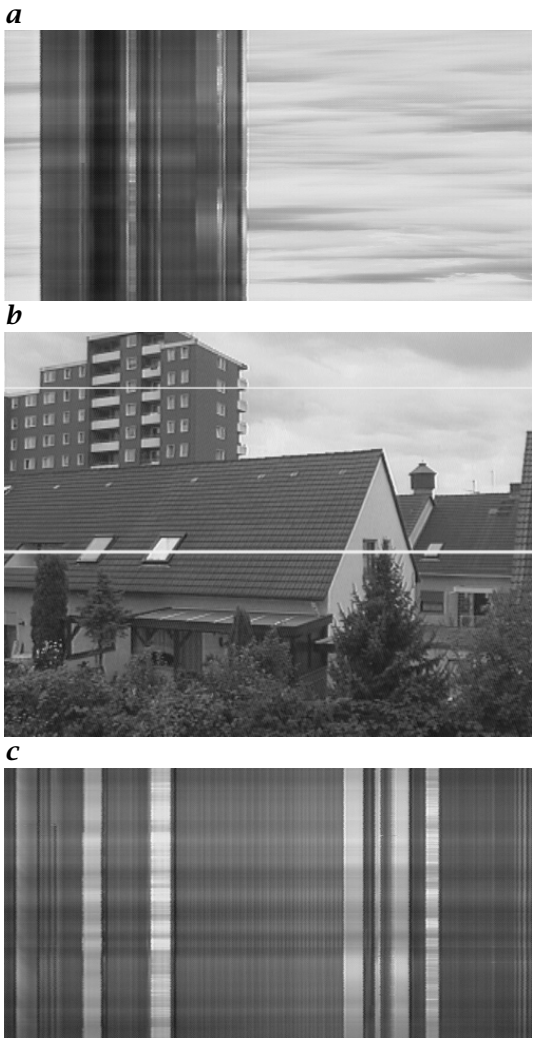

Figure 14.11: *Static scene with illumination changes:* **a** xt *cross section at the upper marked row (sky area) in* **b**; **b** *first image of the sequence;* **c** xt *cross section at the lower marked row (roof area) in* **b**; *the time axis spans 3.4 h, running downwards (from Jähne [88]).*

be seen that it was partly cloudy, but sometimes there was direct solar illumination. The lower line crosses several roof windows, walls, and house roofs.

In both slices the illumination changes appear as horizontal stripes which seem to transparently overlay the vertical stripes, indicating a static scene. As a horizontal patterns indicates an object moving with

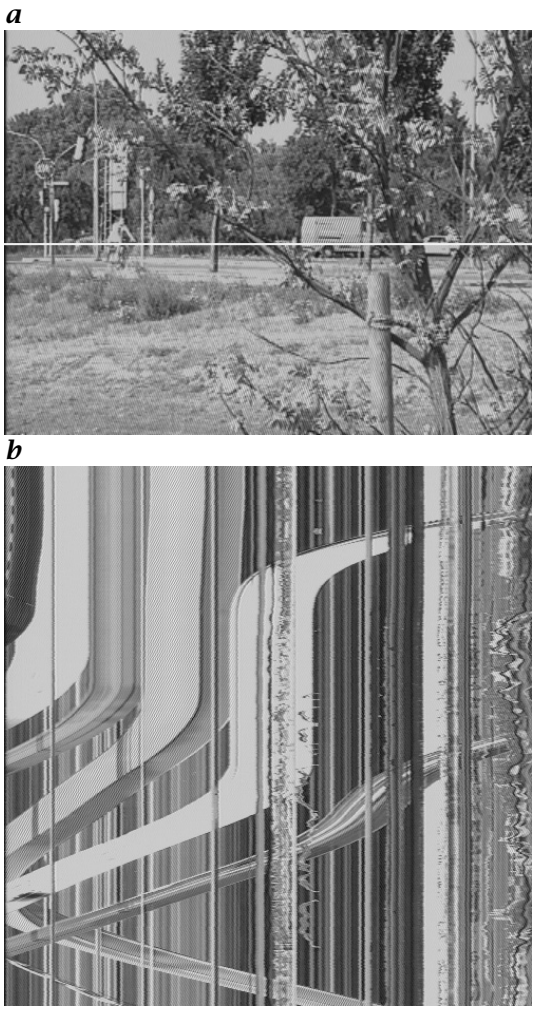

Figure 14.12: *Traffic scene at the border of Hanau, Germany;* ***a*** *last image of the sequence;* ***b*** x*-*t *cross section at the marked line in* ***a****; the time axis spans 20.5 s, running downwards (from Jähne [88]).*

infinite velocity, these patterns can be eliminated, e. g., by directional filtering, without disturbing the motion analysis.

The second example demonstrates that motion determination is still possible in space-time images if occlusions occur and the local illumination of an object is changing because it is turning. Figure 14.12 shows a traffic scene at the city limits of Hanau, Germany. From the last image of the sequence (Fig. 14.12a) we see that a street crossing with a traffic light is observed through the branches of a tree located on the right in

the foreground. One road is running horizontally from left to right, with the traffic light on the left.

The spatiotemporal slice (Fig. 14.12b) has been cut through the image sequence at the horizontal line indicated in Fig. 14.12a. It reveals various occlusions: the car traces disappear under the static vertical patterns of the tree branches and traffic signs. We can also see that the temporal trace of the van shows significant gray value changes because it turned at the street crossing and the illumination conditions are changing while it is moving along in the scene. Nevertheless, the temporal trace is continuous and promises a reliable velocity estimate.

We can conclude that the best approach is to stick to the standard BCCE for motion estimates and use it to develop the motion estimators in this chapter. Because of the wide variety of additional terms this approach still seems to be the most reasonable and most widely applicable, because it contains the fundamental constraint.

14.3 First-Order Differential Methods

14.3.1 Basics

Differential methods are the classical approach to determine motion from two consecutive images. This chapter discusses the question of how these techniques can be applied to space-time images. The continuity equation for the optical flow (Section 14.2.6), in short the *BCCE* or *OFC*, is the starting point for differential methods:

$$\frac{\partial g}{\partial t} + \boldsymbol{f}\nabla g = 0. \tag{14.13}$$

This single scalar equation contains W unknown vector components in the W-dimensional space. Thus we cannot determine the optical flow $\boldsymbol{f} = [f_1, f_2]^T$ unambiguously. The scalar product $\boldsymbol{f}\nabla g$ is equal to the magnitude of the gray value gradient multiplied by the component of $\boldsymbol{f}$ in the direction of the gradient, i. e., normal to the local gray value edge

$$\boldsymbol{f}\nabla g = f_\perp |\nabla g|.$$

Thus we can only determine the optical flow component normal to the edge. This is the well-known *aperture problem*, which we discussed qualitatively in Section 14.2.2. From Eq. (14.13), we obtain

$$f_\perp = -\frac{\partial g}{\partial t} / |\nabla g| \,. \tag{14.14}$$

Consequently, it is not possible to determine the complete vector with first-order derivatives at a *single* point in the space-time image.

14.3.2 First-Order Least Squares Solution

Instead of a single point, we can use a small neighborhood to determine the optical flow. We assume that the optical flow is constant in this region and discuss in this section under which conditions an unambiguous determination of the optical flow is possible. We still have the two unknowns $\boldsymbol{f} = [f_1, f_2]^T$, but we also have the continuity constraint Eq. (14.13) for the optical flow at many points. Thus we generally end up with an overdetermined equation system. Such a system cannot be solved exactly but only by minimizing an error functional. We seek a solution that minimizes Eq. (14.13) within a local neighborhood in a least squares sense. Thus, the convolution integral

$$\|\boldsymbol{e}\|_2^2 = \int_{-\infty}^{\infty} w(\boldsymbol{x}-\boldsymbol{x}', t-t') \left(f_1 g_x(\boldsymbol{x}') + f_2 g_y(\boldsymbol{x}') + g_t(\boldsymbol{x}')\right)^2 \mathrm{d}^2x' \mathrm{d}t' \tag{14.15}$$

should be minimized. Note that $\boldsymbol{f} = [f_1, f_2]^T$ is constant within the local neighborhood. It depends, of course, as $\|\boldsymbol{e}\|$, on $\boldsymbol{x}$. For the sake of more compact equations, we omit the explicit dependency of g_x, g_y, and g_t on the variable $\boldsymbol{x}'$ in the following equations. The partial derivative $\partial g/\partial p$ is abbreviated by g_p.

