Chapter 7
Motion Detection and Digital Polarization

Flying insects have extraordinary visual capabilities. Their ability to detect fast motion in the visual scene and avoid collision using low level image processing and little computational power makes their visual processing interesting for real time motion/collision detection in machine vision applications.

In this chapter, some bio-inspired models of motion and collision detection based on differential imaging and correlation will be presented. Section 7.1 describes the process of obtaining motion from a sequence of images. The conventional models for motion detection using differential imaging, background subtraction and optical flow are also briefly introduced. Motion detection using differential imaging is described in section 7.2. Section 7.3 describes the most popular bio-inspired elementary motion detector. It also discusses on the possible simplification of elementary motion detection using one-dimensional binary optical flow. The spatial summation of differential one-dimensional binary optical flow is shown to be able to detect both horizontal and vertical object motion. The partial charge transfer method to increase the dynamic range and the background illumination invariance are described in section 7.4. A digital representation of polarization is presented in section 7.5, where the one-dimensional binary optical flow is shown to vary with the polarization angle of the incoming light rays.

7.1 Motion Detection

Motion is usually determined from image sequences. The spatiotemporal image sequences can be represented using the plenoptic function. The plenoptic function was introduced by Adelson and Bergen [1] to describe all the information available to an observer at any point in space and time, when a light ray passes through the imaging device. In its most general form, the plenoptic function is a seven-dimensional function given by the equation (7.1)

$$P = P(\theta, \varphi, t, \lambda, V_x, V_y, V_z) \tag{7.1}$$

M. Sarkar and A. Theuwissen: *A Biologically Inspired CMOS Image Sensor*, SCI 461, pp. 215–245.
DOI: 10.1007/ 978-3-642-34901-0_7

where (V_x, V_y, V_z) is the position of the imaging sensor in three-dimensional space, λ is the wavelength of light, (θ, φ) represents the azimuth and elevation angles that index the viewable rays and t is time. In a pinhole camera only one sample of the viewpoint is available at any given instance, thus the plenoptic function can be adapted to standard Cartesian parametrization of the rays *(x,y)*, where *x* and *y* are the spatial coordinates in the image plane.

$$P = P(x, y, t, \lambda) \tag{7.2}$$

Though multispectral imaging has started to become popular, most motion analysis are done for a single wavelength and thus the plenoptic function can be rewritten as

$$P = P(x, y, t) \tag{7.3}$$

The plenotic function of a time sampled set of images or "snapshots" is denoted by

$$P = P(x, y, \{t = i, i + \delta t, ..., i + n\delta t\}) \tag{7.4}$$

where *i* is the time instance of an image capture, δt is the time interval between two consecutive images and *n* is total number of images captured. In machine vision, the variations in the plenotic function across sequences of images are used to calculate the image velocity or motion. One of the measurable parameters of a point object which is assumed not to change with time for changing plenoptic function is the intensity or the brightness of the object given by *I(x,y,t)*, where *(x,y)* represents the spatial *x* and *y* dimensions and *t* denotes time [2]. It is thus useful to combine the geometrical description of the scene with brightness information for motion detection [3]. The brightness constancy equation assuming the intensity of a point to remain constant as it moves δx, δy for a time interval δt can be written as

$$I(x - \delta x, y - \delta y, t - \delta t) = I(x, y, t) \tag{7.5}$$

Taylor series expansion of equation (7.5) can be written as:

$$I(x - \delta x, y - \delta y, t - \delta t) = I(x, y, t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t + \text{higher order terms} \tag{7.6}$$

Neglecting the higher order terms in equation (7.6) and using the brightness constancy equation [1] we get

$$\frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t = 0 \tag{7.7}$$

where $\partial I/\partial x$, $\partial I/\partial y$ and $\partial I/\partial t$ are the derivatives of the image *(x,y,t)* in the corresponding directions and thus can be represented by I_x, I_y and I_t. Equation (7.7) can be rewritten as:

$$I_x \delta x + I_y \delta y + I_t \delta t = 0 \tag{7.8}$$

Dividing all the terms in equation (7.8) by δt we get:

$$I_x u + I_y v + I_t = 0 \tag{7.9}$$

where *u(x,y,t)* and *v(x,y,t)* are the horizontal and vertical components of the motion respectively.

Equation (7.9) relates the image velocity to the spatiotemporal derivative of the image at a particular location and is commonly referred to as the motion constraint equation.

Based on the variation in the intensity obtained from the projection of the plenoptic function on the image sensor from the moving object, three conventional approaches are used to detect motion: temporal differencing [4]; background subtraction [5], [6]; and optical flow [2].

Temporal differencing is based on frame difference, and attempts to detect moving regions by making use of the difference of consecutive frames (two or three) in a video sequence. This method is highly adaptive to dynamic environments, but generally does a poor job of extracting the complete shapes of certain types of moving objects [7].

Background subtraction uses a model of the background and compares the current image with the reference image to separate the background and foreground [8]. The main disadvantage of background subtraction method is that both the background scene and the camera are required to be stationary when this method is applied. They are also extremely sensitive to dynamic scene changes due to background illumination changes.

When either the optical scene or the camera is in motion, optical flow is usually used. An object in motion may exhibit both translation and rotational velocities, which are usually projected as the movement of brightness patterns on the image plane. As per the smoothness constraints, the corresponding points in two successive frames should not move more than a few pixels. Thus in an uncertain environment the camera motion or background changes should be relatively small and thus motion can still be detected with optical flow. Optical flow estimation methods can be classified into three main groups: differential methods; matching-based methods and frequency/phase-based. All the three methods consist of three basic components: pre-filtering, local motion estimation and integration over the field of view. Pre-filtering or smoothing of the image data with a low-pass or band-pass filter allows extracting the signal of interest thus enhancing signal-to-noise ratio.

The local motion estimation is done using spatiotemporal derivatives to measure the velocity component and the integration over the field of view produces a two-dimensional flow field of the moving object.

The differential methods include the gradient based models that determine two-dimensional velocities of the moving object employing first or second order spatiotemporal derivatives of the image sequence. The velocity is obtained by dividing the temporal derivative of local luminance by its spatial derivative. The major problem with gradient based methods is that if the spatial derivative is small, the noise in the temporal derivative is amplified and the velocity is poorly defined. Thus due to the differentiation of the image sequence and their susceptibility to errors under noisy conditions they are usually not preferred in low signal-to-noise ratio regimes [9].

Matching-based methods include feature-based and region-based methods. They are also known as block-based, area-based and correlation based methods. These methods usually use techniques to either maximize cross-correlation or minimize differential error. Feature based methods locate and trace identifiable features of the image over time while region based methods try to locate a delineated region in consecutive frame within a search space. To match the feature or the regions in subsequent images either probabilistic approaches like Kalman filters and Monte Carlo localizations or neural networks are used. Matching based methods are robust to large motion and brightness variations however these methods are found to be accurate only at high velocities and it is difficult to estimate sub-pixel displacements [10]. Further correlation based models are normally very sensitive to the amount of data involved because these algorithms are mainly based on byte-level operations of the whole image. These methods involve an operation of all the pixels in the image even though most pixels may not have changed from one frame to another, severely limiting the bandwidth and speed or motion detection for real-time applications.

Frequency/phase based methods use local energy or phase information to determine the velocity of the moving object. These techniques for determining image motion rely on the phase behaviour of arrays of band-pass filters or changes in the output energy of the velocity tuned filters. These filters decompose the input signal according to scale, speed and orientation. However these methods are still susceptible to noise and discontinuity limitations like the gradient based methods. Additionally the outputs of these methods are limited by the design of the filters used [10].

7.1.1 Motion Detection - Models

The problem of real-time motion detection and tracking is an important issue in artificial vision. The main constraints for real-time implementation of these algorithms are the large amount of data to be processed and the high-computational cost of the algorithms employed. To solve the problems, analog VLSI chips employing early vision processing of the optical scene are becoming popular [11], [12], [13], [14]. They employ simple, low accuracy operations at each pixel in an image or sequence

of images, resulting in a low-level description of a scene useful for higher level machine vision applications. These often results in compact, high speed and low power solutions.

The motion detection can be done either on a pixel level or on a region level. The pixel level analysis shows whether the pixel is stationary or transient by observing its intensity value over time. Region analysis deals with the agglomeration of groups of pixels into moving and stationary regions.

It is usually known that the pixel's intensity value displays three characteristic profiles as shown in the figure 7.1. When an object moves through the pixel it displays a profile that exhibits a step change in intensity, followed by a period of instability, then another step back to the original background intensity. When the object moving through the pixel stops at the pixel location, it exhibits a change in intensity, followed by a period of instability and then settling to the new intensity levels. With variations in the ambient lighting, the intensity changes smoothly with no large steps [15]. To know the state of the object, the nature of the pixel intensity profile is important. To interpret the meaning of a step change (for example object passing through, stopping at, or leaving the pixel), one needs to observe the intensity curve re-stabilizing after the step change. This introduces a time-delay into the process. For detection of fast motion this time delay has to be minimized.

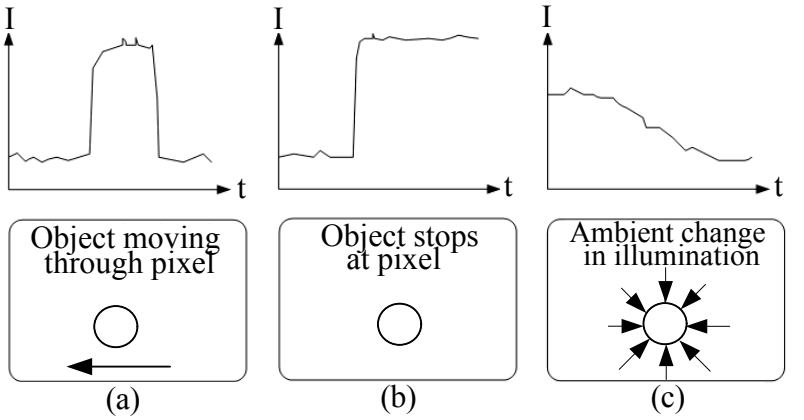

Fig. 7.1 Characteristic pixel intensity profiles for (a) object moving though the pixel, (b) object stopping at the pixel and (c) variations in the ambient illumination changes

In motion detection algorithms where the image motion is analyzed by sending the complete frame out of the image sensor, a very high data rate is required to keep the time delay small. The common method employed to obtain differential images is to use the CCD or CMOS image sensor in high speed frame mode combined with fast readout, a frame buffer to store the frame and a digital subtractor to generate the differential image. The high frame rate is needed to prevent temporal aliasing, which prevents the output signal from being a smooth function of the image parameters thus affecting the efficiency of motion detection algorithms. Furthermore a long time between two image captures will give a difficult correspondence problem in

a dynamic scene. Such an image sensor is highly complex and consumes lot of power. Thus such motion sensing algorithms though in use, for example in MPEG encoders, are not very convenient for real-time motion applications in machine vision. These algorithms for fast motion detection become design overkill for slow motion detections where the time delay can be large.

