NISTIR 5790

Scale-Space-Based Visual-Motion-Cue for Autonomous Navigation

Sridhar R. Kundur
Robotics Center and
Department of Electrical Engineering
Florida Atlantic University
Boca Raton, FL 33431

and

Daniel Raviv
Robotics Center and
Department of Electrical Engineering
Florida Atlantic University
Boca Raton, FL 33431
and Intelligent Systems Division

and

Ernest Kent
Intelligent Systems Division

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Bldg. 220 Rm. B124
Gaithersburg, MD 20899

NISTIR 5790

Scale-Space-Based Visual-Motion-Cue for Autonomous Navigation

Sridhar R. Kundur
Robotics Center and
Department of Electrical Engineering
Florida Atlantic University
Boca Raton, FL 33431

and

Daniel Raviv
Robotics Center and
Department of Electrical Engineering
Florida Atlantic University
Boca Raton, FL 33431
and Intelligent Systems Division

and

Ernest Kent
Intelligent Systems Division

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Bldg. 220 Rm. B124
Gaithersburg, MD 20899

February 1996

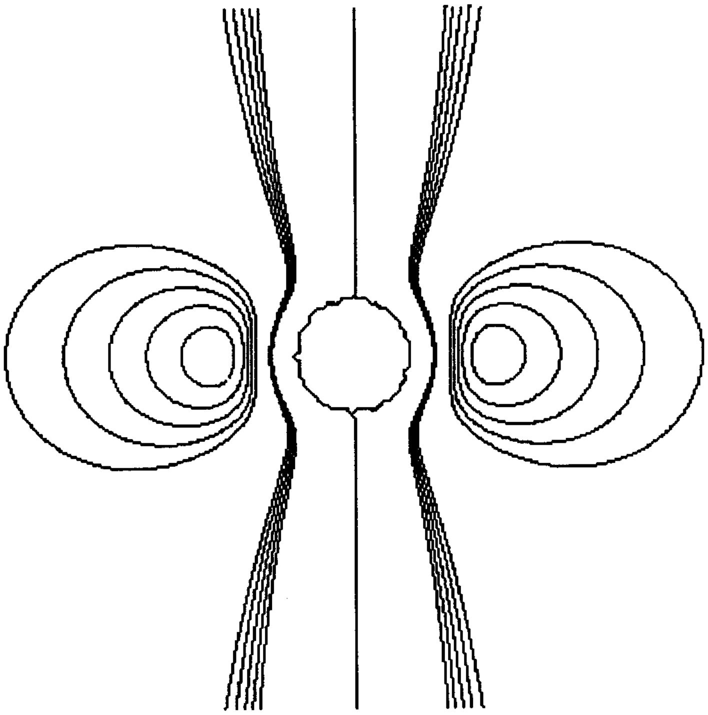
U.S. DEPARTMENT OF COMMERCE
Ronald Brown, Secretary

TECHNOLOGY ADMINISTRATION
Mary L. Good, Under Secretary for Technology

NATIONAL INSTITUTE OF STANDARDS
AND TECHNOLOGY
Arati Prabhakar, Director

Scale-Space-Based Visual-Motion-Cue for Autonomous Navigation*

Sridhar R. Kundur[1], Daniel Raviv[1,2] and Ernest Kent[2]
[1]Robotics Center and Department of Electrical Engineering
Florida Atlantic University, Boca Raton, FL 33431
[2]Intelligent Systems Division,
National Institute of Standards and Technology(NIST)
Bldg. 220, Room B124, Gaithersburg, MD 20899
email: kundur@acc.fau.edu and ravivd@acc.fau.edu



* This work was supported in part by a grant from the National Science Foundation, Division of Information, Robotics and Intelligent Systems, Grant # IRI-9115939

Abstract

This paper presents a new visual motion cue, we call the *Hybrid Visual Threat Cue* (HVTC). The HVTC provides some measure for a change in *relative range* as well as *absolute clearances*, between a 3D surface and an observer when there is a relative motion between them. The visual field associated with the HVTC can be used to demarcate the regions around a moving observer into safe and danger zones of varying degree, which may be suitable for autonomous navigation tasks, in particular *collision avoidance* and *maintenance of clearance*. The HVTC is *independent* of the 3D environment and needs almost no *a-priori* information about it. It is *rotation independent,* and is measured in [time^{-1}] units

When there is a relative motion between a *point of visual fixation* on a 3D surface and an observer, the perceived texture details in the image vary. The rate at which the details vary provides an indication of the observer's relative motion with respect to the 3D surface. *Scale space representation* which is a multiscale approach provides a concrete way to analyze the variations of image details with the image inner scale. We derive a relation between the relative temporal variations of the image inner scale and the HVTC.

A practical method to extract the HVTC from a sequence of images of a 3D textured surface obtained by a *visually fixated, fixed-focus monocular* camera in motion is also presented. A *global dissimilarity measure* is extracted *directly from the raw data of the gray level of textured images* from which the HVTC is obtained. This approach of extracting the HVTC is *independent of the type of 3D surface texture* and needs no *optical flow information, 3D reconstruction, segmentation, feature tracking.* It needs almost *no camera calibration.* This algorithm to extract the HVTC was applied to a set of twelve different texture patterns (of 3D scenes) from the Brodatz's album, where we observed a similar behavior for most of the textures.

Key Words: Active Vision, Visual Navigation, Visual Fields, Collision Avoidance, Scale Space Filtering

1 Introduction

1.1 Vision-Based Navigation: An Overview

The process of driving or flying in a 3D environment usually involves a human operator. The operator acts in part as a sensory feedback in the perception-action closed-loop control system to avoid obstacles, maintain clearance, etc., to ensure safe navigation in real time. It becomes a difficult problem to replace the human operator by a vision-based system to achieve similar tasks for the following reasons: in outdoor navigation the environment is usually a-priori unknown and unstructured, and the same 3D scene may result in many different images due to changes in illumination conditions, relative distances, orientation of the camera, choice of fixation point, etc., as well as various camera parameters such as zoom, resolution, focus, etc. There is a need for an approach, to obtain relevant visual information about relative proximity in the presence of the above mentioned factors.

When dealing with a moving camera-based autonomous navigation system, a huge amount of visual data is captured. For vision-based navigation tasks like obstacle avoidance, maintaining safe clearance, etc., relevant visual information needs to be extracted from this visual data and used in real-time closed-loop perception-action control system. In order to accomplish safe visual navigation several questions need to be answered, including:

1. What is the *relevant* visual information to be extracted from a sequence of images?
2. How does one extract this information from a sequence of 2D images?

3. How to generate control commands to the vehicle based on the visual information extracted?

This paper is focused on the first two of the above mentioned questions.

1.2 The Hybrid Visual Threat Cue: An Overview

This paper presents a new visual motion cue, we call the **Hybrid Visual Threat Cue** (HVTC) that provides some measure for a change in *relative range* as well as *clearance*, between a 3D surface and an observer in motion. It can be shown that the HVTC is a linear combination of the Time-to-Contact [11], the Looming [9] and the Visual Threat Cue (VTC) [50, 64]. The HVTC is *independent* of the 3D environment and needs almost no *a-priori* information about it. It is *rotation independent,* and is measured in [time^{-1}] units. Corresponding to this visual cue there is a visual field associated with the observer in motion. In other words there are imaginary 3D surfaces attached to the observer that move with it. All the points that lie on a particular imaginary surface produce the same value of the cue. The visual field associated with the HVTC can be used to demarcate the regions around an observer in motion into safe and danger zones of varying degree suitable for autonomous visual navigation.

When there is a relative motion between a fixation point on a 3D surface and an observer the perceived texture details vary depending upon the motion. For instance, consider the case of a camera (with fixed parameters) that is gradually moving towards a tree. When the distance between the camera and the scene is very large, details such as leaves and branches are smeared, due to finite spatial sampling. However as the camera moves towards the tree, details start appearing. The *rate* at which the details in the image varies provides an indication of changes in the relative distance between the camera and the observer. In other words, if details start appearing, it indicates that the relative distance between the observer and the camera is decreasing and vice-versa. The concept of scale space filtering introduced by Witkin [40] and Koenderink [44] provides a concrete way to analyze the image details with varying image inner scales.

We derive a relation between the relative temporal variations in the image inner scale and the power spectral density of the images. Also a relationship between the image inner scale and the range between the observer and the 3D point in fixation is derived. These two relations together establish a connection between the power spectral density of images and the HVTC (refer to Figure (1)).

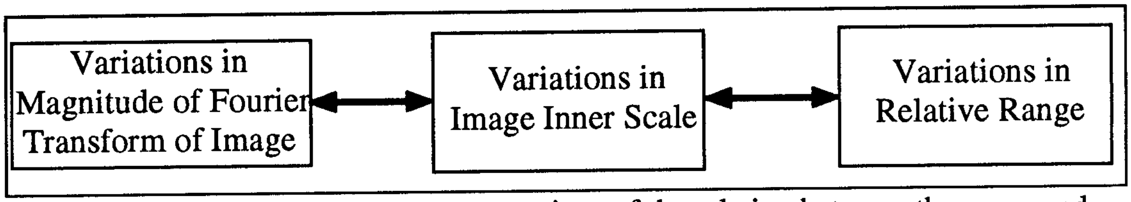

Variations in Magnitude of Fourier Transform of Image
Variations in Image Inner Scale
Variations in Relative Range

Figure (1): Block diagram representations of the relation between the range and image details

Several approaches to extract the HVTC are suggested. A practical method to extract the HVTC from a sequence of images of a 3D textured surface obtained by a *visually fixated* (i.e., observing the same point), *fixed-focus monocular* camera in motion is presented. For each image in such a 2D image sequence of a textured surface, a *global variable* (i.e., a variable that is obtained for each image window) we call the Image Quality Measure (IQM) is obtained *directly from the raw data of the gray level images.* Using the IQM values the HVTC is extracted. This approach of extracting the HVTC is *independent of the 3D surface texture* , i.e., it does not need to know what type of texture is present in the scene. It needs no *optical flow information, 3D reconstruction, segmentation, feature tracking.* The process of extraction can be seen as a *sensory fusion* of *focus, texture* and *motion* at the *raw data level and* needs almost *no camera calibration.* This algorithm works better on images obtained from *natural* scenes including fractal-like images, where more details of the 3D scene are visible in the images as the range shrinks and also can be implemented in parallel hardware. This algorithm to extract the HVTC was applied to a set of 12 different textures from the Brodatz's album [51]. A graphical comparison of the theoretical HVTC and the HVTC extracted from sequences of images is presented.

1.3 Other Approaches to Vision-Based Navigation

First we present a brief overview of autonomous vision-based navigation approaches. The problem of automating vision-based navigation is a challenging one and has drawn the attention of several researchers over the past few years (see for example [1-12]). Usually identifying the surrounding object is not important for such tasks, i.e., is it a tree, mountain or another vehicle; what is more important is whether a particular object is an obstacle or not, i.e., is the observer on a collision course with it, is there enough clearance, etc. For navigation tasks recovering the 3D scene and its attributes may not be necessary as it may contain information which is not relevant for the task at hand. Visual cues such as time-to-contact [11], looming [9], VTC [50] carry important information about the relative proximity. These cues can be obtained without 3D scene reconstruction which is usually computationally intense.

Time-To-Contact (TTC) is an entity that can be extracted from images, provides an indication of time available to the observer in order to make decisions about acceleration/deceleration without measuring the actual depth [11]. The looming effect which is the result of retinal expansion of objects due to change in range has shown to cause defensive reaction in several animals as well as babies [16, 17]. A detailed qualitative as well as quantitative approach to the concept of looming is presented in [10]. The Visual Threat Cue (VTC) provides some measure for relative change in range as well as clearance and is presented in [50].

It is well established in the literature (computer vision as well as psychology) that optical flow plays an important role in the control of human motion behavior in the environment [13-15]. Several researchers have addressed the use of optical flow as a feedback signal for vision based autonomous navigation (see for example [3,5-9]). An optical flow based theory of how a driver visually controls the braking of an automobile is presented in [11] where it is also shown that it is possible to control the braking of a vehicle using visual information without measuring the absolute distance, speed or acceleration/deceleration. A differential invariant of the image flow field-based visual information about time-to-collision is presented in [12]. Application of certain measures of flow field divergence as a qualitative cue for the task of obstacle avoidance is presented in [7]. In [9] the optical flow field is transformed by using a log-polar transformation to extract visual information about time-to-contact [19]. In [3, 8] the variations in peripheral optical flow are employed to guide a mobile robot through obstacles.

Though the optical flow-based approaches provide excellent qualitative approaches to visual navigation, its extraction from a sequence of images may be difficult in certain situations [20]. The extraction of the local optical flow employs a constraint equation between the local brightness gradients and the two components of the optical flow. Additional constraints are needed to evaluate the complete flow field [18-20]. In addition, the extraction of optical flow from a sequence of images needs pre-processing like spatio-temporal smoothing which may be computationally expensive. In such situations where optical flow-based approaches to visual navigation are difficult, alternatives to optical flow information as sensory feedback for obstacle avoidance may be required to increase the reliability of the system. Alternatives to optical flow to accomplish autonomous visual navigation include geometrical properties like size, shape, contour and area of image entities, imaged texture, focus, etc.

In [21] a frequency-based texture operator is employed to classify the characteristics of the Fourier transforms of local image windows, to compute the gradients of texture in the image in order to get depth information. Variations in image statistical parameters are employed to extract the differential invariant of image flow field is presented in [22]. A qualitative view of the use of these components as sensory feedback information for collision avoidance is also presented [22]. In [23] it is shown that the relative changes in edges of visible texture in a unit area are equal to looming , the concept introduced in [9]. This approach of using edge density in an image is an alternative to the use of flow based approach to extract looming which is sensitive to noise.

2 Multiscale Image Analysis: An Overview

A gray level image is a physical entity which is a 2D representation of the 3D scene to which it belongs to. The perceived texture details of the 3D scene mainly depend upon the camera parameters such as zoom, focus, aperture, spatial sampling as well as the distance between the camera and the scene. These parameters collectively represent the two dimensions of the image, namely the *inner* scale and the *outer* scale. The inner scale of an image corresponds to the pixel size converted to the scene dimensions and the outer scale corresponds to the finite size of the image [27].

A multiresolution representation facilitates a simple hierarchical framework for the interpretation of information content of images [27]. At different image inner scales, entities in the image correspond to various entities in the scene. In other words, at coarse resolutions fine details are suppressed. Multiscale image analysis deals with analysis of image entities at various scales. It plays an important role in the analysis of information content in images at various resolutions. The relevant areas of multiscale or multiresolution image analysis include (refer to Figure (2)): 1. Quadtree Approach, 2. Pyramid Representation, 3. Wavelet Representation, 4. Scale Space Filtering.

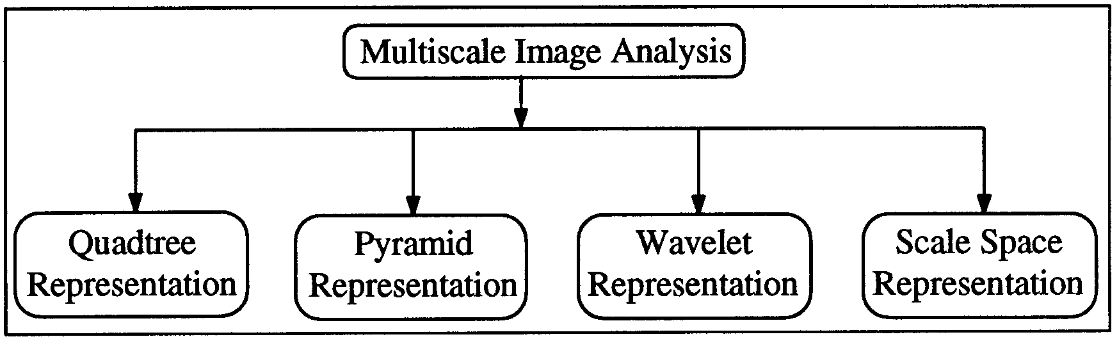


Figure (2): Various Multiresolution Image Analysis Approaches

Early research in this area was reported in [28] in the context of edge detection. The classification of images into edges is a non-trivial problem. Rosenfeld, *et al.* [28] suggested a straight forward combination of outputs of operations, that detect edges of different sizes, it is possible to obtain an output that retains the conspicuous edges in the scene.

Quadtree representation is a multiscale approach introduced by Klinger [29]. In this approach an image is recursively split into smaller regions until certain criteria are met. These criteria could be any function of the image gray level intensity. For example, gray level variance in a window be less than a certain threshold. Quadtree approach has been employed in region splitting and image segmentation algorithms (see for example [30, 31]).

A commonly employed multiscale representation of images is the pyramid approach introduced by Burt [46] and Crowley [47]. This approach facilitates computation of the image details at various resolutions. The concept of pyramids is based on a combination of sub-sampling and smoothing operations. In other words a pyramidal representation of an image is a stack of 2D arrays of exponentially decreasing sizes, for example 2^nx2^n, 2^{n-1}x2^{n-1}, ..., 2x2, 1x1.

Wavelet representation is a multiscale approach based on a family of basis functions (see for example [32]). A wavelet is a two-parameter family of translated and dilated functions [32, 33]. The basic function from which the family of curves is derived, is known as the mother wavelet and has to satisfy certain admissibility conditions [32, 33].

Scale space representation is a multiscale approach employed by many researchers in the recent past (see for example [34-39]). The concept of scale space filtering introduced by Witkin [40] and further developed by other researchers (see for example [41-45]) provides a concrete way to analyze the details in image at various scales. The scale parameter in their approach is a continuous one as opposed to the discrete scales employed in pyramid representations [46-48].

3 Motivation for Using Scale Space Representation

The scale space representation is usually employed to represent intrinsic physical entities (for instance image gray level intensity)that are functions of space, time as well as resolution [27]. We are primarily interested in the changes in the image inner scale. If the camera parameters remain unchanged, then the relative motion between the observer and the 3D scene results in changes in the image dimensions, namely the inner and outer scales. In other words, as the distance between the scene and the observer decreases, the inner scale (i.e., the pixel size converted to the scene dimension) as well as the outer scale decreases. It is the variations of the inner scale that are responsible for variations in the perceived texture details.

