# Direct Extraction of Tau Information for Use in Ego-Motion

---

A thesis

submitted in partial fulfilment

of the requirements for the Degree

of

Master of Science in Psychology

in the

University of Canterbury

by

Stephen J. Gaukrodger

---

University of Canterbury

2004

[A] normal act of visual attention consists of scanning a whole feature of the ambient array, not of fixating a single detail of the array. ... The invariants of structure in an optic array that constitute information are more likely to be gradients than small details, and they are scanned over wide angles.

(Gibson, 1979, p.246)

## Abstract

Avoidance collisions with obstacles is a critical function of any autonomous vehicle. This thesis considers the problem of utilising information about time to contact available in the ambient optic array. Motion-from-smear (W.G. Chen, Nandhakumar, & Martin, 1994; Geisler, 1999) is used to aid judgment of global tau (Kaiser & Mowafy, 1993; D. N. Lee, 1974, 1976). A robotic system employing motion-from-smear was tested in a task requiring judgment of global tau and found to provide adequate accuracy (mean error = -0.52s) but poor precision (SD = 1.52s). Motion-from-smear is also discussed with respect to its application to a novel formulation for composite tau and a use of motion parallax in stair descent.

# Contents

# Chapter 1

# Introduction

## 1.1 The Environment and Ego-Motion

When we navigate through our environment, we rely mainly on visual information for the control of speed, orientation and heading. Not only do we use visual information to identify and avoid obstacles in our path, it is also very difficult to maintain a steady course without it. Try walking in a straight line with your eyes closed, or looking up at the sky, or on a flat, snow covered field.

The environment through which we navigate is made up of surfaces. These surfaces can be as large as the ocean or a flat desert plain, or as small as a grain of sand. Regardless of size, these surfaces reflect or emit light to some degree, and this reflected and emitted light makes up the ambient optic array (J. J. Gibson, 1961). We use our visual system to extract information about the environment and our relation to it from this ambient optic array in order to guide action. Whether the ambient optic array is capable of providing all of the information required for action is an empirical issue. Gibson (1979) proposed that, theoretically, a texture gradient is sufficient for

the provision of layout information. Information processing and modern constructivist theorists claim that, since stimulation is inadequate and the eye is an imperfect instrument, the human brain takes the impoverished information that is available and combines it with previously learned information to represent the environment. Both of these positions may have merit, but as descriptive models of the human visual system, both leave something to be desired. Work by Milner and Goodale (1993, 1995) suggests that the reason may be that one part of the visual cortex might use Gibsonian principles to support perception for action, while another might use information processing methods to support perception for higher cognitive functions such as object recognition and decision-making.

Robots are expected to perform tasks that are usually performed by humans, so it is necessary to determine what visual information is required for a robot system to function in an environment. An artificial visual system, were it to prove capable of mimicking the essential aspects of human perception, could replace guide dogs, improve security systems, drive cars, detect traffic violations and allow autonomous vehicles to function in places where even remote control is difficult. For example, it can take up to 40 minutes for light to get to Mars and back, making it very difficult to operate a vehicle there by remote control. Apart from the obvious utility of a robot capable of autonomous locomotion, such a system could provide candidates for optical information to guide future research into human visual perception.

While vision is by no means the only way of guiding a vehicle capable of autonomous control, it does have enormous advantages over other common techniques. Radar and sonar both use Doppler shifts in emitted signals to specify the layout of the

environment. One major disadvantage of emitting systems is that a large number of sources in close proximity emitting constantly (think of the number of cars at a busy intersection during rush hour) would interfere with each other and make it difficult to calculate the exact layout of the environment without some form of co-ordination between the emitting entities. Secondly, since they emit signals regularly it is difficult to make them stealthy, which is a major consideration for military applications.

We perform many tasks that utilise vision in everyday life, and it would be useful if we were able to incorporate most of these in an autonomous vehicle. Following Milner and Goodale (1992, 1995) we can divide these tasks into two groups: those used for guiding action (e.g. collision avoidance, placement of feet, control of heading direction and posture, and eye tracking) and those used for guiding decision making (object recognition, and other non-active cognitive functions). The main overlap between the two groups is in areas concerning navigation: where do we want to go, how do we get there, and how do we know we have arrived? Such questions require decisions to be made and objects to be recognised, but are clearly tied to the guidance of action. Indeed, it may be useful to view the division of perception and action in a 'fuzzy' way. Milner and Goodale (1995) identified many sites where the two streams interacted, and it would be foolish to assume that the visual system would ignore useful information for the sake of theoretical purity. Different tasks could be placed at different points on a continuum, spread between two extrema (Figure 1.1), with the position on the continuum indicating the relative roles of the two streams.

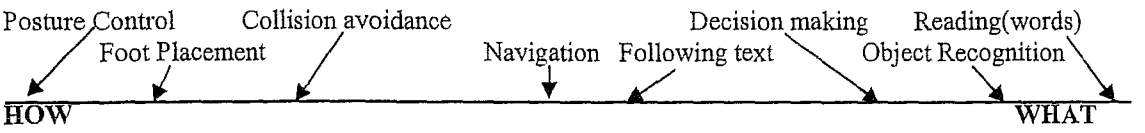



Figure 1.1 A hypothetical example of a fuzzy continuum for describing the roles of the ventral and dorsal streams in visually guided tasks. For example, reading an individual word would only require object recognition, but following text (which can be done without the ability to read a language) would require both object recognition and guidance of eye movements. The names given for the extrema – w*hat* and *how* – follow from Milner and Goodale (1995), but it should be stressed that the positions of the various tasks are purely for the sake of example.

It would be quite possible to build an autonomous automobile that could be provided with navigational information from an external source, such as GPS, while relying on vision to avoid collisions and stay on the road.[1] This thesis will be concerned with those aspects of ego-motion that would concern the visual system of such an autonomous vehicle.

Any actor in the environment has effectivities, those tasks that the actor can perform effectively. For example, the length of my legs allows me to negotiate objects that are not more than a metre above the ground. The ability to negotiate an obstacle one metre in height is an effectivity for me. If another actor were able to negotiate the same object, then that actor and I would have the same effectivity, whether that actor locomoted on its legs, hands, tentacles, or even levitated. A visual system for ego-motion must support detection of properties of surfaces that the actor's effectivities allow it to locomote upon, around or over.

Put simply, any visual system must support detection of properties of surfaces that the

[1] Of course it would also be necessary for the vehicle to identify red lights and the like, but we will consider this to be a type of navigation information.

agent can locomote across. In Gibson's terms such a surface *affords* locomoting upon (Gibson, 1979). These surfaces can include the ground plane, flat-topped steps that are not too high above the ground plane, or holes/ridges that are not so wide that they cannot be negotiated. For a fully functional autonomous vehicle it would also be necessary to perceive surfaces that could **not** be locomoted across, such as quicksand, but this would require differentiation of substances and surfaces (Gibson, 1979) by vision alone. That is a non-trivial task, since the knowledge humans use to solve many of these problems comes from past experience. This additional knowledge is beyond the purview of a simple visual system.

## 1.2 Systems for visual guidance of ego-motion

In order to support visually guided ego-motion the visual system must extract at least two types of information from the ambient optic array(s). One type of information is the bearing from the observer to the surfaces in the environment. It is of little use to know that there is another car on the road without knowing whether or not you are on a collision course with it. There must also be distance information in the ambient optic array. This does not necessarily mean an estimate of distance in space (e.g. in metres), since estimation in terms of time is also adequate, or even superior, for many purposes (i.e. the time to contact (TTC) with an object in front of the observer (Bootsma & Craig, 2002; Kaiser & Mowafy, 1993; Kerzel, Hecht, & Kim, 1999; D. N. Lee, 1974, 1976; Tresilian, 1991). Bearing information is relatively simple to extract, since light usually travels in straight lines and the layout of the retina means that the direction of an object is specified by the point where a given light ray strikes

the retina, thanks to the spherical shape of the eye. Distance information is more difficult to extract, and is the primary goal for many researchers in the field of artificial vision.

Most studies in machine vision have focused on stereopsis (i.e. sensitivity to differences in two ambient optic arrays) for the provision of information about layout and orientation. Their argument is that since the only way to accurately localize points in a static 'floating cloud' of random expansionless spots is to use stereopsis, we should build stereoscopic systems that can function in all situations. This is misguided. We do not live in a floating cloud of points, but rather we are provided with many sources of information about layout and orientation, and we can use these in addition to or instead of stereopsis. In addition, we can function extremely well with one eye closed. One-eyed people can still walk and drive. A one eyed batsman even played county cricket in England!

No one has yet managed to provide us with a fully functioning artificial stereoscopic system, mainly because no one has managed to adequately solve the correspondence problem. The correspondence problem is that when presented with two spatially disparate arrays, how can we know which point in the first array corresponds to which point in the second array? This sounds simple but is not (Palmer, 1999). Typically, the rudimentary artificial stereoscopic systems we do have are heavy users of processing power (Beauchemin & Barron, 1995), which limits their effectiveness. Give the abilities of monocular humans we would expect that as long as pinpoint accuracy is not required, information in a single array should suffice for perception of real environments.

There are numerous potentially informative properties that would allow perception of the environment from a single array. The simplest to understand is the texture gradient. Many surfaces are made up of repeating elements, and the optical angle subtended by these elements gets smaller as the distance from the element to the eye increases. The way the sizes of the elements change over the ambient array indicates the way in which the distances to the elements are changing. While explaining the underlying ideas of texture gradients is simple, implementing algorithms for perceiving them is not. Functions that detect lines and shapes are often slow and make heavy demands upon computational power. This does not rule out the use of texture gradients in human perception, however. Roska and Werblin (2001) showed that the *eye* sends information about line position to the visual cortex, which implies an incredibly fast algorithm for performing one of the most difficult tasks in the analysis of texture gradients. While the algorithm is fast in the eye, this does not imply that it will be so for a computer. Where the nervous system processes in parallel a computer must process serially. While it is possible to recreate any parallel algorithm with a serial processor, it is not necessarily possible to perform that algorithm as quickly. We are limited by computational power when developing an autonomous vehicle, and with current algorithms for the estimation of texture gradients this power seems to be insufficient, it would seem to be more fruitful to explore other sources of optical information.

## 1.3 Motion Perception

Gibson (1961) noted that certain types of motion generated certain patterns of change for elements of the optic array. For instance, motion straight ahead produced radial expansion, in which every point in the optic array seemed to move directly away from the direction of heading. On the other hand, horizontal motion perpendicular to the direction of fixation causes each point in the optic array to appear to move horizontally in the direction opposite to motion. These patterns are described as *optic flow patterns* and refer to the flow of the optic array across the retina. An optic flow pattern can be generated by superimposing two sequential images of a moving scene and drawing an arrow from the position of each feature in the first image to the position of each feature in the second image. The task of a visual system for motion perception is to extract meaningful information from the optic flow field.

To aid experimental control, motion perception is most often studied using simulated environments. A property of these simulations is that real motion is replaced with stroboscopic motion – a series of images (or frames) shown in rapid succession. Add to this the tendency of a publishing author to represent a visual sequence as a series of frames and there is little wonder that motion perception came to be regarded as a type of stereopsis. Instead of arrays displaced laterally in space (as in stereopsis), a flowing array was regarded as two or more images displaced in space-time (for a review see Aggarwal & Nandhakumar, 1988; Beauchemin & Barron, 1995). The link between a pair of images taken from slightly different locations at almost the same time and a true stereoscopic pair makes it very tempting to treat motion perception in much the same way as we treat stereopsis. The mathematical basis of stereopsis is

well developed and it is relatively simple to adapt the techniques involved for use in motion perception.

Once again, however, we are faced with the correspondence problem – which point in the second image corresponds with the point we are examining in the first image? There is a simple way of avoiding this – we can track each point over the *entire* time between the start and finish of image capture. This is very different from stereopsis, in which we must address only two separate points, but nothing in between. Changes of intensity over time can now be observed continuously (as opposed to discretely), so we can track a point as it moves instead of having to find the point in numerous images. Indeed, it is possible to take the quote from Gibson given at the start of this manuscript and make it apply to motion perception as well as to texture gradients. Now however, we are dealing with an intensity gradient over a solid angle in space-time instead of space alone.

A photographic camera works by adding together the energy of all of the photons that strike each point on the film while the shutter is open (think cones on the retina for the eye). This process is known as integration. The longer the integration period, the more photons will be recorded, so the brighter the image will be. This is known as temporal summation and is useful for low light conditions. However, for long integration periods moving objects will appear smeared, so a faster shutter speed is desirable for clear images in normal light. However, most vision scientists have failed to notice that by using a camera with a long integration period, we can record changes in intensity over both space *and* time. The magnitude of the intensity change gradients in the image can then be used to calculate the motion of those intensity

changes. This technique is known as motion-from-smear (Chen, Nandhakumar, & Martin, 1995, 1996; Tull & Katsaggelos, 1996).

## 1.4 Thesis Structure

The next chapter will examine the methods of determining layout. This includes stereopsis, the texture gradient and traditional mathematical approaches to the study of motion perception

Chapter 3 will describe the current extent of investigation into smeared images and the estimation of motion from these images. Evidence from physiology, psychophysiology and computational image processing will be analysed. Important features of a smearing algorithm will be determined and a simple method for producing smeared images will be detailed.

Where Chapter 3 will deal with the algorithm underlying the visual system, Chapter 4 will be concerned with the implementation and evaluation of that algorithm in a real robotic system. I have been fortunate to obtain a Pioneer robot from Industrial Research Limited (IRL) in Christchurch, and have used it to test the feasibility of implementing motion-from-smear in a robotic vehicle, albeit in an extremely limited environment. It will be demonstrated that it is possible to use motion-from-smear to estimate global tau for an approaching point. Tau was first described by Hoyle (1959) in his science fiction story 'The Black Cloud.' He showed that, for a point travelling parallel to the observer, it is possible to use the angular difference from that point to

the direction of heading and the rate of change of this angle to calculate the time until that point reaches the observer (Lee, 1974, 1976; Tresilian, 1991, 1994; Bootsma and Craig, 2002)

Chapter 5 will examine the applicability of motion-from-smear to composite tau (as opposed to global tau), in order to measure time to contact with a moving object. Motion from smear will also be considered with respect to the optical specification of riser height for descending a flight of stairs.

Chapter 6 draws together the various threads described throughout the thesis and considers future research.

Scanners Note: there is no p12 in the original thesis.

# Chapter 2

# Methods of extracting layout from the ambient optic array

## 2.1 Layout information from static images

The task of a visual system for the guidance of ego-motion is to use the ambient optic array(s) to provide information about layout of the environment. The layout of the environment consists of bearing and distance information regarding the surfaces in the environment.

Stereopsis is a powerful method for the derivation of layout information from two ambient optic arrays. This is the result of having two eyes that occupy related but laterally displaced positions in space. The difference between the locations of points in the two arrays (termed binocular disparity) means that if the distance between the eyes is known then simple trigonometry can allow the calculation of the distance to those points. Julesz (1971) showed that the visual cortex could "fuse" two images made up of random dots and produce a 3-dimensional image (Figure 2.1). This suggests that stereoscopic information is processed at an early stage in the visual cortex. Combined with the potential power of stereopsis, this makes a compelling

argument for studying of stereopsis and for its utility in autonomous vehicles.