Temporal differencing at a pixel level rather than at a frame level can help to decrease the data coming from the camera as only the data from the pixel with changed states will be sent out [12], [16]. The focal plane computations are also free from temporal aliasing, which is usually a problem in motion detection algorithms. Further temporal differencing is useful for ambient light suppression. Two methods are discussed, one analog and one digital, in the subsequent sections that can be employed for speeding up low level motion detection without much complexity and power consumption. These algorithms are based on pixel changes instead of full image processing and thus improve performance. The pixel level processing provides an extreme data rate compression; it has a high data rate in (image sequence) and low data rate out (motion).

7.2 Motion Detection - Differential Imaging

In differential imaging two consecutive images are stored in the pixel as an analog voltage and a difference signal is obtained during the readout. Differential imaging can be used for machine vision applications like motion detection, object tracking or object recognition etc. All the stationary objects with constant illumination will be ignored by the imaging system.

An active differential sensing method is described here, with focal plane computations which need minimal external active components. The subtraction of the two images captured at varying exposure time is done pixel by pixel. The process is repeated for each image line and thus a differential image is obtained without a requirement for an off-chip computation.

The designed image sensor explained in chapter 3 can operate in two modes: double differential mode *(DDS)* and differential imaging mode *(DI)* using the two available analog memories in the pixel. The operation of the image sensor in the *DDS* mode was shown in section 3.6.5.1. The pixel diagram with the two analog memories and the operational timing diagram of the *DI* mode of the pixel are shown in the figure 7.2.

In the *DI* mode the sensor takes two samples and subtracts one from the other. The first sample is taken after an integration time T_1 while the second sample is taken after an integration time T_{FT} and are stored in the two analog memories in the pixel. The imaging array operates in global shutter mode as the signals after the integration time is transferred to the memories at the same time for all the pixels of the imaging array. However while reading out, each row is read at a time. The two signals stored in the pixel are subtracted at the column level of the analog signal chain shown in figure 3.33. The subtraction of the two images reduces the *FPN* and since there is no reset between the two image captures the reset noise is the same in both images and it is removed with subtraction.

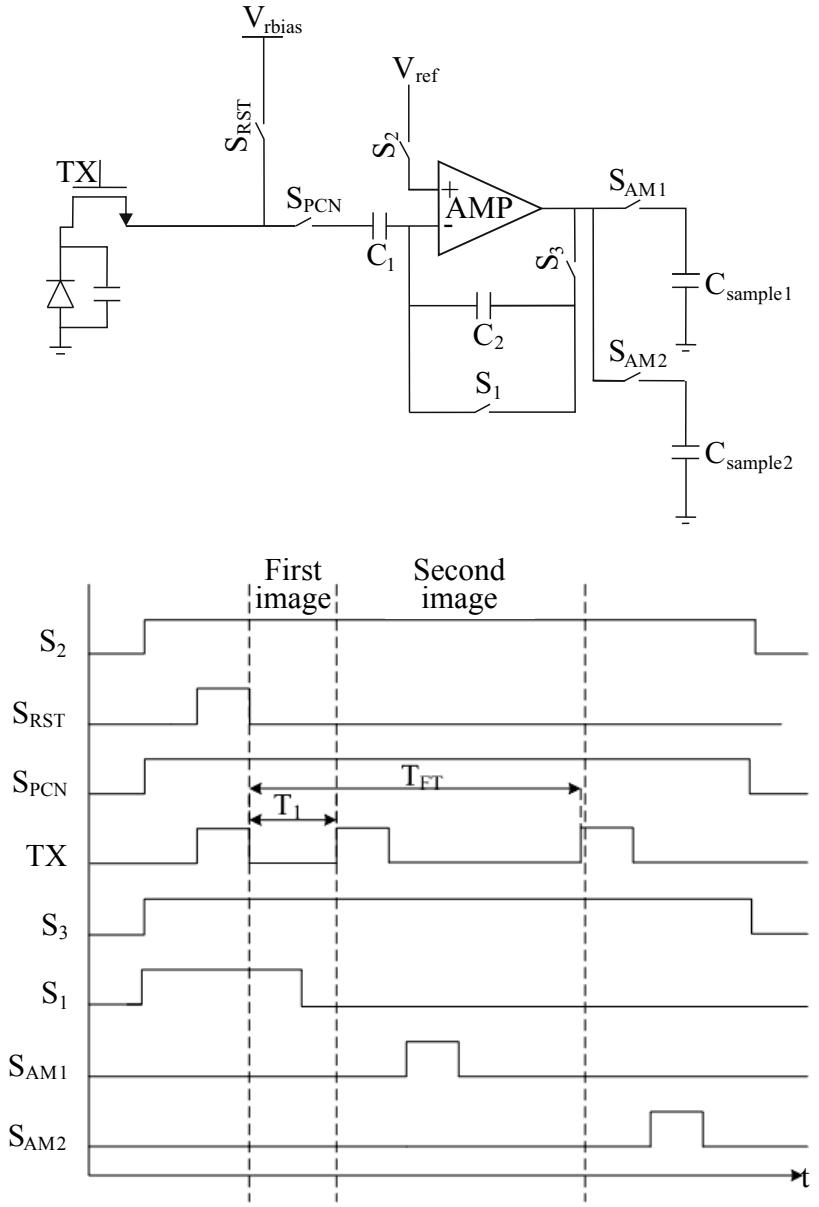

Fig. 7.2 Differential imaging mode

Figure 7.3 shows an example of motion detection, when the object moves to a pixel position and stops, similar to the situation shown in figure 7.1(b).

When the exposure time of the two samples is the same $(T_1 = T_{FT}) = 20ms$, a black image is observed when there is no change in the background as shown in

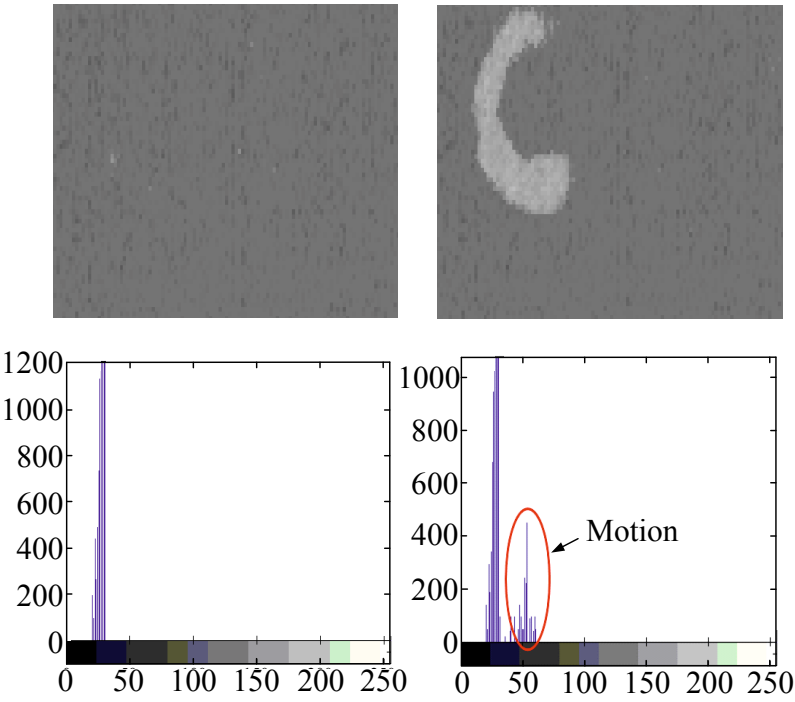

Fig. 7.3 Motion detection using differential mode of operation

figure 7.3(a). When a DC light source is switched on (to resemble motion), there is a change in the background and the change is visible in the differential image 7.3(b). The histograms shows the gray values of the pixels. When there is no motion, most of the pixels have lower gray values while when there is motion pixels start to get higher gray values.

For motion detection all the static objects or DC light sources in the background have to be removed and only the AC change is to be detected. Thus a very high Common-Mode Rejection Ratio *(CMRR)* is desired to remove the common mode or DC signal in subsequent images of a scene. Common-mode rejection ratio of a device is the measure of the tendency of the device to reject common input signal. A high *CMRR* is desired for applications where the signal of interest is a small voltage fluctuation superimposed on a large voltage offset or when the required information is contained in the voltage difference of two signals.

The definition of the *CMRR* of the differential image is simplified from the conventional definition of *CMRR* and defined to be the change in the common mode signal divided by the change in the differential signal due to the change in the common mode signal [17]. A high *CMRR* is desired to have the same effect when two white images or two black images are subtracted from each other. The maximum voltage swing obtained after *DDS* for the designed sensor is 488mV as seen in section 3.7.2.1, for dark to saturation exposure of the sensor. The output

voltage swing obtained when operating in the differential mode ($T_1 = T_{FT}$) is 2.44mV for varying illumination from dark till the sensor saturation. The *CMRR* thus obtained is 200 or 46dB which is higher than the *CMRR* of 100 reported in [17].

The changes in the differential signal when the normal exposure period (T_{FT}) is held constant at 20ms while the short integration time (T_1) is varied is shown in the figure 7.4.