Since the scale space representation is based on a precise definition of *causality*, and deals with a *continuous* scale parameter (image inner scale), we selected it as a tool to study the variations of image details at various inner scales. Also according to [42], the scale space operators closely resemble the receptive field profiles in the front end visual systems of mammalians.

4 Scale Space Filtering: An Overview

The theory of scale space filtering is based on a precise definition of causality (see for example [27, 40-45]), namely: *no spurious detail should be generated with an increasing scale.* The scale space concept is used where this causality condition is satisfied.

This section is organized as follows: in sub-section 4.1 an overview of scale space representation is presented, followed by a relation the scale space images and the variations in the scale parameter in subsection 4.2, sub-section 4.3 presents the relation between the range and the image inner scale parameter.

4.1 Scale Space Representation: An overview

Scale Space filtering allows one to generate a family of derived signals from a given original signal by successively removing details when moving from fine to coarse scales. In other words, given a 2D continuous signal f(x,y), one can obtain a set of smoother versions of f(x,y) by convolving the image with some smoothing filter [40]. Mathematically this operation can be expressed as follows [41]:

$$L(x, y; \sigma) = K(x, y; \sigma) * f(x, y) \tag{1}$$

where $L(x,y;\ \sigma)$ = smoothed version of $f(x,y)$,

σ = smoothing factor or scale parameter,

$f(x,y)$ = original image,

* denotes convolution,

$K(x,y;\ \sigma)$ = smoothing kernel and

x, y are the spatial coordinates of the image.

Among several possible smoothing kernels, Gaussian kernel has been shown to be the unique kernel that satisfies the causality condition for scale space filtering (see for example [41-45]). If $K(x,y;\ \sigma)$ in Equation (1) is replaced by the Gaussian Kernel $G(x,y;\ \sigma)$, Equation (1) can be rewritten as follows:

$$L(x,y;\ \sigma) = G(x,y;\ \sigma) * f(x,y) \tag{2}$$

where

$$G(x, y; \sigma) = \frac{1}{2\pi\,\sigma^2} \exp(-\frac{x^2 + y^2}{2\sigma^2}),$$

and σ = scale parameter, the standard deviation of the smoothing kernel. Note that $\underset{\sigma \to 0}{Lim}\, L(x, y; \sigma) = f(x, y)$, which means that when the scale is zero (i.e., the Gaussian becomes a delta function) one obtains the original image.

This derived one parameter family of images namely *L(x, y; σ)* of the original image *f(x, y)* is known as scale-space images [38]. In the following section we present a relation between the variations in scale-space images and the corresponding scale variations.

4.2 Scale Space Images and Temporal Variations in Image Scale

Given an image sequence L_i that represents the scale space images of an original image *f(x,y)*, with absolutely no information about the scale parameter, it can be shown that the scale space images provide an indication of the scale variations. In this section we derive a relation between the relative variations in the image scale and the scale space images.

Consider an image and its scale space its representation. Let the image be denoted as f(x,y) and its scale space images being denoted as L_i, $i = 1, 2, 3, ...$, where:

$$L_1 = L(x, y; \sigma_1) = G(x, y; \sigma_1) * f(x, y)$$

$$L_2 = L(x,y;\sigma_2) = G(x,y;\sigma_2) * f(x,y)$$

$$L_i = L(x, y; \sigma_i) = G(x, y; \sigma_i) * f(x, y)$$

A relation between the scale space images and the corresponding temporal variations in scale (employing Gaussian Kernel) can be derived as follows (see Appendix A):

$$\frac{\frac{d^2}{dt^2}(M)}{\frac{d}{dt}(M)} = \frac{\frac{d^2}{dt^2}(\sigma^2)}{\frac{d}{dt}(\sigma^2)} \qquad (3)$$

where M = natural logarithm of the Fourier Transform of the image, i.e., ln$(F\{L_i\})$, d(.)/dt is the differentiation of (.) with respect to time, and σ is the scale parameter.

Equation (3) presents a relation between temporal variations of measurable image entities denoted as M and a non-measurable scale parameters denoted as σ. In other words, based on the variations in the measurable image entities it is possible to infer the variations in image scale parameters without measuring the scales. For example, small variations in M correspond to small variations in the scale parameter. Equation (3) is an important relation as it provides a connection between measurable and non measurable image quantities. An important observation is that the left hand side of Equation (3) is independent of the frequency components which is due to the linear shift invariant property of the Gaussian Kernel.

4.3 Relation between Range and Image Scale

Scale Space theory can be employed as a tool when the details in an image disappear with an increase in the scale factor. In other words, scale space filtering can be used as a tool for the analysis of variations in image details whenever the causality condition is satisfied. One such situation is described as follows: Consider the case of a camera that is initially focused to a 3D surface at a very short distance. With this fixed focus setting, as the camera moves away from the surface fixating at approximately the same point on the 3D surface, a sequence of images of the same 3D scene is obtained. In this image sequence the perceived texture details get blurred and eventually disappear as the distance between the surface and the camera increases. In other words, the sequence of images obtained by such a system for ranges greater than the range to which it is focused to initially, is analogous to the scale space images (see Figure (3), since we restrict ourselves to regions $R > R_0$, we do not show regions $R < R_0$). The original image *f(x,y)* in Equation (2) corresponds to the image of the scene in perfect focus and the image sequence $L(x,y,\sigma)$ corresponds to the blurred images. The details in $L(x,y,\sigma)$ get smeared as σ increases which indicates that causality condition is satisfied. Hence for ranges greater than the distance to which the camera is focused to initially, the scale space representation could be appropriate for the image sequence obtained by a fixed-focus visually fixating observer in motion.

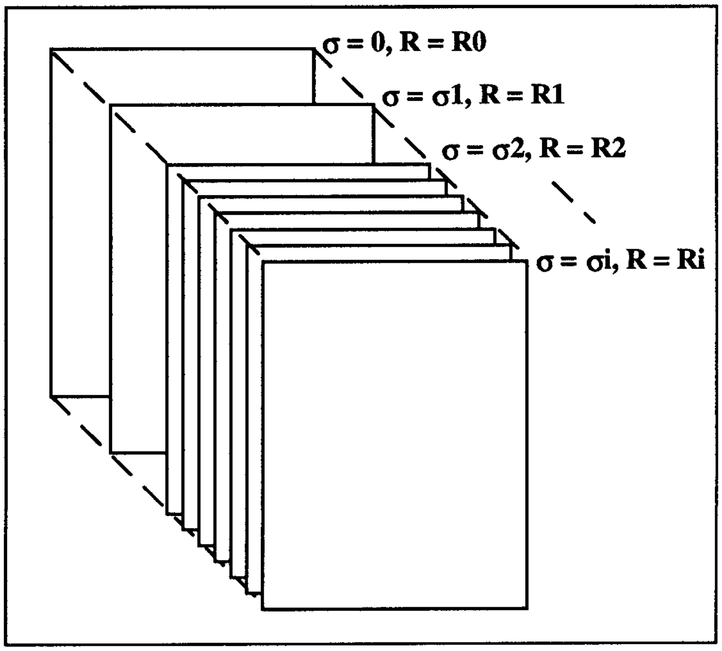


Figure 3: Scale Space Images, $0<\sigma_1< \sigma_2< \sigma_3<...< \sigma_i$, $R_0<R_1<R_2<...<R_i$, where R_0 is the distance to which the camera is focused to initially and R_i is the distance between the camera and the surface and σ_i being the corresponding scale (We restrict $R > R_0$)

A blurred image acquired by a camera can be viewed as the result of convolving a focused image with the Point Spread Function (PSF) of the camera, assuming the camera to be a linear shift invariant system [49]. The PSF of a convex lens is approximated as a 2D Gaussian function (see for example [24,26,49]). Since the Gaussian kernel of the scale space representation is similar to the PSF of the lens, the standard deviation γ of the Gaussian PSF can be seen as analogous to the image inner scale σ of the scale-space kernel and can be written as follows:

$$\sigma = k_1 \gamma \tag{4}$$

where σ = image inner scale, γ = standard deviation of the PSF and k_1 is some positive constant.

The standard deviation of the Gaussian PSF is proportional to the radius of the blur circle (see for example [26, 49]). In other words,

$$\gamma = k_2 \alpha \tag{5}$$

where γ = standard deviation of the PSF, α = radius of the blur circle.

For a fixed focus camera, the relation between the range between a point of visual fixation on a 3D surface and the observer and the radius of the blur circle can be written as follows (refer to Appendix (E)):

$$\alpha = k_3 \left(\frac{1}{R_0} - \frac{1}{R} \right) \tag{6}$$

where α = radius of the blur circle, k_3 = positive constant, R_0 = distance to which the camera is focused to initially, and R is the range between the fixation point on a 3D surface and the observer.

Combining Equations (4-6) the following relation between the image inner scale and the range between the fixation point on a 3D surface and the observer can be derived:

$$\sigma = k_1 k_2 k_3 \left(\frac{1}{R_0} - \frac{1}{R} \right) \tag{7}$$

Since we are *not interested* in the absolute *range*, there is no need to know the proportionality constants namely k_1, k_2, k_3.

The radius of the blur circle α is different for objects in the scene at different distances from the observer. Since the analysis is done for a small portion of the scene (during fixation) we assume that the blur circle for all the elements in the portion of the image near the fixation point is similar [49].

The analogy between the scale space images and the image sequence obtained by a fixed-focus visually fixated camera in motion (for ranges greater than the range to which the camera is focused to initially) can be summarized as follows:

1. The scale space representation is based on causality principle. In otherwords no spurious details must appear as the scale increases.

In the case of defocused image sequence no details appear as the range between the observer and the textured surface increases (see Figure (12)). In otherwords the causality condition is satisfied.

2. The scale space images and the defocused images are the result of convolving the original image with a 2D Gaussian filter of varying standard deviations. In scale space representation the standard deviation of the Gaussian filter is usually referred to as the *scale.* The image inner scale can be seen as analogous to the standard deviation of the Gaussian kernel.

3. The image inner scale is zero for the original image in the scale space images.

The standard deviation of the Gaussian PSF is zero when the image is in perfect focus.

4.4 Scale Space Images and Variations in Range

In Equation (3) a relation between the variations in scale and the corresponding scale-space images is presented and in Equation (7) a relation between the scale parameter σ and the range R between the observer and a fixation point is presented. Under the assumption that the changes in the image outer scale are small for any three consecutive frames of a given sequence, and combining Equations [3-7] the following relation between range and image inner scale can be derived (see Appendix B):

$$\frac{\frac{d^2}{dt^2}(\sigma)}{\frac{d}{dt}(\sigma)} + \frac{\frac{d}{dt}(\sigma)}{\sigma} = \frac{\frac{d^2}{dt^2}(R)}{\frac{d}{dt}(R)} - 2\frac{\frac{d}{dt}(R)}{R} + \frac{R_0}{(R-R_0)}\frac{\frac{d}{dt}(R)}{R} \qquad (8)$$

$$\frac{\frac{d^2}{dt^2}(M)}{\frac{d}{dt}(M)} = \frac{\frac{d^2}{dt^2}(R)}{\frac{d}{dt}(R)} - 2\frac{\frac{d}{dt}(R)}{R} + \frac{R_0}{(R-R_0)}\frac{\frac{d}{dt}(R)}{R} \qquad (9)$$

Equation (9) presents a relation between the temporal variations of measurable image entities namely the natural logarithm of the magnitude of the fourier transform of the image (denoted as *M*), and the temporal variations in the ranges denoted as *R*. Equations (9) is independent of the frequency components and is true for any given frequency (due to the linear shift invariant property of the Gaussian Kernel). The entity on the right hand side of Equation (9) represents a visual motion cue, we call the Hybrid Visual Threat Cue. It is a combination of Time-to-Contact [11], Looming [9] and the Visual Threat Cue [50] (see Appendix F). In later sections a detailed analysis of this cue and how it can be used for autonomous navigation is presented.

5 The Hybrid Visual Threat Cue (HVTC)

5.1 Definition

Following Equation (9) mathematically the Hybrid Visual Threat Cue (HVTC) is defined as follows (for $R > R_0$):

$$\text{HVTC} = \frac{\frac{d^2}{dt^2}(R)}{\frac{d}{dt}(R)} - 2\frac{\frac{d}{dt}(R)}{R} + \frac{R_0}{(R-R_0)}\frac{\frac{d}{dt}(R)}{R} \qquad (10)$$

Where R is the range between the observer and a point of visual fixation on the 3D surface, d(.)/dt is the differentiation of (.) with respect to time and R_0 is the distance to which the camera is focused to initially and has the same units as R. Note that the units of the HVTC are [time^{-1}]. The HVTC is dependent only on the observer's translational velocity component but is independent of relative rotation. The HVTC is a combination of the Time-to-Contact [11], the Looming [9] and the VTC [50] (see Appendix F).

5.2 ISO HVTC Surfaces

In this section we present simulation results to show the location of points, beyond the desired minimum clearance R_0 in 3D space, around the moving observer, that have the same value of the HVTC, for a given motion of the camera. The HVTC corresponds to a visual field surrounding the moving observer, i.e., there are imaginary 3D surfaces attached to the observer that are moving with it. For a given value of the HVTC there is a corresponding imaginary surface around the observer in motion.

Since the HVTC is a linear combination of the following variables:

$$1.\ \frac{\frac{d^2}{dt^2}(R)}{\frac{d}{dt}(R)}, \quad 2.\ -2\frac{\frac{d}{dt}(R)}{R}, \qquad 3.\ \frac{R_0}{(R-R_0)}\frac{\frac{d}{dt}(R)}{R}$$

we present the individual visual fields associated with each of the above mentioned entities.

5.2.1 Iso $\dfrac{\dfrac{d^2}{dt^2}(R)}{\dfrac{d}{dt}(R)}$ Surfaces

There are imaginary surfaces attached to an observer in motion and all the points that lie on a particular surface produce the same value of the $\dfrac{\dfrac{d^2}{dt^2}(R)}{\dfrac{d}{dt}(R)}$, The points that lie in front of the observer produce a negative value and the points that lie in back of the observer produce a positive value of the entity. Points that lie on a relatively closer surface produce a relatively higher value compared to those lying on a farther surface. A qualitative plot of a cross section of the iso $\dfrac{\dfrac{d^2}{dt^2}(R)}{\dfrac{d}{dt}(R)}$ contours is shown in Figure (4).

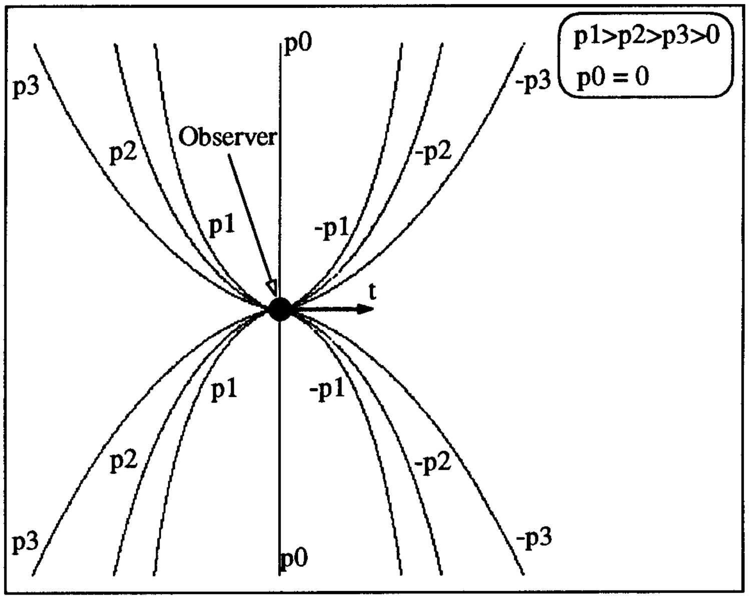


Figure (4): Qualitative Cross Section of the Iso $\dfrac{\dfrac{d^2}{dt^2}(R)}{\dfrac{d}{dt}(R)}$ surfaces

5.2.2 Iso $-2\frac{\frac{d}{dt}(R)}{R}$ Surfaces

The entity $-\frac{\frac{d}{dt}(R)}{R}$ had been defined as looming [9]. The visual field associated with looming had been shown to be a system of spheres whose centers are located as shown in Figure (5) [9]. The points that lie on a surface in front of the observer produce positive value and points in back produce negative values. A qualitative plot of the cross sectional view of the Iso contours is shown in Figure (5).

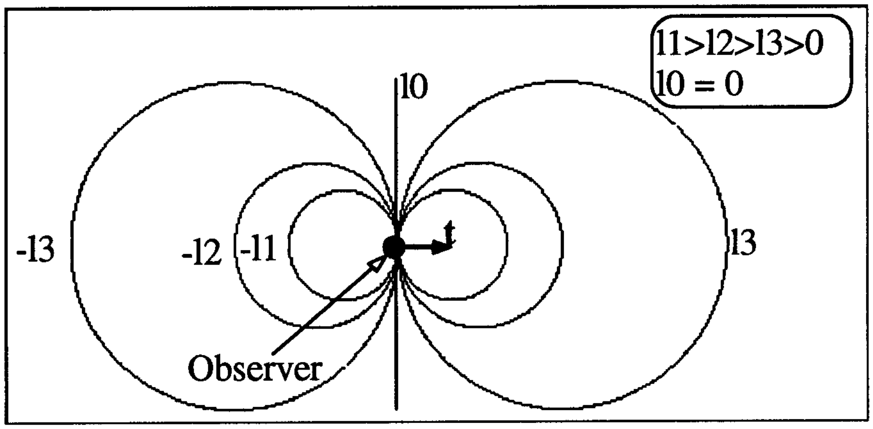


Figure (5): Qualitative cross sectional view of Iso $-2\frac{\frac{d}{dt}(R)}{R}$ surfaces, **t** is the instantaneous Velocity Vector

5.2.3 Iso $-\frac{\mathbf{R_0}}{(\mathbf{R}-\mathbf{R_0})}\frac{\frac{\mathbf{d}}{\mathbf{dt}}(\mathbf{R})}{\mathbf{R}}$ Surfaces

The entity $-\frac{R_0}{(R-R_0)}\frac{\frac{d}{dt}(R)}{R}$ had been defined to be the Visual Threat Cue (VTC) [50].

A cross section of the visual field associated with the VTC is shown in Figure (6).