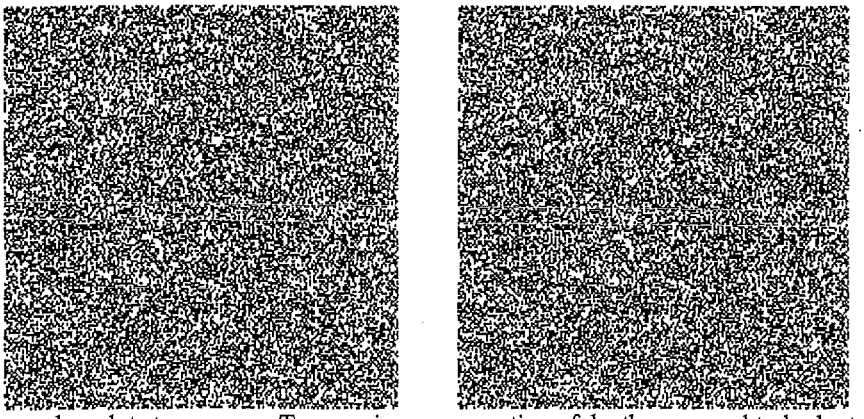


**Figure 2.1** A random dot stereogram. To experience a sensation of depth, you need to look at the left image with your right eye and the right image with your left eye. To do this, first cross your eyes. You should see four squares. Next 'overlap' the middle two squares so that you see only three squares. If you hold this for a few seconds then you should perceive a square section in the middle that lies behind the outer parts of the image.

A problem for stereoscopic models occurs when we examine the methodology more closely. Once we have matched a point in one eye to a point in another, we can easily calculate the distance, but how is such matching to be accomplished? This is no small obstacle and, as noted in Chapter 1, is referred to as the correspondence problem. How are we to decide which feature in one ambient optic array corresponds to which feature in the second ambient optic array? Many attempts have been made to solve the correspondence problem, beginning with the Marr-Poggio algorithm (Marr & Poggio, 1976) and with those models developed since becoming increasingly complex (Jones & Malik, 1992; Marr & Poggio, 1979. For a review see Kumar & Chatterji, 2004). Some modern algorithms may take up to an hour to run on a fast computer, but this is somewhat mitigated by the fact that the parallel nature of the brain may make such algorithms feasible. Those algorithms capable of running in real time often do so by sacrificing the amount of detail that they can process (e.g. Jia et al., 2003).

Texture gradients are an example of Gibson's comment regarding the perception of

gradients over solid angles. When viewing a surface made of regularly shaped, repeating elements, those elements that are farther away will subtend a smaller optical angle than those that are near. Instead of being perceived as smaller, these elements are – correctly – recognised as being farther away (Figure 2.2). No single element, or any part of any element can provide this information. Nor can we view two separate elements and be certain of the gradient. We must view a large number of elements over a reasonably large angle to perceive the gradient. Approaches to vision via sensitivity to texture gradients fall into two main categories. There are statistical methods, which measure factors such as the number of edges per unit of visual angle or differences in intensity levels, and there are structural methods, which measure distortions in the shape of the elements. There is no reason that these methods could not be united into a hybrid function or work in parallel.

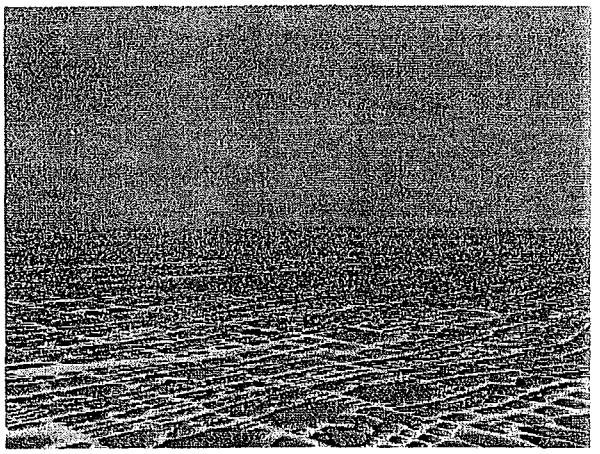


**Figure 2.2** The decrease in the average size of what appear to be approximately uniform cobblestones leads to a sense of increasing distance.

Like stereopsis, however, perception of texture gradients is fraught with difficulty. While it is easy for us to identify and separate different types of texture, it is very difficult to have a computer do the same thing. Likewise, identifying two elements as being part of the same texture pattern can be a difficult exercise in pattern matching.

It appears that computational sensitivity to texture gradient is far from a simple proposition, and not an area that is to be entered lightly. However, there are still other

options available to us, and it is one of these – motion perception – to which the rest of this thesis will be dedicated.

## 2.2 Conventional methods for determining layout information from motion

Motion perception often fails to receive the attention that it deserves. Part of this is due to a mistake in adapting the inverse problem to allow for the inclusion of time. In the classical formulation, the environment around us is three-dimensional, but each of our eyes only produces a two-dimensional image. It is impossible to determine the orientation, size and distance to a surface in the visual field from a two-dimensional image (Figure 2.3) and even binocular disparity cannot compensate completely (Palmer, 1999). Further to this, people with only one eye can perform most tasks quite adequately, so it is necessary to find another source of information. When considering motion perception, the passage of time can be considered a fourth dimension; observation over time means that enough additional information is available that a solution of the inverse problem becomes possible. However, as Palmer (1999) points out, the passage of time allows another degree of freedom to the surfaces in the environment, and as such a solution is still not possible. The fault in Palmer's reasoning is that, as Gibson (1962) has pointed out, we are active rather than passive observers. As we move through the environment we are aware of our own motion, at least approximately, and have some control over it. The motion of our eyes through space is rarely in a straight line – or even at a constant velocity – so it would take a very special type of circumstance for the solution of the inverse problem provided by motion perception to be incorrect – the environment would have to be

moving in such a way as to compensate for our motion. This would be much like hiding from a person behind a pillar. If you are careful, you can ensure that the pillar is always between you and the person, which would suggest, at least visually, that you are not behind the pillar. In the same way, there are an infinite number of ways that the environment could produce the changes in the ambient optic array that an active observer perceives, but only a tiny fraction of those possible configurations would not involve an actively deceitful environment.

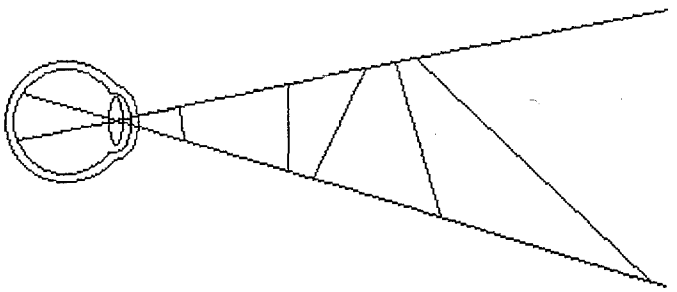


**Figure 2.3** An illustration of the inverse problem for a static situation. The retinal projection does not allow for specification of distance or orientation.

Stappers (1992) used a tele-operation system to establish that an active observer was more effective at performing a spatial task than was a passive observer. The active observer viewed a television screen. The screen showed a room with spatial information necessary for completing the task. A camera was mounted in the room, and a sensor behind the head of the active observer caused the camera to move in accordance. This meant that the active observer change his view of the room using head movement over which he had full control. The passive observer, on the other hand, did not have control over his environment. He sat in another room, but watched the same images as the active observer. This meant that he received exactly the same visual information as the active observer, but was not able to control the information. The active observer was far more accurate in task performance than the passive observer.

Although motion perception has not received its fair share of attention, it would be wrong to say that it has been ignored. There is a large literature dealing with optic flow (see Aggarwal & Nandhakumar, 1988; Beauchemin & Barron, 1995 for a review), but most of this work was not readily understood by psychologists because of its mathematical nature.

The analysis of optic flow is divided into two types – correlational and differential. Correlational methods are essentially an adaptation of stereoscopic analysis to the study of changes in both space and time. Two separate images produced while in motion are treated as a stereoscopic pair, but displaced in space-time rather than space. As well as resurrecting the correspondence problem, this type of methodology also requires extreme precision in the measurement of the passage of time and of the movement of the observer by external sources. Experiments such as Lee and Lishman (1975) in which movements of the walls induce automatic compensatory movements by the subject would seem to indicate that we do not have such precise knowledge of our position supplied to us by non-optical sources such as our inner ear.

Correlational methods are also subject to temporal aliasing, which occurs when the sampling rate is too low. As can be seen in Figure 2.4, a low sampling rate can mean that two distinct situations can produce indistinguishable sequences of frames.

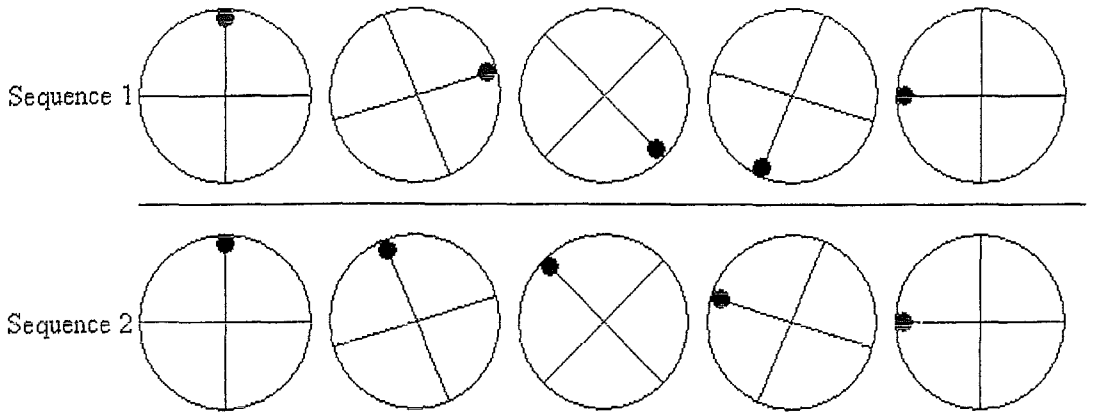



**Figure 2.4** The wagon wheel effect. Temporal aliasing occurs when the sampling rate is too low. The sequences are consecutive frames of a spoked wagon wheel rotating. The wheel in sequence one is moving clockwise at one rotation per 8 frames, while the wheel in sequence two is moving counter-clockwise at one rotation per 24 frames. However, the two sequences would be indistinguishable without the dot on the top spoke.

By treating the images as samples of a differentiable function in time and space differential methods avoid the correspondence problem, essentially treating the data as a discrete approximation to a continuous function. Differential methods are generally more effective than correlational methods, but the mathematics required for these methods is extremely complex, which makes development difficult. These systems are also subject to temporal aliasing, especially when dealing with a small number of images.

Both correlational and differential methods are place extremely heavy demands upon processing power and, like many systems developed in psychology (e.g. Marr, 1982), they normally assume a stationary environment. While most of the environment is stationary, those parts that are in motion are often those that are most related to the survival of the animal: predators and prey. Although this constraint does simplify the problem dramatically, it does so by eliminating many of the elements of the environment that may be of greatest importance.

It should also be noted that there are several systems that allow successful guidance of

aspects of ego-motion in real environments, but most of these systems rely on some form of object recognition to eliminate those parts of the optic array that are not of interest (e.g. Kehtarnavaz, Griswold, & Lee, 1991) or by implementing neural networks (e.g. Jochem & Pomerleau, 1996);. While neural networks are a viable solution, they often function in ways that we do not understand, and as such are less likely to contribute to our understanding of human vision. Those systems that do function in real environments are not yet perfected, and it is hoped that the method of motion perception developed in this thesis will eventually aid accurate real-time processing of the entire environment.

# Chapter 3

# Motion-from-Smear

## 3.1 THE DEVELOPMENT OF MOTION-FROM-SMEAR

Smear is the blurring of an image of an object due to motion. It is different from conventional blur in that its point-spread function is elongated in the direction of travel. Thus for an image of a horizontally translating square we would expect to see a rectangle rather than a square (Figure 3.1). Smear has traditionally been considered a nuisance, present only because of the need for a camera to have a finite integration period. Image processing programs have been designed to eliminate smear, but it was not until Chen et al. (1994; 1996) that the possibility of using smear as a cue was considered.

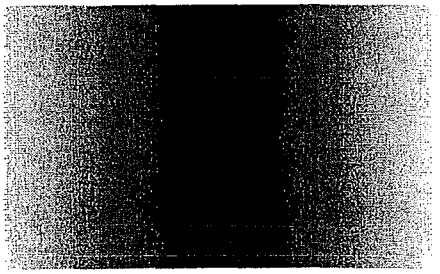


**Figure 3.1** A square translating left to right. Instead of the original square we see a rectangle. The blurring at the left and right edges is known as smear.

There are numerous reasons that may have contributed to the neglect of this potentially informative cue. As mentioned in the introduction, movies are a discrete method of replicating a continuous process. The success of movies in fooling the

human observer makes it very tempting to conclude that vision works by integrating a series of static images, as in a movie film.

The parallel streams hypothesised by Milner and Goodale (1993; 1995) may also be largely to blame for the lack of interest in smear. Mounting physiological and psychological evidence (Boussaoud, Ungerleider, & Desimone, 1990; Mishkin, 1972; Mishkin, Ungerlieder, & Macko, 1983; Pohl, 1973; Ungerleider & Mishkin, 1982) indicated that visual information followed two distinct pathways in the brain (Figure 3.2). It had originally been suggested that the dorsal stream provided 'where' information while the ventral stream provided 'what' information. However, Milner and Goodale (1993, 1995) suggested that the dorsal stream was in fact responsible for 'how' information. Put simply, the ventral stream decided that a task needed to be performed and the dorsal stream provided the specific information necessary for performance of that task. Milner and Goodale (1995) reported an experiment they had conducted in which subjects reported falling for an optical illusion yet when asked to interact with the illusion were able to perform as if no illusion were present. However, the ability to interact with the illusion was transitory. If subjects viewed the object while making a decision about the relative size of two illusory objects, then closed their eyes while interacting with the objects, they behaved as if they had fallen for the illusion. This was taken to mean that the precise knowledge of spatial relations fades quickly.

The experimental evidence was supported by examples of patients with brain damage and the resulting effects upon their performance. The combination of behavioural and psychophysiological data was taken as evidence that the dorsal stream provides

detailed but transitory information about the environment, while the ventral stream provides information that is less detailed but more easily remembered. It is thought that only information from the ventral stream is available to consciousness, which means that information and methods used by the dorsal stream might not be open to phenomenological investigation.

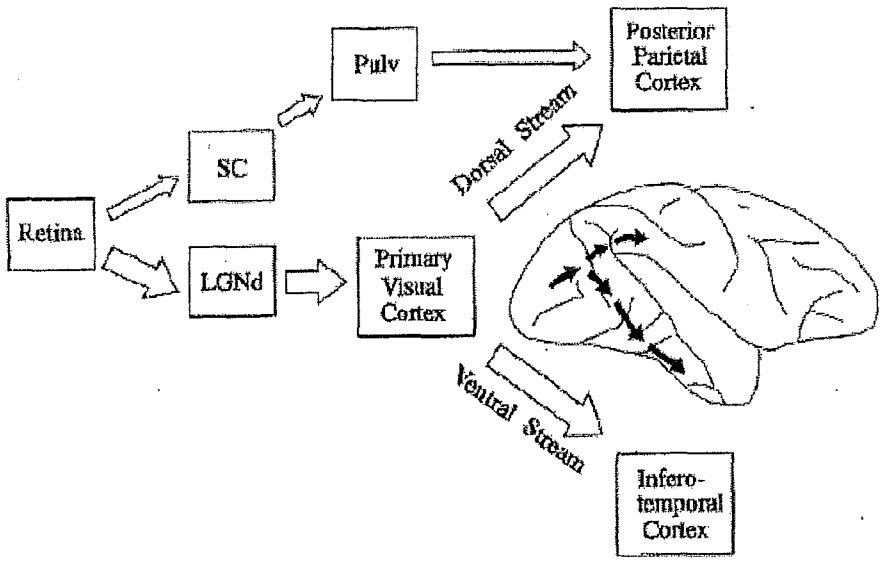



**Figure 3.2** The parallel streams hypothesised by Milner and Goodale. The dorsal stream is hypothesised to be used for detailed but transient 'how' processing while the ventral stream is used for simpler but resilient 'what' information. (Milner & Goodale, 1995; reprinted with permission.)