In this integral, the square of the residual deviation from the continuity constraint is summed up over a region determined by the size of the window function w. In order to simplify the equations further, we use the following abbreviation for this weighted averaging procedure:

$$\boxed{\|\boldsymbol{e}\|_2^2 = \overline{\left(f_1 g_x + f_2 g_y + g_t\right)^2} \rightarrow \text{minimum.}} \tag{14.16}$$

The window function w determines the size of the neighborhood. This makes the least-squares approach very flexible. The averaging in Eq. (14.15) can be but must not be extended in the temporal direction. If we choose a rectangular neighborhood with constant weighting for all points, we end up with a simple *block matching* technique. This corresponds to an averaging with a *box filter*. However, because of the bad averaging properties of box filters (Section 11.3), an averaging with a weighting function that decreases with the distance of the point $[\boldsymbol{x}, t]^T$ from $[\boldsymbol{x}, t]^T$ appears to be a more suitable approach. In continuous space, averaging with a Gaussian filter is a good choice. For discrete images, averaging with a *binomial filter* is most suitable (Section 11.4).

Equation (14.16) can be solved by setting the partial derivatives

$$\begin{aligned} \frac{\partial \|\boldsymbol{e}\|_2^2}{\partial f_1} &= \overline{2 g_x \left(f_1 g_x + f_2 g_y + g_t\right)} \stackrel{!}{=} 0, \\ \frac{\partial \|\boldsymbol{e}\|_2^2}{\partial f_2} &= \overline{2 g_y \left(f_1 g_x + f_2 g_y + g_t\right)} \stackrel{!}{=} 0 \end{aligned} \tag{14.17}$$

to zero. From this condition we obtain the linear equation system

$$\boxed{\begin{bmatrix} \overline{g_x g_x} & \overline{g_x g_y} \\ \overline{g_x g_y} & \overline{g_y g_y} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = - \begin{bmatrix} \overline{g_x g_t} \\ \overline{g_y g_t} \end{bmatrix}}, \tag{14.18}$$

or more compact in matrix notation

$$\boldsymbol{G}\boldsymbol{f} = \boldsymbol{g}. \tag{14.19}$$

The terms $\overline{g_p g_q}$ represent regularized estimates that are composed of convolution and nonlinear point operations. In the operator notation, we can replace it by

$$\mathcal{B}(\mathcal{D}_p \cdot \mathcal{D}_q), \tag{14.20}$$

where $\mathcal{D}_p$ is a suitable discrete first-order derivative operator in the direction p (Chapter 12) and $\mathcal{B}$ an averaging operator (Chapter 11). Thus, the operator expression in Eq. (14.20) includes the following sequence of image processing operators:

1. Apply the convolution operators $\mathcal{D}_p$ and $\mathcal{D}_q$ to the image to obtain images with the first-order derivatives in directions p and q.
2. Multiply the two derivative images pointwise.
3. Convolve the resulting image with the averaging mask $\mathcal{B}$.

Note that the point operation is a nonlinear operation. Therefore, it must not be interchanged with the averaging.

The linear equation system Eq. (14.18) can be solved if the matrix can be inverted. This is the case when the determinant of the matrix is not zero:

$$\det \boldsymbol{G} = \overline{g_x g_x}\,\overline{g_y g_y} - \overline{g_x g_y}^2 \neq 0. \tag{14.21}$$

From this equation, we can deduce two conditions that must be met:

1. Not all partial derivatives g_x and g_y must be zero. In other words, the neighborhood must not consist of an area with constant gray values.
2. The gradients in the neighborhood must not point in the same direction. If this were the case, we could express g_y by g_x except for a constant factor and the determinant of $\boldsymbol{G}$ in Eq. (14.21) would vanish.

The solution for the optical flow $\boldsymbol{f}$ can be written down explicitly because it is easy to invert the 2×2 matix $\boldsymbol{G}$:

$$\boldsymbol{G}^{-1} = \frac{1}{\det \boldsymbol{G}} \begin{bmatrix} \overline{g_y g_y} & -\overline{g_x g_y} \\ -\overline{g_x g_y} & \overline{g_x g_x} \end{bmatrix} \quad \text{if} \quad \det \boldsymbol{G} \neq 0. \tag{14.22}$$

With $\boldsymbol{f} = \boldsymbol{G}^{-1}\boldsymbol{g}$ we then obtain

$$\begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = -\frac{1}{\det \boldsymbol{G}} \begin{bmatrix} \overline{g_x g_t}\,\overline{g_y g_y} - \overline{g_y g_t}\,\overline{g_x g_y} \\ \overline{g_y g_t}\,\overline{g_x g_x} - \overline{g_x g_t}\,\overline{g_x g_y} \end{bmatrix}. \tag{14.23}$$

The solution looks still quite complex. It can be simplified considerably by observing that $\boldsymbol{G}$ is a symmetric matrix. Any symmetric matrix can be brought into diagonal form by a rotation of the coordinate system into the so-called *principle-axes* coordinate system. Then the matrix $\boldsymbol{G}$ reduces to

$$\boldsymbol{G}' = \begin{bmatrix} \overline{g_{x'}g_{x'}} & 0 \\ 0 & \overline{g_{y'}g_{y'}} \end{bmatrix}, \tag{14.24}$$

the determinant $\det \boldsymbol{G}' = \overline{g_{x'}g_{x'}}\,\overline{g_{y'}g_{y'}}$, and the optical flow is

$$\begin{bmatrix} f_{1'} \\ f_{2'} \end{bmatrix} = -\begin{bmatrix} \dfrac{\overline{g_{x'}g_t}}{\overline{g_{x'}g_{x'}}} \\ \dfrac{\overline{g_{y'}g_t}}{\overline{g_{y'}g_{y'}}} \end{bmatrix}. \tag{14.25}$$

This equation reflects in a quantitative way the qualitative discussion about the *aperture problem* discussed in Section 14.2.2. The principal axes are oriented along the directions of the maximum and minimum mean square spatial gray value changes, which are perpendicular to each other. Because the matrix $\boldsymbol{G}'$ is diagonal, both changes are uncorrelated. Now, we can distinguish three cases:

1. $\overline{g_{x'}g_{x'}} > 0$, $\overline{g_{y'}g_{y'}} > 0$: spatial gray value changes in all directions. Then both components of the optical flow can be determined.
2. $\overline{g_{x'}g_{x'}} > 0$, $\overline{g_{y'}g_{y'}} = 0$: spatial gray value changes only in x' direction (perpendicularly to an edge). Then only the component of the optical flow in x' direction can be determined (aperture problem). The component of the optical flow parallel to the edge remains unknown.
3. $\overline{g_{x'}g_{x'}} = \overline{g_{y'}g_{y'}} = 0$: no spatial gray value changes in both directions. In the case of a constant region, none of the components of the optical flow can be determined at all.

It is important to note that only the matrix $\boldsymbol{G}$ determines the type of solution of the least-squares approach. In this matrix only spatial and no temporal derivatives occur. This means that the spatial derivatives and thus the spatial structure of the image entirely determines whether and how accurately the optical flow can be estimated.