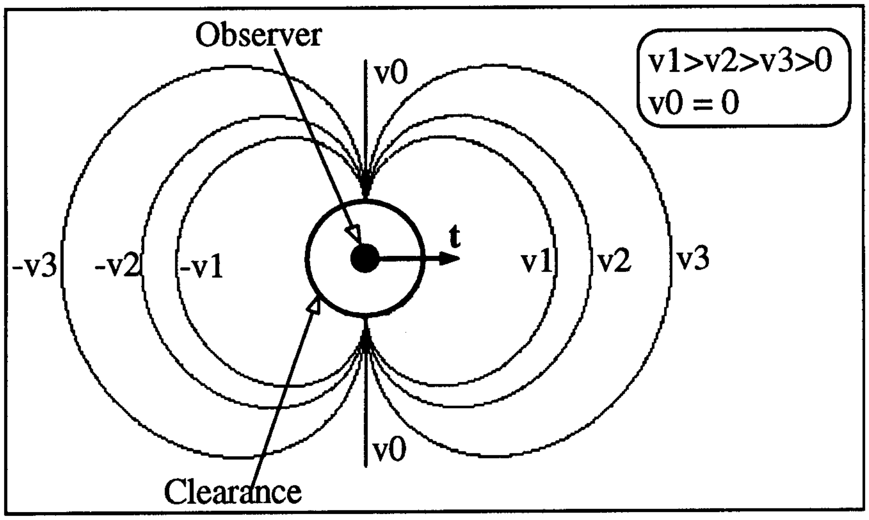


Figure (6): Qualitative Cross sectional view of the Iso VTC surfaces, **t** is the instantaneous Velocity Vector

Even though the HVTC is a linear combination of the TTC, the Looming and the VTC, the visual field associated with it is not as simple as the individual fields associated with the TTC, the looming or the VTC (see Figures [2-6]). In the following sub-section the visual field associated with the HVTC is described.

5.2.4 Iso HVTC Contours

In this section we provide simulation results to show the location of points in 3D space around an observer in motion, that have same value of the HVTC (see Equation (10)) for a given motion of the observer. The HVTC corresponds to a visual field surrounding the moving observer, i.e., there are imaginary 3D surfaces attached to the observer that are moving with it, each of which corresponds to a value of the HVTC.

There is one region in front of the observer and one region in back of the observer that produce positive values of the HVTC and also there is one region in front of the observer and also the back of the observer that produce negative values of the HVTC. In other words for the region in front of the observer there are two sub-regions one corresponding to positive values of the HVTC denoted as FP (Front Positive) and the other corresponding to negative values of the HVTC denoted as FN (Front Negative) (see Figures (7a) and (7b)).

Similarly for the region in back of the observer there are two sub-regions, one corresponding to positive values of the HVTC denoted as BP (Back Positive) and the other region corresponding to negative values of the HVTC denoted as Back Negative (BN) (see Figures (7a) and (7b)).

All the points that lie on a particular surface in the FP region produce the same value of the HVTC. The points that lie on a relatively smaller surface produce a relatively greater value of the HVTC (see Figure (7c)). There is a point in the FP region (denoted as S in Figure (7d)) on the instantaneous translational velocity t, where the HVTC in the FP is the maximum.

Similarly all the points that lie on a particular surface in the BP region produce the same value of the HVTC. Points that lie on a relatively smaller surface produce a relatively greater value of the HVTC (see Figure (7c)). There is a point in the BP region (denoted as S' in Figure (7d)) on the instantaneous translational velocity t, where the HVTC is minimum in the BP region.

The point on the instantaneous translational vector **t**, where the HVTC is the maximum in the FP region lies on the instantaneous translational vector t, at a distance of $2.35R_0$, where R_0 is the desired clearance.

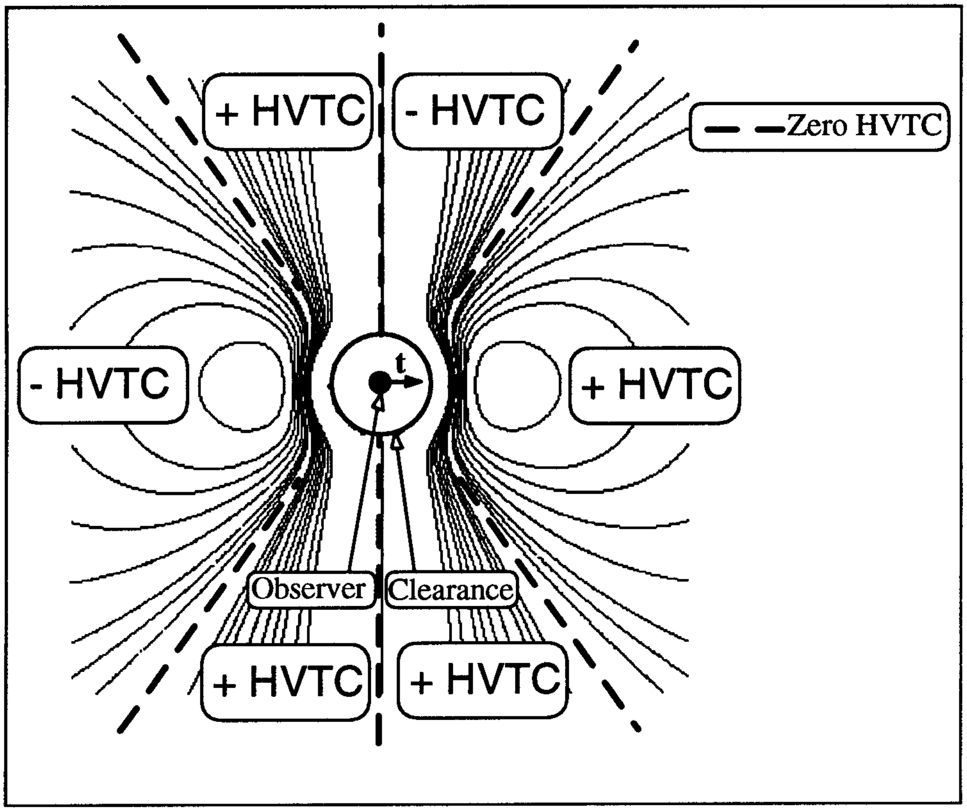


Figure (7a): Qualitative Iso HVTC Surfaces, **t** is the instantaneous Velocity Vector

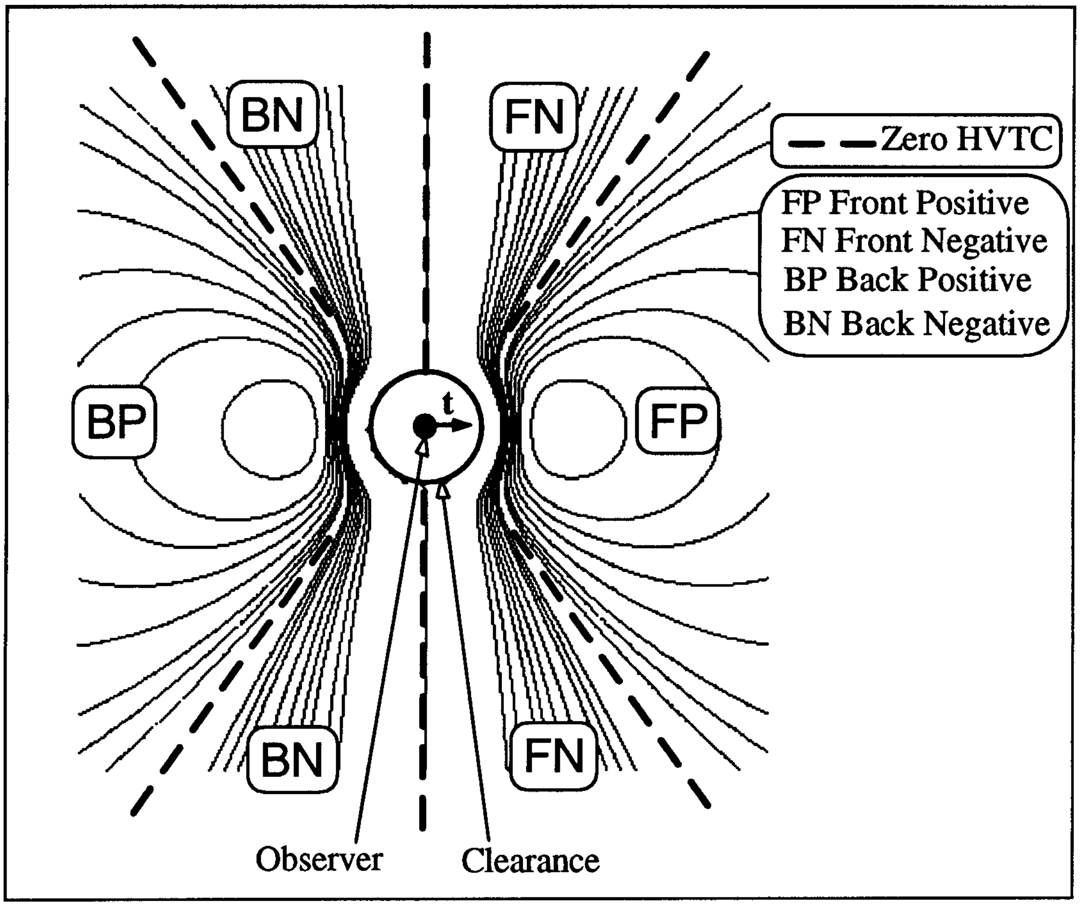


Figure (7b): Qualitative Iso HVTC Surfaces, **t** is the instantaneous Velocity Vector

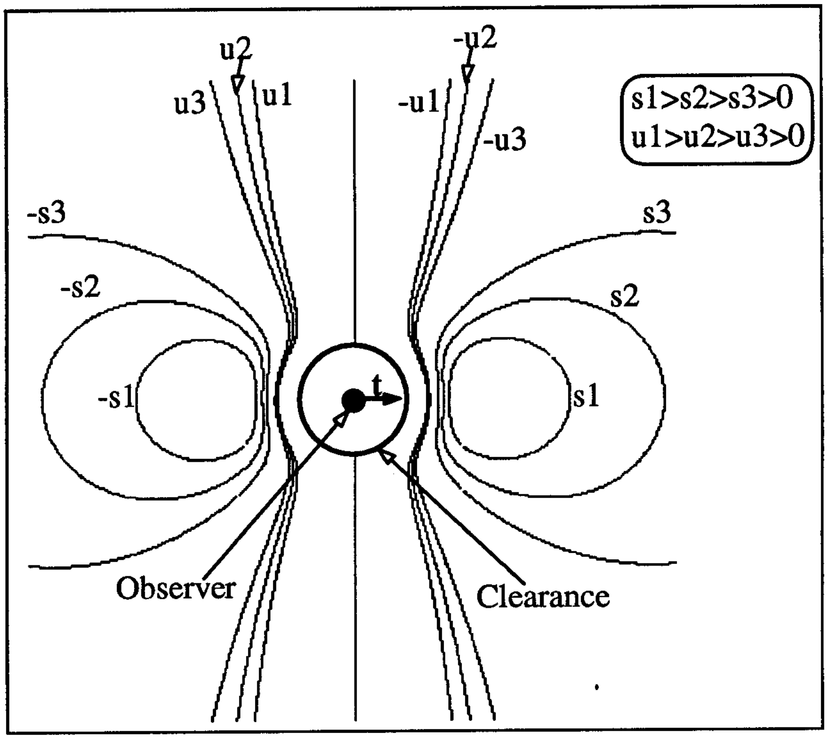


Figure (7c): Qualitative Iso HVTC contours

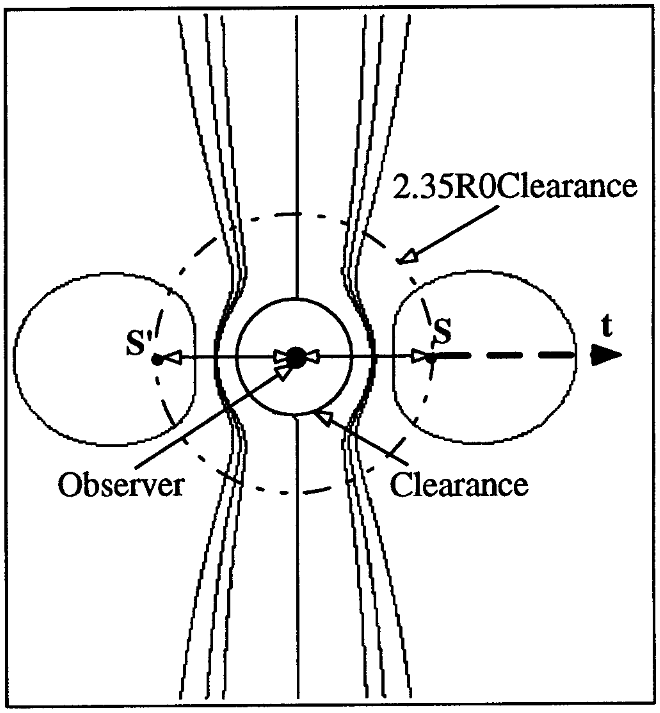


Figure (7d): Qualitative Iso HVTC Surfaces, **t** is the instantaneous Velocity Vector

5.3 The HVTC as a Sensory Feedback Signal

The HVTC divides the 3D space around the observer in motion into four 3D sub-regions as shown in Figures (7a-7d). For local navigation decisions, i.e., to obtain information about obstacles in the observer's surrounding, the HVTC information alone is not sufficient. For instance consider an obstacle in the FP region, it produces a positive value of the HVTC. Without any a-priori knowledge about the observer's motion or the obstacle's location in the surrounding it becomes very difficult to judge the location of the obstacle's location as obstacles in FP as well as BP regions produce positive values. This difficulty can be overcome using the VTC (which measurable from images [50]) and its temporal derivative [60].

5.3.1 Iso VTC Field

The visual field associated with the VTC is shown in Figure (6). The VTC divides the space around an observer in motion into two different regions namely one region in front where the VTC is positive and one region in the back of the observer where the VTC is negative (see Figure 8a). The problem in employing the VTC information alone for navigation tasks is explained as follows (refer to Figure (8b)): The points 1,2, ...,7 lie on the same VTC surface, hence produce the same value of the VTC. But for navigation purposes, point 4 poses the maximum threat as it lies on the instantaneous translational velocity vector. Points 3 and 5 pose a relatively high threat as they are closer to the instantaneous translational vector. Points 1,2, 6 and 7 pose low threat. Using the VTC alone (without any information about the heading vector) it is not possible to distinguish whether the point is closer to the instantaneous velocity or far from it. However, this problem can be overcome by using the temporal variations of the VTC (TVTC) which is described in the following section. Note that the VTC is measurable (as will be shown in later sections).

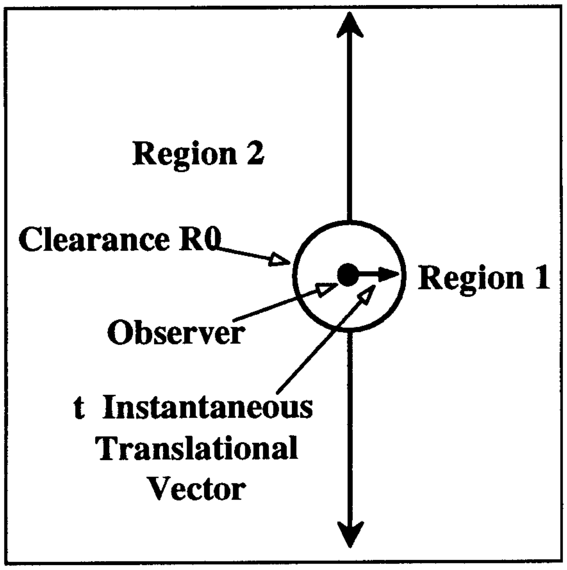


Figure (8a) : The VTC and the space around an observer in motion

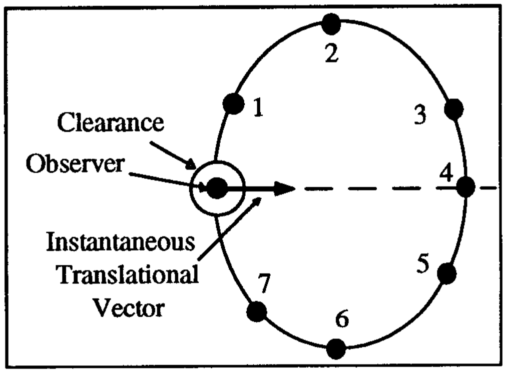


Figure (8b): Qualitative VTC

5.3.2 Iso TVTC Field

The Temporal VTC [60] corresponds to a visual field surrounding the moving observer, i.e., there are imaginary 3D surfaces attached to the observer that are moving with it, each of which corresponds to a value of the TVTC. The points that lie on a relatively smaller surface corresponds to a relatively larger value of the VTC, indicating a relatively higher threat of collision. The VTC value on the minimum clearance hemi-sphere of radius R_0 centered at the location of the observer is the maximum which is infinity, indicating that the absolute distance between the observer and the camera is the minimum clearance. Note that this field is symmetric about the instantaneous translational vector **t**. The visual field associated with the TVTC is shown in Figure (9a). There are regions in front and in the back of the observer that produce a positive values as well as negative values of the cue as shown in the Figure (3). It has been shown that for $R >> R_0$, the angle between the direction of motion and the zero TVTC is about 54.74° [60].