Our ventral stream is hypothesised to facilitate tasks such as object recognition and navigation. Recognising smeared objects can be quite difficult, so it would make sense for the ventral stream to reduce smear as much as possible. Also, due to our limited attention abilities, we are rarely aware of anything outside the fovea, especially when we perform a tracking task or are busy walking. By moving our eyes we keep the image on the fovea relatively still and as such are even less aware of the smear in our environment.

Despite the findings of Milner and Goodale (1995) the lack of attention for smear as a cue to motion remains somewhat paradoxical. Cartoonists have used speed lines

behind an object to induce motion for years (Figure 3.3). It is also known by those in the data compression industry that a lower frame rate can be used if one out of every three images is slightly smeared. An emerging technology is frameless rendering, which increases motion smear and consequently increases the quality of moving images.


**Figure 3.3** An example of speed lines employed by cartoonists. It is clear that the purse-snatcher is moving quickly from left to right. "Dilbert" cartoon reprinted with permission of Scott Adams.

Frameless rendering works by refreshing the pixels on the screen in a near-random order. Consider the following imaginary TV screen as an example. The entire screen on a normal TV is refreshed once every 20ms. The screen is divided into ten sets of pixels, with the members of each set evenly distributed around the screen. These sets are refreshed one after the other. This means that out of any 10x10 square of pixels, 10 have just been refreshed, while 10 are about 18ms old. The rest are distributed between these extremes. The technique is known as frameless rendering because there is no point at which an image on the screen can be considered a distinct frame from the previous image. Refreshing a distributed set of pixels takes a little longer than refreshing a compact set, but this is more than made up for by the motion smear that is produced.

3.1 The development of motion from smear

### 3.1.1 Smear and the common toad (*Bufo bufo*)

Nocturnal animals are forced to use smeared images to catch their prey at low light levels. Aho, Donner, Helenius, Larsen, and Reuter (1993) showed that the common toad (*Bufo bufo*) was able to accurately snap at wood lice under lighting conditions so low that neither the toad nor the lice would be visible to the human eye. The toad achieved this by using temporal summation, essentially the same as lengthening the exposure time on a camera. While temporal summation has the advantage of amplifying weak sources of light, it has the side effect of producing smeared images when viewing moving objects. Aho et al. (1993) did not consider smear when discussing their results, but it has the ability to improve upon their explanation of toad performance.

Aho et al. (1993) used single cell recording under four lighting conditions and found that the amount of temporal summation used by the toad increased from 0.25s in moonlight to 2.9s at the lowest lighting condition. They tested the snapping accuracy of toads under the four lighting conditions by presenting them with dummies resembling woodlice moving across the toad's visual field. It was noted that the toad's tongue hit near to the heads of the dummy lice under the highest lighting level, but that the snap position moved further and further back from the heads of the targets under lower illumination, eventually missing entirely (Figure 3.4). They postulated that the toad aimed what it thought was the head of the target. It was further assumed that, due to the response latency, the toad would aim at the point where the head of the target would have been when summation began. Aho et al. (1993) also pointed out that after acquisition there were additional delays brought about by the time taken to

act upon this information. They added 0.2s to the response latency when predicting the position that a toad would snap to. Thus if a toad had a response latency of $K$ and the target was moving at a speed of 13mm/s then the snap position would be $\Delta x$mm behind the head of the target.

$$\Delta x = 13 \times (K + 0.2)$$ Equation 3.1

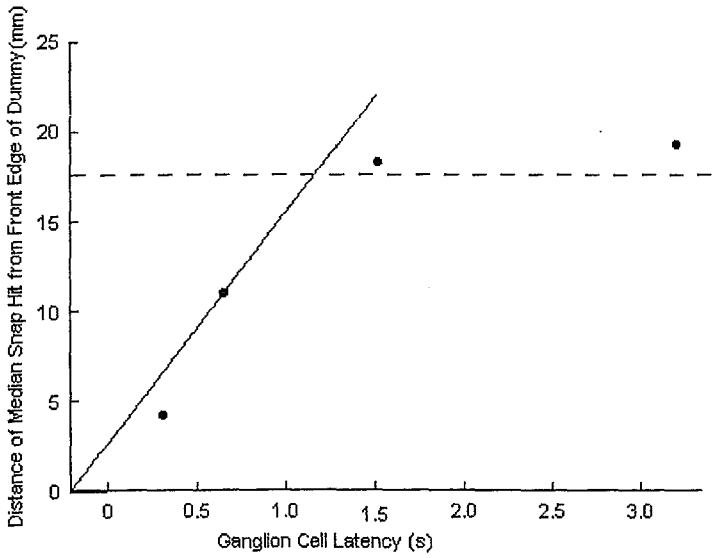



**Figure 3.4** Toad accuracy decreases at increasing latencies (Aho et al., 1993). Filled circles denote the longitudinal distance of toad snaps from the head of a 15mm long wooden dummy moving at 13mm/s. The solid line is the distance predicted by equation 3.1. The dashed line is an approximate maximum distance predicted by Equation 3.2.

While this seemed to apply to the three higher lighting conditions, it failed to account for the performance at the lowest lighting level. Performance at a level just above threshold was as good as performance at the next lowest level. Given that the response latency had doubled, this was considered somewhat surprising. Aho et al.'s (1993) explanation was that the toad must use some form of representation of the lice's motion at extremely low light levels. This is a *post hoc* solution, since if the toad uses a representation at all then it should use a representation at all light levels. However, if we examine the images that would be produced using the four levels of temporal summation employed by the toad, a solution presents itself (Figure 3.5).

3.1 The development of motion from smear

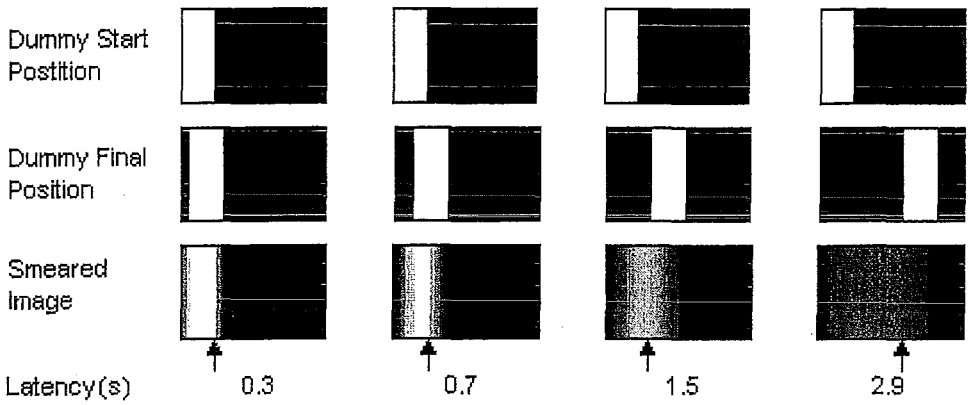



**Figure 3.5** Simulated woodlouse motion. A band 15 pixels wide moves to the right with a constant velocity of 13 pixels per second with different levels of response latency. Arrows indicate the brightest point in the direction of motion, i.e. the bright point that lies furthest to the right.

If we hypothesise that the toad aims at the brightest part of the image in the direction of travel, it would aim at the points indicated by the arrows in Figure 3.5. As long as the response latency is short enough that the target cannot move more than one body length, the brightest part of the image in the direction of travel will be the head of the target at the start of motion. However, if the target traverses more than one body length, there will be a large, dull smear area. Aiming at the brightest point in the direction of travel in this case will produce a snap at the tail of the target, no matter how long the response latency of the ganglial cells. Thus we would expect that performance in the snapping task would have a maximum error of one body length, which in Aho et al. was 15mm. If we add the delay of 0.2s to the snap, we have a maximum value for $\Delta x$ given by

$$\Delta x_{max} = 15 + 13 \times 0.2 \qquad \text{Equation 3.2}$$

As can be seen from the dashed line in Figure 3.4, this is an extremely accurate

explanation that does not require a representation on the part of the toad. Although this does not provide evidence for detecting motion-from-smear, it is support for the availability of smeared images to the toad brain.

### 3.1.2 Motion estimation from smeared images

It was not until recently that the potential utility of motion from smear was seriously considered. It is indicative of the potential utility of motion-from-smear that this research is now occurring in several different fields, but as is so often typical of new developments, those performing this research seem to be largely unaware of each other's work. Ballard (1986) mentioned using motion-from-smear briefly in a paper on neural networks and cortical structure, but this was not explored further at the time. Martin and Marshall (Martin & Marshall, 1993) produced a neural method for unsmearing a moving image, although they never took the logical step of using the unsmearing process to indicate motion.

Chen et al. (1994, 1996) showed that by using two smeared images it would be possible to determine motion. Since they were using a Fourier transform to analyse the two images, they could not simply use a slower shutter speed for both images because the pair of images would not have a stable inverse system. As such, one of the images has to use a non-uniform filter (a filter that allows differing amounts of light to pass over the time taken for a single image to be captured). If both images are captured using the same filter then there will be an identifiability problem for the Fourier transform, so the second image is acquired conventionally. They compared

motion-from-smear with correlational methods and showed that where correlational methods were subject to temporal aliasing, motion from smear was not. Chen et al. (1996) also compared a differential method of motion estimation with motion-from-smear and found that the differential method was very much inferior, producing both spurious and inaccurate motion estimates. Additionally, since conventional methods of motion estimation and motion-from smear use different information in an image, the two methodologies are complementary rather than competitive.

Shortly after Chen et al. began publishing their work, Tull and Katsaggelos (1996) published a paper that approached motion from smear with a methodology similar to the differential approaches to motion estimation from sequences of images. They used blur as information about the image motion then used this information to assist in one of the more traditional techniques for motion estimation. While this approach is effective, it still has many of the limitations imposed by differential and correlational approaches to motion estimation.

### 3.1.3 Motion from smear in the visual cortex

More recently, neurobiologists have found that cells in the visual cortex appear to be sensitive to a form a blur (Burr, 2000; Burr & Rose, 2002; Geisler, 1999; Geisler, Albrecht, Crane, & Stern, 2001). It has long been known that there are cells in V1 of the primate brain selective for spatial orientation (Hubel & Wiesel, 1968), in that they respond strongly to stationary contours at their preferred orientation but not at all to the same contour rotated by 90 degrees. As well as contour detection cells there are cells sensitive to orientation *and* motion. These cells are sensitive to a contour of one specific orientation moving in one particular direction (Sekuler & Ganz, 1963; Sharpe & Tolhurst, 1973; Tolhurst, 1973). For example, such a cortical cell could respond to a vertical line, and to the same line moving to the right, but not to a horizontal line or to a vertical line moving leftward. Such a cell is direction-selective (DS), whereas a cell that is sensitive to orientation but not motion is non-direction-selective (NDS).

Geisler (1999) noted that since humans use temporal summation over a period of about 100ms, a moving shape should appear slightly smeared and this smear could make the moving shape appear to be a stationary line (Figure 3.6a). If this were the case, then a cell sensitive to contour orientation but not motion direction should be strongly activated. Geisler supported this hypothesis by showing that detection of a dim, moving spot of light was worse when moving parallel to a mask of dynamic lines than when moving in the perpendicular direction as long and the spot was moving at least 1 spot width per 100ms. In a situation with no mask, the spot should smear at high speeds, meaning that NDS cells oriented to the direction of motion should be activated, aiding detection of the spot (although this will be countered by a lessening

in overall spot intensity because the same amount of illumination will be spread over a greater area. This should mean a gradual worsening of performance with increasing dot speed). If a mask has no effect on these NDS cells (Figure 3.6b) then there should be less effect on spot detection than if the mask activates these cells regardless of whether a spot is present or not (Figure 3.6c). This is exactly what Geisler (1999) found. That is, he found that once spot motion exceeded about 12 spot widths per second, the threshold for spot detection increased dramatically for the parallel mask, but not for the perpendicular mask. This implies that for higher spot velocities the visual cortex is relying on the activation of NDS cells to aid in detection of the dim moving spots. This strongly suggests that smear information is available to the visual cortex. Whether this information is *used* is another question.

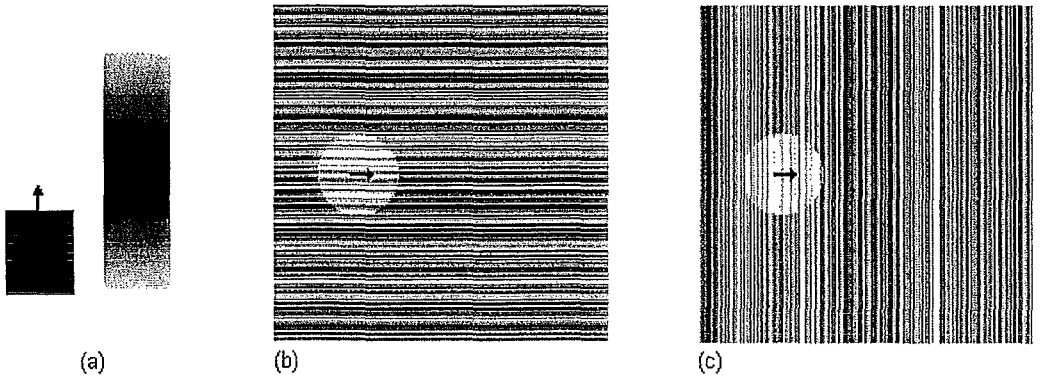



**Figure 3.6 (a)** A moving square becomes a virtual contour under temporal summation. A dim spot moving to the right is harder to see with a dynamic parallel mask **(b)** than with a dynamic perpendicular mask **(c)** when spot speed exceeds 12 spot widths per second (Geisler, 1999).

These virtual contours produced by smear would of course have to be distinguished from real contours produced by edges in the world. To solve this problem, Geisler suggested combining an NDS cell with a DS cell oriented perpendicular to the direction of motion using a multiplicative function. If a feature moves in the selected direction of the DS cell, the pair will respond strongly, whereas a stationary feature –

or a feature moving in the opposite direction – would produce no response. As well as experimental evidence, Geisler (1999) used a series of simulations to support this hypothesis, successfully applying his motion detector to a natural scene. He also noted that the information derived from smear would be complementary to other forms of information, rather than competitive, since smear becomes more informative at higher velocities, while the information from DS cells becomes less reliable.

Burr and Rose (2002) found further psychophysical support for the use of 'speed lines' in the visual cortex. Since motion-from-smear depends on separating real contours from virtual contours, it is to be expected that in situations where this separation is difficult the perception of motion will be difficult. Burr and Rose (2002) presented sequential dot pairs that produced an appearance of motion. Short (length), briefly presented, real contours oriented up to 20° from the direction of motion served to inhibit discrimination of the direction of the apparently moving dots.

Geisler et al. (2001) showed that many cells in the visual cortex of the cat and of the monkey, previously believed to be only selective for motion perpendicular to contour orientation, were in fact also selective for motion parallel to preferred orientation as well. They had predicted that these cells would respond to parallel motion, since the moving spots would form virtual contours. However, they did not expect that it would be possible for these cells to differentiate between the directions of movement since these virtual contours should provide ambiguous information regarding direction – a line drawn left to right should be equal to a line drawn right to left. To explain this, Geisler et al. (2001) took existing models for detection of perpendicular motion and adapted them to parallel motion. Most of these models rely on a neural

equivalent of the correlational methods applied by engineers, and as such are somewhat antithetical to the use of motion-from-smear. While it cannot be assumed that the visual cortex will use a particular method of estimating motion for the sake of theoretical purity, it is possible for smear to provide additional information that will eliminate the need for correlational techniques from the detection of parallel motion. The remainder of this chapter will be dedicated to producing smear that provides direction information in a computer.