14.3.3 Error Analysis

Noise may introduce a systematic error into the estimate of the optical flow. Here we show how we can analyze the influence of noise on the determination of optical flow in a very general way. We assume that the image signal is composed of a structure moving with a constant velocity $\boldsymbol{u}$ superimposed by zero-mean isotropic noise:

$$g'(\boldsymbol{x}, t) = g(\boldsymbol{x} - \boldsymbol{u}t) + n(\boldsymbol{x}, t). \tag{14.26}$$

This is a very general approach because we do not rely on any specific form of the gray value structure. The expression $g(\boldsymbol{x} - \boldsymbol{u}t)$ just says that an arbitrary spatial structure is moving with a constant velocity $\boldsymbol{u}$. In this way, a function with three parameters $g(x_1, x_2, t)$ is reduced to a function with only two parameters $g(x_1 - u_1 t, x_2 - u_2 t)$. We further assume that the partial derivatives of the noise function are not correlated with themselves or the partial derivatives of the image patterns. Therefore we use the conditions

$$\overline{n} = 0, \quad \overline{n_p n_q} = \sigma_n^2 \delta_{p-q}, \quad \overline{g_p n_q} = 0, \tag{14.27}$$

and the partial derivatives are

$$\nabla g' = \nabla g + \nabla n \quad g'_t = -\boldsymbol{u}\nabla g + \partial_t n_t. \tag{14.28}$$

These conditions result in the optical flow estimate

$$\boldsymbol{f} = \boldsymbol{u}(\overline{\nabla g \nabla g^T} + \overline{\nabla n \nabla n^T})^{-1} \overline{\nabla g \nabla g^T}. \tag{14.29}$$

The key to understanding this matrix equation is to observe that the noise matrix $\overline{\nabla n \nabla n^T}$ is diagonal in any coordinate system, because of the conditions set by Eq. (14.27). Therefore, we can transform the equation into the principal-axes coordinate system in which $\nabla g \nabla g^T$ is diagonal. Then we obtain

$$\boldsymbol{f} = \boldsymbol{u} \begin{bmatrix} \overline{g_{x'}{}^2} + \sigma_n^2 & 0 \\ 0 & \overline{g_{y'}{}^2} + \sigma_n^2 \end{bmatrix}^{-1} \begin{bmatrix} \overline{g_{x'}{}^2} & 0 \\ 0 & \overline{g_{y'}{}^2} \end{bmatrix}.$$

When the variance of the noise is not zero, the inverse of the first matrix always exists and we obtain

$$\boldsymbol{f} = \boldsymbol{u} \begin{bmatrix} \dfrac{\overline{g_{x'}{}^2}}{\overline{g_{x'}{}^2} + \sigma_n^2} & 0 \\ 0 & \dfrac{\overline{g_{y'}{}^2}}{\overline{g_{y'}{}^2} + \sigma_n^2} \end{bmatrix}. \tag{14.30}$$

This equation shows that the estimate of the optical flow is biased towards lower values. If the variance of the noise is about the squared magnitude of the gradient, the estimated values are only about half of the true values. Thus the differential method is an example of a nonrobust technique because it deteriorates in noisy image sequences.

If the noise is negligible, however, the estimate of the optical flow is correct. This result is in contradiction to the widespread claim that differential methods do not deliver accurate results if the spatial gray value structure cannot be adequately approximated by a first-order Taylor series (see, for example, [189]). Kearney et al. [105], for instance, provided

an error analysis of the gradient approach and concluded that it gives erroneous results as soon as second-order spatial derivatives become significant.

These contradictory findings resolve if we analyze the additional errors in the estimation of optical flow that are introduced by an inadequate discretization of the partial derivative operators (see the discussion on optimal derivative filters in Section 12.4). The error in the optical flow estimate is directly related to the error in the direction of discrete gradient operators (compare also the discussion on orientation estimates in Section 13.3.6). Therefore accurate optical flow estimates require carefully optimized derivative operators such as the optimized regularized gradient operators discussed in Section 12.7.5.

14.4 Tensor Methods

The tensor method for the analysis of local orientation has already been discussed in detail in Section 13.3. Since motion constitutes locally oriented structure in space-time images, all we have to do is to extend the tensor method to three dimensions. First, we will revisit the optimization criterion used for the tensor approach in Section 14.4.1 in order to distinguish this technique from the differential method (Section 14.3).

14.4.1 Optimization Strategy

In Section 13.3.1 we stated that the optimum orientation is defined as the orientation that shows the least deviations from the direction of the gradient vectors. We introduced the squared scalar product of the gradient vector and the unit vector representing the local orientation $\bar{\boldsymbol{n}}$ as an adequate measure:

$$(\boldsymbol{\nabla} g^T \bar{\boldsymbol{n}})^2 = |\boldsymbol{\nabla} g|^2 \cos^2 \left(\angle(\boldsymbol{\nabla} g, \bar{\boldsymbol{n}})\right). \tag{14.31}$$

This measure can be used in vector spaces of any dimension. In order to determine orientation in space-time images, we take the spatiotemporal gradient

$$\boldsymbol{\nabla}_{xt} g = \left[\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial g}{\partial t}\right]^T = \left[g_x, g_y, g_t\right]^T \tag{14.32}$$

and write

$$(\boldsymbol{\nabla}_{xt} g^T \bar{\boldsymbol{n}})^2 = |\boldsymbol{\nabla}_{xt} g|^2 \cos^2 \left(\angle(\boldsymbol{\nabla}_{xt} g, \bar{\boldsymbol{n}})\right). \tag{14.33}$$

For the 2-D orientation analysis we maximized the expression

$$\int w(\boldsymbol{x} - \boldsymbol{x}') \left(\boldsymbol{\nabla} g(\boldsymbol{x}')^T \bar{\boldsymbol{n}}\right)^2 \mathrm{d}^W x' = \overline{(\boldsymbol{\nabla} g\, \bar{\boldsymbol{n}})^2} \tag{14.34}$$

in order to find the optimum orientation. For analysis of motion in space-time images, we are not interested in the direction of maximal gray value changes but in the direction of minimal gray value changes. We denote this orientation by the unit vector $\bar{\boldsymbol{e}}_3 = [e_{31}, e_{32}, e_{33}]^T$. This 3-D vector is, according to the considerations in Section 14.2.4 Eq. (14.2), related to the 2-D velocity vector by

$$\boldsymbol{f} = \frac{1}{e_{33}} \begin{bmatrix} e_{31} \\ e_{32} \end{bmatrix}. \tag{14.35}$$

By analogy to Eq. (14.34) we therefore *minimize*

$$\int w(\boldsymbol{x} - \boldsymbol{x}', t - t') \left(\nabla_{xt} g(\boldsymbol{x}', t')^T \bar{\boldsymbol{e}}_3 \right)^2 \mathrm{d}^W x' \mathrm{d}t' \tag{14.36}$$

or, in more compact notation,

$$\boxed{\overline{(\nabla_{xt} g^T \bar{\boldsymbol{e}}_3)^2} \rightarrow \text{minimum.}} \tag{14.37}$$

The window function w now also extends into the time coordinate and determines the size and shape of the neighborhood around a point $[\boldsymbol{x}, t]^T$ in which the orientation is averaged. Equation (14.37) has to be compared with the corresponding expression that is minimized with the differential method (Eq. (14.16)):

$$\overline{(\boldsymbol{f} \nabla g + g_t)^2}. \tag{14.38}$$

Note the subtle difference between the two optimization strategies Eqs. (14.37) and (14.38). Both are least squares problems for determining the velocity in such a way that the deviation from the continuity of the optical flow becomes minimal. The two methods differ, however, in the parameters to be estimated. The estimation of a 3-D unit vector turns out to be a so-called *total least squares* problem [84]. This method is more suitable for the problem because all components of the space-time gradient are though to have statistical errors and not only the temporal derivative as in Eq. (14.38).

By analogy to the discussion in Section 13.3.1 we can conclude that the determination of the optical flow in a space-time image is equivalent to finding the *eigenvector* $\bar{\boldsymbol{e}}_3$ of the *smallest* eigenvalue λ_3 of the structure tensor $\boldsymbol{J}$:

$$\boxed{\boldsymbol{J} = \begin{bmatrix} \overline{g_x g_x} & \overline{g_x g_y} & \overline{g_x g_t} \\ \overline{g_x g_y} & \overline{g_y g_y} & \overline{g_y g_t} \\ \overline{g_x g_t} & \overline{g_y g_t} & \overline{g_t g_t} \end{bmatrix}}, \tag{14.39}$$

where $\overline{g_p g_q}$ with $p, q \in \{x, y, t\}$ is given by

$$\overline{g_p g_q}(\boldsymbol{x}, t) = \int w(\boldsymbol{x} - \boldsymbol{x}', t - t') g_p(\boldsymbol{x}', t') g_q(\boldsymbol{x}', t') \mathrm{d}^2 x' \mathrm{d}t'. \quad (14.40)$$

At this point, we can compare the tensor method again with the differential technique. While the tensor method essentially performs an eigenvalue analysis of a symmetric tensor with six regularized products of spatial and temporal derivatives, the differential method uses the same products but only five of them. Thus the differential technique misses $\overline{g_t g_t}$. We will see in the next section that this additional term enables the tensor method to detect whether a local neighbor shows a constant velocity or not. This is not possible with the differential method.