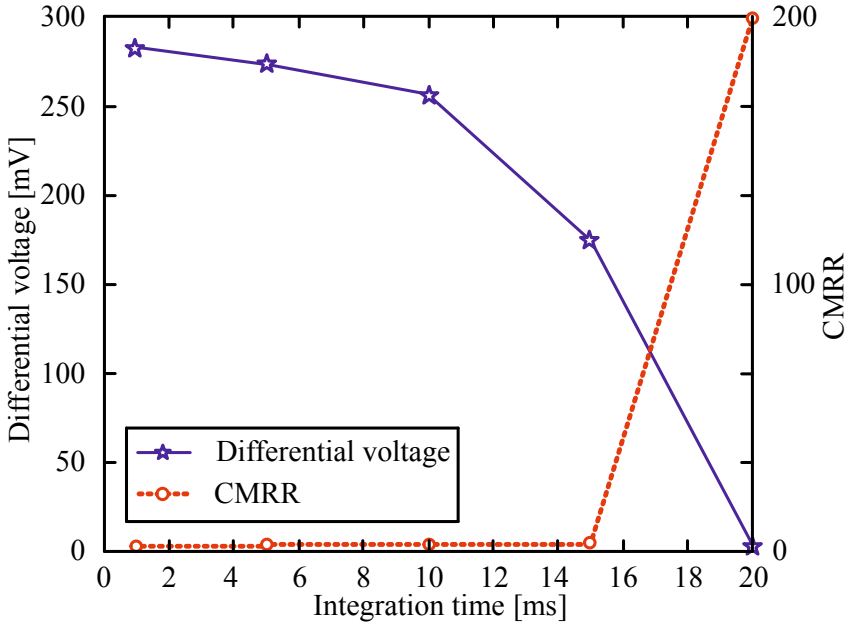

Fig. 7.4 Variation in the differential swing and CMRR with changes in the exposure time

A non-linear behavior is seen when T_1 is very low compared to T_{FT}. If the ratio of the two integration periods is defined as $R_{DE} = T_{FT}/T_1$, then it is observed that for larger values of R_{DE} the non-linearity is quite high. The accumulated electrons for a given illumination condition decrease with the increase in R_{DE} due to shorter accumulation period (The accumulated electrons are proportional to $1/R_{DE}$). When the accumulated electrons are low, the error in transferring these accumulated electrons from the photodiode to the floating diffusion node is higher. Thus the non linearity increases with increase in R_{DE}. The results are very much in agreement with the theoretical studies of [18]. The *CMRR* practically remains low for all values of T_1. When T_1 is equal to T_{FT}, it reaches the maximum value of 200.

7.3 Motion Detection - Optical Flow

A one-dimensional motion can either be in a horizontal direction or in a vertical direction. Vertical motion of an object towards the sensor leads to a collision

and thus needs to be prevented. The visual guided collision avoidance has been extensively studied using conventional cameras and digital processing devices. However conventional cameras are not suitable for real time applications as they rely on motion estimation using sequential images.

It is well known that flying insects are able to detect obstacles in their flying path efficiently with little computational power by using optical flow. Optical flow is the pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer and the scene. Optical flow usually contains information about self motion and distance to potential obstacles and thus it is very useful for navigation [19]. In section 7.1, the three methods of obtaining optical flow for motion detection have been summarized. These methods are however broadly equivalent. In biological motion vision, the correlation based model or the gradient detector model are often used to account for the direction selectivity [20]. The correlation based model is shown to have significant advantages over the gradient based model in regimes where signal-to-noise ratio is low and detector noise is of concern.

The circuit of each pixel of the designed image sensor contains a comparator to detect the difference between the integrated charge from the photodiode and an external threshold voltage (chapter 3). This allows the generation of binary optical flow similar to the effect of *"flickering"* in the eyes of insects. The binary optic flow is the change in the digital pixel values in response to objects motion in the focal plane of the image sensor. The generated optical flow can be used to detect motion both in vertical (section 7.3.1) and horizontal (section 7.3.2) direction with minimal processing and hardware.

7.3.1 Motion in Vertical Direction - Collision Detection

In this section vertical motion in the focal plane is considered to estimate the time to collision of a moving object with the image sensor using binary optic flow. Reliable estimation of the time to collision between two moving objects is very important in many applications such as autonomous agent navigation or predictive crash sensors for autonomous safety systems. Currently existing non-biologically inspired collision avoidance systems use a CCD/CMOS camera and digital processing devices to detect the approaching object. Such a collision detection system is not suitable for compact real-time computations as it requires large amount of computations.

7.3.1.1 Elementary Motion Detector (EMD)

The most popular bio-inspired visual guided collision avoidance approach uses the correlation-type elementary motion detector *(EMD)*, first proposed by Hassenstein and Reichardt to compute the optical flow [21]. They have been used to explain directional selective motion in a wide variety of insects, birds and mammals including humans [10]. This model is very well established and often used in

bio-inspired robots [22], [23] and [24]. The *EMD* correlates the response of one photoreceptor to the delayed response (inhibitory response) of an adjacent photoreceptor, both looking in the same direction as shown schematically in figure 7.5.

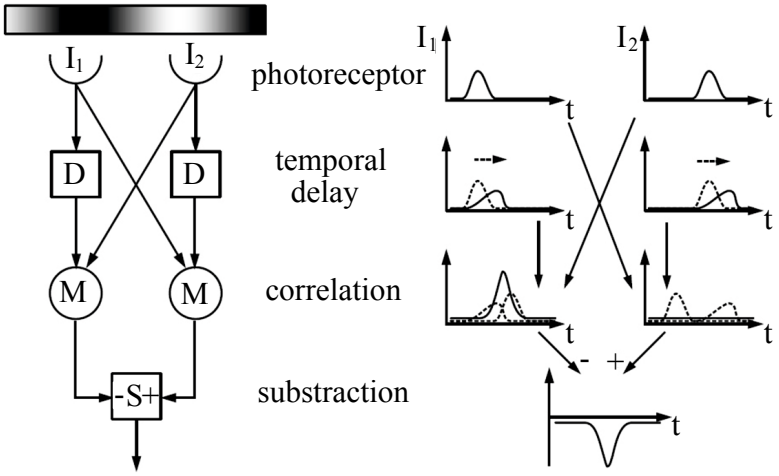

Fig. 7.5 The Hassenstein-Reichardt or correlational elementary motion detector (EMD) [25]

The elementary motion detector consists of two spatially separate inputs to measure the changes across space, temporal filters (delay) to measure the changes across time and a comparator to evaluate spatial and temporal changes. The coincidence of the original signal from one point in space and the delayed signal from the neighboring point in space leads to a positive output signal. The final output signal of the *EMD* is a transient response obtained by subtracting two images. The output image is sensitive to temporal changes in intensity of light, enhancing the directional properties and motion detection by rejecting the effects of temporal contrast not resulting from motion. The motion detection by subtracting two images is better suited than the motion detection using derivative which amplify the noise.

The two separate inputs of the *EMD* model achieve what is referred to as "temporal decorrelation" or "transient enhancement" [23], which is a way of removing the redundant information from the photoreceptor signal before further processing to determine motion. The temporal decorrelation is a filtering mechanism, in which the DC light (mean ambient light) which doesn't have any motion information will be subtracted and AC light (which corresponds to motion) would be allowed to the next stage for further processing.

Existing implementations of the *EMD* use a complex circuit with many active and passive components in order to obtain the inhibition of the signal and the correlation. Further because the basic collision detectors try to measure the velocity of the approaching object, they need dedicated processing blocks for velocity calculation from optical flow field. Additionally, the outputs of the *EMD* are not invariant to the changes in the background illumination and their responses are not only

proportional to velocity changes but are strongly affected by the contrast and the spatial frequency components of the scene [26]. Because the *EMD* multiples its two input, it has quadratic dependence on the contrast and is thus not suitable for low contrast motions. This problem of quadratic dependence can be solved by introducing static expansive non-linearity in the channels [23] however this makes the model more complex. Further because the time delay in the two channels is constant, the *EMD* may infer a wrong conclusion if the time intervals of the moving object between the two points is not the same as the delay.

The standard output of a collision detector using an *EMD* is a response which stays low when the object is far enough and peaks before collision for an approaching object and then collapses to low values again [23]. The peaking of the collision detection algorithm depends on the delay in the delaying channel and the time an object takes to appear in the neighboring channel as well as the variations in the background illuminations [27]. Thresholding of this response to ascertain collision is thus difficult.

7.3.1.2 Proposed EMD Model for Collision Detection

The correlation based models are usually computationally intensive and have difficulties in selecting reliable estimates. In order to have compact and low-power real time autonomous motion detection systems, the estimation of the optical flow has to be simplified. This means that rough qualitative properties of the optical flow are more desirable for efficient collision detection than accurate target distance estimation [19]. The collision avoidance maneuvers in insects can be explained in terms of perception of looming stimuli or expanding images. The landing behavior of insects and the saccade (rapid turns) exhibited by example flies are believed to be triggered by image expansion as detected by an array of local motion detectors [28]. Collision detection using expanding images is also relatively independent of the spatial structure of the object being approached.

To understand the plenoptic function of expansion of images the motion constraint equation (7.9) is used. Since in collision detection the motion is assumed to be only in the vertical direction of the focal plane, the motion constraint equation can be rewritten as

$$I_y v + I_t = 0 \tag{7.10}$$

Equation (7.10) shows that the time of impact or time of collision of the object is directly proportional to the reciprocal of velocity of the object. Equation (7.10) can be better understood using figure 7.6, which shows an image of an approaching object of diameter *D* at a constant velocity *V* along the optical axis. The distance between the lens and the object is *d(t)* while the focal length of the lens is *f*.