## 3.2 A MODEL FOR SMEAR PERCEPTION

### 3.2.1 Producing a smeared image

One major difference between the eye and a camera is that the cones in the retina do not produce individual frames. As mentioned in Chapter 2 (Figure 2.5), temporal aliasing occurs when the sampling rate is too low, and resulting in under specification of motion. The cells in the retinal ganglion constantly receive information from cells in their receptive fields so the eye is a perfect environment for supersampling to occur. Though we are still limited by the constraints of a camera we can, however, replace the conventional camera with one that implements frameless imaging. Recall that frameless rendering updates distributed sets of pixels with current information, rather than using whole frames. Frameless imaging would work by taking the information from ten distributed groups of receptors in a digital camera at different times. This would provide a constantly updated image to the processor, rather than images that are already out of date when they reach the processor. While this is a form of spatial summation, and as such means a loss of resolution, this resolution

would only be lost in the motion processing domain. The images could still be combined conventionally to facilitate tasks such as object recognition.

Since we are using a two-dimensional image to represent a three-dimensional array (x, y and time), there must be some method for discerning the early parts of the image from the later parts. Just as in many graphing utilities colour is used to indicate the z-coordinate of a value, so in this technique time is modelled by the contribution that the parts of the image are allowed to make at different times. Thus, like Chen et al. (1996) this algorithm will employ a non-uniform integration filter.

For instance, a square moving from left to right (Figure 3.7a) is indistinguishable from one moving right to left when captured with a uniform integration filter. However, a filter that allows more light through as time passes would make the two directions of motion easily distinguishable (Figure 3.7b).

As in Chen et al. (1996) this algorithm will also be using an end-state image. This image allows elimination of those parts of the image that are not moving. By subtracting the final state image from the smeared image, any stationary objects will have zero intensity. Likewise, any (approximately) homogeneous surface will also have zero intensity. The only parts of the image with non-zero intensities will be those places where a moving intensity boundary was present. By examining a non-zero region of the image we can calculate the velocity of that part of the image.

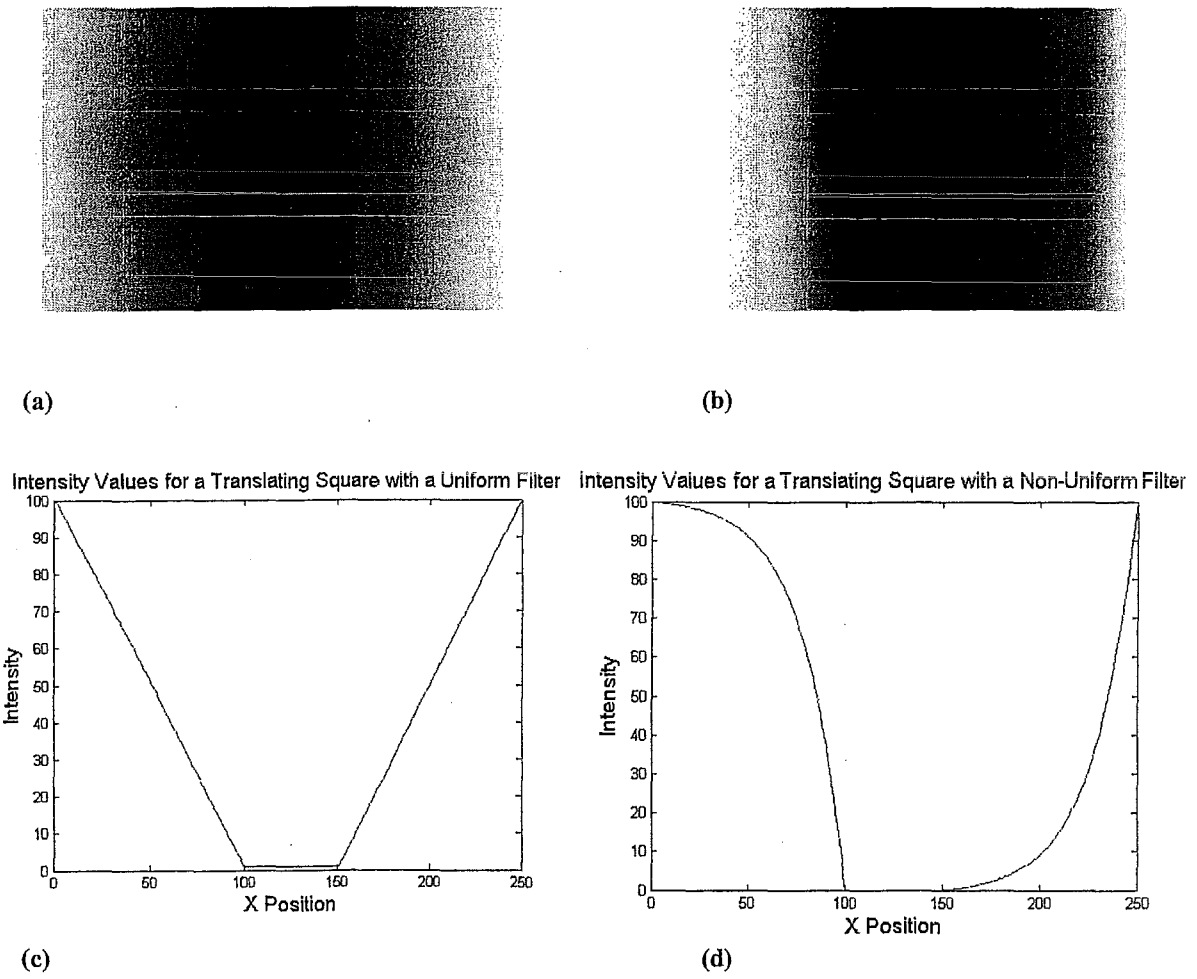



**Figure 3.7** **(a)** Uniform and **(b)** non-uniform filters applied to a square translating to the right. In the intensity graphs **(c, d)** it can be seen how the intensity gradients are symmetrical for a uniform filter but not for a non-uniform filter.

Since it is necessary to know what the end state is anyway, it could be argued that we could apply a uniform integration filter and simply use the final image to calculate direction. This would, however, require identification of which objects in the final image belong to which parts of the smeared image, which returns us to the correspondence problem.

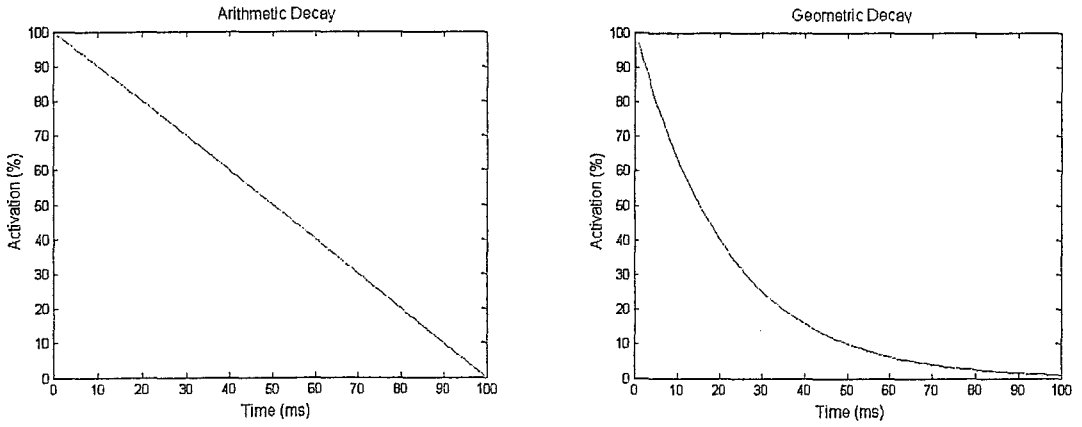
What sort of non-uniform filter could be applied? Before we think about this we need

to reconsider the common toad. Aho et al. (1993) seemed surprised that the toad was able to snap at the point that the wood louse would be occupying, not the position it had occupied at the start of image capture. This result is not at all unexpected given the use of a non-uniform filter and a continuously sampling eye. Often the eye is considered to be analogous to a camera, and one of the primary properties of cameras is the simultaneous integration of all points on the image. A toad that worked in this way would see a still image every 3 seconds and have to derive the motion of the louse from that, meaning at least a 6 second wait for a chance to snap, but, a *continuously sampling* eye would be continuously updated, thus providing the toad with up-to-date and reasonably accurate information about its environment. Yet for the toad to establish the direction of louse motion from only one integration period it would be necessary to either solve some form of correspondence problem or to use a non-uniform filter.

The potential utility of a non-uniform filter can be derived from introspection as well as mathematics. Imagine someone standing at the far end of a darkened room with a small torch. They face to one side and begin to rotate their entire arm at the shoulder, like a bowler in cricket. The torch would seem to be a point of light moving in a circle. There would also, if the person were moving their arm fast enough, be an apparent blur of light trailing behind the torch. If the person were able to rotate their arm ten times per second the light would appear to form a complete circle. For a mere mortal, rotating their arm at less than ten hertz, we would see a bright point followed by a trail that darkens as the distance from the point source increases. As described previously, this type of phenomenon is a result of temporal summation. The eye integrates over a period of about 100ms so our image of the torch is smeared over

those positions that the torch occupied over that 100ms period. Additionally, the parts of the motion streak that are closer to the torch are brighter than the parts of the streak furthest away. Events that occurred more recently appear brighter, thus we would expect the filter to apply less weight to information from early in the integration period, which is the definition of a non-uniform filter.

How is this weighting to be determined? There are two ways immediately obvious possibilities. The first is arithmetical decay – a particular event's contribution to the image decays at a fixed rate (e.g. 1% of the original intensity per millisecond) until it reaches zero after 100ms (Figure 3.8a.) The second possibility is geometric decay – the ganglion's contribution to the image decays by a fixed proportion over a fixed period of time, say 50% every 15ms (Figure 3.8b). This geometric decay has the advantage that the only thing the cell needs to "know" is the current level of activation. There is no need to "remember" an initial value of which 1% is lost per period of time, or to keep track of how long it has been since the event occurred.




**(a)** **(b)**

**Figure 3.8** **(a)** Activation of a neuron with arithmetical decay and **(b)** Activation of a neuron with geometric decay. Either would be able to produce an image similar to Figure 3.7 (b) but a geometric decay curve is simpler to apply in both a neuron and a computer.

When we examine computer analogues of motion from smear, the geometric decay option becomes even more useful. Although it should be relatively easy to set up frameless imaging through hardware, especially with a C-MOS camera, the technology is not available as yet. Therefore, for the purposes of this thesis it was necessary to produce smeared images using a series of still frames.

This was achieved by combining five frames taken at regular intervals using a non-uniform filter. The use of separate frames rather than true temporal summation means that a form of the correspondence problem would be problematic in complex environments, but it is useful for demonstration purposes in simple situations. In order to combine five frames arithmetically it would be necessary to keep the five frames in memory and add a portion of each to the whole. This is heavy on computational power, requiring a large amount of RAM as well as a series of quite time-consuming calculations. On the other hand, a geometric filter can operate by storing only two images (a composite image and the current image) and by using only a few simple calculations.

3.2 A model for smear perception

These are the steps for producing a smeared image to emulate a geometric decay filter (Figure 3.9). It will hence be referred to as the *discrete smear algorithm.*

1. An initial 'composite' image is produced. This can be the first image from the camera, or a random array. (Figure 3.9a)
2. Image 1 is acquired. (Figure 3.9b)
3. Image 1 is added to the composite image.
4. All intensity values in the new composite image are halved. (Figure 3.9c)
5. Image 2 is acquired. (Figure 3.9d)
6. Image 2 is added to the composite image.
7. All intensity values in the new composite image are halved. (Figure 3.9e)
8. Image 3 is acquired. (Figure 3.9f)
9. Image 2 is added to the composite image.
10. All intensity values in the new composite image are halved. (Figure 3.9g)

This process can be continued indefinitely. Due to the geometric decay of earlier images, the effect of the initial image will be lost amongst noise after about 5 images (its contribution will be $(½)^5$ or 1/32 of its original intensity). Using this algorithm, after the first image has faded to noise there will always be an up-to-date version of the smeared image available. To find motion we need only subtract the current image from the current version of the composite. Areas containing moving boundaries have non-zero values in the resulting image (Figure 3.10).

Image from camera Composite image

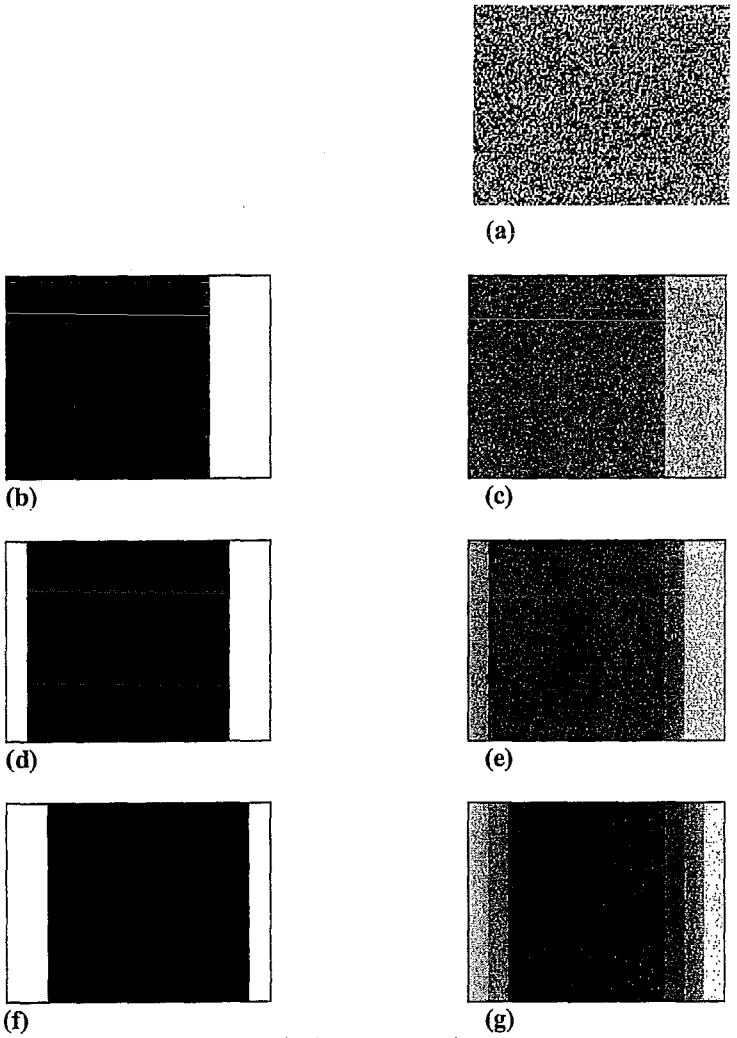



**Figure 3.9** Steps for producing a smeared image from a sequence of images by using a non-uniform filter. Starting with a random composite image **(a)**, the first image from the camera is acquired **(b)** and the values of image 1 and the composite are averaged to provide a new composite **(c)**. The second image from the camera **(d)** is averaged with this composite **(e)**, which is averaged with the third camera image **(f)** to provide a final composite image **(g)**. Note that the image in (g) is beginning to approximate Figure 3.7(b).