14.4.2 Eigenvalue Analysis

Unfortunately, the *eigenvalue analysis* of a symmetric 3×3 tensor is not as simple as for a symmetric 2×2 tensor. In two dimensions we could solve the eigenvalue problem directly. In Section 13.3.3 we transformed the three independent components of the symmetric 2×2 tensor into three parameters: the orientation and the rotation-invariant certainty and coherency measures (Section 13.3.4).

The symmetric 3×3 tensor now contains six independent components and we need to find a corresponding number of parameters that describe the local structure of the space-time image adequately. Again it is useful to decompose these six parameters into rotation-variant and rotation-invariant parameters.

As already mentioned, the solution of the eigenvalue problem cannot be written down readily. It requires a suitable numerical algorithm. We will not detail this problem since it is a nontrivial but standard problem of numerical mathematics for which a number of efficient solutions are available [*Press et al.*, 1992 and *Golub and van Loan*, 1989]. Thus we assume in the following that we have solved the eigenvalue problem and have obtained a set of three orthonormal eigenvectors and three eigenvalues. With the solution of the eigenvalue problem, we have essentially obtained the principal-axes coordinate system in which the structure tensor is diagonal and contains the eigenvalues as diagonal elements:

$$\boldsymbol{J}' = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}. \quad (14.41)$$

Without restricting generality, the eigenvalues are sorted in descending order:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0. \quad (14.42)$$

The principal-axes coordinate system is spanned by the three eigenvectors. The rotation into this coordinate system requires three independent parameters as we have discussed in Section 7.2.2. Thus, three of the six parameters are used up to describe its orientation in the space-time domain. This information is contained in the three orthonormal eigenvectors.

The remaining parameters are the three rotation-invariant eigenvalues. We will now show how the different classes of local structures in space-time images can be differentiated by the values of the three eigenvalues. This approach will also help us find an efficient implementation of the tensor-based motion analysis. Four different classes of neighborhoods can be distinguished in a space-time image, corresponding to a rank from 0 to 3 for the symmetric tensor:

Constant gray value. All elements and eigenvalues of the structure tensor are zero.

$$\lambda_1 = \lambda_2 = \lambda_3 = 0. \tag{14.43}$$

The rank of the tensor is also zero. Therefore no velocity at all can be obtained. This condition is easy to recognize. The sum of the eigenvalues must be below a critical level, determined by the noise level in the image sequence. As the sum of the eigenvalues is equal to the trace of the tensor, no eigenvalue analysis is required to sort out this condition:

$$\text{trace}(\boldsymbol{J}) = \sum_{p=1}^{3} \overline{g_p g_p} < \gamma, \tag{14.44}$$

where γ is a suitable measure for the noise level in the image sequence. For all points where the condition Eq. (14.44) is met, the eigenvalue analysis can be skipped completely.

Spatial orientation and constant motion. In this case two eigenvalues are zero since the gray values only change in one direction:

$$\lambda_1 > 0 \quad \text{and} \quad \lambda_2 = \lambda_3 = 0. \tag{14.45}$$

The rank of the tensor is one. The spatial gray value structure shows *linear symmetry*. This condition can easily be detected again without performing an eigenvalue analysis, because the determinant of the leading 2×2 subtensor should be below a threshold γ^2:

$$\overline{g_x g_x}\ \overline{g_y g_y} - \overline{g_x g_y}^2 < \gamma^2. \tag{14.46}$$

The eigenvector $\bar{\boldsymbol{e}}_1$ belonging to the only nonzero (i. e., largest) eigenvalue points in the direction of maximum change of gray values. Thus it gives both the spatial orientation and the velocity in this direction. Note that only the *normal velocity*, i. e., the velocity in the direction of

the spatial gradient, can be obtained because of the *aperture problem* (Section 14.2.2). The spatial orientation is given by the two spatial coordinates of the eigenvector $\bar{e}_1$ of the largest eigenvalue. As the normal optical flow is in this direction, it is given by

$$\boldsymbol{f}_\perp = -\frac{e_{1t}}{e_{1x}^2 + e_{1y}^2} \begin{bmatrix} e_{1x} \\ e_{1y} \end{bmatrix} \tag{14.47}$$

and the magnitude of the normal optical flow reduces to

$$|\boldsymbol{f}_\perp| = \sqrt{\frac{e_{1t}^2}{e_{1x}^2 + e_{1y}^2}} = \sqrt{\frac{e_{1t}^2}{1 - e_{1t}^2}}. \tag{14.48}$$

Distributed spatial structure and constant motion. In this case only one eigenvalue is zero:

$$\lambda_1, \lambda_2 > 0 \quad \text{and} \quad \lambda_3 = 0. \tag{14.49}$$

As the motion is constant, the principal-axes coordinate system is moving with the scene. The eigenvector $\bar{e}_3$ with the zero eigenvalue points in the direction of constant gray values in the space-time domain. Thus the optical flow is given by

$$\boldsymbol{f} = \frac{1}{e_{3t}} \begin{bmatrix} e_{3x} \\ e_{3y} \end{bmatrix} \tag{14.50}$$

and its magnitude by

$$|\boldsymbol{f}| = \sqrt{\frac{e_{3x}^2 + e_{3y}^2}{e_{3t}^2}} = \sqrt{\frac{1 - e_{3t}^2}{e_{3t}^2}}. \tag{14.51}$$

Distributed spatial structure and non-constant motion. In this case all three eigenvalues are larger than zero and the rank of the tensor is three:

$$\lambda_1, \lambda_2, \lambda_3 > 0. \tag{14.52}$$

No useful optical flow estimate can be obtained in this case.

After this detailed classification, we turn to the question of which three rotation-invariant parameters can be extracted from the structure tensor in order to obtain a useful description of the local structure independent of the velocity and the spatial orientation of the gray scale parameters.

Certainty measure. The first parameter is again a certainty measure that gives a measure for the gray value changes. We have two choices. Either we could take the mean square spatial gradient (trace of the upper 2×2 subtensor) or the mean square spatiotemporal gradient. From a practical point of view the mean square spatial gradient is to be preferred because the spatial gradient does not change in a sequence if the

velocity is increasing. The mean square spatiotemporal gradient, however, increases with increasing velocity because higher temporal gradients are added. Thus, surprisingly, the mean square spatial gradient is the better certainty measure:

$$c_c = \overline{g_x g_x} + \overline{g_y g_y}. \tag{14.53}$$

Spatial coherency measure. As a second measure we take the already known coherency measure from the analysis of 2-D local neighborhoods (Section 13.3.4) and denote it here as the spatial coherency measure:

$$c_s = \frac{(\overline{g_x g_x} - \overline{g_y g_y})^2 + 4\overline{g_x g_y}^2}{(\overline{g_x g_x} + \overline{g_y g_y})^2}. \tag{14.54}$$

Its value is between 0 and 1 and decides whether only the normal optical flow or both components of the optical flow can be determined.

Total coherency measure. Finally, we need an additional measure that tells us whether we encounter a local neighborhood with a constant velocity or not. This measure should be independent of the spatial coherency. The following measure using the largest and smallest eigenvalues meets this condition:

$$c_t = \left(\frac{\lambda_1 - \lambda_3}{\lambda_1 + \lambda_3}\right)^2. \tag{14.55}$$

The total coherency measure is one as soon as the eigenvalue λ_3 is zero. The other two eigenvalues may then take any other values. The total coherency approaches zero if all three eigenvalues are equal. In contrast to the other two measures c_c and c_s, the total coherency requires an eigenvalue analysis since the smallest and largest eigenvalues are needed to compute it. There is one caveat with this measure. It is also one with a spatially oriented pattern and a non-constant motion. This special case can be recognized, however, from the condition that both the spatial and total coherency are one but that only one eigenvalue is zero. Another simple criterion is that the eigenvector to the zero eigenvalue lies in the xy plane. This implies that $e_{33} = 0$, so according to Eq. (14.50) we would obtain an infinite value for the optical flow vector.