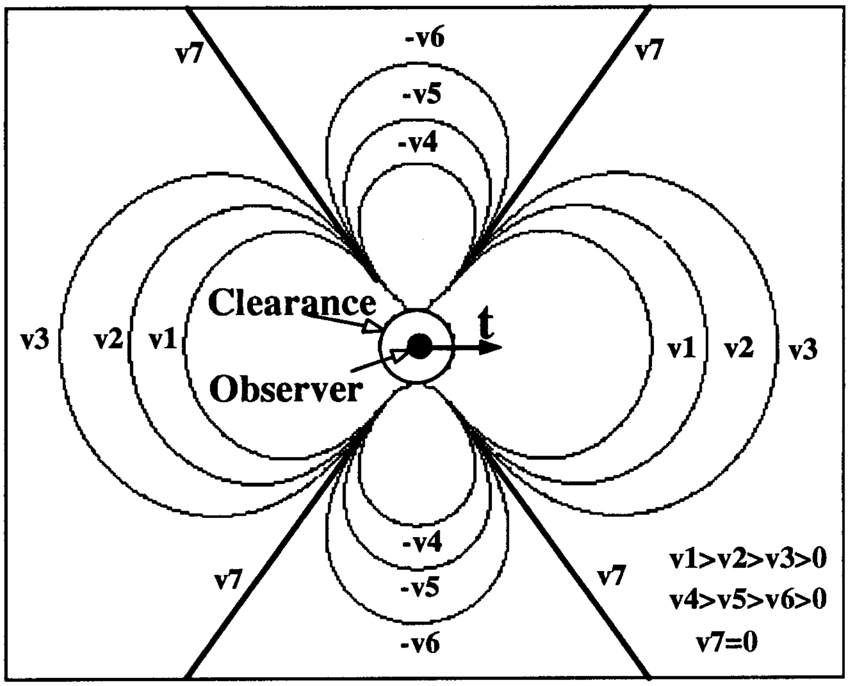


Figure (9a): Cross section of the TVTC, **t** is the instantaneous translational vector

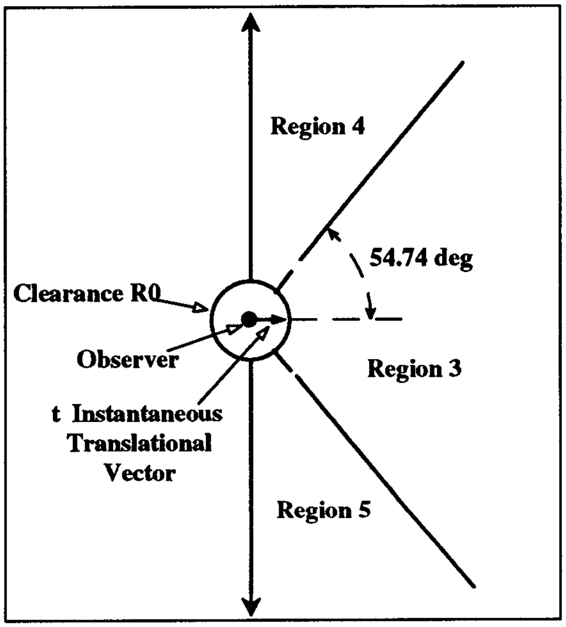


Figure (9b) : TVTC and the space around a moving observer, **t** is the instantaneous translational vector

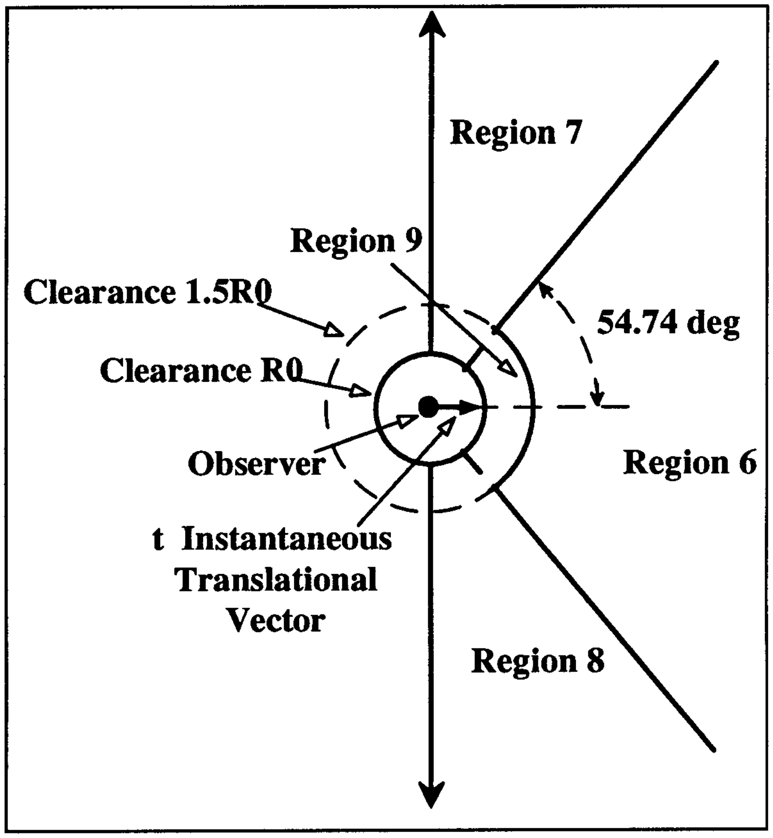


Figure (10): The HVTC and Space around a moving observer, **t** is the instantaneous translational vector

The VTC, TVTC and the HVTC divide the region around an observer in motion into different regions as shown Figures (8-10). Based on these visual motion cues it is possible to demarcate region around an observer in motion into several regions. Using this information about the space around the observer it is possible to generate appropriate control commands to the autonomous observer to avoid collisions with obstacles.

Visual Cue	Positive Regions	Negative Regions
VTC	Region 1	Region 2
TVTC	Region 3	Region 4, Region 5
HVTC	Region 6	Regions 7-9

Table (1): Demarcation Table

6 Extraction of the HVTC

This section describes several possible approaches to extract the HVTC from a sequence of monocular images.

The HVTC can be extracted by measuring the radius of the blur circle and its temporal variations. Several researchers have suggested various approaches to extract the radius of the blur circle (σ) for 3D scene reconstruction tasks. These approaches usually involve the Fourier transform and some times may require special purpose hardware to extract to measure σ.

This section describes a practical approach to extract the VTCs from images. The approach is based on measuring a global image variable called the Image Quality Measurement (IQM) and the visual cues can be extracted from the relative temporal variations of the IQM.

Since the visual cues can be extracted in several ways, namely by measuring the radius of blur circle, by employing the variations in perceived texture details, etc., a brief overview of 3D surface reconstruction approaches from defocused images is presented.

6.1 Related work on 3D surface reconstruction using defocused images

Pentland [61] is one of the pioneers to investigate approaches for extracting depth information from defocused images. He proposed two approaches to extract depth information from defocused images using σ the radius of the radius of blur circle. One approach is based on measuring the slope of edges in blurred images (in focused images they correspond to a step discontinuity). The approach requires *a-priori* knowledge of the location and magnitude of the step edges in the focused images (which is difficult to obtain in real situations). He also suggested a second approach in which the same scene is viewed with two different aperture. Based on the focal gradient in the image due to varying aperture widths, he formulated an expression for σ in terms of the Fourier transforms of the images. A special purpose hardware is suggested to obtain two images of the same scene at two different width of apertures.

Subbarao [62] is another researcher who is actively involved in depth reconstruction approaches using defocused images. In [62] he described three approaches for 3D depth-map recovery. The approaches are based on variations in the image of a scene due to a small known variation in one of the three intrinsic camera parameters namely, distance between the lens and the image plane, focal length of the lens and the diameter of the lens aperture.

In order to extract the HVTC from a sequence of images using Equation (9), we need to extract the Fourier transform of the image window. Extraction of the Fourier transform for the image window of our choice turned out to be computationally intensive for a 486-based Personal Computer-based imaging system. However if hardware implementation of Fourier transforms are available, extraction of the HVTC from images is possible using Equation (9).

In this section we present an alternate practical way to extract the HVTC described in Equation (9) from a sequence of images, using temporal variations of the Image Quality Measure (IQM). It is extracted directly from the raw gray level data without measuring the Fourier transform or σ.

A practical and robust method to extract the VTC from a sequence of images of a 3D textured surface obtained by a *visually fixated, fixed-focus monocular* camera in motion has been presented in [16]. This approach is *independent* of the type of 3D surface texture and needs almost no camera calibration. For each image in such a 2D image sequence of a textured surface, a *global variable* (which is a measure for dissimilarity) called the Image Quality Measure (IQM) is obtained *directly* from the raw data of the *gray level images*. The VTC is obtained by calculating relative temporal changes in the IQM. This approach by which the VTC is extracted can be seen as a *sensory fusion* of focus, texture and motion at the *raw data level*. The algorithm to extract this cue works *better* on natural images including fractal-like images, where more details of the 3D scene are visible in the images as the range shrinks and also can be implemented in parallel hardware. In order to minimize the depth of field of the camera, the aperture is open wide (see Appendix H)).

6.2 Image Quality Measure (IQM)

Local spatial gray tone variations in an image give rise to a visual pattern in the image known as texture. These spatial gray level variations are due to the visual characteristics of the 3D scene being imaged, the illumination, the range between the scene and the observer, as well as due to camera parameters like zoom, aperture, resolution, focus, etc. When there is a relative motion between a textured surface and a visually fixated, fixed-focus moving observer, the perceived texture in the 2D image varies. For instance, consider the case of a camera that is initially focused to a 3D surface at a very short distance and gradually moves away from this surface. As a result, the perceived 2D image texture varies from one image to another, mainly due to focus, i.e., the image of the scene in perfect focus is very sharp and has many details, then as the camera moves away from the scene, fine details gradually get smeared and eventually disappear (see Figure 12). When the image is in perfect focus, the dissimilarity, i.e., spatial gray level variations is very high, and as the details get smeared the dissimilarity gets smaller and smaller. We describe an IQM to measure the dissimilarity of the image. Using the relative temporal variations in this IQM we extract the HVTC.

Among several possible approaches to describe the quality of texture in an image, we employed a measure, we call the Image Quality Measure (IQM) that is based on city block metric, to describe the dissimilarity of images [58,59]. The advantages of using this approach over the other approaches are:

1. It gives a *global* measure of quality of the image, i.e., *one* number which characterizes the image dissimilarity is obtained.

2. It does not need any preprocessing, i.e., it works directly on the raw gray level data without any spatial or temporal smoothing.

3. It does not need a model of the texture and is suitable for many textures.

4. It is simple and can be implemented in real time on parallel hardware.

Mathematically, the IQM is defined as follows [50](see Appendix F):

$$IQM = \frac{1}{|D|} \sum_{x=x_i}^{x_f} \sum_{y=y_i}^{y_f} \left(\sum_{p=-L_c}^{L_c} \sum_{q=-L_r}^{L_r} \left| I(x,y) - I(x+p,y+q) \right| \right) \tag{10}$$

where I(x,y) is the intensity at pixel (x,y) and x_i and x_f are the initial and final x-coordinates of the window respectively ; y_i and y_f are the initial and final y-coordinates of the window in the image respectively and Lc and Lr are positive integer constants; and D is a number defined as $D=(2L_c+1)\times(2L_r+1)\times(x_f-x_i)\times(y_f-y_i)$. One can see from Equation (10) the IQM is a measure for the dissimilarity of gray level intensity in the image. In our experiments we arbitrarily chose a window of size 50 x 50 pixels in the center of the image and Lc = 5 and Lr = 4.

6.2 Extraction of the HVTC from relative variations of the IQM

The IQM described in Equation (5) was applied to a set of 12 different textures from Brodatz's album [51]. The experimental details are provided in the following section. Based on our experimental results, we observed that the IQM mentioned above is almost a constant when the range between the surface and the camera is very large and it increases non linearly as the camera approached the distance to which it is focused to. We observed that the radius of the blur circle varies inversely with the IQM, i.e., when the texture details are sharp, IQM is very high and the radius of blur circle is almost zero, and vice-versa. Hence for ranges greater than the initial distance to which the camera is focused to, we modeled the radius of the blur circle in terms of IQM as follows:

$$IQM \propto \frac{1}{\sigma} \tag{11}$$

or:

$$IQM = \frac{\sigma_0}{\sigma} \tag{12}$$

where σ_0 is some proportionality constant and σ is the radius of the blur circle.

From Equation (7):

$$\frac{\frac{d^2}{dt^2}(\sigma)}{\frac{d}{dt}(\sigma)} + \frac{\frac{d}{dt}(\sigma)}{\sigma} = \frac{\frac{d^2}{dt^2}(IQM)}{\frac{d}{dt}(IQM)} - 3\frac{\frac{d}{dt}(IQM)}{IQM} \tag{13}$$

By combining Equation (4) with Equation (13), using $u_0 = R_0$, we obtain the following relation

$$HVTC = \frac{\frac{d^2}{dt^2}(IQM)}{\frac{d}{dt}(IQM)} - 3\frac{\frac{d}{dt}(IQM)}{IQM} \tag{14}$$

The HVTC obtained by using Equation (14) does not need knowledge about the camera parameters like the focal number f or the focal length F and is independent of the magnitude of the IQM.

7 Experimental Details

Several experiments were performed to study the variations in the IQM of image sequences in order to extract the HVTC. The system used in the experiments include a Coordinate Measuring Machine (CMM), a CCD video camera, a 486 based personal computer, ITEX PC-VISION

PLUS image processing system and several texture plates from Brodatz's album [51]. A block diagram of the connections is shown in Figure (10a).

7.1 Procedure

A CCD camera is attached to the CMM and the texture surface is placed in front of the camera as shown in the Figure (10b). The maximum distance between the surface and the camera is 900 mm and the minimum distance being 200 mm. The camera is focused to the closest possible distance which in the case of the camera used is 200 mm, i.e., texture details are sharp when the distance between the camera and the surface is 200 mm. The error in the initial setting is about 1 mm. Once this is set, the measurements in relative ranges (for obtaining the ground truth values) were as accurate as the CMM. With this focus setting, the distance between the camera and the surface is varied from 900 mm to 200 mm in steps of 10 mm.

The CCD camera attached to the CMM as shown in Figure (10a) and Figure (10b) captures the images of the texture. These images are then digitized by the PC-based image processor PC-VISION PLUS. These digitized images are then processed by a 486-based personal computer, to extract the IQM and the VTC. For a given texture, we computed these measures at 71 different distances and this was repeated for 12 different textures (shown in Figure (11)) from Brodatz's album [42]. Figure (12) shows an imaged texture (D18 from [42]) as a function of range for 20 different ranges. This set of images shows intuitively the evolution of details in the image as a function of range for a fixed focus camera. The experimental results along with the textures used are presented in the following section.

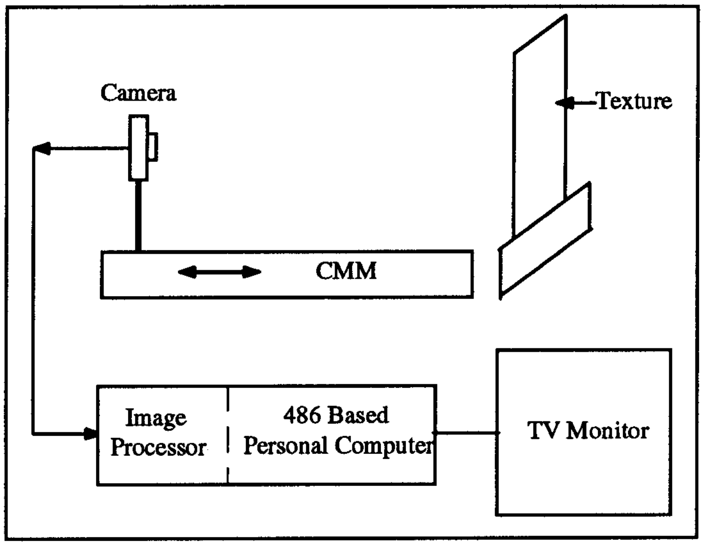


Figure (10a) : Block diagram of the Experimental setup

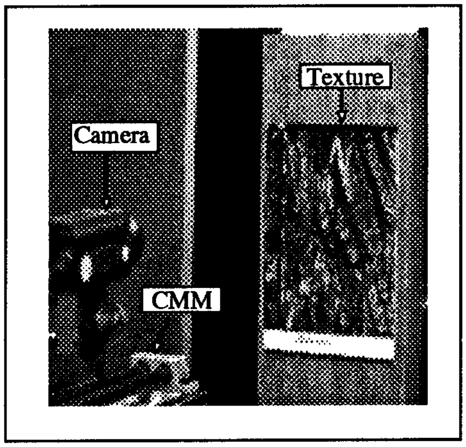
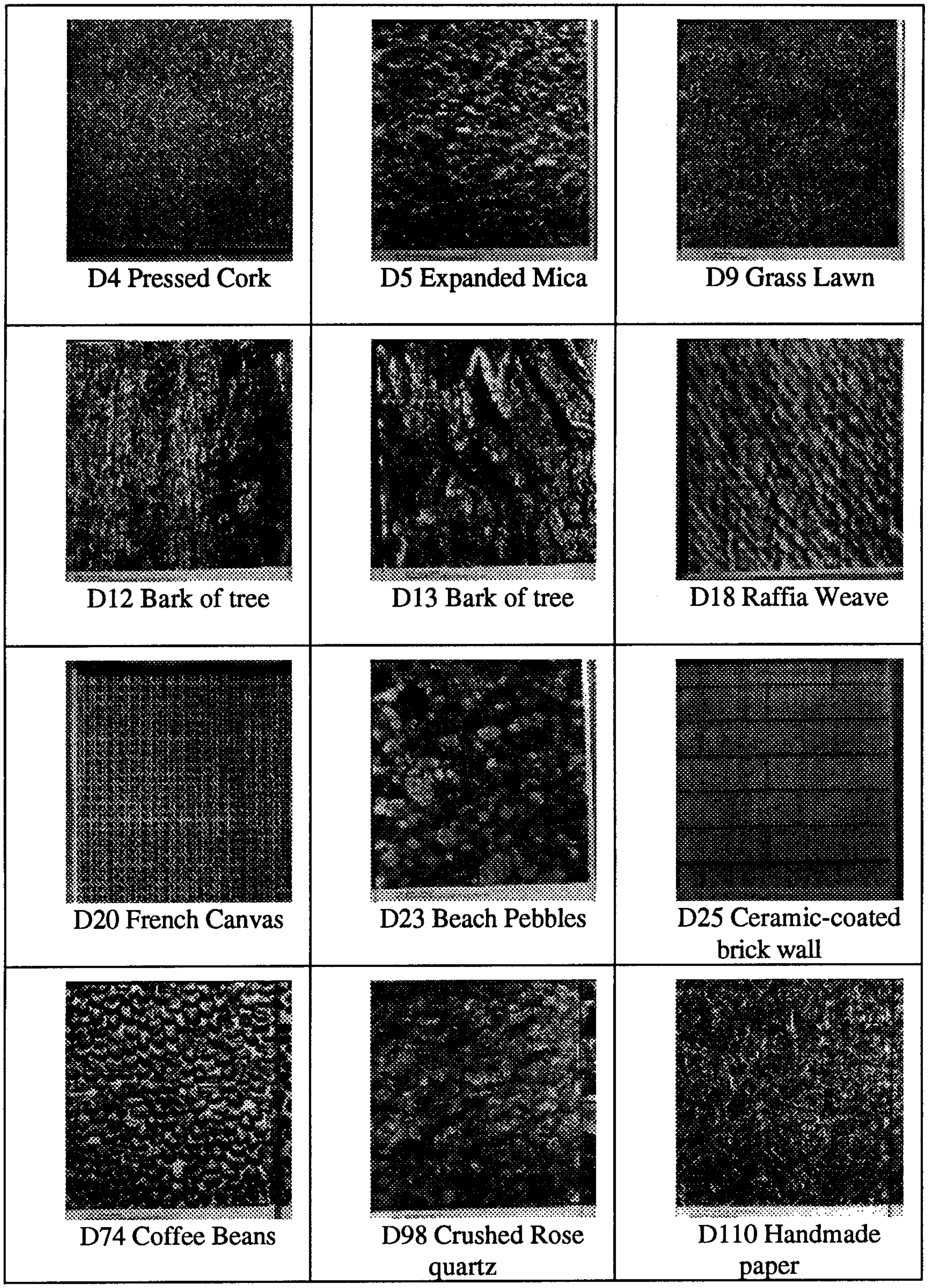