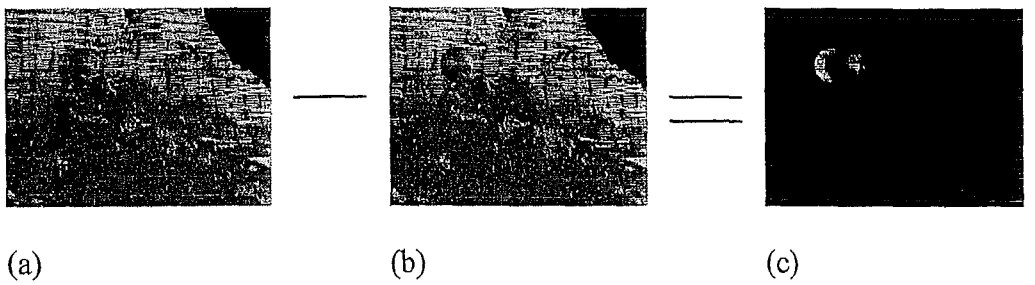
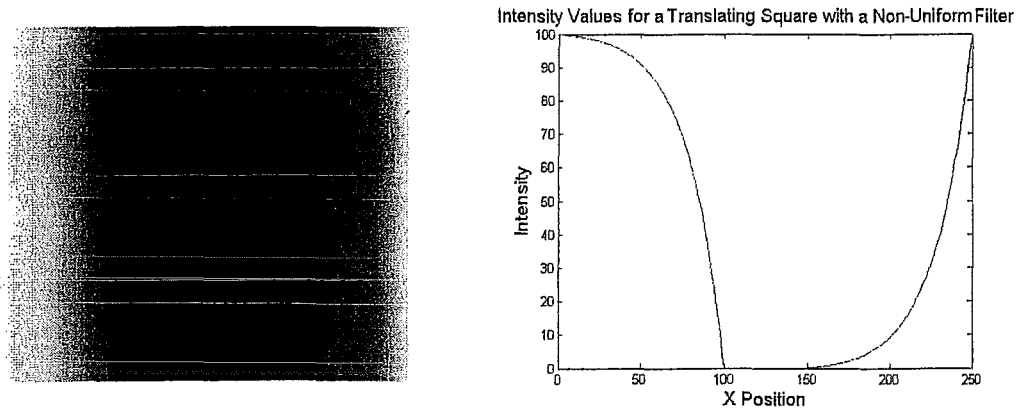



**Figure 3.10** The composite image **(a)** minus the final image **(b)** gives nonzero values in those areas with motion. Note that the smeared sphere is barely visible in (a) but becomes clear in (c)

The discrete smear algorithm also has the advantage of being exceptionally fast. Although multiplication and division are usually computationally expensive, division by two is simple, due to the architecture of the computer. Since a computer works in base two, dividing by two works in the same way as dividing by ten does in the decimal system. Any base 10 number divided by ten is the same number but with the decimal place shifted one place to the left. In the same way, any base-two number divided by two need only drop its final digit. This process is known as a binary shift. In this way, the discrete smear algorithm can produce motion-smeared images easily, reliably, and continuously.

### 3.2.2 Extracting motion information from a smeared image

Having produced images that contain the information necessary for the calculation of motion information, it be comes necessary to extract that information. The nature of the geometric non-uniform filter described above means that the best measure of the motion of a feature is the half-life of the smear in the direction of motion (Figure 3.11).




(a) (b)

**Figure 3.11** **(a)** A translating square recorded with a non-uniform filter. **(b)** A horizontal cross section of (a).

Of course, this relies upon the x-axis being parallel to the axis of motion. This is not a trivial assumption. If the x-axis is not parallel to the axis of motion then the half-life will be overestimated. As such, motion from smear is still subject to the aperture problem. When a straight translating line is viewed through a small aperture it is impossible to be sure of the direction of motion (Figure 3.12). Only the motion component perpendicular to the orientation of the line can be determined.

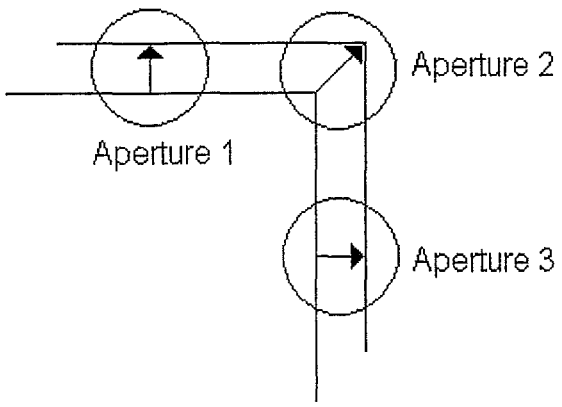



**Figure 3.12** The aperture problem. When a moving line is viewed through an aperture (as in apertures 1 and 3) only motion perpendicular to the orientation of the line can be determined. A distinctive feature, such as a corner (aperture 2) allows estimation of total motion.

If our examination of smear is restricted only to virtual edges smeared parallel to the orientation of the edge then the problem is eliminated. Instead of a smeared shape, we can instead consider two smeared edges with a coherently moving surface between

them. Note that the two smeared edges need not correspond to the edges of the surface (Figure 3.13).

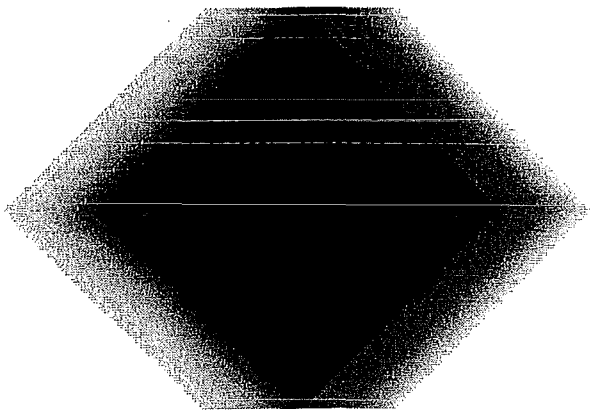


**Figure 3.13** A horizontally translating diamond. The virtual edges at top and bottom indicate the direction of motion. The edges are "virtual" because no such edge exists in the real object.

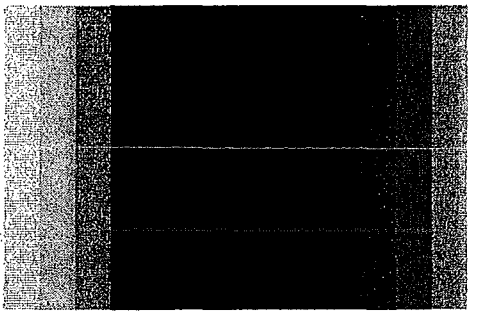
This means that a system employed to measure smeared edges parallel to the direction of motion (e.g. Geisler, 1999; Geisler et al. 2001; Burr and Rose, 2002) would be ideally suited for the extraction of motion information from smeared images. It should again be noted that although edge detection is not a trivial matter, the visual cortex is extremely adept at this task. As such, this thesis will only consider situations in which the axis of motion is known.

Another assumption is that only one moving edge passes through the given space at a time. This is because two smeared edges will interfere with each other as if they were waveforms. This is another non-trivial problem, but it is to be hoped that a form of Fourier analysis can solve this. Since Chen et al. (1994, 1996) were able to solve a similar problem in this way this would seem to be a fruitful avenue for research.

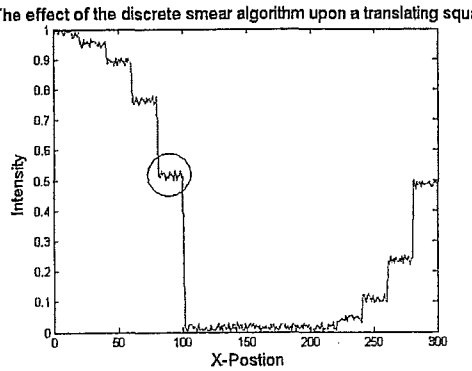
Extraction of the half-life of smear produced by the discrete smear algorithm is slightly different to images produced using true temporal summation. Since such

images are discrete approximations to a continuous curve (Figure 3.14a, b), our real interest is in the length of the steps in intensity (Figure 3.14c). Exactly how this information should be extracted is beyond the scope of this discussion. A suitable algorithm would be resistant to noise and also able to be combined with an algorithm able to separate two superimposed moving edges.

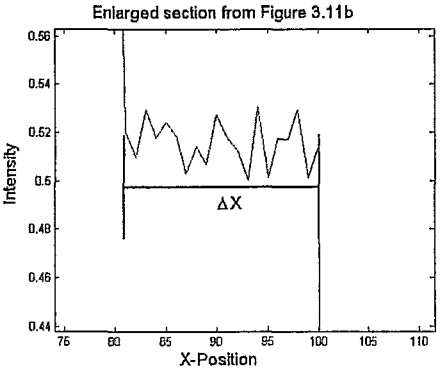
**Figure 3.14 (a)** A translating rectangle produced by the discrete smear algorithm. **(b)** The intensity of a horizontal section of (a). **(c)** An enlargement of a single step.



**(a)**




**(b)**




**(c)**

Although it does not produce truly smeared images, the discrete smear algorithm provides an approximation that is sufficient for demonstrating several potential applications of motion-from-smear using readily available hardware.

# Chapter 4

# Implementing and evaluating motion-from-smear in a real robotic system

## 4.1 Time to contact: local and global tau

Most studies in artificial vision focus on measurement of layout in terms of distance. Combined with velocity information this would allow avoidance of obstacles and interaction with goals. This methodology relies on precise knowledge of the velocity of the observer and of the parameters used in image acquisition. It also assumes a stationary environment.

As already noted, it is most likely that human agents do not have precise knowledge of their velocity without the use of optical information (D.N. Lee & Lishman, 1975). While it is true that most elements of the environment are stationary, those elements that are of most interest are usually moving. What is needed is a measure that incorporates both velocity and distance and is easily derived from visual information. Lee (1974; 1976) popularised just such a measure – tau ($\tau$), which allowed

calculation of the time to contact (TTC) with an object – which was described in a short story by Hoyle (1959). He noted that the angle from the focus of expansion to an object ($\theta$) and the rate of optical expansion for that object ($\dot{\theta}$) could be used to approximate the amount of time until that object reaches the observer (Figure 4.1).

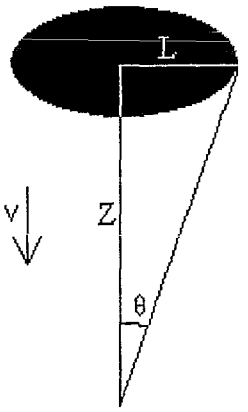



**Figure 4.1** A black cloud approaches earth. *V*, *Z*, and *L* are unknown, but TTC (*Z*/*V*)can be determined by and its rate of change.

TTC is a sensible way to scale the distance to objects in the environment. Research shows that we may use the height of our eyes above the ground to scale distance information (e.g. Warren, 1984) and TTC scales the distance in terms of time instead of space. The distance between an agent and an object is really relevant to the speed at which the agent is moving - a concrete block ten metres distant is of little concern to a snail, but of immediate concern to a speeding vehicle. TTC neatly encapsulates this relationship.

Since Lee's (1974; 1976) seminal work, the field has expanded greatly. While experimental and theoretical debate about human sensitivity and attention to tau is ongoing (e.g. Bootsma & Craig, 2002; Kaiser & Mowafy, 1993; Kerzel et al., 1999), neural structures for the detection of looming have been found (Sun & Frost, 1998;

Wang & Frost, 1992) providing strong support for the relative rate of optical expansion as a source of information in humans.

One major point in the development of tau was the distinction between local and global tau. In its original formulation tau was considered to be the inverse of the rate at which the angle subtended by an object expanded about the centre of expansion (Figure 4.2a). However, it was later recognised (Tresilian, 1991) that this original formulation was really relating the angle from the focus of expansion to the point and the rate of change of that angle. It did not matter whether the focus of expansion and the point were joined or not, as long as the distance between them did not change (Figure 4.2b). The term 'time to passage' was differentiated from TTC (Kaiser & Mowafy, 1993), since if the point was displaced more than half of a body width from the focus of expansion then it would pass the observer rather than making contact. This new description was named global tau.

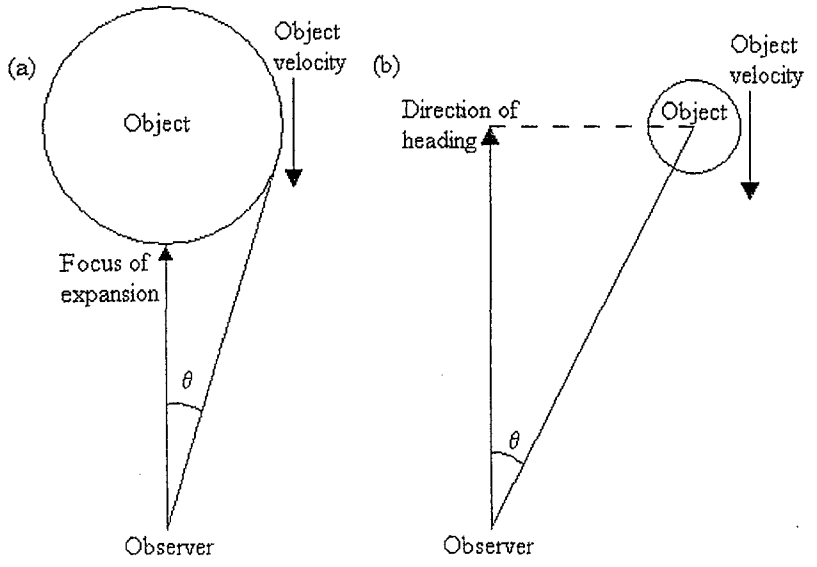



**Figure 4.2** (a) Local tau and (b) global tau. Local tau refers to the rate of expansion of a solid object. Global tau refers to the rate of increase of the angle from the direction of heading to the object.

Local tau refers to the inverse of the rate at which the angle subtended by an object grows over time. The complexity involved in analysing local tau is such that researchers have attempted to eliminate it from experiments regarding tau, focussing almost exclusively on global tau by using expansionless dots (e.g. Kaiser & Mowafy, 1993; Kerzel et al., 1999). Recently, Bootsma and Craig (2002) combined local and global tau to form composite tau. This will be examined in Chapter 5, along with a novel way of incorporating the two forms of tau information.

The TTC is the actual time before the object reaches the observer. Global tau is a ratio that allows TTC to be approximated. Where $L$ is the optical lateral distance and $\dot{L}$ is its rate of change, global tau is given by

$$\tau = \frac{L}{\dot{L}} \qquad \text{Equation 4.1}$$

As long as $L$ is not too big, then TTC will be approximately the inverse of global tau. Although global tau is an approximation that is not strictly necessary (Bootsma & Craig, 2002), it does have some utility. Only the optical lateral distance is needed so tau can be calculated from a pair of still images without any knowledge of any other parameters, such as the focal length of the camera that produced them.

Use of global tau as an approximation to TTC does mean that the environment must be constrained. The object for analysis must have no motion component perpendicular to the direction of observer motion and the path of the observer's eyes

must be linear or else $\dot{L}$ would not be an accurate measure of optical expansion. Observer motion must be at a constant speed relative to the object or else time to collision will not be equivalent to tau. A rolling robot moving through a stationary environment satisfies these constraints.

The utility of tau and the computational simplicity of its production make it an ideal property with which to evaluate a robotic system implementing the discrete smear algorithm.