14.5 Correlation Methods

14.5.1 Principle

As with the differential method, the correlation technique is an approach which originates from analyzing the displacement between two consecutive images. To find a characteristic feature from the first image within the second, we take the first image $g(t_1) = g_1$ and compare it with the

second image $g(t_2) = g_2$ within a certain search range. Within this range we search for the position of optimum similarity between the two images. When do we regard two features as being similar? The similarity measure should be robust against changes in the illumination. Thus we regard two spatial feature patterns as equal if they differ only by a constant factor α which reflects the difference in illumination. In the language of inner product vector spaces, this means that the two feature vectors g_1 and g_2 are parallel. This can be the case if and only if an equality occurs in the *Cauchy–Schwarz inequality*

$$\left| \int_{-\infty}^{\infty} g_1(\boldsymbol{x}) g_2(\boldsymbol{x}-\boldsymbol{s}) \mathrm{d}^2x \right|^2 \leq \int_{-\infty}^{\infty} g_1^2(\boldsymbol{x}) \mathrm{d}^2x \int_{-\infty}^{\infty} g_2^2(\boldsymbol{x}-\boldsymbol{s}) \mathrm{d}^2x. \tag{14.56}$$

In other words, we need to maximize the *cross-correlation coefficient*

$$r(\boldsymbol{s}) = \frac{\int_{-\infty}^{\infty} g_1(\boldsymbol{x}) g_2(\boldsymbol{x}-\boldsymbol{s}) \mathrm{d}^2x}{\left(\int_{-\infty}^{\infty} g_1^2(\boldsymbol{x}) \mathrm{d}^2x \int_{-\infty}^{\infty} g_2^2(\boldsymbol{x}-\boldsymbol{s}) \mathrm{d}^2x \right)^{1/2}}. \tag{14.57}$$

The cross-correlation coefficient is a useful similarity measure. It is zero for totally dissimilar (orthogonal) patterns, and reaches a maximum of one for similar features.

In a similar way as for the differential method (Section 14.3), the correlation method can be performed by a combination of convolution and point operations. The first step is to introduce a window function w into the definition of the cross-correlation coefficient. This window is moved around the image to compute the local cross-correlation coefficient. Then Eq. (14.57) becomes

$$r(\boldsymbol{x},\boldsymbol{s}) = \frac{\int_{-\infty}^{\infty} w(\boldsymbol{x}-\boldsymbol{x}') g_1(\boldsymbol{x}') g_2(\boldsymbol{x}'-\boldsymbol{s}) \mathrm{d}^2x'}{\left(\int_{-\infty}^{\infty} w(\boldsymbol{x}-\boldsymbol{x}') g_1^2(\boldsymbol{x}') \mathrm{d}^2x' \int_{-\infty}^{\infty} w(\boldsymbol{x}-\boldsymbol{x}') g_2^2(\boldsymbol{x}'-\boldsymbol{s}) \mathrm{d}^2x' \right)^{1/2}} \tag{14.58}$$

or in the more compact notation already used in Sections 14.3.2 and 14.4.1

$$\boxed{r(\boldsymbol{x},\boldsymbol{s}) = \frac{\overline{g_1(\boldsymbol{x}) g_2(\boldsymbol{x}-\boldsymbol{s})}}{\left[\overline{g_1^2(\boldsymbol{x}) g_2^2(\boldsymbol{x}-\boldsymbol{s})} \right]^{1/2}} \rightarrow \text{maximum.}} \tag{14.59}$$

The resulting cross-correlation coefficient is a four-dimensional function, depending on the position in the image $\boldsymbol{x}$ and the shift $\boldsymbol{s}$.

14.5.2 Fast Iterative Maximum Search

It is obvious that the correlation method as discussed so far is a very costly operation. A considerable speed-up can be gained if we restrict the computation to a fast approach to search the position of the maximum of r because this is all we are interested in.

One way for a direct computation of the position of the maximum is the approximation of the cross-correlation function in a Taylor series. We expand the cross-correlation coefficient in a second-order Taylor expansion at the position of the maximum $\check{\boldsymbol{s}}$,

$$\begin{aligned} r(\boldsymbol{s}) &\approx r(\check{\boldsymbol{s}}) + \tfrac{1}{2} r_{xx}(\check{\boldsymbol{s}})(s_1 - \check{s}_1)^2 + \tfrac{1}{2} r_{yy}(\check{\boldsymbol{s}})(s_2 - \check{s}_2)^2 \\ &\quad + r_{xy}(\check{\boldsymbol{s}})(s_1 - \check{s}_1)(s_2 - \check{s}_2) \\ &= r(\check{\boldsymbol{s}}) + \tfrac{1}{2}(\boldsymbol{s} - \check{\boldsymbol{s}})^T \boldsymbol{H}(\check{\boldsymbol{s}})(\boldsymbol{s} - \check{\boldsymbol{s}}), \end{aligned} \tag{14.60}$$

where $\boldsymbol{H}$ is the Hessian matrix introduced in Eq. (12.6).

We do not know the position of the maximum correlation coefficient. Thus we assume that the second-order derivatives are constant sufficiently close to the position of the maximum and compute it at the position of the previous iteration $\boldsymbol{s}^{(i)}$. If we have no other information, we set the initial estimate to zero: $\boldsymbol{s}^{(0)} = \boldsymbol{0}$. As long as we have not yet found the position of the maximum correlation coefficient, there will be a residual slope at $\boldsymbol{s}^{(i)}$ that can be computed by derivating Eq. (14.60)

$$\nabla r(\boldsymbol{s}^{(i)}) = \boldsymbol{H}(\boldsymbol{s}^{(i)})(\boldsymbol{s}^{(i)} - \check{\boldsymbol{s}}). \tag{14.61}$$

Provided that the Hessian matrix is invertible, we obtain the following iteration

$$\boldsymbol{s}^{(i+1)} = \boldsymbol{s}^{(i)} - \boldsymbol{H}^{-1}(\boldsymbol{s}^{(i)}) \nabla r(\boldsymbol{s}^{(i)}) \quad \text{with} \quad \boldsymbol{s}^{(0)} = \boldsymbol{0}. \tag{14.62}$$

This type of iteration is known as Newton-Raphson iteration [158]. In order to estimate the shift, we need to compute only the first- and second-order partial derivatives of the cross-correlation coefficient.

14.5.3 Evaluation and Comparison

In contrast to the differential methods, which are based on the continuity of the optical flux, the correlation approach is insensitive to intensity changes between the two images. This makes correlation-based techniques very useful for stereo-image processing where slight intensity variations always occur between the left and right image because of the two different cameras used. Actually, the fast maximum search described Section 14.5.2 is the standard approach for determining the stereo disparity. Quam [160] used it with a coarse-to-fine control strategy, and Nishihara [146] used it in a modified version, taking the sign of

the Laplacian of the Gaussian as a feature. He reports a resolution accuracy of about 0.1 pixel for small displacements. Gelles et al. [58] measured movements in cells with a precision of about 0.02 pixel using the correlation method. However, they used a more costly approach, computing the centroid of a clipped cross-correlation function. The model-adapted approach of Diehl and Burkhardt [35] can be understood as an extended correlation approach as it allows also for rotation and other forms of motion.

The correlation method deviates from all other methods discussed in this work in the respect that it is conceptually based on the comparison of only two images. Even if we extend the correlation technique by multiple correlations to more than two frames, it remains a discrete time-step approach. Thus it lacks the elegance of the other methods, which were formulated in continuous space before being implemented for discrete images. Furthermore, it is obvious that a multiframe extension will be computationally quite expensive.

14.6 *Phase Method*

14.6.1 *Principle*

Except for the costly correlation method, all other methods that compute the optical flow are more or less sensitive to temporal illumination changes. Thus we search for a rich feature which contains the essential information in the images with regard to motion analysis. Fleet and Jepson [51] and Fleet [48] proposed using the phase for the computation of optical flow. We have discussed the crucial role of the phase already in Sections 2.3.5 and 13.4.1. In Section 2.3.5 we demonstrated that the phase of the Fourier transform of a signal carries the essential information. An image can still be recognized when the amplitude information is lost, but not when the phase is lost [124]. Global illumination changes the amplitude of a signal but not its phase.