The diameter of the obtained image and its derivative with respect to time is given as:

$$a(t) = \frac{f \times D}{d(t)} \tag{7.11}$$

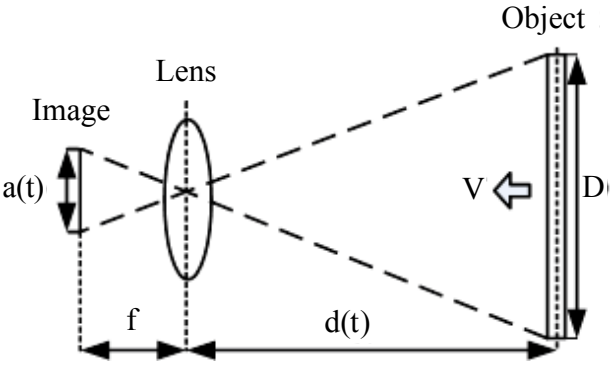

Fig. 7.6 Perception of approaching objects

$$\frac{d(a(t))}{dt} = -\frac{f \times D}{d(t)^2} \tag{7.12}$$

The size of the image *a(t)* in equation (7.11) increases with a decreasing distance *d(t)* and vice versa. The change in the image size affects the optical flow perceived by the image sensor.

To simplify the optical flow generation algorithm, the binary output of the pixel is used. The binarization of the data in the pixel is explained in section 2.6.2.3. In the collision detection experiments, the image sensor is held stationary so that the optical flow is always generated by the motion of the object in the visual field. Figure 7.7 shows the variation in the light spot (approaching object). As the object moves closer to the image sensor, the image size (spot size on the imaging plane) grows or the optical flow expands. With the expanding of the optic flow the intensity profile of the pixel will also increase (figure 7.1(c)), and more pixels will have an output voltage higher than the reference voltage, and a digital '1' will be stored in the *SRAM* cells.

Fig. 7.7 Perception of approaching objects

From the digital images, the percentage of active high pixels for a given illumination condition can be computed as the ratio of total active high pixel to the total number of pixels in the array:

$$\%\ \text{of active high pixels} = \frac{\text{Total active high pixels}}{\text{Total number of pixels}} \tag{7.13}$$

Equation (7.13) represents the one-dimensional binary optical flow. The percentage of the active high pixels will increase with the approaching bright object as predicted by equation (7.11). The measured variations in the percentage of active high pixels with the variation in the distance of the light source for single image capture is shown in figure 7.8. It shows that when the light source approaches the image sensor, the optical flow, which is the variation in the intensity with motion, causes more pixels to become active thus increasing the percentage of active high pixels.

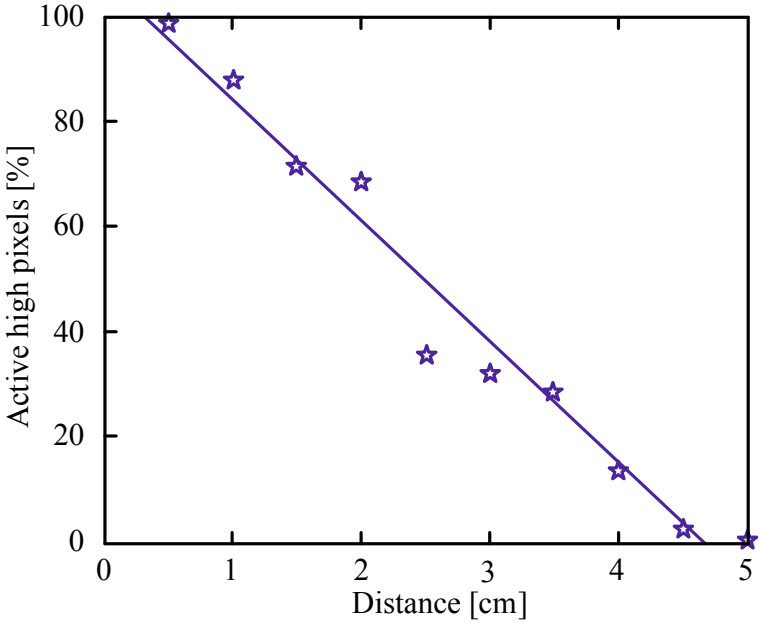

Fig. 7.8 One-dimensional binary optical flow variation with approaching object

One of the major requirements of motion detection using correlation models is temporal decorrelation. Temporal decorrelation refers to the process of reducing the autocorrelation with a signal in time domain, this helps in removing redundant information in images separated in time. Temporal decorrelation can be obtained using differential imaging, where two samples are spaced in time. The differential image is generated using partial charge transfer, where the integrated charge at the photodiode capacitance is transferred to the *FD* node multiple times in one frame (chapter 3). Figure 7.9 shows the algorithm used.

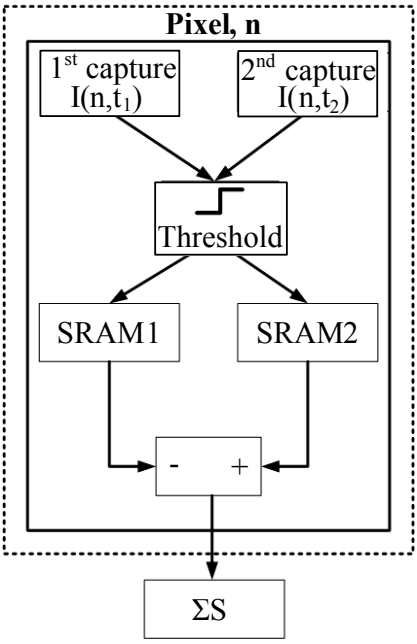

Fig. 7.9 Modified EMD model for collision detection

The first image capture is at a time instance of T_1 and the second capture is at T_{FT}. The captured samples are then compared with the reference voltage and the digital output is stored in the two *SRAM* cells available in the pixel. The differential image of the two spatially integrated digital images obtained from *SRAM1* and *SRAM2* was computed off chip for this version of the sensor.

The temporal decorrelation of the optic flow obtained using the partial charge transfer is shown in figure 7.10. The figure shows the variation in the percentage of active high pixels of two image captures for varying distance of the object from the imager. The first image is captured after an integration time of T_1 and the second after the total frame time T_{FT}. Two measurements for T_1 when 10ms and 1.5s are shown in the figure 7.10. The variation of the time instances allows to generate varying decorrelated one-dimensional binary optical flow.

The difference of the two temporally decorrelated optical flows is plotted in the figure 7.11. It can be observed that as the object moves towards the image sensor, it has a certain threshold of percentage of pixels with changed states below which the object is very near to collision. In this case the collision detection mechanism does not need to use dedicated motion processing blocks. The collision can be detected to a very good degree of reliability using the percentage of changed pixels with the varying one-dimensional differential optical flow.

Harrison [23] uses a collision detection algorithm based on the basic *EMD* model and the algorithm peaks at 230ms before collision for an object with a velocity of 17cm/s. Thus the collision alert is generated at a distance of approximately 4cm.

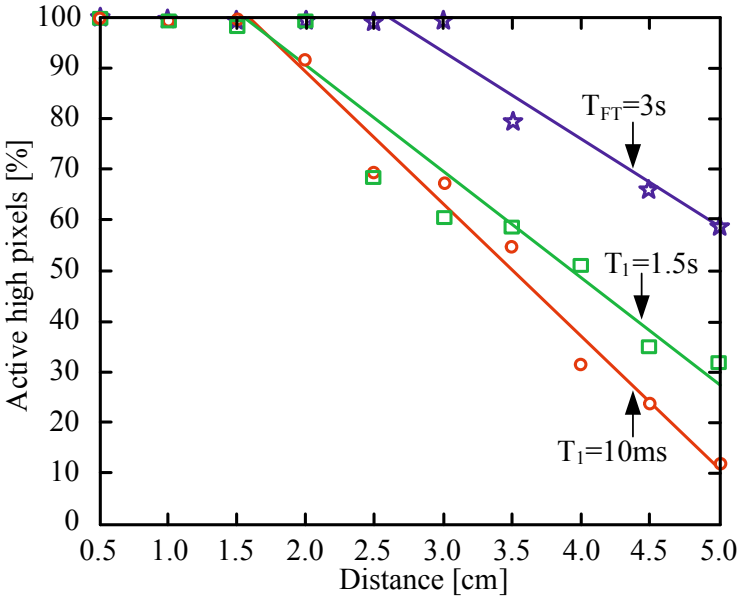

Fig. 7.10 Temporally decorrelated optical flow with approaching objects

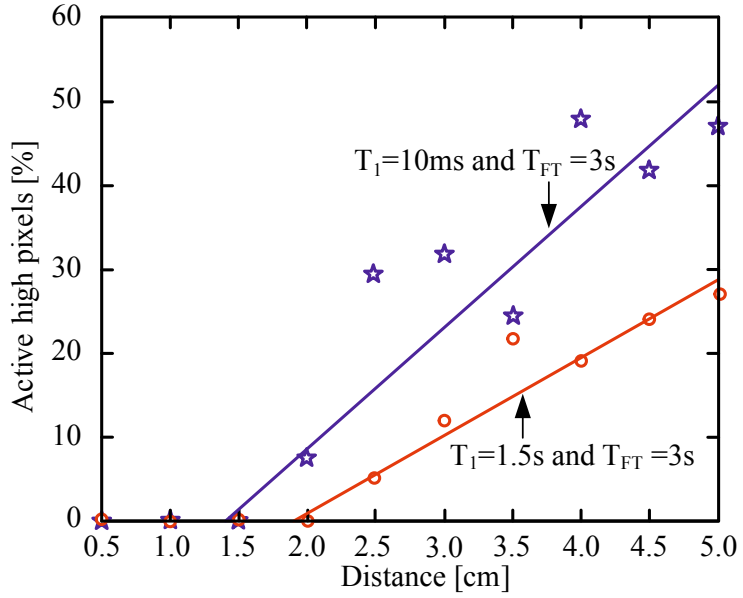

Fig. 7.11 Temporal differential decorrealated optical flow with approaching object

A similar method used by Okuno and Yagi [22] produces a collision detection alert at a distance of 63m. The differential optical flow is shown to generate a collision alert for distances less than 2cm. For narrow path autonomous agent navigations, a collision alert for very small distance is desired. For example in applications like endoscopy, a collision alert distance of less than a cm would be ideal. The distance of collision alert can also be modulated by varying the T_1 and T_{FT}, depending on the application and situation.