Figure (10b): CMM Set up



D4 Pressed Cork

D5 Expanded Mica

D9 Grass Lawn

D12 Bark of tree

D13 Bark of tree

D18 Raffia Weave

D20 French Canvas

D23 Beach Pebbles

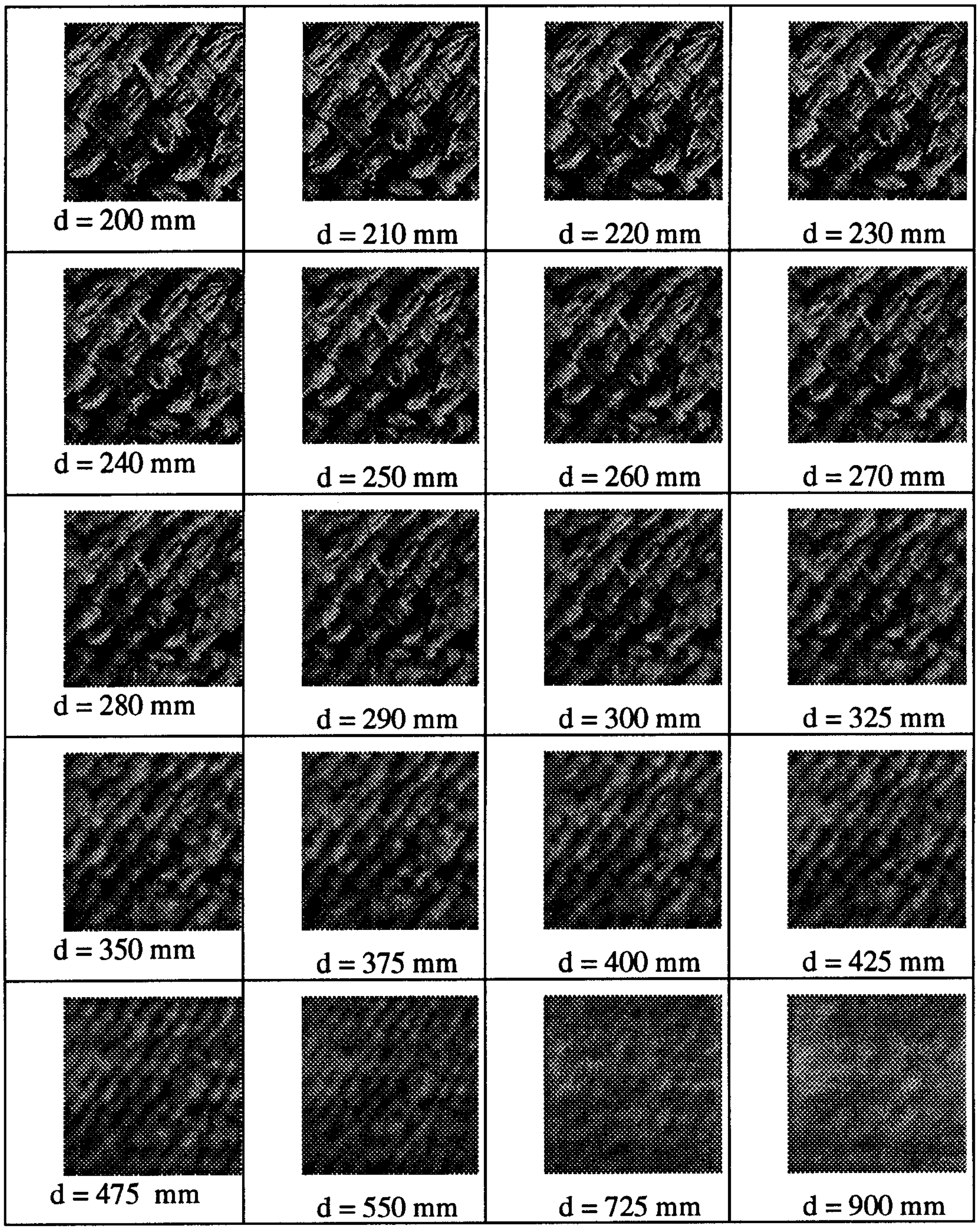
D25 Ceramic-coated brick wall

D74 Coffee Beans

D98 Crushed Rose quartz

D110 Handmade paper

Figure (11): Various texture patterns used in the experiments



d = 200 mm

d = 210 mm

d = 220 mm

d = 230 mm

d = 240 mm

d = 250 mm

d = 260 mm

d = 270 mm

d = 280 mm

d = 290 mm

d = 300 mm

d = 325 mm

d = 350 mm

d = 375 mm

d = 400 mm

d = 425 mm

d = 475 mm

d = 550 mm

d = 725 mm

d = 900 mm

Figure (12): Sequence of images depicting the evolution of details in the image a decrease in the relative range, d is the range between surface (D18) and the camera

7.2 Results and Analysis

The IQM described in Equation [10] is extracted according to the procedure described in section (7.1). Since the extraction of the HVTC is a nonlinear function of the second derivatives, its straight forward extraction using IQM becomes problematic. To overcome this problem, we employed a curve fitting strategy described in Appendix [G]. Using atleast six values of the measured IQM values in the past we fit a sixth order polynomial to the IQM values of the past to estimate the current IQM. Then using the estimated IQM values we compute the HVTC according to Equation [8].

The IQM described in Equation (10) is extracted according to the procedure described in the section 3.1, and the HVTC is extracted from the relative temporal variations of the IQM.

For each of the texture patterns employed, we present the following:

1. Five sample images (out of a total 71 images) relative ranges 200 mm, 280 mm, 400 mm, 550 mm, 900 mm (Figures 13(a)-24(a)).

2. The normalized measured IQM as function of the distance between the camera and the surface (It is normalized since the extraction of the HVTC is independent of the absolute magnitude of the IQM. Figures 13b-24b).