As an introduction to the phase method, we consider a planar 1-D wave with a wave number k and a frequency ν, traveling with a phase speed $u = \nu/k$:

$$g(x,t) = g_0 \exp[-2\pi \mathrm{i}(\phi(x,t))] = g_0 \exp[-2\pi \mathrm{i}(kx - \nu t)]. \tag{14.63}$$

The position and thus also the displacement is given by the *phase*. The phase depends on both the spatial and temporal coordinates. For a planar wave, the phase varies linearly in time and space,

$$\phi(x,t) = 2\pi(kx - \nu t) = 2\pi(kx - ukt), \tag{14.64}$$

where k and ν are the wave number and the frequency of the pattern, respectively. Computing the temporal and spatial derivatives of the phase, i. e., the spatiotemporal gradient, yields both the wave number and the frequency of the moving periodic structure:

$$\nabla_{xt}\phi = \begin{bmatrix} \phi_x \\ \phi_t \end{bmatrix} = 2\pi \begin{bmatrix} k \\ -\nu \end{bmatrix}. \tag{14.65}$$

Then the velocity is given as the ratio of the frequency to the wave number:

$$u = \frac{\nu}{k} = -\partial_t \phi \,/\partial_x \phi \;. \tag{14.66}$$

This formula is very similar to the estimate based on the optical flow (Eq. (14.11)). In both cases, the velocity is given as a ratio of temporal and spatial derivatives.

Direct computation of the partial derivatives from the phase signal is not advisable because of the inherent discontinuities in the phase signal (restriction to the main interval $[-\pi, \pi[$). As we discussed in Section 13.4.6, it is possible to compute the phase gradients directly from the output of a quadrature filter pair. If we denote the quadrature filter pair with $p(x,t)$ and $q(x,t)$, the spatiotemporal phase gradient is given by (compare Eq. (13.52)):

$$\nabla_{xt}\phi(x,t) = \frac{p(x,t)\,\nabla_{xt}p(x,t) - q(x,t)\,\nabla_{xt}q(x,t)}{p^2(x,t) + q^2(x,t)}. \tag{14.67}$$

Using Eq. (14.66), the phase derived optical flow f is

$$\boxed{f = -\frac{p\,q_t - -q\,p_t}{p\,q_x - -q\,p_x}.} \tag{14.68}$$

14.6.2 *Evaluation and Comparison*

At first sight the phase method appears to offer nothing new. Replacing the gray value by the phase is a significant improvement, however, as the phase is much less dependent on the illumination than the gray value itself. Using only the phase signal, the amplitude of the gray value variations may change without affecting the velocity estimates at all.

So far, we have only considered an ideal periodic gray value structure. Generally, images are composed of gray value structures with different wave numbers. From such a structure we cannot obtain useful phase estimates. Consequently, we need to decompose the image into a set of wave number ranges.

This implies that the phase method is not appropriate to handle two-dimensional shifts. It is essentially a 1-D concept which measures the motion of a linearly oriented structure, e. g., a planar wave, in the direction of the gray value gradients. From this fact, Fleet and Jepson [50] derived a new paradigm for motion analysis. The image is decomposed with directional filters and in each of the components *normal velocities* are determined.

The 2-D motion field is then composed from these normal velocities. This approach has the advantage that the composition to a complete motion field is postponed to a second processing step which can be adapted to the kind of motion occurring in the images. Therefore this approach can also handle more complex cases such as motion superimposition of transparent objects.

Fleet and Jepson [50] use a set of *Gabor filters* (Section 13.4.5) with an angular resolution of 30° and a bandwidth of 0.8 octaves for the directional decomposition.

Alternatively, a bandpass decomposition and a *Hilbert filter* (Section 13.4.2) can be used. The motivation for this idea is the fact that the decomposition with a set of Gabor filters, as proposed by *Fleet and Jepson*, does not allow easy

reconstruction of the original image. The transfer functions of the Gabor filter series do not add up to a unit transfer function but show considerable ripples, as shown by Riemer [169].

A bandpass decomposition, for example using a *Laplacian pyramid* [18, 19], does not share this disadvantage (Section 5.2.3). In addition, it is computationally more efficient. However, we are faced with the problem that no directional decomposition is gained.

Jähne [86, 87] showed how the concept of the Laplacian pyramid can effectively be extended into a *directiopyramidal decomposition*. Each level of the pyramid is further decomposed into two or four directional components which add up directly to the corresponding isotropically filtered pyramid level (see also Section 5.2.4).

14.6.3 *From Normal Flow to 2-D Flow*

As the phase method gives only the normal optical flow, a technique is required to determine the two-dimensional optical flow from the normal flow. The basic relation between the normal and 2-D flow is as follows. We assume that $\boldsymbol{f}_\perp$ is a normal flow vector. It is a result of the projection of the 2-D flow vector $\boldsymbol{f}$ in the direction of the normal flow. Thus we can write:

$$f_\perp = \bar{\boldsymbol{f}}_\perp \boldsymbol{f}, \tag{14.69}$$

where $\bar{\boldsymbol{f}}_\perp$ is a unit vector in the direction of the normal flow. From Eq. (14.69), it is obvious that we can determine the unknown 2-D optical flow in a least squares approach if we have more than two estimates of the normal flow in different directions. In a similar way as in Section 14.3.2, this approach yields the linear equation system

$$\begin{bmatrix} \overline{\bar{f}_{\perp x}\bar{f}_{\perp x}} & \overline{\bar{f}_{\perp x}\bar{f}_{\perp y}} \\ \overline{\bar{f}_{\perp x}\bar{f}_{\perp y}} & \overline{\bar{f}_{\perp y}\bar{f}_{\perp y}} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} \overline{\bar{f}_{\perp x} f_\perp} \\ \overline{\bar{f}_{\perp y} f_\perp} \end{bmatrix} \tag{14.70}$$

with

$$\overline{\bar{f}_{\perp p}\bar{f}_{\perp q}} = \int w(\boldsymbol{x}-\boldsymbol{x}', t-t')\bar{f}_{\perp p}\bar{f}_{\perp q}\mathrm{d}^2x'\mathrm{d}t' \tag{14.71}$$

and

$$\overline{\bar{f}_{\perp p}\bar{f}_\perp} = \int w(\boldsymbol{x}-\boldsymbol{x}', t-t')\bar{f}_{\perp p} f_\perp \mathrm{d}^2x'\mathrm{d}t'. \tag{14.72}$$

14.7 *Additional Methods*

14.7.1 *Second-Order Differential Methods*

The first-order differential method has the basic flaw that the continuity of the optical flow gives only one constraint for the two unknown components of the velocity (Section 14.3.1). So far we could only make up for this deficit by modeling the velocity and thus using a whole neighborhood to determine one estimate of the optical flow vector (Section 14.3.2).

An alternative approach is to use multiple feature or multichannel images. Then we can have two or more independent constraints at the same location and thus may be able to determine both components of the optical flow at a single point. The essential point, however, is that the new feature must bring really new information into the image. It does not help at all if a new feature is closely related to those already used.

In this way we come to an important generalization of the differential method. We can apply any preprocessing to image sequences, or can extract arbitrary feature images and apply all the methods discussed so far. If the continuity of the optical flow is preserved for the original image, it is also for any feature image derived from the original image. We can both apply nonlinear point operators as well as any neighborhood operator.

We first discuss the technique of Girosi et al. [59]. He applied the continuity of the optical flow to two feature images, namely the horizontal and vertical spatial derivative:

$$\begin{aligned} \boldsymbol{f} \nabla g_x + g_{xt} &= 0 \\ \boldsymbol{f} \nabla g_y + g_{yt} &= 0. \end{aligned} \tag{14.73}$$

The usage of horizontal and vertical derivative images thus results into a second-order differential method with the solution

$$\boxed{\boldsymbol{f} = -\boldsymbol{H}^{-1} \nabla g_t \quad \text{if} \quad \det \boldsymbol{H} \neq 0,} \tag{14.74}$$

where $\boldsymbol{H}$ is the *Hessian matrix* as defined in Eq. (12.6).

If we also include the standard optical flow equation, we end up with an overdetermined linear equation system with three equations and two unknowns:

$$\begin{bmatrix} g_x & g_y \\ g_{xx} & g_{xy} \\ g_{xy} & g_{yy} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = - \begin{bmatrix} g_t \\ g_{xt} \\ g_{yt} \end{bmatrix}. \tag{14.75}$$

In this respect, fusion of images gained from different sensors may be a promising method. Markandey and Flinchbaugh [130], for example, used multispectral imagery, one visible and one IR image. Image sequence processing of scenes illuminated with light sources from different directions has been studied by Woodham [220]. This approach is especially interesting since it has the potential to detect specular reflexes and thus to exclude an important source of errors in motion estimation.