## 4.2 Producing smeared images with a robotic system.

For the purposes of testing motion-from-smear in the real environment, instead of mathematical simulation, a robot was borrowed from Industrial Research Limited. The robot was a Pioneer 2, operating via Saphira version 6.2 (ActivMedia, 1996).

The Pioneer robot was controlled by a laptop attached by a COM port. The laptop was a Compaq Armada E500 with 196MB of RAM. The laptop controlled the robot via MATLAB version 6 realease 12.

The camera was mounted at the front of the Pioneer robot. A *Genius webCAM Live* was used, with a streaming frame rate of 30 frames per second.

Images were acquired using a custom program based on *Microsoft Windows AVI* software. A stream of images was taken from the camera at 30 frames per second and images from that stream were sampled once every 200ms. It was not possible to make the sampling interval exactly equal to the interval specified by the user so a log was kept of the actual sampling intervals. Images were stored in RAM and the smeared image was produced using the discrete smear algorithm following the image capture process. The smeared image and the final clear image were saved to disk, along with the log of actual sampling intervals. The distance from the robot to its target was recorded at the end of image capture.

### 4.2.1 Processing smeared images

Recall that for a continuously smeared image the half-life indicates the rate of change of the optical angle. This equation cannot be applied to an image produced using the discrete smear algorithm, since the images are stepped (Figure 3.14b) instead of smooth (3.11b). Hence, the following steps were followed (Figure 4.3):

1 Take the smeared image and subtract the end-state image from it.

2 Take the intensity values of the horizon line from that image.

3 Identify the first point on the first step.

4 Find the first point on the last step. This is $L$.

5 Measure the distance in pixels between the first and last points. This is $\Delta L$

6 Divide this distance by the time taken to capture the images. This is $\dot{L}$.

7 Divide $L$ by $\dot{L}$ to obtain global tau.

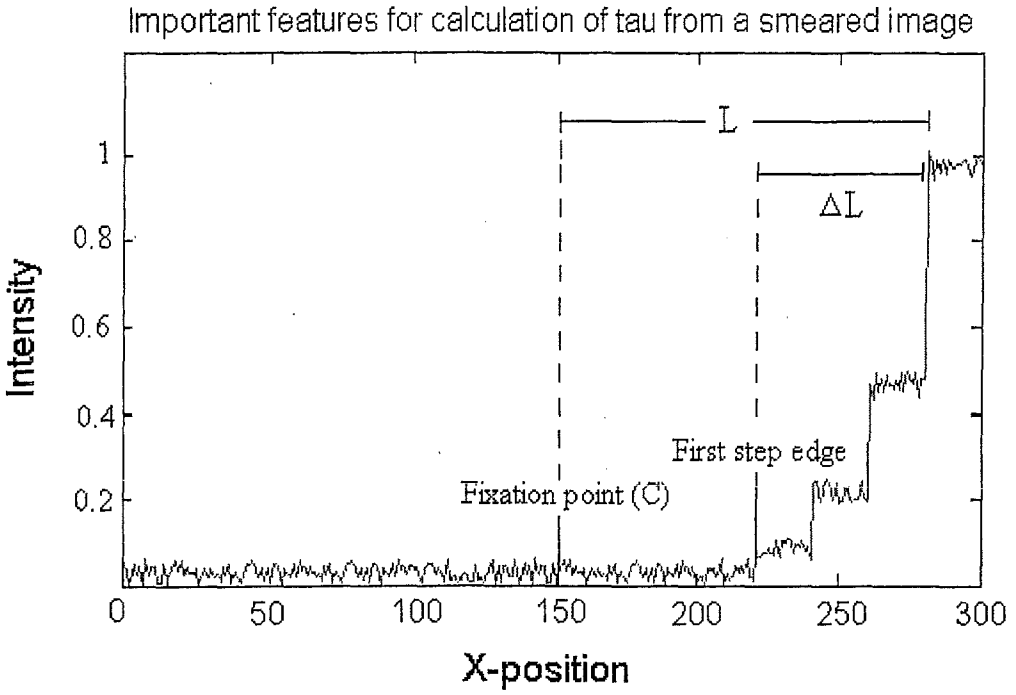



Figure 4.3 $L$ is the distance from the edge of the last step to the fixation point. Dividing the distance between the edges of the first and last steps, $L$, by the time taken to acquire the images gives the differential of $L$.

## 4.3 Evaluating the robotic system

### 4.3.1 Method

The task of the robot was to judge the TTC with a white wall marked by horizontally offset black patch. Correct judgment would indicate that the information necessary for wall avoidance is present in the images produced by the discrete smear algorithm.

As shown in figure 4.4, the robot was positioned 4m from the white wall. The black patch on the wall was offset horizontally from the point directly in front of the robot ($C$) by a distance, $D$. The robot moved toward the point $C$ at a constant speed, $S$. The

distance from the wall at which image capture ended ($X$) was recorded. The actual TTC ($X/S$) was compared to the estimated value for tau. There were three levels of distance offset (0.25m, 0.5m and 0.75m) and three nominal levels of distance (1.5m, 2m, 2.5m). The architecture of the camera control meant that exact control of the distance was difficult, but actual distance was accurately measured. Speed was set to $0.3ms^{-1}$ for all trials. Two trials was recorded at each distance. A problem with the robot meant that some trials had to be discarded, in which case they were repeated.

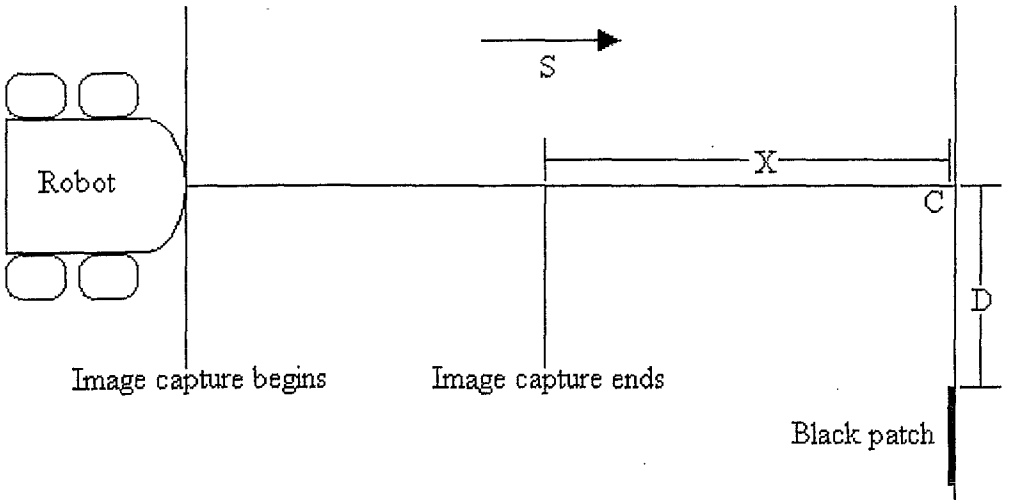



**Figure 4.4** The robot moves towards the point $C$ at speed $S$. Image capture ends at distance $X$ from the wall. A black patch on the wall is horizontally offset by distance $D$ from point $C$. The time to collision is $X/S$.

### 4.3.2 Results

It was found that overall accuracy was good, but that precision was not. The mean difference between TTC and tau was -0.52s, but the standard deviation of tau scores was 1.52s.

### 4.3.3 Discussion

The finding that the mean error was close to zero is strong support for the potential utility of motion from smear. The low precision can be accounted for by the low resolution of the camera and the poor quality of the images produced, while the increase in precision with increasing rates of optical expansion can be explained by a decrease in relative experimental uncertainty. The low resolution of the camera meant that localising the edge of each step is more difficult than would be the case if a higher resolution were used. Figure 4.5 illustrates how a higher resolution decreases the uncertainty about the exact position of the edge of each step. The *Genius webCAM* also produces a blurry image, which can make it difficult to even localise the edge of a step to within a single pixel. The imprecision can be eliminated by using a higher resolution camera, but it would not be as serious a problem for an autonomous vehicle as it might appear. The nature of the discrete smear algorithm means that the estimate of tau can be continuously updated. This means that the small, non-systematic, variations brought about by uncertainty can be "averaged out" over a number of samples. This would allow imprecision to be corrected, and the full potential accuracy of motion-from-smear to be realised.

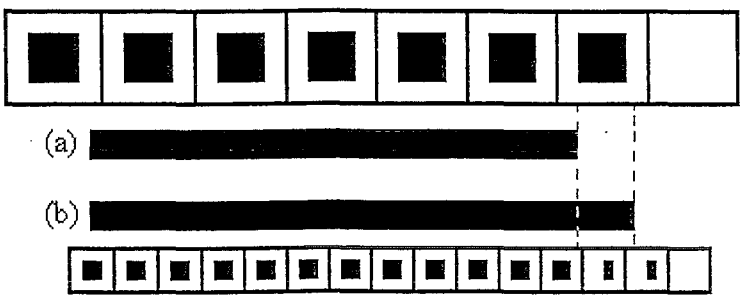
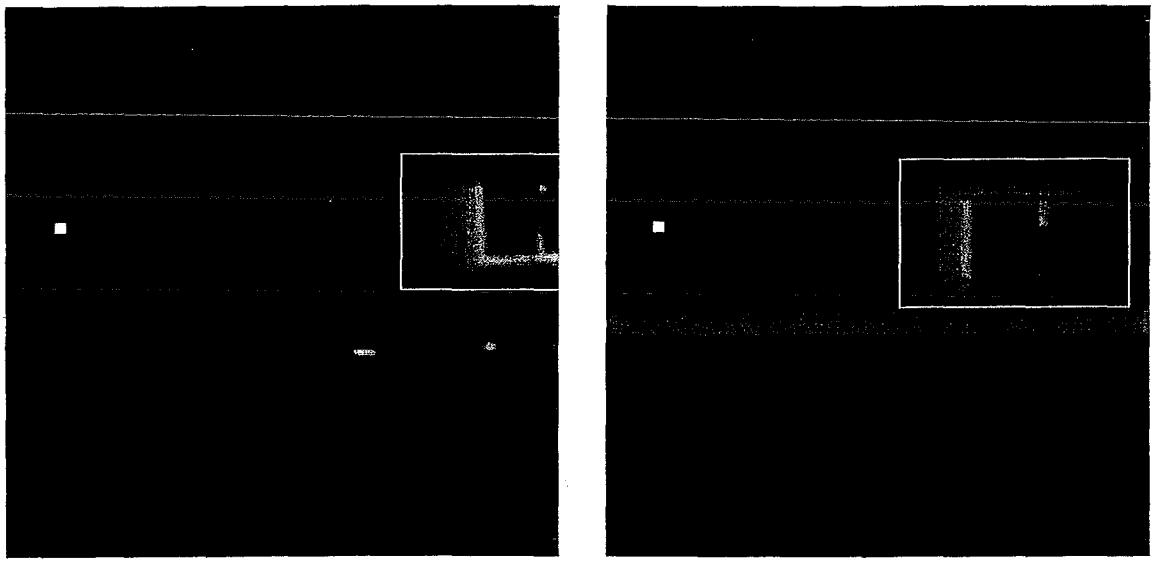



**Figure 4.5** The effect of resolution upon uncertainty. The resolution of the upper row of pixels is worse than that of the lower row. As such, the upper row cannot discriminate between lines (a) and (b). The lower row of pixels has a higher resolution and is thus able to make that distinction.

The most noticeable outcome of this experiment was that most of the experimental trials had to be discarded. The problem was not that the algorithm was incorrect, but that one of the constraints was not met. The robot had a warped wheel, which meant that the speed could not be measured accurately in many of the trials. Some trials had a vertical or horizontal component to their motion, and this caused distortion to the way in which the image was blurred (Figure 4.6).



**(a)** **(b)**

**Figure 4.6** Two smeared images (after end-state subtraction) produced during the experiment. The solid white squares indicate the fixation point, *C*. The hollow white rectangles are the black patches. **(a)** An image in which all elements were produced with a relatively constant velocity, allowing good judgment of tau. **(b)** An image with a significant perpendicular velocity, which made tau judgement difficult.

The overall conclusion of this test is that the discrete smear algorithm does provide information necessary for wall avoidance, but that the constraints on observer motion must be met more closely than was achieved in this experiment.

# Chapter 5

# Applications of motion-from-smear

## 5.1 Composite tau and time to contact

As described above, global tau was an approximation to the actual TTC. However, Bootsma and Craig (2002) showed that such an approximation was not strictly necessary. While this reformulation allowed precise calculation of TTC, it was still a requirement that the object not have a motion component perpendicular to the direction of observer motion. To address this, Bootsma and Craig (2002) devised composite tau, which incorporated a form of local tau. The size of the optical angle subtended by an object ($\phi$) and the rate at which it grew ($\dot{\phi}$) were combined with the angular displacement from the direction of heading and the rate of change of that angle ($\theta$ and $\dot{\theta}$ respectively, see Figure 5.1). This allowed TTC to be calculated for an object moving freely in three dimensions (Equation 5.1). This equation works well for moving spheres, but begins to break down when applied to objects such as cubes. As an object moves toward or away from the focus of expansion the optical shape of the object is transformed in ways that can confound calculation of composite tau

(Kaiser & Mowafy, 1993). The need to measure $\phi$ also makes this method more difficult than is strictly necessary.

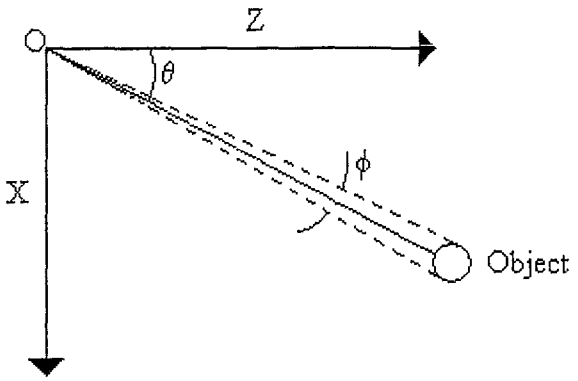
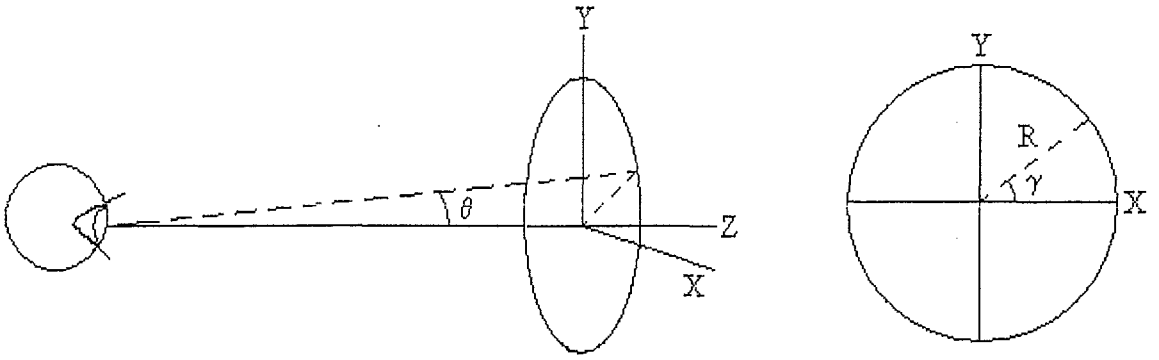



**Figure 5.1** The TTC for an object free to move in three dimensions can be calculated by using $\theta$ and $\phi$.

$$\tau = \frac{\dot{\theta}}{\tan\theta} + \frac{\dot{\phi}}{\sin\phi}$$ Equation 5.1

Any point on a two-dimensional image can be specified in terms of two co-ordinates – in Cartesian form these co-ordinates are positions on the x- and y-axes. There are other ways to define position, however, the most common of which is in polar co-ordinates. Each point is defined in terms of the distance, $R$, and bearing, $\gamma$, from an origin point (Figure 5.2(a)). If the focal length is known, the distance can be replaced by a second angle, $\theta$. In this way, any point on an image can be defined in terms of a pair of angles. These co-ordinates do not provide adequate information for calculation of composite tau when an object is free to move in three dimensions, but there is enough to extract composite tau in a constrained environment. While we have already decided that a stationary environment is too constrained to be of use, there is a compromise solution. The best option is to constrain motion to two dimensions, such as motion over a flat ground plane. If an object is constrained to move only in two dimensions then two co-ordinates, $\theta$ and $\gamma$, are sufficient for calculation of composite

tau. Motion over a ground plane is also a constraint that is often satisfied in the real world. People, cars and animals all move parallel to the ground – even birds usually fly horizontally. If motion is constrained to the horizontal direction then composite tau can be calculated by Equation 5.2 (see Appendix 1 for proof).