Table 7.1 Performance comparison

	[22]	[23]	This work
Method used	*EMD* model based on locust	Delay and correlate *EMD*	Differential optical flow imaging
Collision alert distance	63m	$\sim$4cm	$\leq$2cm

Furthermore the proposed algorithm doesn't suffer from the inherent disadvantage of the *EMD* model in the accurate detection of the output peak. The output of the differential optical flow continues to stay low near collision allowing thresholding and thus is more stable. By modulating the differential time, it would be possible to prevent collision in very narrow paths thus helping navigation of the autonomous agents. Further as most of the computations are done on focal plane there is no image transfer bottleneck, which is usually present in the conventional approach of using an image sensor together with a digital signal processor.

So far the collision detection algorithm is employed in an environment with an illuminated object moving towards the image sensor in a dark background. However since the temporally decorrelated signal is a differential signal, it would also work in an environment condition where a dark object moves vertically with respect to the image sensor in a light background. In such a case initially the binary optic flow would be composed off all digital '1' and as the dark object approaches the image sensor, the optic flow would toggle to digital '0'. The temporally decorelated difference image would thus subtract two images whose optic flow would be dominated by digital '0'. The difference image would thus show the % percentage of active high pixels to exhibit a behavior similar to figure 7.11.

7.3.2 Motion in Horizontal Direction

In this section only horizontal motion is considered. From the pixel array only the binary output of the comparators are used for the motion detection. The 7-bit counter counts the number of '1' in each row of the pixel for each frame. The algorithm for motion detection then compares the counter outputs to decide if there is motion. If the difference of the counter outputs for two exposures is higher than a certain threshold, motion occurrence is flagged.

For the designed sensor the brightness control voltage is the reference voltage to which the analog signal obtained after each exposure is compared. The two *SRAM*

cells in the pixels serve as frame latches and offer both past and current data. The pixel converts the image data into a one-bit data stream by the comparators.

To verify the proposed model two consecutive frames of a light source moving over the image sensor are shown in figure 7.12.

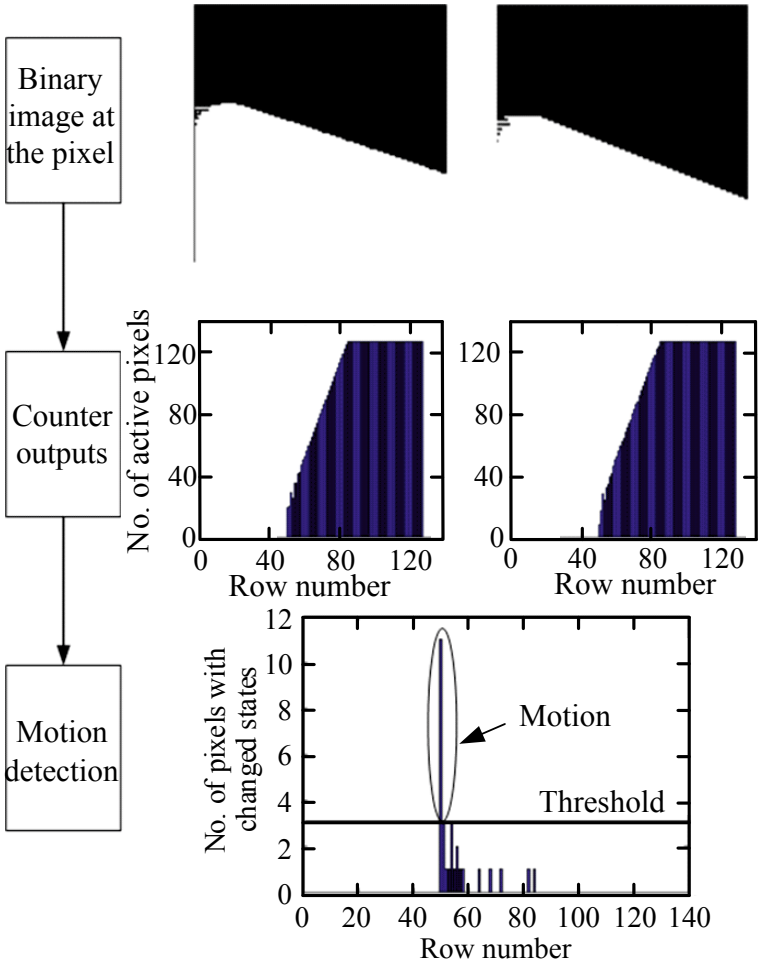

Fig. 7.12 Horizontal motion detection using spatially integrated binary optical flow

The left image shows the light source at its initial position and the right image shows it after a slight movement. The two images look very similar, as only a very small motion was introduced. The histograms of the two images are shown in figure 7.12. The subtraction of the two images results in a difference image, and the histogram of the pixels which changed states are shown at the bottom of the figure. By selecting a proper threshold, accurate detection of motion can be done.

7.4 Illumination Invariant and High Dynamic Range Motion Detection

The outputs of the *EMD* are not invariant to the changes in the background illuminations as discussed in section 7.3. This problem is illustrated in figure 7.13 which shows the increase in the *EMD* output with the increase in the background illumination.

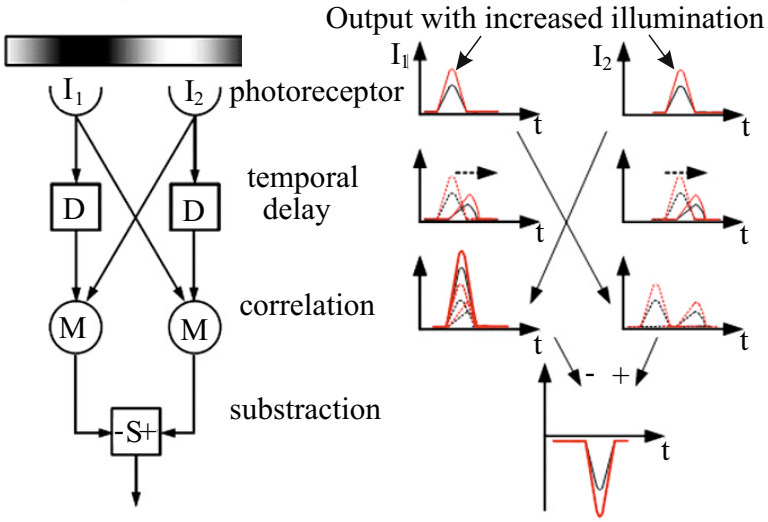

Fig. 7.13 The correlational elementary motion detector showing increased motion for change in illumination

An increase in the background illumination increases the output of the *EMD* and this increased output can be confused with motion. Also it is shown than the optical flow computed on the image doesn't match actual motion with non-uniform illumination [29]. The estimation of motion parameters from the spatiotemporal patterns of visual stimuli or optical flow implicitly requires responses that are invariant with respect to absolute illumination levels, contrast and the spatial structure of the scene.

For the *EMD* model to have a good performance over varying illumination conditions, the dynamic range of the sensor also needs to be high. A real-world scene is composed of varying level of brightness within it. The dynamic range of the scene is typically always higher than the dynamic range of the sensor used to capture the scene. A typical image sensor has a dynamic range of about 65 to 75dB while a scene can have a dynamic range of over 120dB. In high illumination conditions the photodiode of the pixel saturates very quickly. For a very bright object most of the photodiodes will be saturated and thus finer details and motion cannot be captured.

To increase the dynamic range either the maximum signal at the photodiode node (well capacity) has to be increased or the read noise has to be decreased

(equation (3.15)). The relationship between output, integration time and photocurrent at the floating diffusion node of a linear image sensor is given by:

$$V_{signal} = \frac{t_{int} \times I_{photo}}{C_{eff}} \tag{7.14}$$

The photocurrent produced in the photodiode is proportional to the irradiance and the light collection area or the photodiode area. The larger the pixel exposed surface, the more light it can collect during the exposure period and thus the higher is the dynamic range. The photodiode area is limited by pixel size constraints. For high resolution a small pixel size is desired. This leaves us with two parameters in equation (7.14) that can be modified to increase the dynamic range: the integration time t_{int} and the effective capacitance to store the collected charges during integration C_{eff}. The total capacitance C_{eff} on the photodiode sense node is the sum of the junction capacitance proportional to the diode area and the parasitics due to the sensing and amplifying circuitry.

From equation (7.14) it is clear that the slope of the pixel response can be changed by the scaling of the effective capacitance. Changing the effective capacitance varies the well capacity. The effective capacitance can be varied using either smart reset pixels [30], using overflow MOSFET capacitors [31] or by using multiple shorter exposure periods [32], [33]. In the smart reset pixels, the reset gate voltage is monotonically decreased during integration causing the well capacity (charge capacity) of the sensor to monotonically increase. Using a lateral overflow capacitor, the overflowed charges from a fully depleted photodiode during an exposure can still be integrated. The disadvantage of such a scheme is the effective increase of the floating node capacitance which reduces the charge conversion gain and also the fill factor of the pixel. In the multiple shorter integration period methods, several images with different exposure time are captured. The images with a shorter exposure time captures the brightest areas of the scene, while the images with longer exposure time capture the darker areas of the scene. A high dynamic range image is then synthesized from the multiple captures.

Since differential imaging is used in the proposed model of the *EMD* to obtain the temporal decorrelation necessary for motion detection, a partial charge transfer method is used to increase the dynamic range of the model. This method provides a signal after a short integration time in addition to a signal after a long integration time [18]. The difference of the two charge accumulation times in one frame extends the dynamic range of the sensor by removing the static or DC light.

The photo-conversion characteristics for the complete charge transfer *(DDS)* and the differential signal for the partial charge transfer with various integration is shown in figure 7.14. T_1 is varied from 5ms to 20ms.

Figure 7.15 shows the variation in the saturation intensity with varying $T_1 : T_{FT}$ ratio. The saturation intensity is the intensity of light causing the photodiode to saturate. It is observed that the saturation intensity increases with an increase in $T_1 : T_{FT}$ ratio or with an increase in the accumulation time T_1. By appropriately changing the $T_1 : T_{FT}$ ratio, the saturation intensity can be doubled thus increasing the sensitivity to the background brightness.