14.7.2 *Differential Geometric Modeling*

The discussion in the last sections clearly showed that the spatial structure of the gray values governs the determination of motion. The first-order differential method does not adequately account for this basic fact as we have just relied on first-order spatial derivatives. Second-order differential methods provide a direct solution, provided the Hessian matrix can be inverted (Eq. (14.73)). In this section, we approach differential methods from a different point of view using *differential geometry*. We assume that the gray value structure in the two consecutive images differs only by a constant displacement $\boldsymbol{s}$:

$$g(\boldsymbol{x} - 1/2\boldsymbol{s}, t_1) = g(\boldsymbol{x} + 1/2\boldsymbol{s}, t_2). \tag{14.76}$$

This approach is another formulation of the continuity equation assuming only a translation of the image and neglecting any rotation or deformation of surface elements. We simply assume that the velocity field does not change in a small neighborhood. For the sake of symmetry, we divide the displacement evenly among the two images. With the assumption that the displacement vector $\boldsymbol{s}$ and the size of the surface element are small, we can expand the gray value in both images at the point $\boldsymbol{x} = \boldsymbol{0}$ in a Taylor expansion. First we consider a first-order expansion, i. e., we approximate the gray value distribution by a *plane*:

$$g(\boldsymbol{x} \pm 1/2\boldsymbol{s}) = g_0 + \nabla g \cdot (\boldsymbol{x} \pm 1/2\boldsymbol{s}). \tag{14.77}$$

The planes in both images must be identical except for the displacement $\boldsymbol{s}$. We sort the term in Eq. (14.77) in increasing powers of $\boldsymbol{x}$ in order to be able to perform a coefficient comparison

$$g(\boldsymbol{x} \pm 1/2\boldsymbol{s}) = \underbrace{g_0 \pm 1/2\nabla g\, \boldsymbol{s}}_{\text{offset}} + \underbrace{\nabla g}_{\text{slope}}\, \boldsymbol{x}. \tag{14.78}$$

The first and second term contain the offset and slope of the plane, respectively. We can now estimate the displacement $\boldsymbol{s} = (p, q)^T$ from the condition that both planes must be identical. Consequently, the two coefficients must be identical and we obtain two equations:

$$\begin{aligned} g_0(t_1) - g_0(t_2) &= 1/2\left(\nabla g(t_1) + \nabla g(t_2)\right)\boldsymbol{s}, \\ \nabla g(t_1) &= \nabla g(t_2). \end{aligned} \tag{14.79}$$

The second equation states that the gradient must be equal in both images. Otherwise, a plane fit of the spatial gray value does not seem to be a useful representation. The first equation corresponds to the continuity of the optical flow Eq. (14.9). In Eq. (14.79) only the temporal derivative is already expressed in a discrete manner as the difference of the mean gray values in both images. Another refinement is also due to the digitization in time. The gradient is replaced by the mean gradient of both images. Moreover, we use the displacement vector field (DVF) $\boldsymbol{s}$ instead of the optical flow $\boldsymbol{f}$. As expected, a plane fit of the gray value distribution does not yield anything new. We are still only able to estimate the velocity component in the direction of the gray value gradient.

Therefore a Taylor series expansion of Eq. (14.76) to the second order gives

$$\begin{aligned} g(\boldsymbol{x} \pm 1/2\boldsymbol{s}) &= g_0 \\ &+ g_x \cdot (x \pm 1/2s_1) + g_y \cdot (y \pm 1/2s_2) \\ &+ 1/2 g_{xx} \cdot (x \pm 1/2s_1)^2 + 1/2 g_{yy} \cdot (y \pm 1/2s_2)^2 \\ &+ g_{xy} \cdot (x \pm 1/2s_1)(y \pm 1/2s_2). \end{aligned}$$

Nagel [142] performed a very similar modeling of the gray value geometry, expanding it in a Taylor series up to second order. However, he ended up with quite complex nonlinear equations, which could be solved easily only for special conditions. He termed them *gray value corner* and *gray value extreme*. The reason for the different results lies in the approach towards the solution. Nagel

compares the Taylor expansion in the two images in a least square sense, while here a direct coefficient comparison is performed.

A comparison of the coefficients of the second-order expansion yields six equations in total. The quadratic terms yield three equations which state that all second-order spatial derivatives must coincide in both images:

$$\begin{aligned} g_{xx}(t_1) &= g_{xx}(t_2), \\ g_{yy}(t_1) &= g_{yy}(t_2), \\ g_{xy}(t_1) &= g_{xy}(t_2). \end{aligned}$$

If this is not the case, either the second-order expansion to the gray value distribution does not adequately fit the gray value distribution or the presumption of a constant displacement in the neighborhood is not valid. The coefficient comparison of the zero- and first-order terms results in the following three equations:

$$\begin{aligned} -(g_0(t_2) - g_0(t_1)) &= \tfrac{1}{2}\left(g_x(t_1) + g_x(t_2)\right) s_1 \\ &\quad + \tfrac{1}{2}\left(g_y(t_1) + g_y(t_2)\right) s_2, \\ -(g_x(t_2) - g_x(t_1)) &= g_{xx} s_1 + g_{xy} s_2, \\ -(g_y(t_2) - g_y(t_1)) &= g_{yy} s_2 + g_{xy} s_1. \end{aligned} \tag{14.81}$$

Surprisingly, the coefficient comparison for the zero-order term (offset) yields the same result as the plane fit Eq. (14.79). This means that the DVF is computed correctly by a simple plane fit, even if the gray value distribution is no longer adequately fitted by a plane but by a second-order polynomial.

The two other equations constitute a simple linear equation system with two unknowns:

$$\begin{bmatrix} g_{xx} & g_{xy} \\ g_{xy} & g_{yy} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = - \begin{bmatrix} g_x(t_2) - g_x(t_1) \\ g_y(t_2) - g_y(t_1) \end{bmatrix}. \tag{14.82}$$

We can easily invert the left 2×2 matrix like we inverted the matrix, provided that $g_{xx}g_{yy} - (g_{xy})^2$ does not vanish. Therefore it is possible to estimate the displacement between two images from a local neighborhood if we take into account the *curvatures* of the gray value distribution. We have not yet discussed the conditions the gray value distribution must meet, for Eq. (14.81) to be invertible. This is the case if either a *gray value extreme* or a *gray value corner* is encountered. As already mentioned, these terms were coined by Nagel [142]. At a gray value extreme (as well as at a saddle point) both principal curvatures are non-zero. Thus Eq. (14.82) can be solved. At a gray value corner, only one principal curvature is zero, but the gradient in this direction is not. Thus the first and second equation from Eq. (14.81) can be used to determine both components of the optical flow vector.

With the differential geometric method no smoothing is required since second-order derivatives at only one point are used. However, for a more robust estimate of derivatives, often regularized derivative operators are used as discussed in Section 12.7. Since convolution operations are commutative, this smoothing could also be applied after computing the derivatives.

The difference in the first-order spatial derivatives between the two images at times t_2 and t_1 in the right-hand vector in Eq. (14.82) is a discrete approximation of a temporal derivative which can be replaced by a temporal derivative operator. Then, the displacement vector has to be replaced by the optical flow vector. Thus a continuous formulation of the differential geometric method results in

$$\begin{bmatrix} \overline{g_{xx}} & \overline{g_{xy}} \\ \overline{g_{xy}} & \overline{g_{yy}} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = - \begin{bmatrix} \overline{g_{xt}} \\ \overline{g_{yt}} \end{bmatrix}. \tag{14.83}$$

14.7.3 *Spatiotemporal Energy Models*

Models using Gabor-like *quadrature filters* (Section 13.4.5) are common in biological vision. They are the basis for so-called *spatiotemporal energy* models [1, 2, 75]. This term can easily be misunderstood. It is not the kinetic energy of the moving objects that is referred to but the energy (squared amplitude) of a signal at the sensor in a certain $\boldsymbol{k}\omega$ interval. Here we want to compare this type of model with the differential method discussed previously.