3. A plot depicting the theoretical HVTC and the HVTC extracted from the images (Figures 13c-24c).

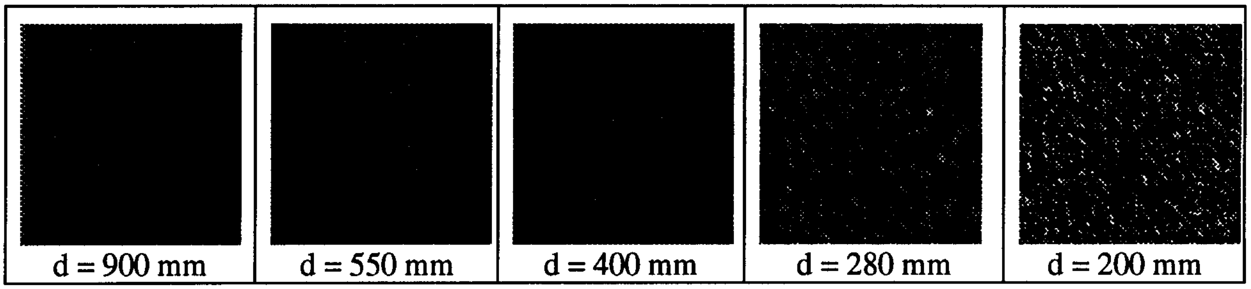


Figure (13a): Image Sequence for Texture D4, d is the relative distance

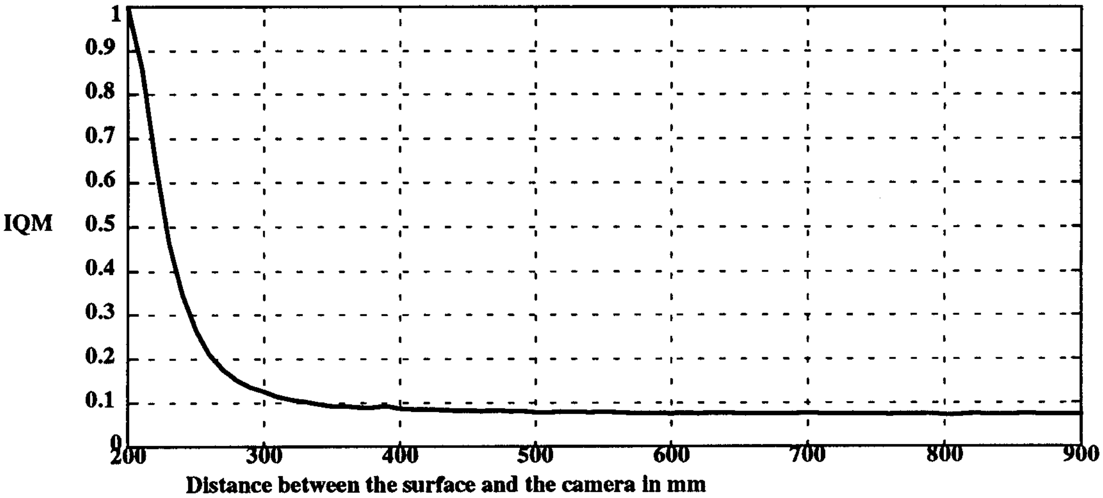


Figure (13b): Measured IQM vs. Distance between the camera and surface for D4

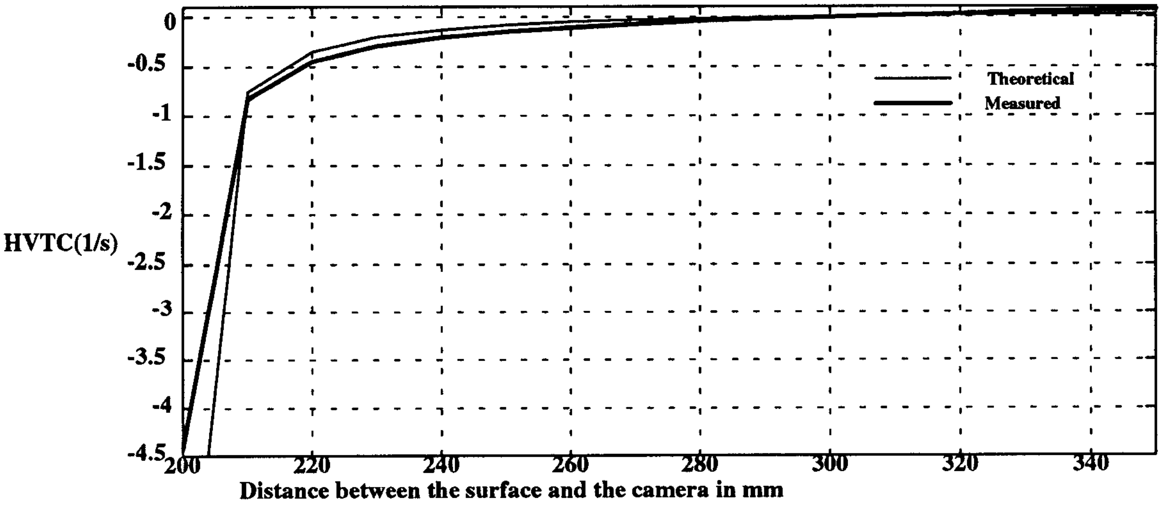


Figure (13c): HVTC vs. Distance between the camera and surface for D4

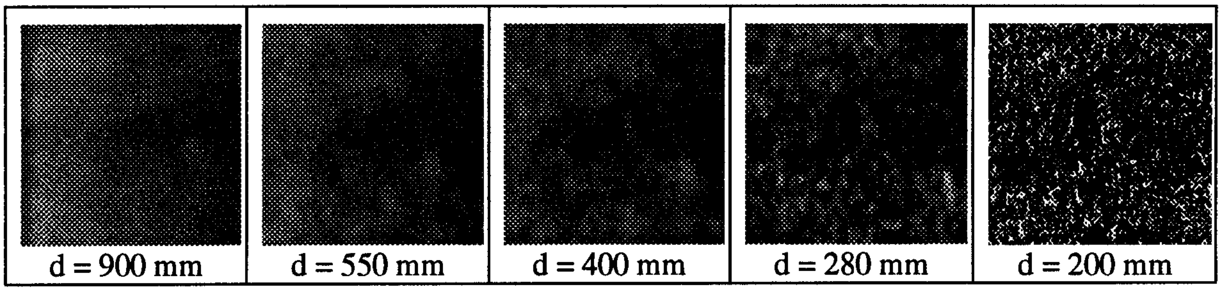


Figure (14a): Image sequence for Texture D9, d is the relative distance

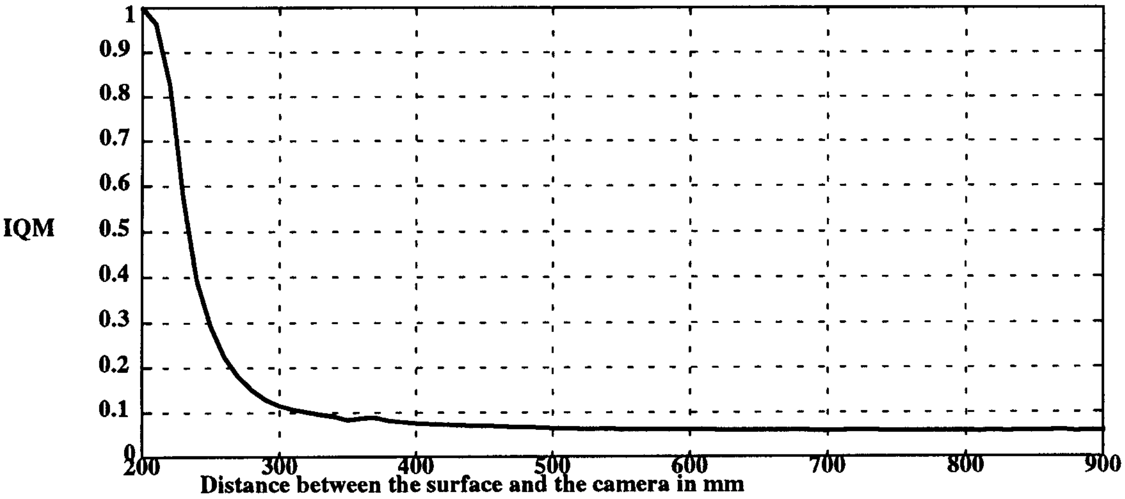


Figure (14b): Measured IQM vs. Distance between the camera and surface for D9

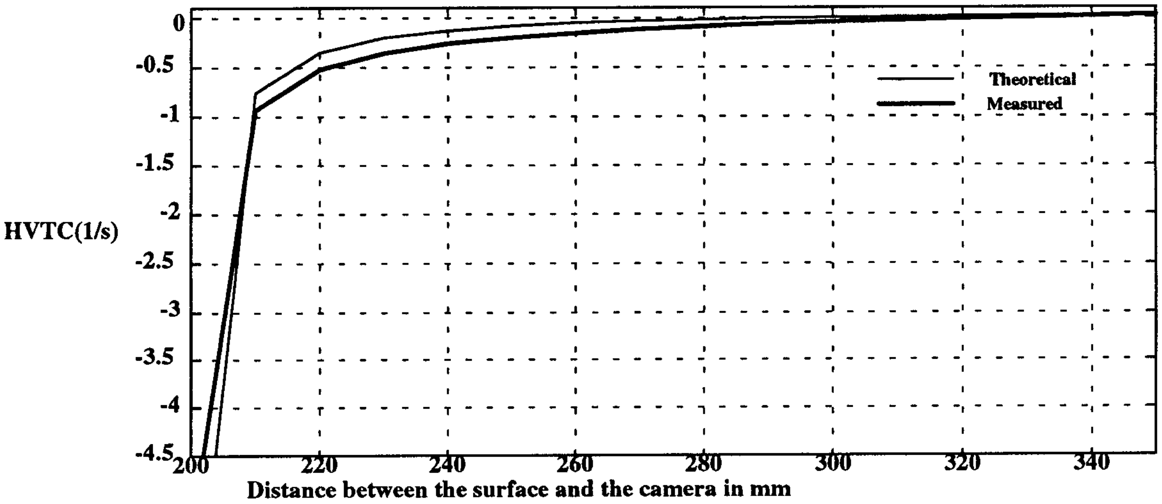


Figure (14c): HVTC vs. Distance between the camera and surface for D9

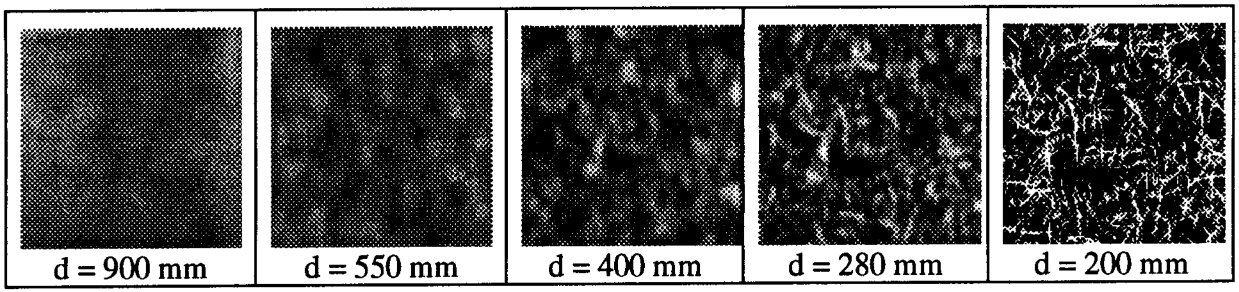


Figure (15a): Image sequence for texture D110, d is the relative distance

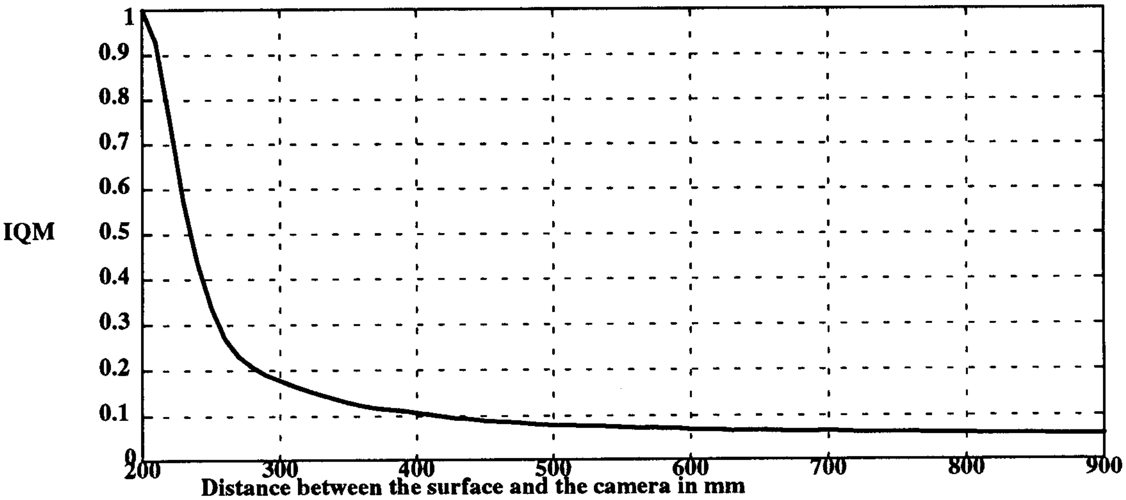


Figure (15b): Measured IQM vs Distance between the camera and surface for D110

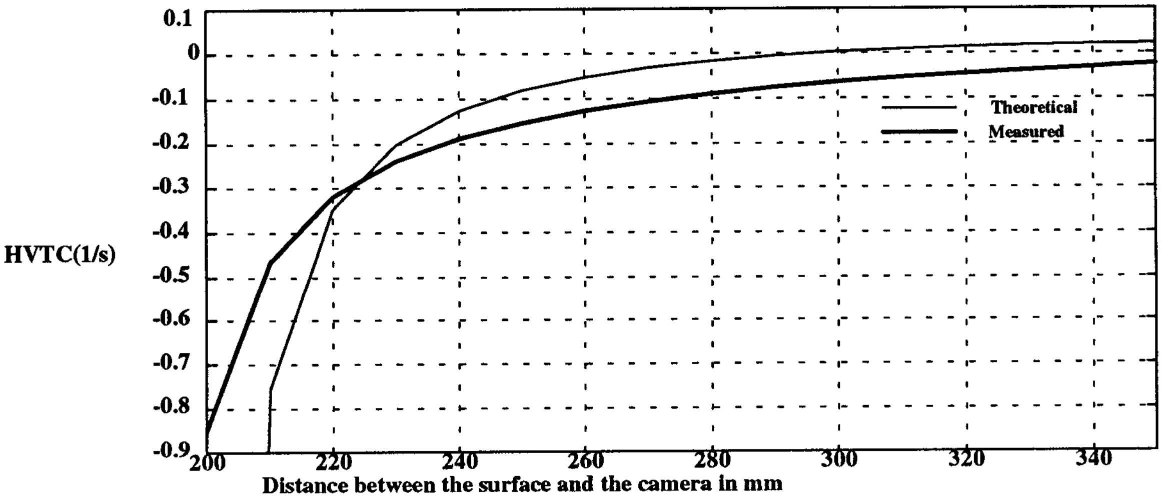


Figure (15c): HVTC vs Distance between the camera and surface for D110

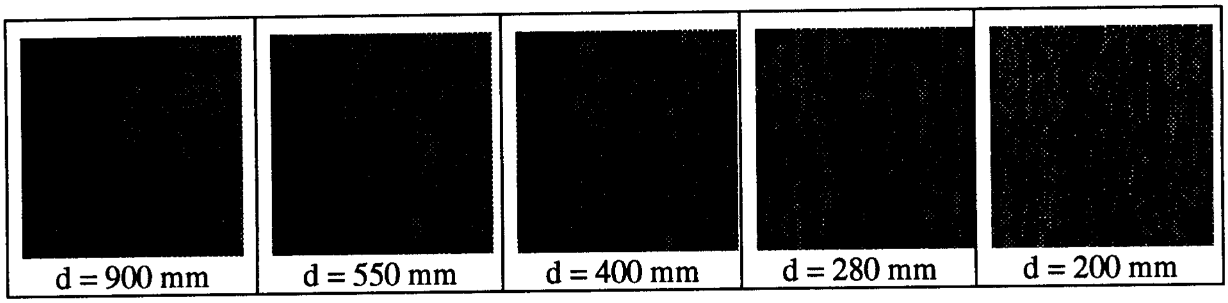


Figure (16a): Image sequence for Texture D12, d is the relative distance

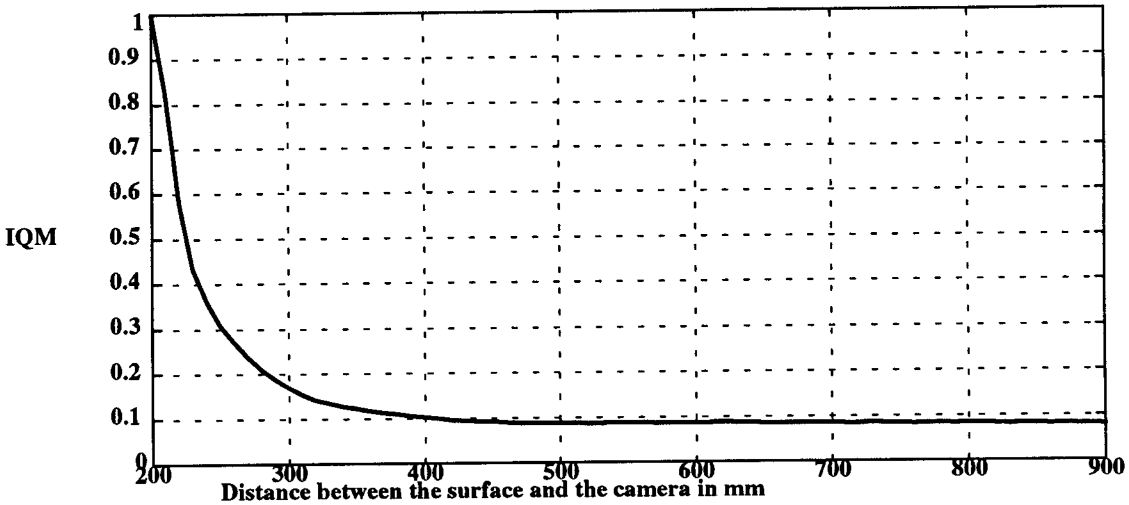


Figure (16b): Measured IQM vs Distance between the camera and surface for D12

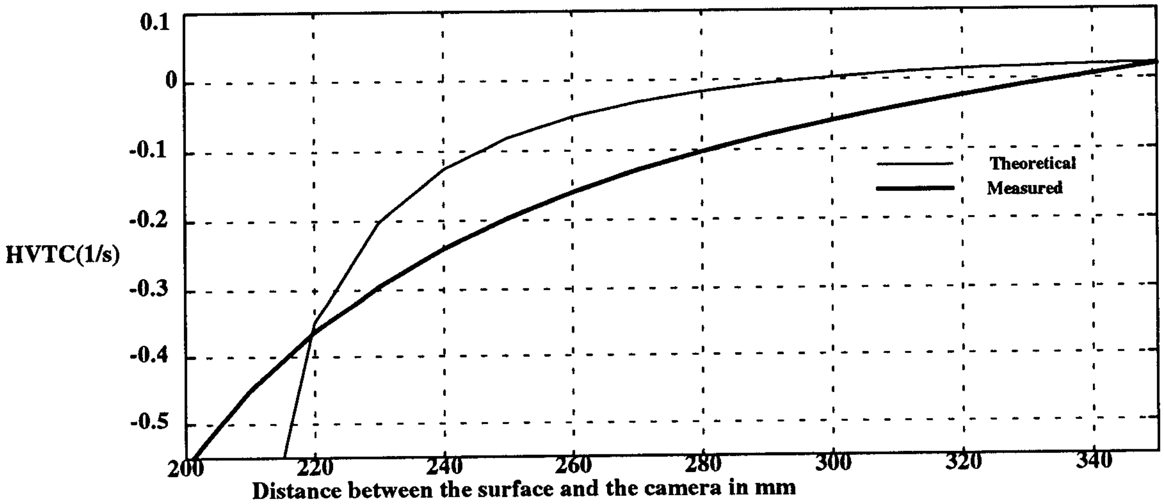


Figure (16c): HVTC vs Distance between the camera and surface for D12

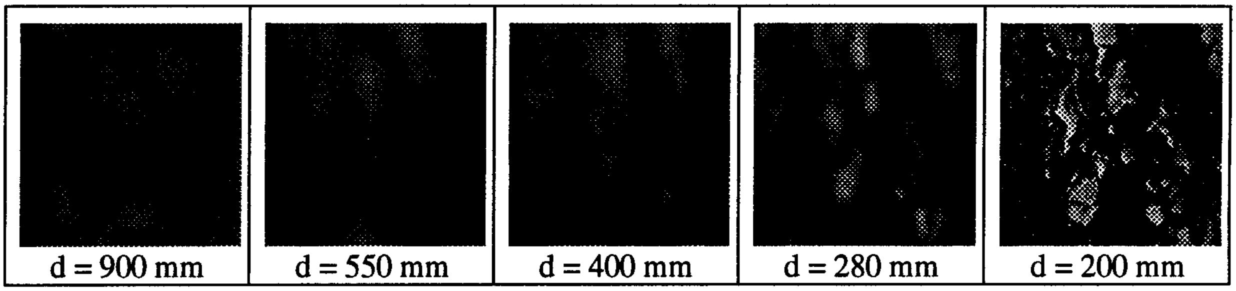


Figure (17a): Image Sequence for Texture D13, d is the relative distance

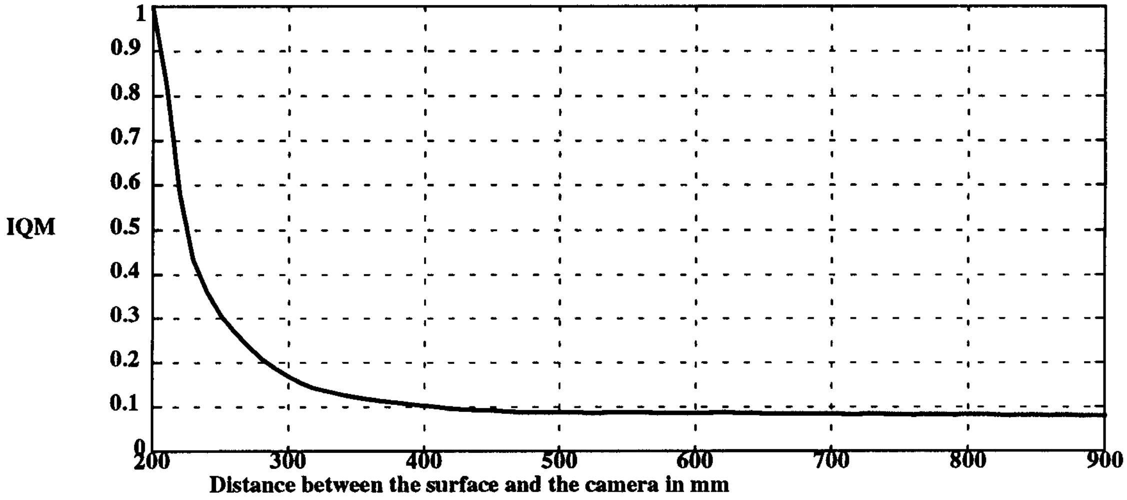


Figure (17b): Measured IQM vs Distance between the camera and surface for D13

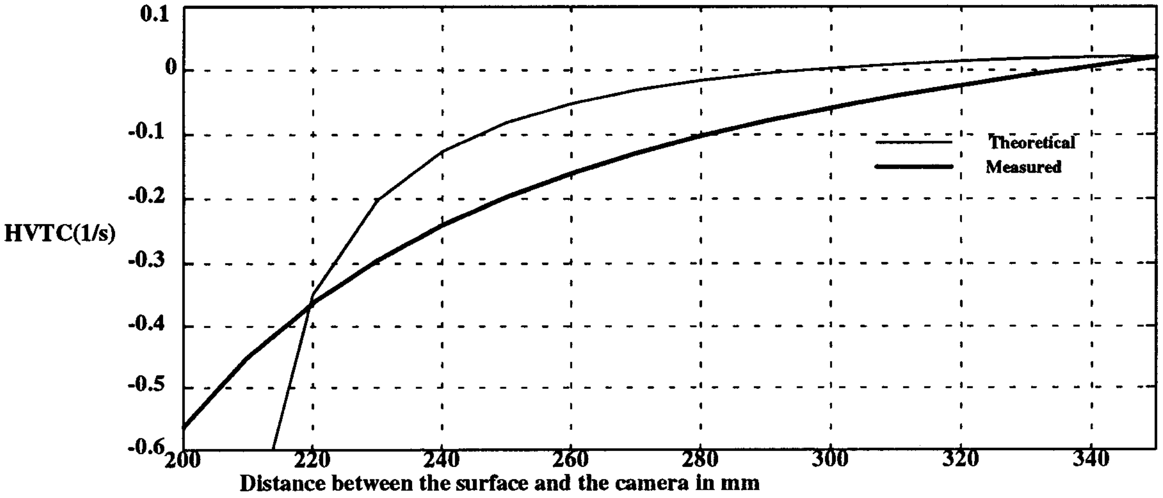


Figure (17c): HVTC vs Distance between the camera and surface for D13

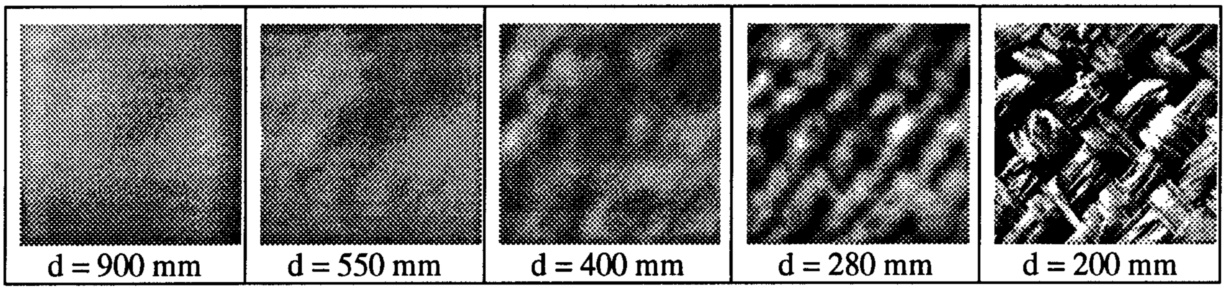


Figure (18a): Image Sequence for Texture D18, d is the relative distance

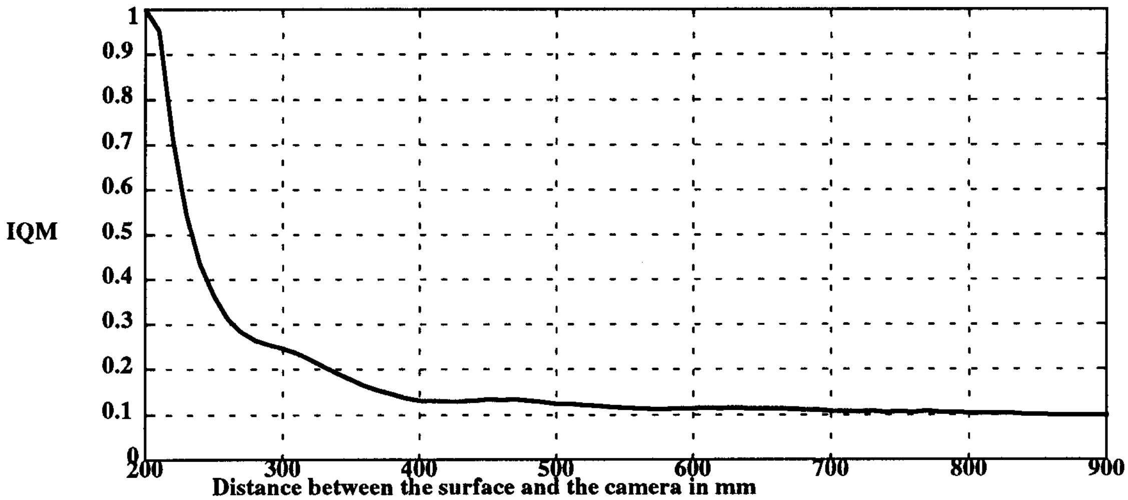


Figure (18b): Measured IQM vs Distance between the camera and surface for D18

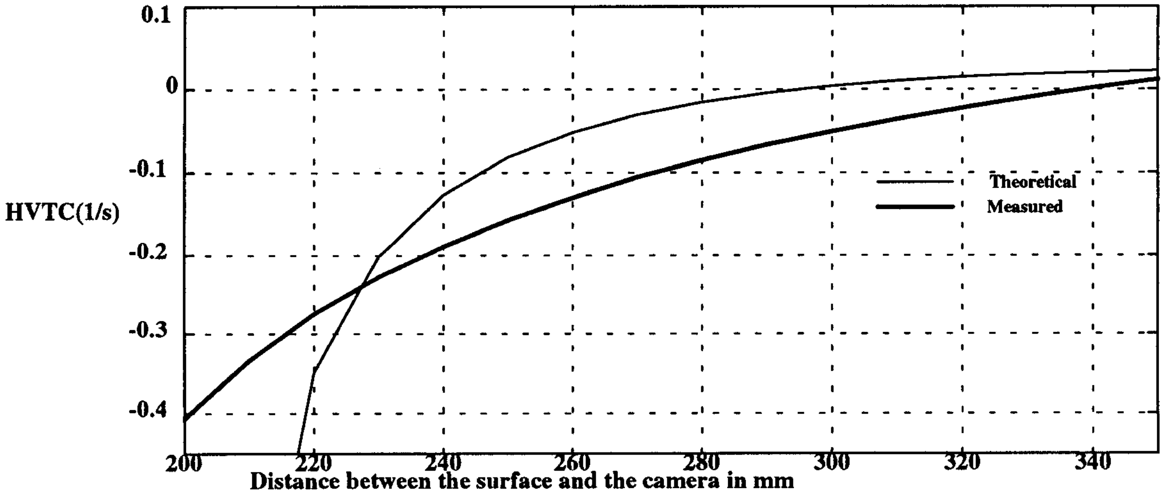


Figure (18c): HVTC vs Distance between the camera and surface for D18

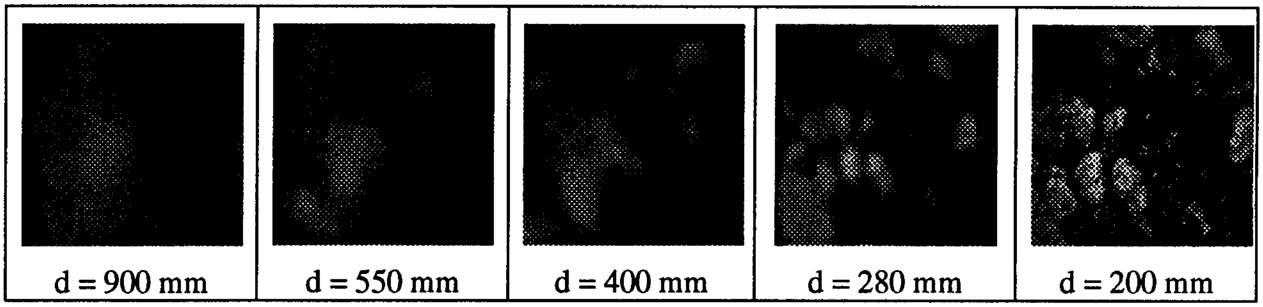


Figure (19a): Image Sequence for Texture D23, d is the relative range

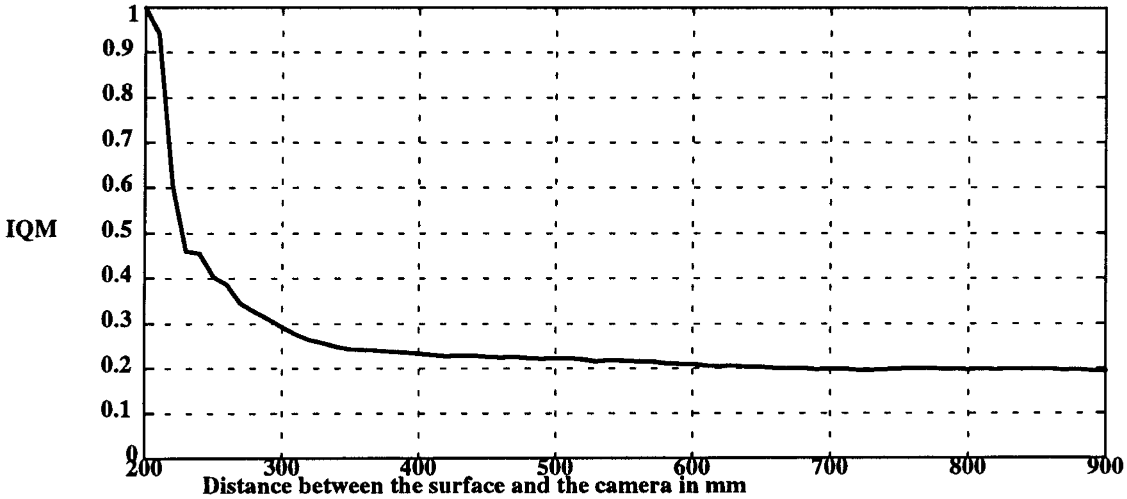


Figure (19b): Measured IQM vs Distance between the camera and surface for D23

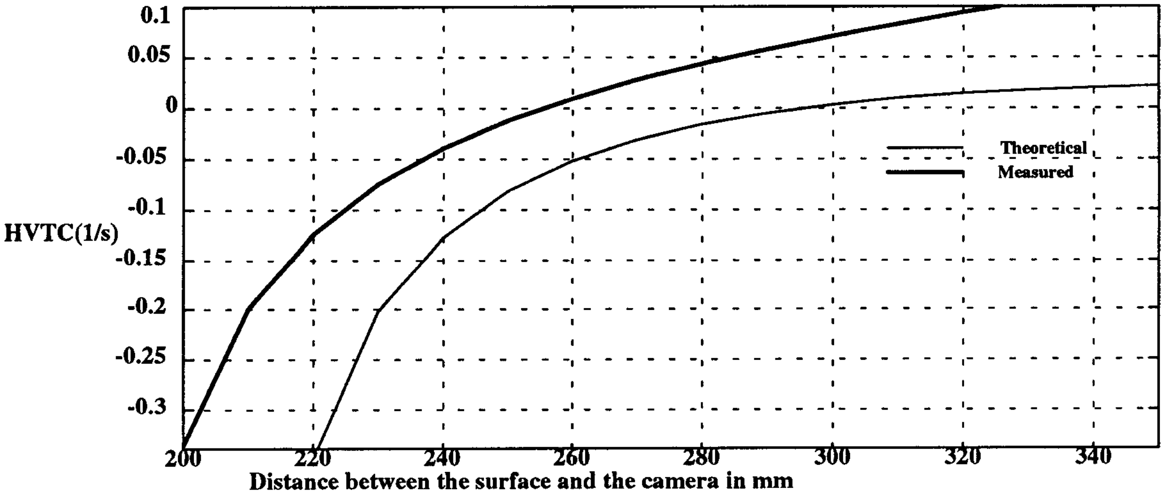


Figure (19c): HVTC vs Distance between the camera and surface for D23

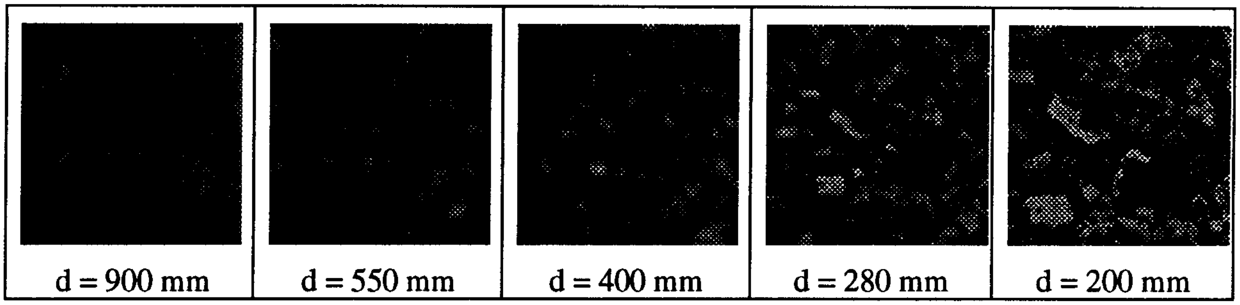


Figure (20a): Image Sequence for Texture D5, d is the relative distance

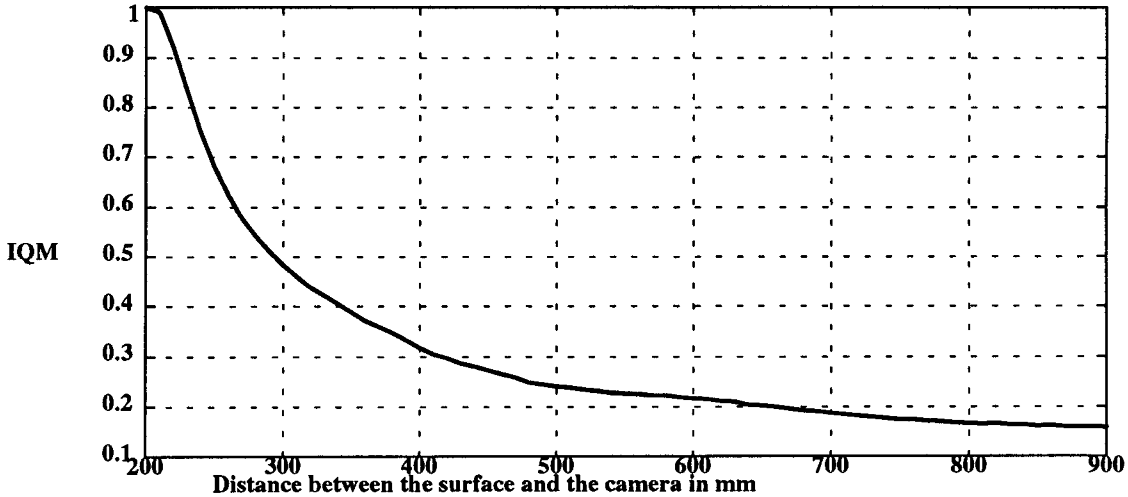


Figure (20b): Measured IQM vs Distance between the camera and surface for D5

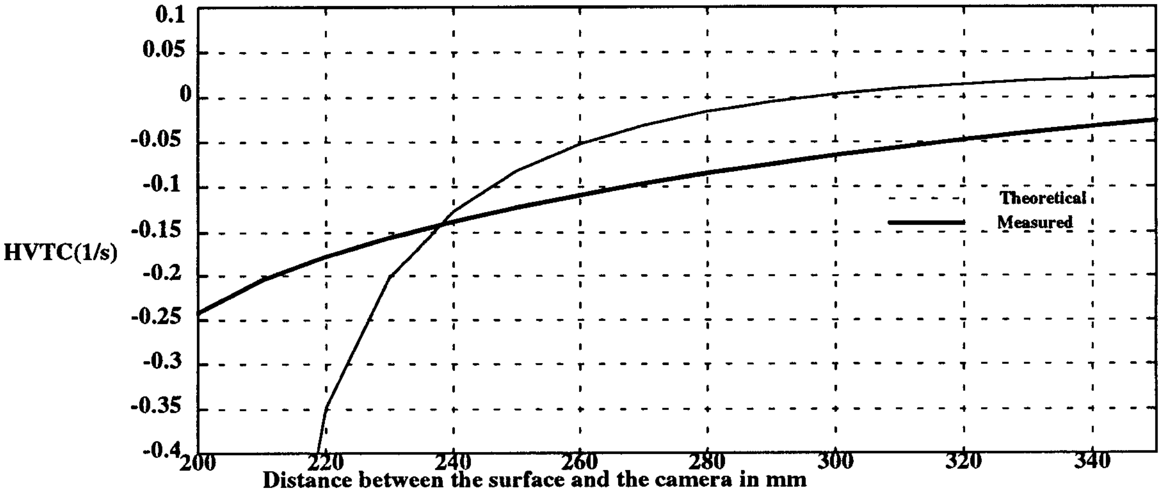


Figure (20c): HVTC vs Distance between the camera and surface for D5

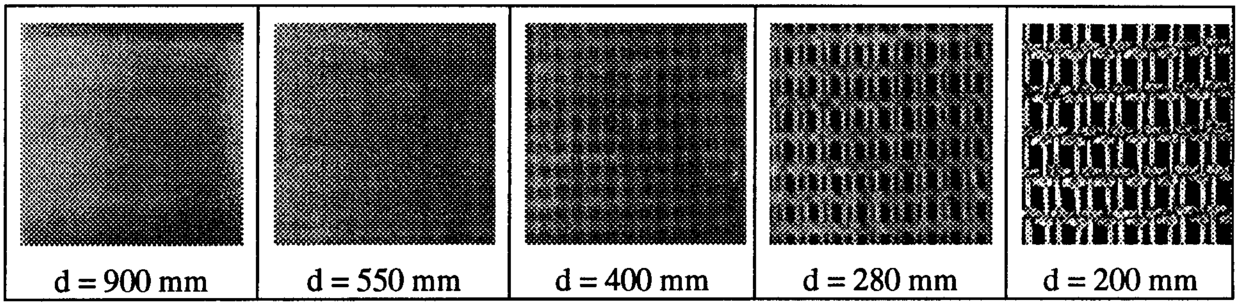


Figure (21a): Image Sequence for Texture D20, d is the relative range

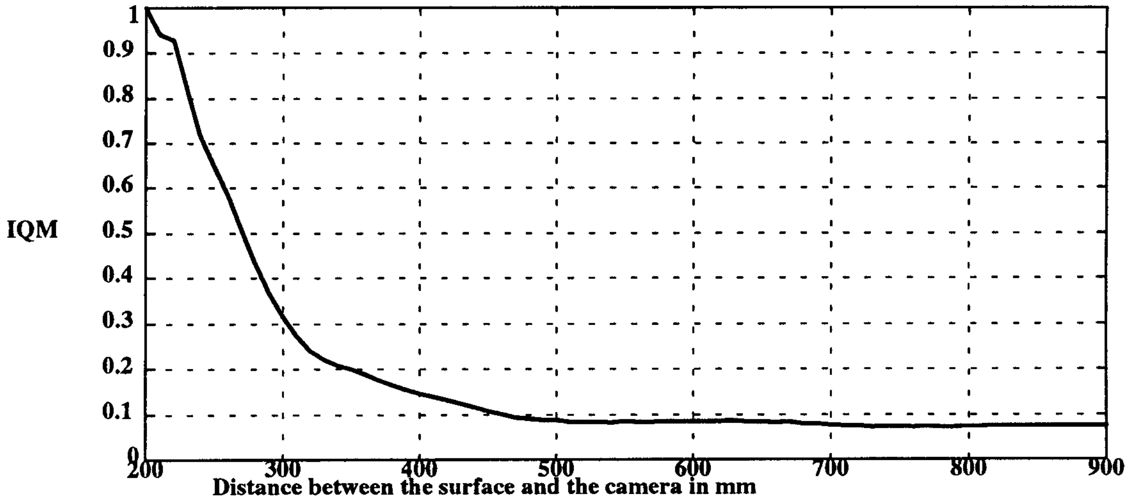


Figure (21b): Measured IQM vs Distance between the camera and surface for D20

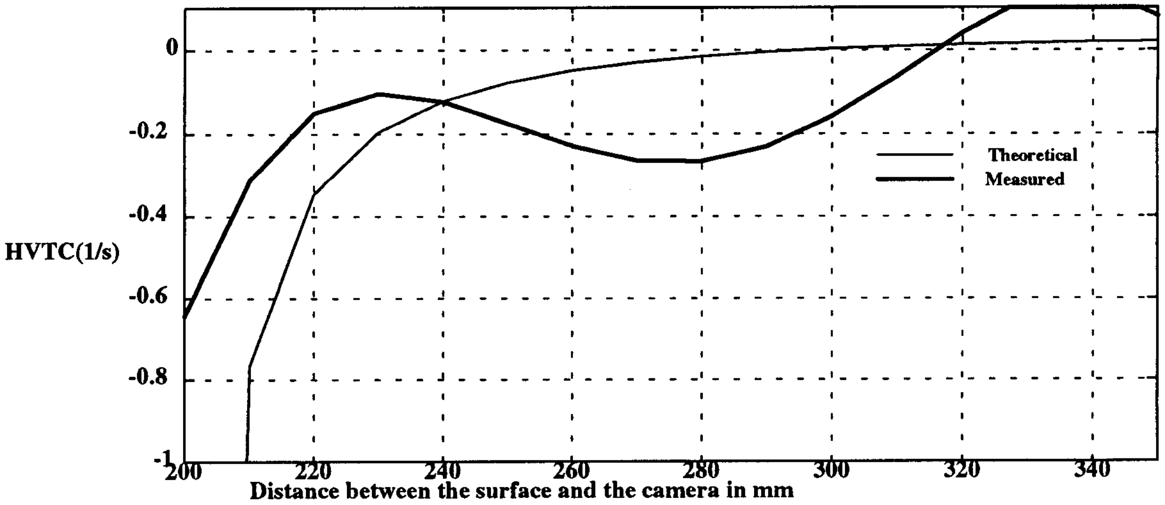


Figure (21c): HVTC vs Distance between the camera and surface for D20

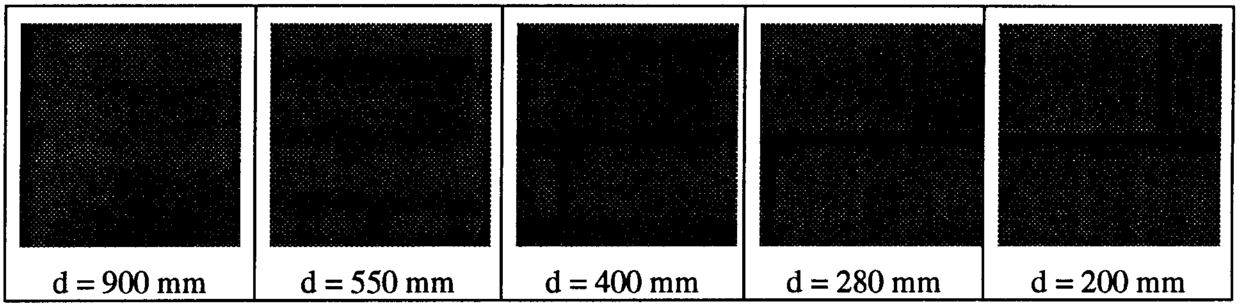