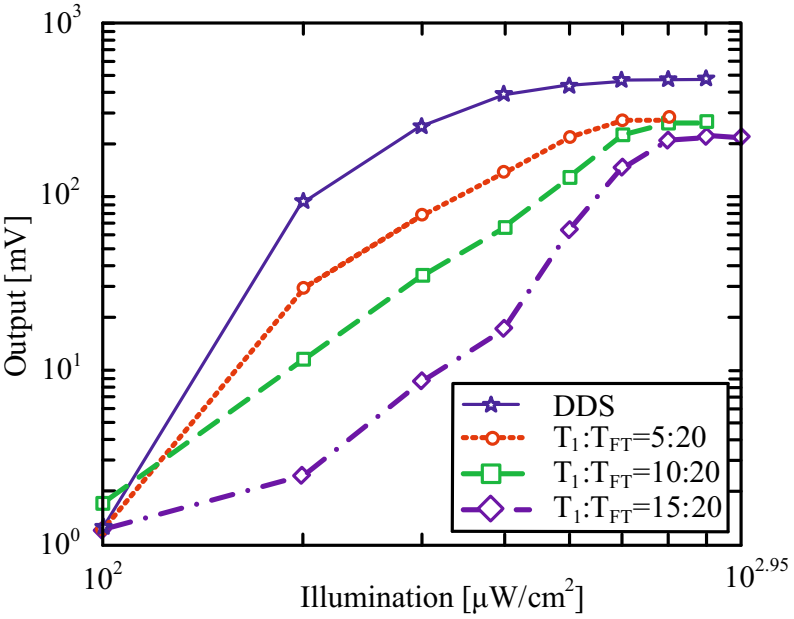

Fig. 7.14 Photo-conversion characteristics at $\lambda = 550nm$

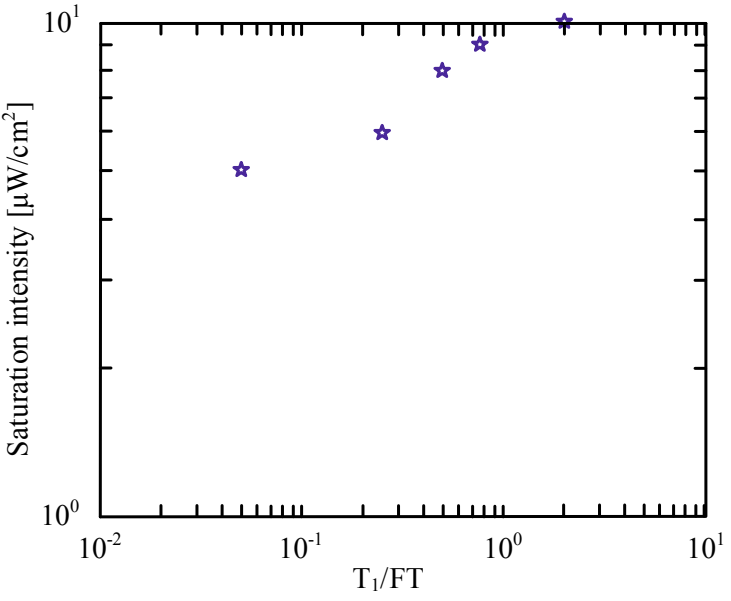

Fig. 7.15 Saturation intensity variation with variation in $T_1 : T_{FT}$ ratio

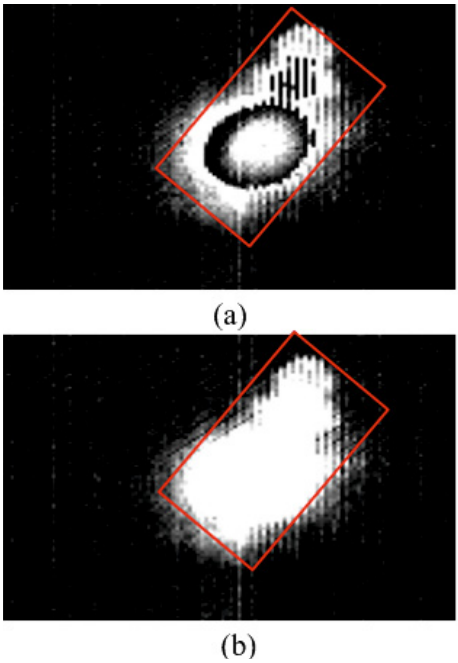

Fig. 7.16 DC light source images with $T_1 : T_{FT}$ of (a) 0.0005 and (b) 0.05

Figure 7.16 shows an example of increasing the saturation level using partial charge transfer. Images of a DC light source were taken with different $T_1 : T_{FT}$ ratios. Figure 7.16(a) and (b) show the images with $T_1 : T_{FT}$ values of 0.0005 and 0.05 respectively. In figure 7.16(a) the pixels at the bright portions of the DC light source saturate, causing an overflow with a very short accumulation time T_1. The differential signal with increased T_1 is shown in figure 7.16(b) and is able to capture the bright portion of the DC light source.

In partial charge transfer, when the charges are transferred from the photodiode to the sense node, some amount of charge is left in the photodiode, as shown in figure 3.24(d) (chapter 3). This introduces image lag and can be removed by resetting the photodiode. However, a reset introduces *kT/C* noise. The amount of residual charge in the photodiode depends on the accumulated charge. Since the accumulated charge is proportional to the exposure time, the residual charge increases with the exposure time. The increase in the residual charge increases in turn the non-linearity of the photo-conversion characteristic and further decreases the sensitivity. The accumulated signal obtained after partial charge transfer with a shorter (T_1) period is also non-linear with respect to the incident light. The non-linearity is believed to be caused by the carrier diffusion and the initial condition of the photodiode [18].

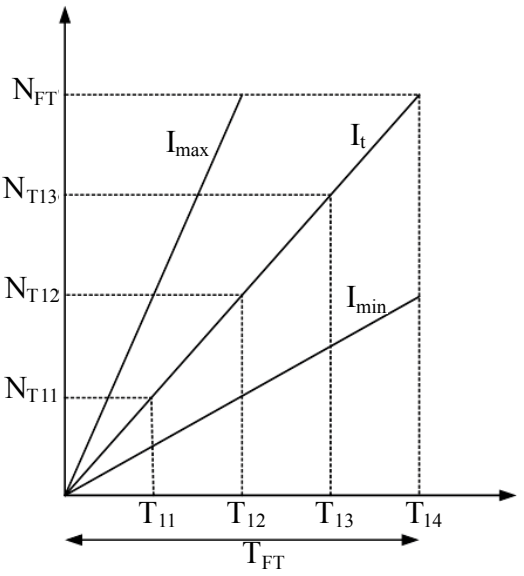

Fig. 7.17 Charge accumulation for various illumination and integration periods

The increase in non-linearity at higher illumination levels for the partial charge transfer can be modeled using figure 7.17. It shows the accumulated charges for various intensities of light I_t (normal), I_{max} (maximum) and I_{min} (minimum) and varying T_1 (T_{11}, T_{12}, T_{13} and T_{14}).

N_{FT} is the number of electrons accumulated when the sensor is operated in conventional or complete charge transfer mode with a frame time of T_{FT}. For the partial charge transfer mode T_1 can vary and four different time instances are shown in figure 7.17. For illumination I_t, the following equations hold

$$q \times N_{T12} = I_t \times T_{12} \tag{7.15}$$

If N_{a1} is the number of residual charges after the first charge transfer at T_{12} for a illumination of I_t then

$$q \times (N_{FT} - N_{a1}) = I_t \times T_{FT} \tag{7.16}$$

$$N_{T12} = \frac{T_{12}}{T_{FT}} \times (N_{FT} - N_{a1}) \tag{7.17}$$

For the highest illumination I_{max}, the photodiode saturates very quickly. N_{a2} is the residual charge after the first charge transfer at T_{12} for an illumination of I_{max}

$$q \times N_{T12} = I_{max} \times T_{11} \tag{7.18}$$

$$q \times (N_{FT} - N_{a2}) = I_{max} \times T_{12} \tag{7.19}$$

$$N_{FT}{}^2 - a_1 \times N_{FT} + b_1 = 0 \tag{7.20}$$

where

$$a_1 = N_{a1} + N_{a2} \tag{7.21}$$

$$b_1 = N_{a1} \times N_{a2} - \frac{T_{FT}}{T_{11}} N_{T12}{}^2 \tag{7.22}$$

Figure 7.18 shows the photo-conversion characteristics of the synthesized wide dynamic range signals with the dynamic range extension ratio T_1 to T_{FT} set to 1:20, 5:20 and 10:20 for higher illuminations. It is observed that for decreasing T_1 to T_{FT} ratios the non-linearity increases at high illumination. It is further observed that with the decrease in the T_1 to T_{FT} ratio the sensitivity increases.

The sensitivities obtained for $T_1 : T_{FT}$ ratio of 10:20, 5:20 and 1:20 are 11.46 mV/ms, 12.89 mV/ms and 14.37 mV/ms respectively.

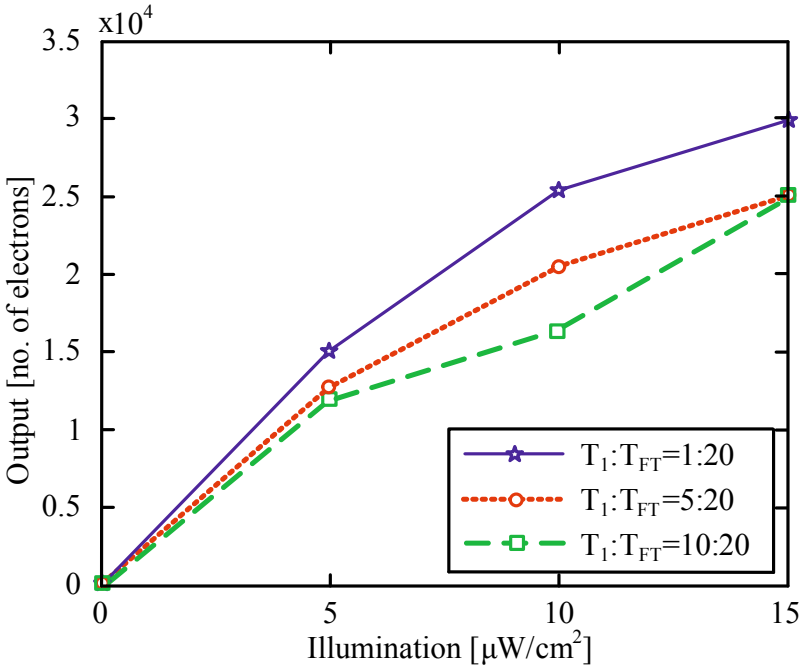

Fig. 7.18 Non-linearity at high illumination for partial charge transfer

The associated non-linearity at high illumination with partial charge transfer is interesting, as in motion detection a non-linear interaction between two spatially separated signals is usually desired [34]. In insects the photoreceptors adapt to the mean luminance in the enviornment and give an approximately logarithmic response to changes in light intensity [35]. This compressive non-linearity allows a better response to contrast than absolute luminance. Non-linearity in the Hassenstein-Reichardt *EMD* model is introduced either by thresholding or rectification to obtain a saturating contrast response curve for motion adaptation [36], [37]. The effect of the non-linearities associated with high illumination obtained with partial charge transfer on motion detection is yet to be explored.