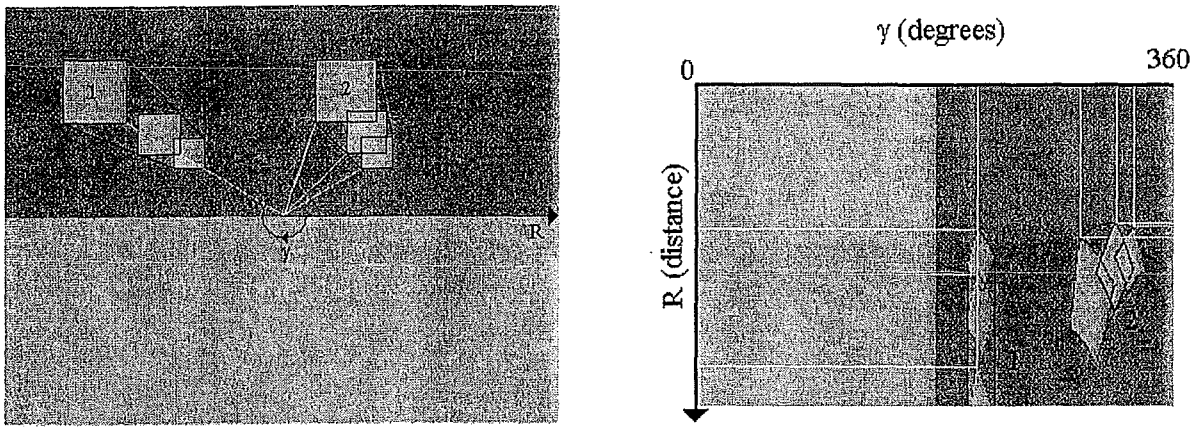

**(a)** **(b)**

Figure 5.2 An object moving parallel to the ground can be calculated using , , and their differentials. (a) Side view. (b) Frontal view.

$$\tau = \frac{2\dot{\theta}}{\sin 2\theta} + \frac{\dot{\gamma}}{\tan \gamma} \qquad \text{Equation 5.2}$$

It is logical, when confronted with Equation 5.2, to ask why it is not necessary to measure expansion. In fact, expansion is taken into account by the parameter $\gamma$. The points on an object on a parallel approach to the observer will have a constant $\gamma$, no matter how close it gets (Figure 5.3(a)). If $\dot{\gamma}$ is zero then tau can be calculated using only $\theta$ and its derivative. However, a point with a motion component in the $X$ direction will have a non-zero $\dot{\gamma}$, requiring the use of composite tau.

Changes in $\gamma$ can be demonstrated more easily by performing a polar transformation upon the image. Each point on the image is plotted in a Cartesian fashion with $\gamma$ on the horizontal axis and $R$ on the vertical axis. The centre of the original image is at the top of the transformed image.




(a) (b)

**Figure 5.3** **(a)** A Cartesian representation of two approaching squares. Square 1 is on a parallel course, while square 2 has a motion component from left to right. **(b)** A polar representation of the same two squares. Square 1 has a constant , so it travels straight down in this form. Square 2 has a changing , and thus has a horizontal component to its motion here. White lines indicate , while black lines indicate the path taken.

This has the effect of making motion perpendicular to the radius appear as translation along the horizontal axis (Figure 5.3). An object following a path parallel to the observer will now appear to move straight down while an object moving on a path perpendicular to the observer will now appear to have an angular component to is motion. The implications for the calculation of TTC are enormous. If motion is straight down after applying a polar transformation then $\dot{\gamma}$ is zero and the TTC can be calculated using normal methods. While the use of polar transformations to detect translational motion is not new (Schwartz, 1977; Weiman, 1990), it has not previously been used for detection of expansion in TTC calculation.

Like other formulations of tau, Equation 5.2 is *first order*. It requires that the object not be accelerating or decelerating. An additional constraint is that the point under observation cannot be located *on* the horizon. This is because the proof requires that Equation 5.3 be satisfied. While we have already established that the rate of change of $Y$ is equal to zero, the equation will be undefined if $Y$ is equal to zero. In spatial terms, if the point is located on the horizon then parallel motion becomes confounded with perpendicular motion. A solitary point is dimensionless so it is both invisible and harmless. Any real surface with which we may collide will have points that are not located exactly on the horizon, meaning that for any real surface $Y$ will not equal zero for all points on the surface.

$$\frac{\dot{Y}}{Y} = 0 \qquad \text{Equation 5.3}$$

The requirement that the point be displaced from the horizon line might explain why this method for calculation of composite tau has not been discovered before. Most research on TTC uses objects located on the horizon line (e.g. Bootsma & Craig, 2002; Kaiser & Mowafy, 1993; Kerzel et al., 1999), perhaps because the calculation of position for rendering is simpler. If so this added simplicity should have been enough to suggest that perhaps less information was being encoded by the experimenters.

It should be relatively simple to determine whether the assumption of horizontal motion is employed by the visual cortex. The perception of TTC for an expansionless

dot displaced from the horizon and moving across the ground plane should be equal to perception for a similarly constrained, but optically expandable, sphere. Performance should be far worse for an expansionless dot than an expandable sphere if motion in three dimensions is permitted. Future research will examine this area.

Lines were used to depict motion of points in Figure 5.3. Smearing would mean that these lines would become virtual contours in the image. While no new information is been added, the value of $R$, $\gamma$ and their respective rates of change are measured easily, facilitating the calculation of composite tau.

The ease with which x-axial motion can be detected means that those objects with a motion component perpendicular to the observer are more easily detected, simplifying the task of identifying those objects most likely to become hazards. In order to detect probable hazards, a detection algorithm need only detect parts of the image with a horizontal smear component.

## 5.2 Stair descending – judging riser height

There have been many studies of human perception of stairs and the ways in which the height of each step (riser height) is specified (Cesari, Formenti, & Olivato, 2003; Kerr, Eves, & Carroll, 2001; Konczak, Meeuwsen, & Cress, 1992; Maraj, 2003; Maraj & Domingue, 1999; McKenzie & Forbes, 1992; Warren, 1984). Most studies agree

that riser height is specified in terms of the eye-height for the normal observer (but see Konczak et al., 1992 for comparison with older adults). This makes sense, since an observer needs to know if their effectivities, particularly their leg length, are sufficient for climbing the stairs (Warren, 1984). If an observer knows their leg length in terms of eye-heights – which would seem reasonable – then it is logical to expect that riser height would need to be scaled in eye-heights as well. However there is little, if any, study of the mathematical principles underlying the optical specification of riser height. In addition, most study has centred upon stair climbing when in fact stair descending is more easily described mathematically. An equation for the specification of riser height while descending a flight of stairs will now be given and the ability of the discrete smear algorithm to aid implementation of this equation will be discussed.

When approaching the top of a flight of stairs we see a series of disoccluding edges; as we get closer we can see more and more of each step.

Equation 5.4 shows that the rate of change of the angle to the riser edge ($\dot{\psi}$) and the angular rate of change of the point on the surface currently being disoccluded by the riser edge ($\dot{\alpha}$) can specify the riser height (*RH*) in terms of eye height (*EH*) without knowledge of observer velocity (Figure 5.4, see Appendix 1 for proof). This equation holds even if the velocity of the observer has a vertical motion component, which is important since the eyes of a bipedal human follow a sinusoidal path through space.

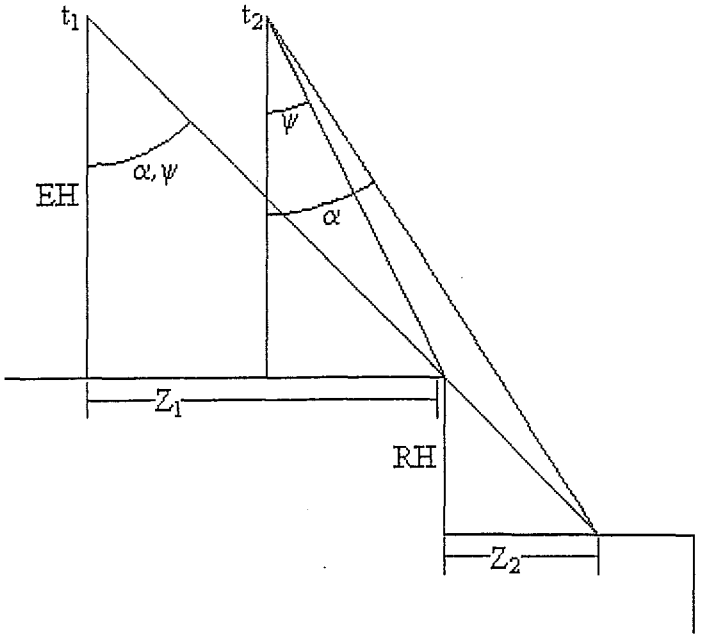



**Figure 5.4** Angles and distances involved in stair descending. Combining $\dot{\psi}$ and $\dot{\alpha}$ can determine *RH* in terms of *EH* regardless of $Z_1$

$$\frac{RH}{EH} = \frac{\dot{\psi}}{\dot{\alpha}} - 1$$

Equation 5.4

Equation 5.4 is really an application of motion parallax to the case of stair descending. If an observer watches two points at different distances then the bearings to those points will change at different rates. The difference in these rates can indicate relative depth. Stair descending is different from classical formulations of motion parallax in two main respects. Firstly, the perpendicular distance from the occluding edge to the observer is already specified (*EH*). Secondly, the second "point" is actually a surface. This allows us to equate the two angles ($\psi$ and $\alpha$), which simplifies the equation dramatically (see Appendix 1).

Equation 5.4 relies on evaluating two separate rates of change at the same point. The visual cortex might be able to do this using purely optical means, but it is not really

necessary to find out how. Instead we can apply efferent perception of eye tracking to the problem. Traditional views of eye-movements hold that "[t]he primary purpose of saccadic eye movements is to bring the image of visual objects of interest onto the fovea" (Edelman, Cherkasova, & Nakayama, 2002, pg 572a).

By tracking the edge of the riser, however, we can use the rotational velocity of the viewpoint to specify $\dot{\psi}$. This means that it is only necessary to determine $\dot{\alpha}$ optically. The use of the instructions for the control of movement as a source of information is known as *efferent perception* (Brindley & Merton, 1960; Helmholtz, 1867/1925). There is debate about whether efferent perception is actually used in human perception (c.f. Gibson 1966), but if an autonomous artificial agent can track a moving object (using either hardware or software) it may be simpler to use efferent perception than to analyse the entire optic flow field.

It is for this purpose that the motion-from-smear is useful in this context. In the case of the smeared image in Figure 5.5, the disoccluding edge is held in the centre of the image. The ground between the observer and the image becomes progressively more smeared at points nearer to the observer, but the points near the edge of the top step are clear. Each sudden increase in the amount of smear is a clear indicator of a step edge. Those points on the lower step are noticeably more smeared than those on the step above. Likewise, the points on the third step are even more smeared and so it carries on, until eventually the smear is so great as to be unintelligible. This confusion is not an overly serious problem, since if an agent is concerned about the lower steps they need only change their fixation point.

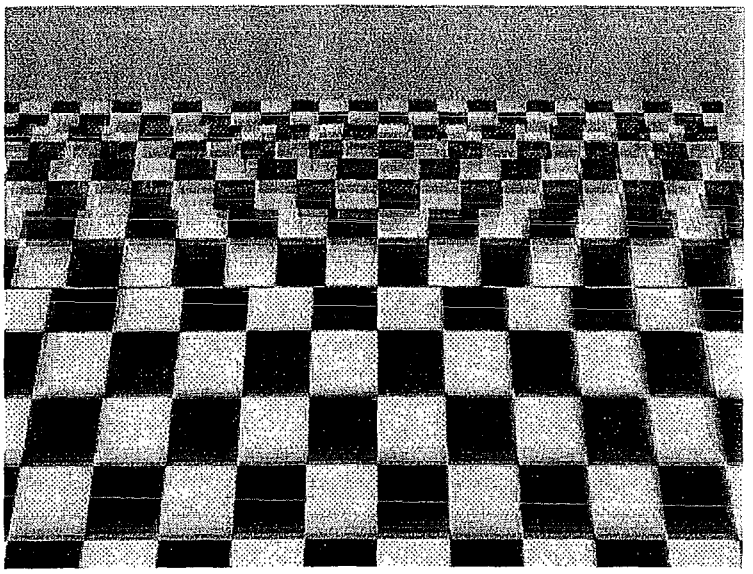


**Figure 5.5** A set of steps viewed from above. An agent approaches the stairs and fixates upon a point on the edge of the top step. The second step is more blurred than the ground leading up to the steps. The third step is more blurred than the second and so on. The fourth and fifth steps are moving so fast that the smear is unintelligible for those steps. The amount of smear allows judgment of $\dot{\alpha}$

All of this assumes that the observer is aware that they are looking at a disoccluding edge. Sun and Frost (Frost & Sun, 1997) found cells in the visual cortex of the pigeon that are selective for occluding or disoccluding edges, and it would not be surprising to find such cells in the human brain as well. It might well be that the rate of disocclusion contributes to – or is sufficient for – estimation of $\dot{\alpha}$.

The main direction of future research should be toward the optical specification of riser height when climbing stairs. To examine the optical specification of riser height further, it would first be necessary to determine the tolerance for error in a person descending stairs. How much can the perceived riser height and the actual riser height differ before injury occurs? Research into the actual point of fixation also needs to be carried out, in order to discern whether fixation is on the edge of the top step or on the surface of the second step. Only after error tolerance and the fixation points have been established can research into human determination of riser height and the possible role of motion-from-smear proceed properly.

# Chapter 6

# Conclusions

The work of Aho et al. (1993) makes it clear that toads use extremely long integration periods under low light levels. Even under normal lighting humans use temporal summation over intervals up to 100ms (Hamilton, Albrecht, & Geisler, 1989; Hawken, Shapley, & Grosof, 1996; Movshon, Thompson, & Tolhurst, 1978). Although smear has traditionally been regarded as a unwanted side effect of this temporal summation, recent work has begun to consider its use as a cue for the estimation of optic flow (Burr, 2000; Burr & Rose, 2002; W.G. Chen et al., 1994; W. G. Chen et al., 1996; Geisler, 1999; Geisler et al., 2001; Tull & Katsaggelos, 1996).

Where methods comparing separate images must solve a correspondence problem, a smeared image enables a point to be tracked continuously, eliminating the need for the correspondence problem to be solved at all.

Although first suggested by Ballard (1986), it was not seriously considered until Chen et al. (1995, 1996) and Tull and Katsaggellos (1996) individually began work upon

motion-from-smear. Recent physiological experiments have found neurons in the visual cortex of cats and monkeys that would appear to be responsive to smeared features (Geisler et al, 2001), while psychophysical experiments have found support for the use of smear by humans (Geisler, 1999; Burr & Rose, 2002).