Figure (22a): Image Sequence for Texture D25, d is the relative range

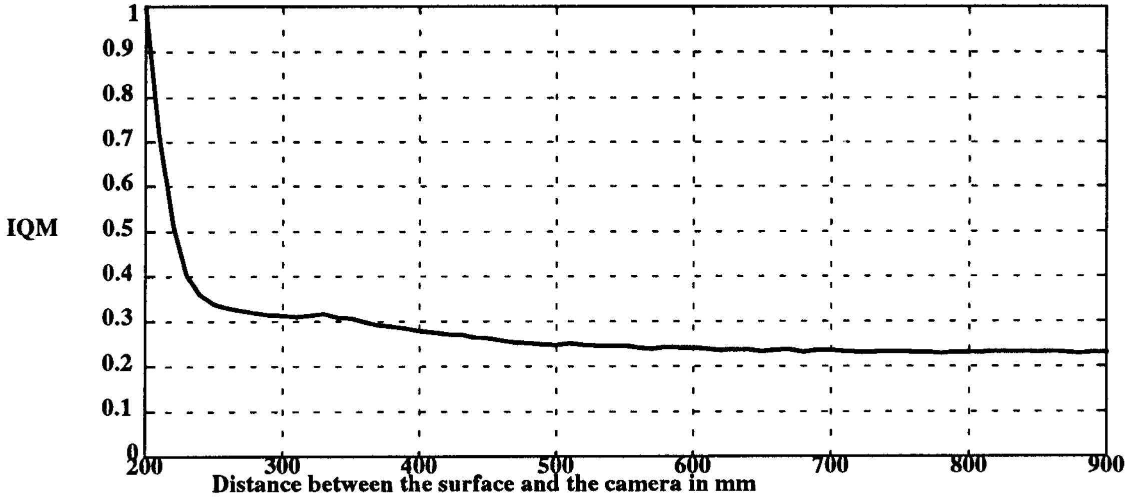


Figure (22b): Measured IQM vs Distance between the camera and surface for D25

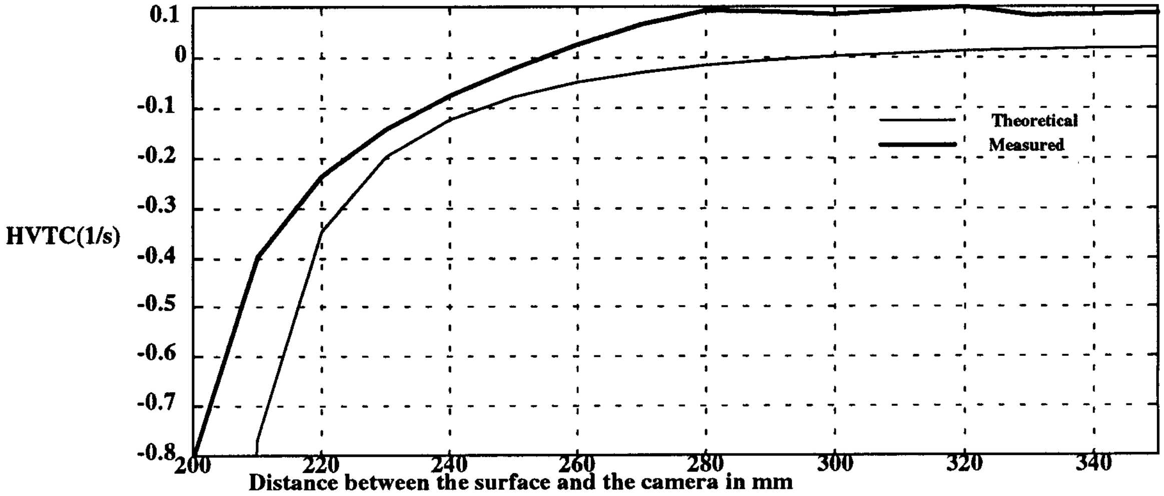


Figure (22c): HVTC vs Distance between the camera and surface for D25

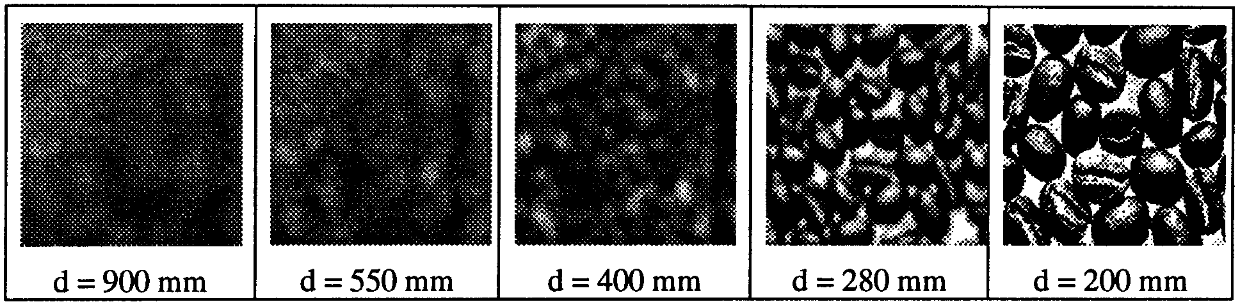


Figure (23a): Image Sequence for Texture D74, d is the relative range

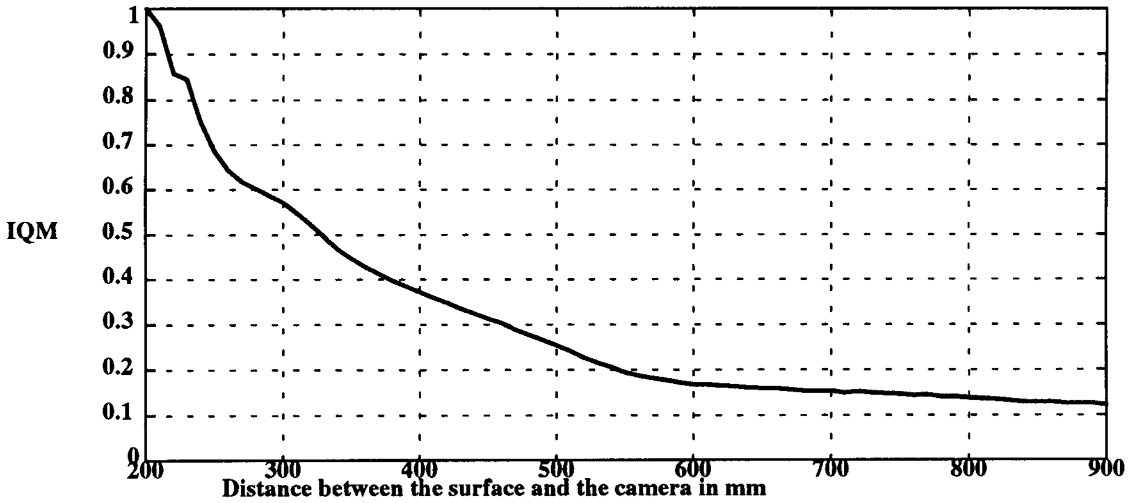


Figure (23b): Measured IQM vs Distance between the camera and surface for D74

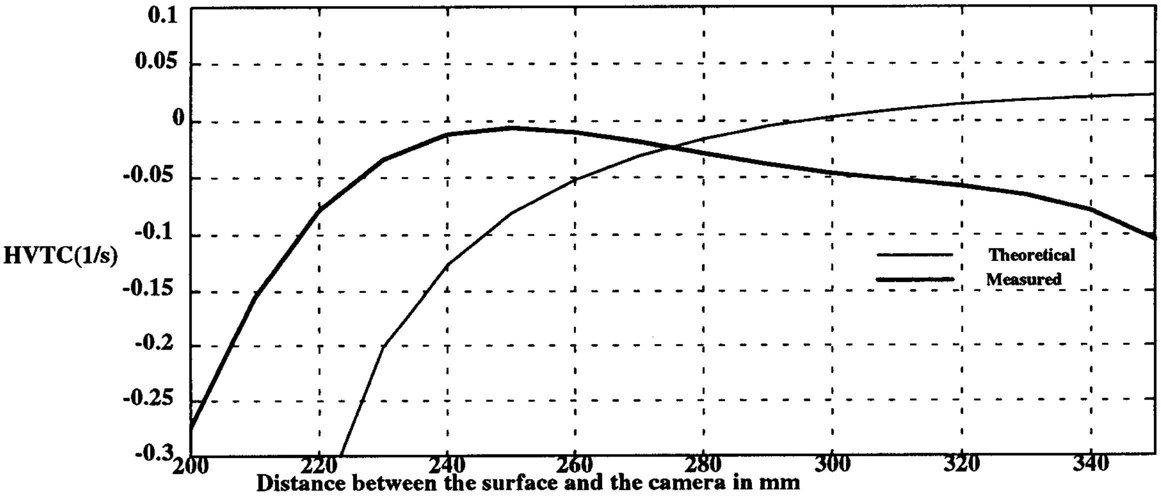


Figure (23c): HVTC vs Distance between the camera and surface for D74

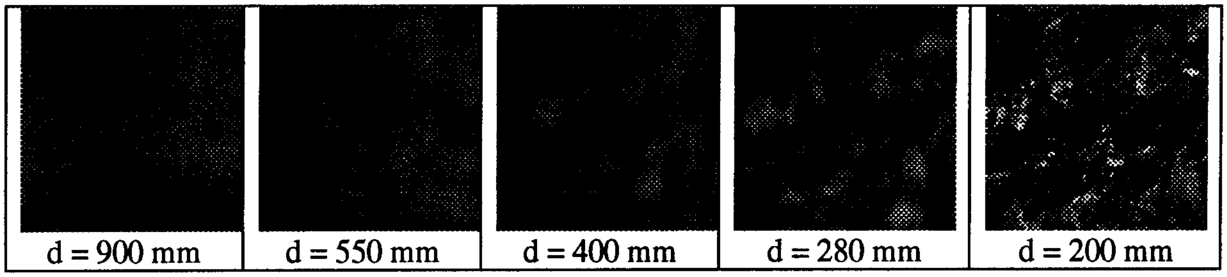


Figure (24a): Image Sequence for Texture D98, d is the relative distance

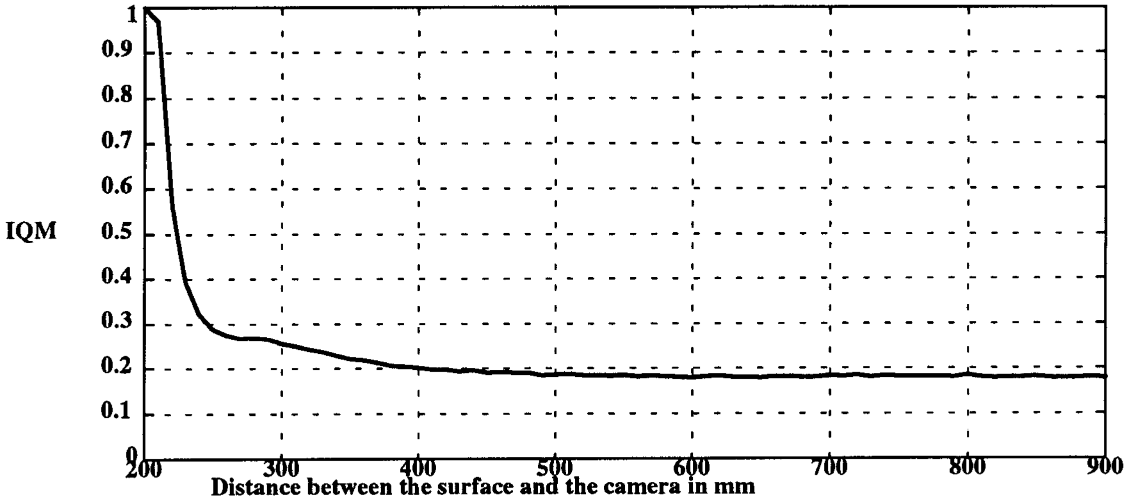


Figure (24b): Measured IQM vs Distance between the camera and surface for D98

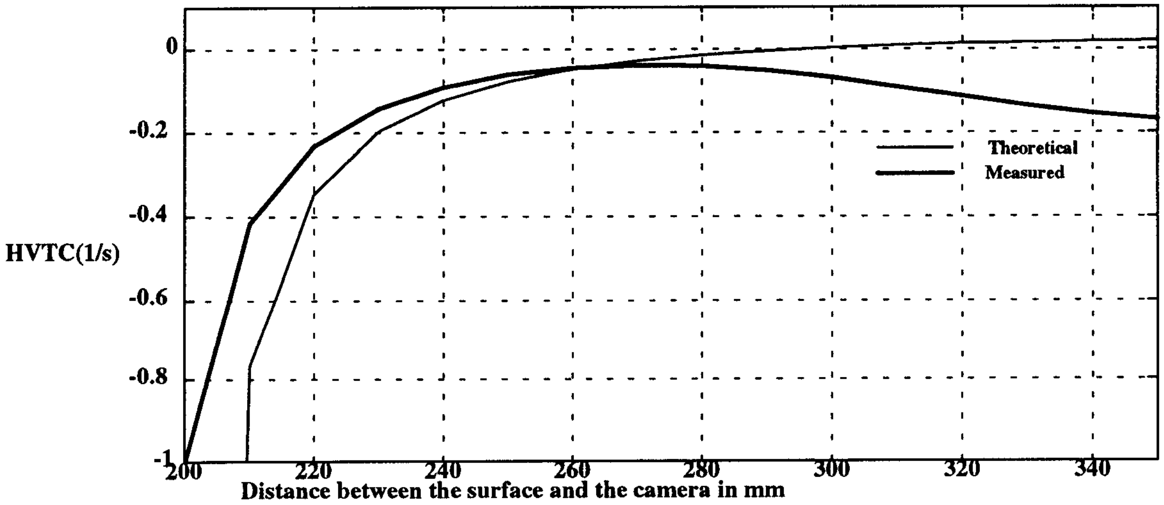


Figure (24c): HVTC vs Distance between the camera and surface for D98

7.3 Extension to Textureless Surfaces

The approach mentioned in the previous section to extract the HVTC is passive and is suitable for textured surfaces only. In otherwords, passive approaches to extract the HVTC are limited to textured surfaces and are independent of the type of the texture in the environment. In this section we describe an alternative active approach to extract the HVTC from textureless surfaces.

The idea is based on projecting an *a-priori* unknown texture pattern on to the scene in a small region around the fixation point. There are no constraints on the type of the texture pattern employed. This active texture projection approach to extract the HVTC is suitable for both textureless as well as textured surfaces, as the extraction of the HVTC is independent of the type of the texture in the scene.

The idea of active texture projection to extraction depth information from textureless surfaces has been suggested by Pentland *et al.* [56] and Nayar *et al.* [57]. Pentland *et al.* [56] describes an approach in which an a-priori known texture pattern is projected on to the scene and the depth information is obtained by comparing the blurred picture with the known original one. A simple texture pattern composed of parallel lines is employed as the a-priori known texture in the experiments [56]. A relation between the width of the blurred line and the depth of the scene is also derived. This idea active texture projection to extract depth information works very well in structured environments where there is absolutely no texture in the scene. It might lead to erroneous results in the presence of texture in the scene, as the resulting blurred image is different from the *a-priori* known image.

Nayar *et al.* [57] describe a depth from defocus approach in which an illumination pattern is projected on the scene using a high power light source and a telecentric lens identical to the one used to image the scene. Due to this reason they assume that the projected illumination is the primary cause for the surface texture and is assumed to be stronger than the natural texture of the surface [57]. An optimal pattern is sought that would ensure all the scene points have the same dominant texture, one that maximizes the spatial resolution and accuracy of computed depth. Online derivation of the optimal projected pattern is posed as an optimization in Fourier domain [57]. This approach needs modification for outdoor implementation [57].