The collision detection algorithm in section 7.3 is designed using one-dimensional binary optical flow. Figure 7.19 shows the percentage of active high pixels when the photodiode is operated in complete charge transfer $(T_1 = 0ms)$ and partial charge transfer $(T_1 = 10ms)$ respectively. It is observed that with partial charge transfer higher illumination levels are captured, thus increasing the digital dynamic range of the image sensor.

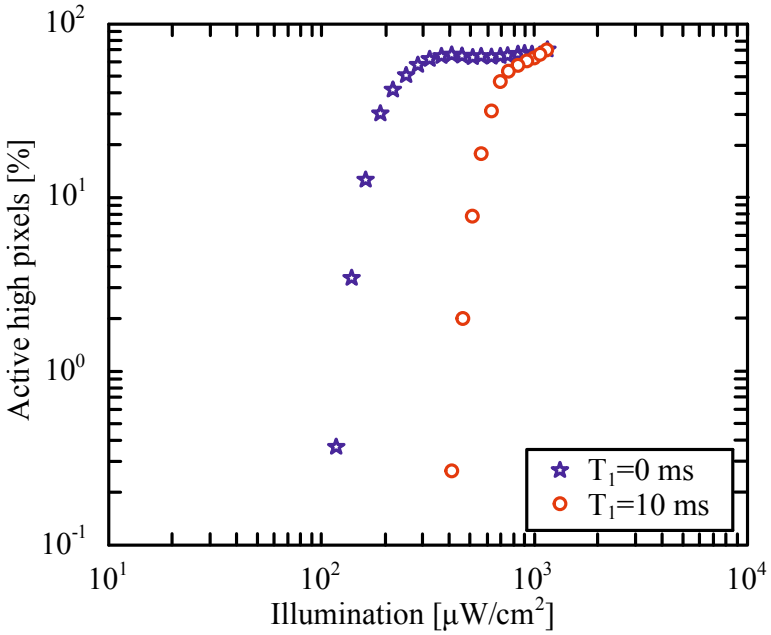

Fig. 7.19 Variations in the optical flow for partial charge transfer

7.5 Digital Polarization

In chapter 4 a CMOS polarization sensor with embedded metallic wire gird was discussed. The image sensor was shown to respond to the changes in the polarization angle. This polarization sensitivity was further shown to be useful in machine vision

applications like material classification and autonomous agent navigation in chapter 5 and 6 respectively. However the proposed solutions for these applications were based on using the analog signals from the pixels. In section 7.3, a simple way to calculate one-dimensional-correlational optical flow using the number of active high pixels at a given time was introduced. The number of active high pixels will vary depending on the horizontal and vertical motion of the object source and its relative intensity variations on the focal plane. Here only motion in vertical direction to the imaging plane is considered. Thus depending on the object moving close or far away from the image sensor, the spot on the focal plane increases or decreases. The size of the spot will be further dependent on the polarization angle of the incoming light ray and the one-dimensional binary optical flow can thus be used to represent polarization in digital format.

The expected theoretical behaviour of the one-dimensional binary optical flow for the two polarization sense regions (chapter 4) is shown in figure 7.20.

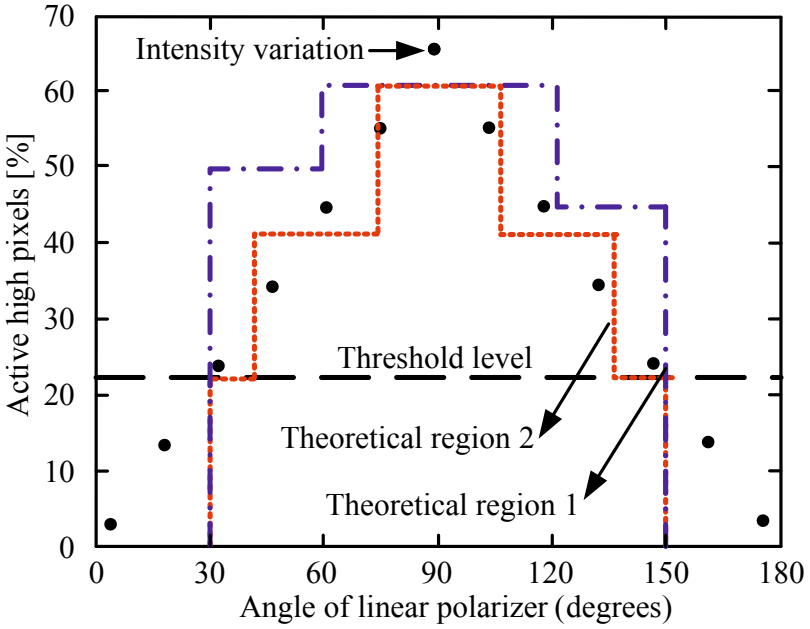

Fig. 7.20 Theoretical behaviour of one-dimensional optical flow with variation in linear polarizer angle

For the polarization sense region 1, when the object is moving towards the imaging plane the intensity of the pixels slowly increases. As the light intensity reaches the threshold level, one expects the two intensity sensitive pixels to store a digital '1', the output of the comparator, in the *SRAM* cells. The percentage of active high pixels shows a step rise to around 50%. As the intensity is further increased around 90°, the 90° sensitive pixels will slowly start to have a have a digital '1',

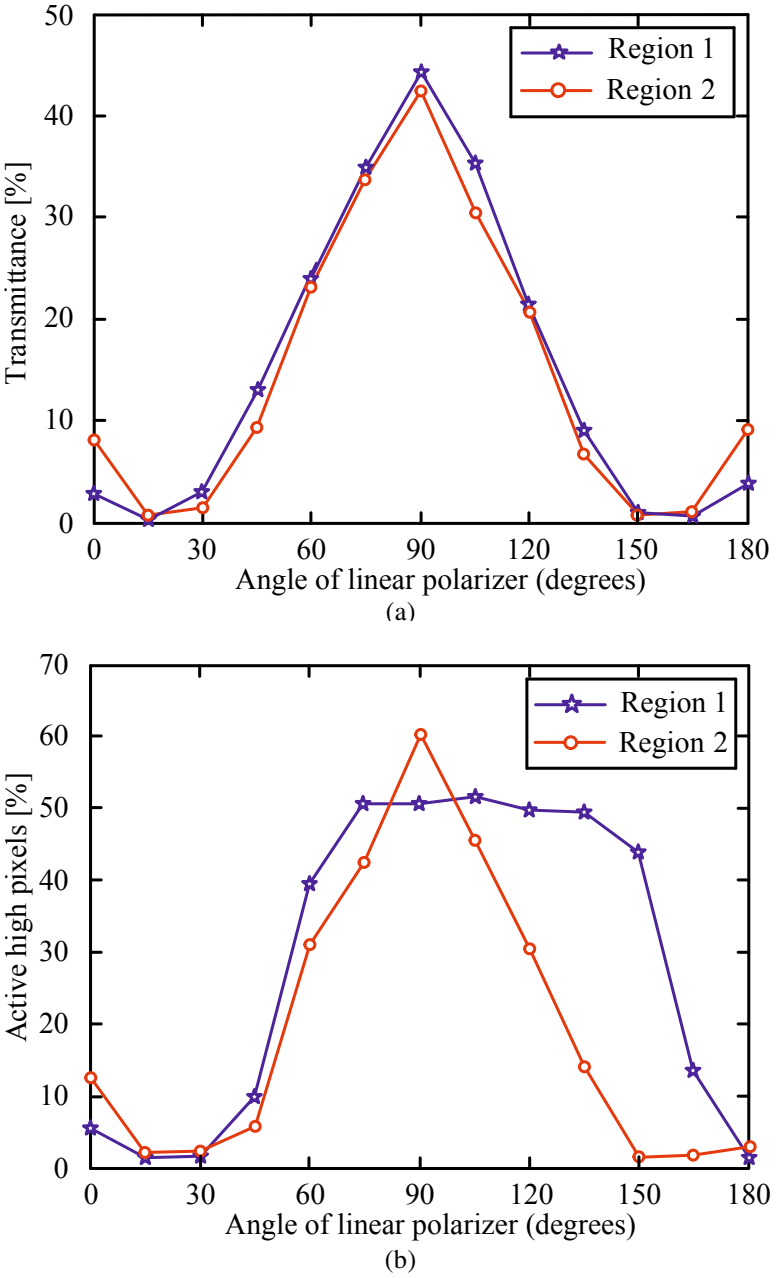

Fig. 7.21 90° polarization in sense regions 1 and 2: (a) analog (b) digital (optical flow variation)

as their comparators' outputs, and hence the percentage of active high pixels will increase above 50%. Ideally, when all the 90° sensitive pixels turn high, 75% of the pixels in region 1 will be active high. However due to the attenuation of the light by the external polarizer not all 90° sensitive pixels turn high. In the polarization sense region 2, the theoretical behavior of the percentage of active high pixels would be the same as region 1, except that as the intensity increases, the 45° sensitive pixels start to turn high. Thus there is an additional step rise when the polarizer angle is 45°. In the experiments, the optical flow is obtained by increasing the light spot gradually from the center to the periphery of the polarization sense regions, using a linear polarizer. This results in a linear increase in the percentage of active high pixels with the variations in the linear polarizer angle, instead of the expected step rise.