Although current camera technology does not allow direct production of smeared images, this thesis has introduced the discrete smear algorithm for the simple production of smeared images from a series of stills. The images produced by the discrete smear algorithm contain the information necessary for the estimation of global tau in a rolling robot, as shown by experimentation. Additionally, simulated images have suggested that cues to composite tau and riser height while descending stairs can be found in smeared images.

Future research will focus on the calculation of motion-from-smear from the images produced by the discrete smear algorithm. It is expected that a form of Fourier transform should be capable of determining the optic flow for complex fields. Future exploration of the application of motion-from-smear to perception of time to contact and determination of the riser height for stairs is also likely.

The most important point to remember is that motion-from-smear uses cues that are different to those used in conventional methods of optic flow estimation. As such it acts as a complement to these techniques, rather than a form of competition. It is hoped that incorporation of the disparate information types will increase the power of artificial visual systems and aid in the understanding of human vision.

## Acknowledgements

I would like to express my gratitude to Dean Owen and Steven Lamb for proof reading this thesis. You should too; otherwise you would be reading a solid mass of typing errors. I would especially like to show my appreciation for Dean's relentless, but futile, attempts to turn me into an accurate and precise writer. The fault is mine, not his. I would also like to thank my parents, my brothers and my friends. Most of all I would like to thank Sharmila, who cleaned up after me for a month. Other, than that thanks to everyone who helped me, listened to me, pretended to understand my garbled ranting and otherwise put up with me.

Scanners Note: there is no p68 in the original thesis.

# References


ActivMedia. (1996). *Pioneer 1 Software Manual.* Jaffrey, NH: RWI.

Aggarwal, J., & Nandhakumar, N. (1988). On the computation of motion from sequences of images - a review. *Proceedings of the IEEE, 76*(8), 917-935.

Aho, A. C., Donner, K., Helenius, S., Larsen, L. O., & Reuter, T. (1993). Visual Performance of the Toad (Bufo-Bufo) at Low-Light Levels - Retinal Ganglion-Cell Responses and Prey-Catching Accuracy. *Journal of Comparative Physiology a-Sensory Neural and Behavioral Physiology, 172*(6), 671-682.

Ballard, D. H. (1986). Cortical connections and parallel processing: Structure and function. *Behavioral & Brain Sciences, 9*(1), 67-120.

Beauchemin, S., & Barron, J. (1995). The computation of optical flow. *ACM Computing Surveys, 27*(3), 433-467.

Bootsma, R. J., & Craig, C. M. (2002). Global and local contributions to the optical specification of time to contact: Observer sensitivity to composite tau. *Perception, 31*(8), 901-924.

Boussaoud, D., Ungerleider, L. G., & Desimone, R. (1990). Pathways for motion analysis - cortical connections in the medial superior temporal and fundus of the superior temporal visual areas in the macaque. *Journal of comparative Neuropsychology, 296*(3), 462-495.

Brindley, G. S., & Merton, P. A. (1960). The absence of position sense in the human eye. *Journal of Physiology, 153*, 127-130.

Burr, D. C. (2000). Motion vision: are "speed lines" used in human visual motion? *Current Biology, 10*, R440-R443.

Burr, D. C., & Rose, J. (2002). Direct evidence that "Speedlines" influence motion mechanisms. *The Journal of Neuroscience, 22*(19), 8661-8664.

Cesari, P., Formenti, F., & Olivato, P. (2003). A common perceptual parameter for stair climbing for children, young and old adults. *Human Movement Science, 22*(1), 111-124.

Chen, W. G., Nandhakumar, N., & Martin, W. N. (1994). *Investigating a new visual cue for image motion estimation: "motion-from-smear".* Paper presented at the Proceedings of the 1994 IEEE International Conference on Image Processing, Austin, Texas, USA.

Chen, W. G., Nandhakumar, N., & Martin, W. N. (1996). Image motion estimation from motion smear - A new computational model. *Ieee Transactions on Pattern Analysis and Machine Intelligence, 18*(4), 412-425.

Edelman, J. A., Cherkasova, M. V., & Nakayama, K. (2002). A spatial memory system for the guidance of eye movements in crowded visual scenes. *Journal of Vision, 2*(7), 572a.

Frost, B. J., & Sun, H. (1997). Visual motion processing for figure/ground segregation, collision avoidance, and optic flow analysis in the pigeon. In M. V. Srinivasan & S. Venkatesh (Eds.), *From living eyes to seeing machines* (pp. 80-103).

Geisler, W. S. (1999). Motion streaks provide a spatial code for motion direction. *Nature, 400*(6739), 65-69.

Geisler, W. S., Albrecht, D. G., Crane, A. M., & Stern, L. (2001). Motion direction signals in the primary visual cortex of cat and monkey. *Visual Neuroscience, 18*(4), 501-516.

Gibson, J. J. (1961). Ecological optics. *Vision Research, 1*(253-262), 23-30.

Gibson, J. J. (1962). Observations on active touch.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Hamilton, D. B., Albrecht, D. G., & Geisler, W. S. (1989). Visual Cortical Receptive-Fields in Monkey and Cat - Spatial and Temporal Phase-Transfer Function. *Vision Research, 29*(10), 1285-&.

Hawken, M. J., Shapley, R. M., & Grosof, D. H. (1996). Temporal-frequency selectivity in monkey visual cortex. *Visual Neuroscience, 13*(3), 477-492.

Helmholtz, H. v. (1867/1925). *Treatise on physiological optics* (Translated from 3rd German ed. Vol. III). New York: Dover Publications.

Hoyle, F. (1959). *The Black Cloud.* New York: Signet.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology (London), 195*, 215-243.

Jia, Y., Xu, Y., Liu, W., Yang, C., Zhu, Y., Zhang, X., et al. (2003). *A miniature stereo vision machine for real-time dense depth mapping.* Paper presented at the Computer Vision Systems: Third International Conference, Graz, Austria.

Jochem, T., & Pomerleau, D. (1996). Life in the fast lane: The evolution of an adaptive vehicle control system. *AI Magazine, 17*(2), 11-50.

Jones, D. G., & Malik, J. (1992). Computational Framework for Determining Stereo Correspondence from a Set of Linear Spatial Filters. *Image and Vision Computing, 10*(10), 699-708.

Julesz, B. (1971). Foundations of cyclopean perception.

Kaiser, M. K., & Mowafy, L. (1993). Optical Specification of Time-to-Passage - Observers Sensitivity to Global Tau. *Journal of Experimental Psychology-Human Perception and Performance, 19*(5), 1028-1040.

Kehtarnavaz, N., Griswold, N. C., & Lee, J. S. (1991). Visual control of an autonomous vehicle (BART)-the vehicle-following problem. *IEEE Transactions on Vehicular Technology, 40*(3), 654 - 662.

Kerr, J., Eves, F., & Carroll, D. (2001). Six-month observational study of prompted stair climbing. *Preventive Medicine, 33*(5), 422-427.

Kerzel, D., Hecht, H., & Kim, N. G. (1999). Image velocity, not tau, explains arrival-time judgments from global optical flow. *Journal of Experimental Psychology-Human Perception and Performance, 25*(6), 1540-1555.

Konczak, J., Meeuwsen, H. J., & Cress, M. E. (1992). Changing Affordances in Stair Climbing - the Perception of Maximum Climbability in Young and Older Adults. *Journal of Experimental Psychology-Human Perception and Performance, 18*(3), 691-697.

Kumar, S. S., & Chatterji, B. N. (2004). Two-frame stereo matching algorithms - A review. *IETE JOURNAL OF RESEARCH, 50*(1), 49-61.

Lee, D. N. (1974). Visual Information During Locomotion. In R. B. MacLeod & H. L. Pick Jr. (Eds.), *Perception: Essays in honor of J. J. Gibson* (pp. 250-267). Ithaca, NY: Cornell University Press.

Lee, D. N. (1976). Theory of Visual Control of Braking Based on Information About Time-to-Collision. *Perception, 5*(4), 437-459.

Lee, D. N., & Lishman, J. R. (1975). Visual Proprioceptive Control Of Stance. *Journal of Human Movement Studies, 1*, 87-95.

Maraj, B. K. V. (2003). Perceptual judgments for stair climbing as a function of pitch angle. *Research Quarterly for Exercise and Sport, 74*(3), 248-256.

Maraj, B. K. V., & Domingue, J. A. (1999). Standing distance in climbability of stairs. *Perceptual and Motor Skills, 88*(2), 682-684.

Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman and Company.

Marr, D., & Poggio, T. (1976). Cooperative Computation of Stereo Disparity. *Science, 194*(4262), 283-287.

Marr, D., & Poggio, T. (1979). Computational Theory of Human Stereo Vision. *Proceedings of the Royal Society of London Series B-Biological Sciences, 204*(1156), 301-328.

Martin, K. E., & Marshall, J. A. (1993). Unsmearing visual motion: development of long-range horizontal intrinsic connections. In S. J. Hanson, J. D. Cowan & C. L. Giles (Eds.), *Advances in Neural Information Processing Systems*. San Mateo, California: Morgan Kaufmann.

McKenzie, B. E., & Forbes, C. (1992). Does Vision Guide Stair Climbing - a Developmental-Study. *Australian Journal of Psychology, 44*(3), 177-183.

Milner, A. D., & Goodale, M. A. (1993). Visual Pathways to Perception and Action. *Progress in Brain Research, 95*, 317-337.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford ; New York: Oxford University Press.

Mishkin, M. (1972). Visual mechanisms beyond the striate cortex. In R. Russell (Ed.), *Frontiers in Physiological Psychology*. New York: Academic.

Mishkin, M., Ungerlieder, L. G., & Macko, K. A. (1983). Object vision and spatial vision - 2 cortical pathways. *Trends in Neurosciences, 6*(10), 414-417.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978). Spatial and Temporal Contrast Sensitivity of Neurons in Areas-17 and Areas-18 of Cats Visual-Cortex. *Journal of Physiology-London, 283*(OCT), 101-120.

Palmer, S. E. (1999). *Vision Science: Photons to Phenomenology*. Cambridge, Massachusetts: The MIT Press.

Pohl, W. (1973). Dissociation of spatial discrimination deficits following frontal and parietal lesions in monkeys. *Journal of Comparative & Physiological Psychology, 82*, 227-239.

Roska, B., & Werblin, F. (2001). Vertical interactions across ten parallel, stacked representations in the mammalian retina. *Nature, 410*(6828), 583-587.

Schwartz, E. (1977). Spatial mapping in the primate sensory projection : Analytic structure and relevance to perception. *Biological Cybernetics, 25*, 181-194.

Sekuler, R. W., & Ganz, L. (1963). After effect of seen motion with a stabilized retinal image. *Science, 139*, 419-420.

Sharpe, C. R., & Tolhurst, D. J. (1973). Orientation and spatial frequency channels in peripheral vision. *Vision Research, 13*, 2103-2112.

Stappers, P. J. (1992). *Scaling the visual consequences of active head movements*.Unpublished manuscript, Delft University of Technology.

Sun, H. J., & Frost, B. J. (1998). Computation of different optical variables of looming objects in pigeon nucleus rotundus neurons. *Nature Neuroscience, 1*(4), 296-303.

Tolhurst, D. J. (1973). Separate channels for the analysis of the shape and the movement of a moving visual stimulus. *Journal of Physiology, 231*, 385-402.

Tresilian, J. R. (1991). Empirical and Theoretical Issues in the Perception of Time to Contact. *Journal of Experimental Psychology-Human Perception and Performance, 17*(3), 865-876.

Tull, D. L., & Katsaggelos, A. K. (1996). *Regularized blur-assisted displacement field estimation.* Paper presented at the Proceedings of the 1996 IEEE International Conference on Image Processing, Lausanne, Switzerland.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale & R. J. W. Mansfield (Eds.), *Analysis of Visual Behavior* (pp. 549-586). Cambridge, Massachusetts: MIT Press.

Wang, Y. C., & Frost, B. J. (1992). Time to Collision Is Signaled by Neurons in the Nucleus Rotundus of Pigeons. *Nature, 356*(6366), 236-238.

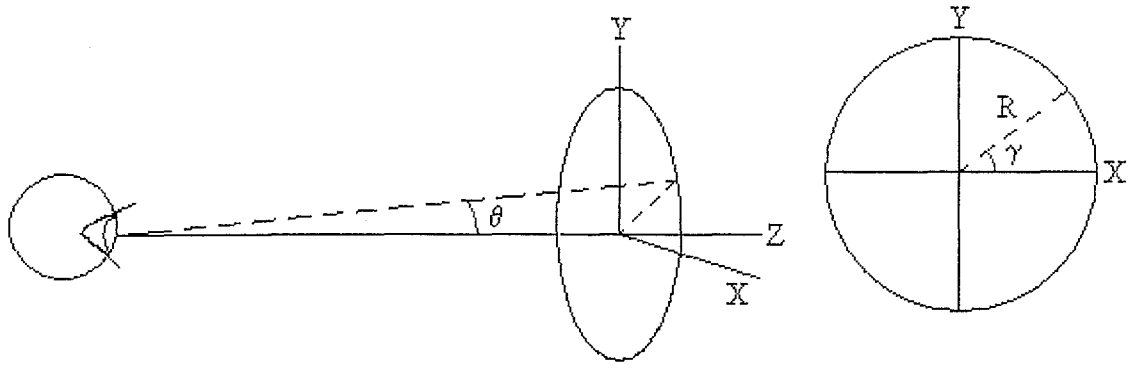
Warren, W. H. (1984). Perceiving affordances: Visual guidance of stair climbing. *Journal of Experimental Psychology: Human Perception & Performance, 10*(5), 683-703.

Weiman, C. (1990). *Log-polar vision for mobile robot navigation.* Paper presented at the Electronic Imaging Conference, Boston, USA.

# Appendix

Proofs

## A.1 EQUATION 5.2




(a) (b)

**Figure A.1** (a) Side view of the angles involved in viewing a point. (b) Frontal view.

By definition time to collision along the $X$ axis $TC(Z)$ is specified by

$$[TC(Z)]^{-1} = -\frac{\dot{Z}}{Z} \tag{1}$$

From Figure A.1(b)

$$R = \sqrt{X^2 + Y^2} \tag{2}$$

$$\tan\theta = \frac{R}{Z} \tag{3}$$

Differentiating (3) with respect to time

$$\frac{\dot{\theta}}{\cos^2\theta} = \frac{\dot{R}Z - \dot{Z}R}{Z^2} \tag{4}$$

Rearranging (4) and substituting in (3)

$$\frac{\dot{\theta}}{\sin\theta\cos\theta} - \frac{\dot{R}}{R} = -\frac{\dot{Z}}{Z} \tag{5}$$

$$\sin\gamma = \frac{Y}{R} \tag{6}$$

$$\gamma\cos\gamma = \frac{\dot{Y}R - Y\dot{R}}{R^2} \tag{7}$$

Rearranging (7) and substituting in (6)

$$\frac{\dot{\gamma}}{\tan\gamma} - \frac{\dot{Y}}{Y} = -\frac{\dot{R}}{R} \tag{8}$$

If $\dot{Y} = 0$ and $Y \neq 0$

$$\frac{\dot{\gamma}}{\tan\gamma} = -\frac{\dot{R}}{R} \tag{9}$$

Substituting into (5) and equating with (1)

$$\underline{\underline{[TC(Z)]^{-1} = -\frac{\dot{Z}}{Z} = \frac{2\dot{\theta}}{\sin 2\theta} + \frac{\dot{\gamma}}{\tan\gamma}}}$$