One of the simplest models for 1-D motion vision uses just three quadrature filters. This set of directional filters detects objects moving to the right or to the left, and those which are not moving. We denote the squared magnitude of these quadrature operators by $\mathcal{R}$, $\mathcal{L}$, and S. Then we can obtain an estimate of the 1-D optical flow by using the operator [1, 2]:

$$\mathcal{U} = \frac{\mathcal{R} - \mathcal{L}}{S}. \tag{14.84}$$

An interesting interconnection of this approach with the differential method (Section 14.3.2) can be found, so that the differential method can also be understood as an energy extraction method.

We perform this comparison here for the analysis of 1-D motion, i. e., in a 2-D space-time image. In this case, the solution of the differential method can be written in operator notation according to Eq. (14.25) as

$$\mathcal{U}' = -\frac{\mathcal{B}_{xt}\,(\mathcal{D}_t \cdot \mathcal{D}_x)}{\mathcal{B}_{xt}\,(\mathcal{D}_x \cdot \mathcal{D}_x)}. \tag{14.85}$$

We rewrite this equation with a slight modification to smooth the images with the binomial mask $\mathcal{B}_{xt}$ before we apply the derivative operators, i. e., we use a regularized derivative operator (Section 12.7):

$$\mathcal{U}' = -\frac{\mathcal{B}_{xt}\,[(\mathcal{D}_t\mathcal{B}_{xt}) \cdot (\mathcal{D}_x\mathcal{B}_{xt})]}{\mathcal{B}_{xt}\,[(\mathcal{D}_x\mathcal{B}_{xt}) \cdot (\mathcal{D}_x\mathcal{B}_{xt})]}. \tag{14.86}$$

Smoothing with $\mathcal{B}_{xt}$ means a regularization of the derivative operator. The indices xt indicate that the smoothing is performed along both the temporal and spatial axes. Using the operator identity

$$\mathcal{A}\mathcal{B} = \frac{1}{4}\left[(\mathcal{A} + \mathcal{B})^2 - (\mathcal{A} - \mathcal{B})^2\right] \tag{14.87}$$

and the abbreviations

$$\mathcal{R}' = (\mathcal{D}_x + \mathcal{D}_t)\mathcal{B}_{xt}, \quad \mathcal{L}' = (\mathcal{D}_x - \mathcal{D}_t)\mathcal{B}_{xt}, \quad S' = 2\mathcal{D}_x\mathcal{B}_{xt}, \tag{14.88}$$

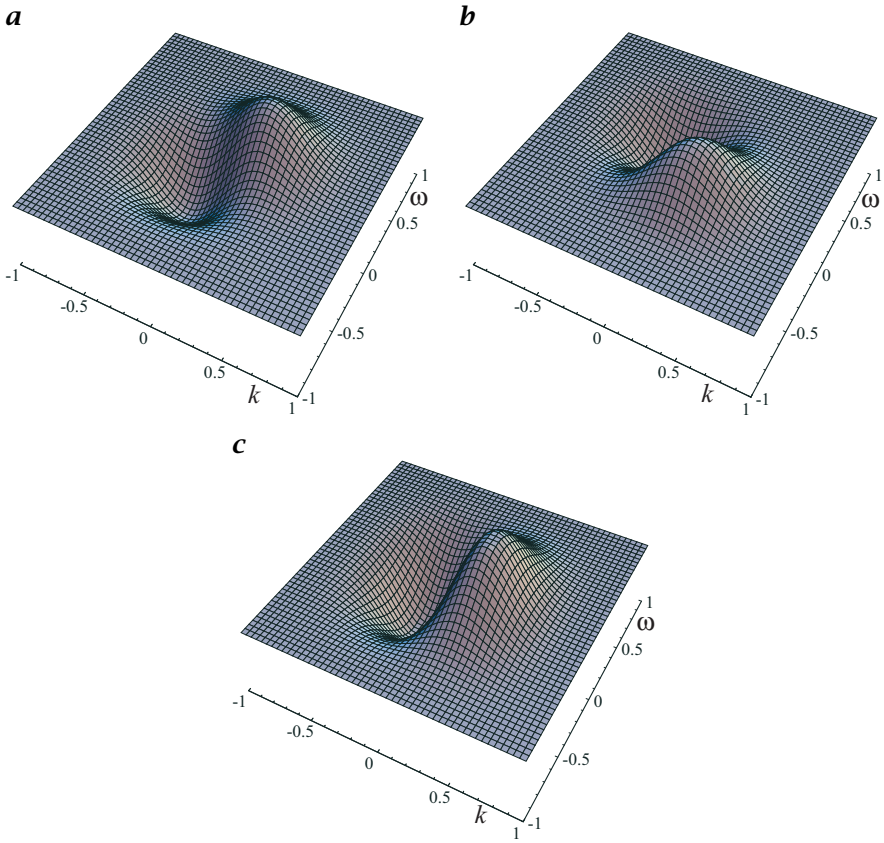

Figure 14.13: *Transfer functions for the convolution operators Eq. (14.88) to detect objects moving right or left, or at rest:* **a** $\mathcal{R}'$, **b** $\mathcal{L}'$, *and* **c** S'.

we can rewrite Eq. (14.86) and obtain a very similar expression as Eq. (14.84):

$$\mathcal{U}' = \frac{\mathcal{B}_{xt}(\mathcal{R}' \cdot \mathcal{R}' - \mathcal{L}' \cdot \mathcal{L}')}{\mathcal{B}_{xt}(S' \cdot S')}. \tag{14.89}$$

The filters $\mathcal{R}'$, $\mathcal{L}'$, and S' are regularized derivative filters. The transfer functions show that objects moving to the right, to the left, and at rest are selected (Fig. 14.13). These filters are *not* quadrature filters. The squaring of the filter responses and further smoothing with $\mathcal{B}_{xt}$, however, approximately results in a phase-independent detection of the squared amplitude as with a quadrature filter under certain conditions. Let us assume a fine-scale periodic structure. The derivative filters will preserve these structures but remove the mean gray value. The subsequent squaring of the zero-mean filter results yields a mean gray value with half of the squared amplitude of the gray value variations and a rapid spatial oscillation of the gray values with the double wave number (half the wavelength). If the subsequent smoothing removes these fast oscillations, a phase-independent response to the filter is obtained just as with a quadra-

ture filter. In contrast to quadrature filters, this result can only be achieved in regions where the scales of the structures are so fine that the doubled wave number can be removed with the smoothing filter.

14.8 *Exercises*

14.1: Accuracy of motion analysis

Interactive demonstration of the accuracy of several methods to determine the motion field using test sequences with known velocity values; output of errors; investigation of the influence of noise and temporal undersampling (dip6ex14.01)

14.2: Motion analysis

Interactive demonstration of various methods for motion analysis with real image sequences (dip6ex14.02)

14.3: **Accelerated morion

With accelerated motion, the continuity equation of the optical flow can be extended as follows:

$$(\boldsymbol{f} + \boldsymbol{a}t)\nabla g + g_t = 0$$

1. Formulate the overdetermined linear equation system for the optical flow $\boldsymbol{f}$ and the acceleration $\boldsymbol{a}$ (4 parameters in 2-D images) with an approach similar to that in Section 14.3.2.
2. Show that it is impossible to determine the acceleration if the sequence contains only two images.

14.4: **Second-order differential method

The second-order differential method determines optical flow without further averaging from Eq. (14.74). At which gray value structures it is possible to determine optical flow from Eq. (14.74) without ambiguity? Does this cover all types of second-order gray value structures at which it is principally possible to determine the complete optical flow vector?

14.9 *Further Readings*

The following monographs on motion analysis are available: Singh [189], Fleet [49], and Jähne [88]. A good survey of motion analysis can also be found in the review articles of Beauchemin and Barron [7] and Jähne and Haußecker [93, Chapter 10]. The latter article also includes the estimation of higher-order motion fields. Readers interested in visual detection of motion in biological systems are referred to the monograph edited by Smith and Snowden [190]. The extension of motion analysis to the estimation of parameters of dynamic processes and illumination variation is described in Haußecker and Fleet [73], Haußecker [72], and Jähne [91]. Methods to analyze various types of complex motion are discussed in Jähne et al. [92].