The experimental setup for the measurement is the same as described in chapter 4 for analog polarization measurements. The analog performance for the 90° polarizer filter in the two polarization sense regions 1 and 2 is compared with the one-dimensional binary optical flow for varying angle of linear polarizer in figure 7.21.

The measured one-dimensional binary optical flow is shown to have an angular dependence on the angle of the linear polarizer and is very similar to the theoretically predicted behavior. The optical flow and analog representations of polarization in region 2 match closely. It can be predicted that by increasing the number of metallic wire grid orientations over the photodiode a digital representation of polarization very similar to the analog representation can be obtained [38].

A generalized algorithm to represent polarization information will have multiple advantages in low level polarization based machine vision applications. Based on the one-dimensional binary optical flow variations with the polarization angle, a way to determine the Stokes parameters, degree of polarization and polarization Fresnel ratio in binary format can be formulated which will allow focal plane processing of applications like material classification and autonomous agent navigation. Such a sensor would be miniaturized, bandwidth compressor and low power which are highly desirable in the future generations of sensors in machine vision applications.

7.6 Summary

- Motion is usually determined from image sequences. The spatiotemporal image sequences can be represented using the plenoptic function.
- For changing plenoptic function, the intensity of the point object is assumed to be constant.
- Three conventional approaches are used to detect motion: temporal differencing; background subtraction; and optical flow.
- Temporal differencing is based on frame difference. Background subtraction separates the background and foreground and optic flow detects the movement of the brightness patterns on the image plane.
- Optical flow estimation methods can be classified into three main groups: differential methods; matching-based methods and frequency/phase-based.

- The differential methods determines two-dimensional velocities of the moving object employing first or second order spatiotemporal derivatives of the image sequence.
- Matching based methods usually use techniques to either maximize cross-correlation or minimize differential error.
- Frequency/phase based methods use local energy or phase information to determine the velocity of the moving object.
- In motion detection algorithms where the image motion is analyzed by sending the complete frame out of the image sensor, a very high data rate is required to keep the time delay small.
- Temporal differencing at a pixel level rather than at a frame level can help to decrease the data coming from the camera as only the data from the pixel with changed states will be sent out.
- The most popular bio-inspired visual guided motion detector is the correlation-type elementary motion detection (EMD). The optic flow is computed by correlating the response of one photoreceptor to the delayed response of an adjacent photoreceptor, both looking in the same direction.
- The compound eyes of insects are better suited for detection of motion. Inspired by the them, a CMOS image senor operating in temporal differential mode and spatial integration of one-dimensional binary optical flow to detect motion/collision of moving objects was designed.
- The dynamic range of a sensor can be enhanced using partial charge transfer, which also provides for a background illumination invariant motion detection system.
- The one-dimensional binary optical flow also have an angular dependence on the angle of the linear polarizer. By increasing the number of wire grid orientations, a digital representation of the polarization very similar to the analog can be obtained.

References

[1] Adelson, E., Bergen, J.: The plenoptic function and the elements of early vision. In: Landy, M., Movshon, J. (eds.) Computational Models of Visual Processing, pp. 3–20. MIT Press, Cambridge (1991)

[2] Barron, J., Fleet, D., Beauchemin, S.: Systems and experiment performance of optical flow techniques. International Journal of Computer Vision 12(1), 43–77 (1994)

[3] Negahdaripour, S., Horn, B.: Direct passive navigation. IEEE Transactions on Pattern Analysis and Machine Intelligence 9(1), 168–176 (1987)

[4] Anderson, C., Burt, P., van der Wal, G.: Change detection and tracking using pyramid transformation techniques. In: Proceedings of SPIE-Intelligent Robotics and Computer Vision, vol. 579, pp. 72–78 (1985)

[5] Haritaoglu, I., Davis, L., Harwood, D.: W4 who? when? where? what? a real time system for detecting and tracking people. In: International Conference on Automatic Face and Gesture Recognition, pp. 222–227 (1998)

[6] Wren, C., Azarbayejani, A., Darrell, T., Pentland, A.: Pfinder: Real-time tracking of the human body. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 780–785 (1997)
[7] Doretto, G., Soatto, S.: Editable dynamic textures. In: IEEE Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 137–142 (2003)
[8] Elgammal, A., Harwood, D., Davis, L.: Non-parametric Model for Background Subtraction. In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1843, pp. 751–767. Springer, Heidelberg (2000)
[9] Borst, A.: How do flies land? from behavior to neuronal circuits. Bio-Science 40(4), 292–299 (1990)
[10] Brinkworth, R., O'Carroll, D.: Robust models for optic codeing in natural scences inspired by insect biology. PLoS Computational Biology 5(11), 1–14 (2009)
[11] Tanner, J., Mead, C.: An integrated analog optical motion sensor. In: VLSI Signal Processing II, pp. 59–76. IEEE Press, New York (1986)
[12] Delbrück, T.: Silicon retina with correlation-based, velocity-tuned pixels. IEEE Transactions on Neural Networks 4(3), 529–541 (1993)
[13] Kramer, J.: Compact integrated motion sensor with three-pixel interaction. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(4), 455–460 (1996)
[14] Moini, A., Bouzerdoum, A., Eshraghian, K., Yakovleff, A., Nguyen, X., Blanksby, A., Beare, R., Abbott, D., Bogner, R.: An insect vision based motion detection chip. IEEE Journal of Solid-State Circuits 32(2), 279–283 (1997)
[15] Collins, R.T., Lipton, A.J., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., Tolliver, D., Enomoto, N., Hasegawa, O., Burt, P., Wixson, L.: A system for video surveillance and monitoring. CMU VSAM Final Report, 1–69 (1999)
[16] Lichtsteiner, P., Posch, C., Delbrück, T.: A A 128×128 120db 30mw asynchronous vision sensor that responds to relative intensity change. In: IEEE ISSCC Digest of Technical Papers, pp. 508–509 (2006)
[17] Innocent, M., Meynants, G.: Differential image senor with high common mode rejection. In: Proceedings of European Solid-State Circuits Conference, pp. 483–486 (2005)
[18] Shafie, S., Kawahito, S., Halin, I., Hasan, W.: Non-linearity in wider dynamic range CMOS image sensors utilizing a partial charge transfer technique. Sensors 9, 9452–9467 (2009)
[19] Srinivasan, M.V., Zhang, S.W., Chahl, J.S., Barth, E., Venkatesh, S.: How honeybees make grazing landings on flat surfaces. Biological Cybernetics 83, 171–183 (2000)
[20] Borst, A.: Correlation versus gradient type motion detectors: the pros and cons. Philosophical Transactions of the Royal Society, B: Biological Sciences 362(1479), 369–374 (2007)
[21] Reichardt, W.: Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. Principles of Sensory Communication, 303–317 (1961)
[22] Okuno, H., Yagi, T.: Bio-inspired real-time robot vision for collision avoidance. Journal of Robotics and Mechatronics 20(1), 68–74 (2008)
[23] Harrison, R.: A biologically inspired analog IC for visual collision detection. IEEE Transactions on Circuits and Systems-I 52(11), 2308–2318 (2005)
[24] Reiser, M., Dickinson, M.: A test bed for insect-inspired robotic control. Philosophical Transactions of the Royal Society, A: Mathematical, Physical and Engineering Sciences 361, 2267–2285 (2003)
[25] Neumann, T.R., Bülthoff, H.H.: Behavior-oriented vision for biomimetic flight control. In: Proceedings of the International Workshop on Biologically Inspired Robotics, pp. 196–203 (2002)

[26] Krapp, H.: Neuronal matched filters for optic flow processing in flying insects. In: Neurological Processing of Optic Flow, pp. 93–120. Academic Press, San Diego (2000)
[27] Tao, S., Zeng, L.: An elementary-motion-detector based on hardware lateral inhibition network. Journal of Physics: Conference Series 48, 212–216 (2006)
[28] Lee, D.: A theory of visual control of braking based on information about time-to-collision. Perception 5, 437–459 (1976)
[29] Horn, B., Schunck, P.: Determining optical flow. Artificial Intelligence 17, 185–203 (1981)
[30] Decker, S., McGrath, R., Brehmer, K., Sodini, C.: A 256 x 256 CMOS imaging array with wide dynamic range pixels and column- parallel digital output. IEEE Journal of Solid-State Circuits 33, 2081–2091 (1998)
[31] Akahane, N., Sugawa, S., Adachi, S., Mori, K., Ishiuchi, T., Mizobuchi, K.: A sensitivity and linearity improvement of a 100db dynamic range CMOS image sensor using a lateral overflow integration capacitor. In: Symposium on VLSI Circuits, Digest of Technical Papers, pp. 62–65 (2005)
[32] Yadid-Pecht, O., Fossum, E.: Wide intrascene dynamic range cmos aps using dual sampling. IEEE Transactions on Electron Devices 44, 1721–1723 (1997)
[33] Yang, D., Gamal, A.E., Fowler, B., Tian, H.: A 640 x 512 cmos image sensor with ultra-wide dynamic range floating-point pixel level ADC. IEEE Journal of Solid-State Circuits 34, 1821–1834 (1999)
[34] Buchner, E.: Elementary movement detectors in an insect visual system. Biological Cybernetics 24, 85–101 (1976)
[35] van Hateren, J., Snippe, H.P.: Information theoretical evaluation of parametric models of gain control in blowfly photoreceptor cells. Vision Research 41(14), 1851–1865 (2001)
[36] Buchner, E.: Behavioural analysis of spatial vision in insects. In: Photoreception and Vision in Invertebrates. A: Life Sciences, New York, London, vol. 74, pp. 561–621 (1984)
[37] Borst, A., Egelhaaf, M.: Principles of visual motion detection. Trends in Neurosciences 12(8), 297–306 (1989)
[38] Sarkar, M., Segundo, D.S., van Hoof, C., Theuwissen, A.: An analog and digital representation of polarization using CMOS image sensor. In: Proceedings of 5th European Optical Society Tropical Meeting on Advanced Imaging Techniques (2010) ISBN: 9783000305030