We propose an alternate approach to extract the HVTC information from textureless surfaces as opposed to the depth reconstruction approaches as mentioned earlier. The proposed system consists of a fixed-focus camera and a projection system held very close to the camera and moves along the camera as shown in Figure (1). The camera is initially focused to the desired clearance R_0. The projection system projects an *a-priori* unknown texture pattern on to the scene in a small region around the fixation point. The amount of details in the images obtained by the fixed focus camera depends mainly upon the range between the observer and the fixation point in the 3D scene. In otherwords if the range between the camera and the fixation point is large the corresponding image is blurred and is smooth. We extract the IQM mentioned in the previous section for each image and the HVTC is extracted by using the relative temporal variations of the IQM values. Though the IQM values are dependent upto a certain degree upon the illumination power of the projection source, the scene illumination between any two consecutive frames is assumed to be almost constant. Since the extraction of the HVTC from relative temporal variations of the IQM, the HVTC is almost independent of the scene illumination.

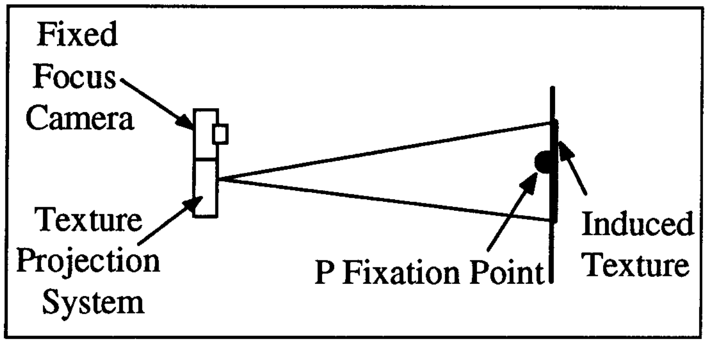


Figure (25): An Active Texture Projection System

This approach of active projection of a-priori unknown texture on to scenes works well for textureless as well as textured surfaces as we are interested in the relative temporal variations in the image smear and not on the texture in the scene. Since the approach works well on textured as well as textureless environments, absolutely no *a-priori* information about the environment in which the observer is traversing is necessary.

8 Conclusion and Future Work

This paper presents a new visual motion cue, called the *Hybrid Visual Threat Cue* (HVTC) that provides some measure for a change in *relative range* as well as *absolute clearances*, between a 3D surface and a visually fixating observer in motion. The visual field associated with the HVTC can be used to demarcate the regions around a moving observer into safe and danger zones of varying degree, which may be suitable for autonomous navigation tasks, in particular *collision avoidance* and *maintenance of clearance*. The HVTC is *independent* of the 3D environment and needs almost no *a-priori* information about it. It is *rotation independent,* and is measured in [time^{-1}] units. Based on scale space representation, we establish a link between the HVTC and the image inner scale.

A practical method to extract the HVTC from a sequence of images of a 3D textured surface obtained by a *fixated, fixed-focus monocular* camera in motion is also presented. A *global dissimilarity measure* is extracted *directly from the raw data of the gray level of textured images* from which the HVTC is obtained. This approach of extracting the HVTC is *independent of the type of 3D surface texture* and needs no *optical flow information, 3D reconstruction, segmentation, feature tracking*. It needs almost *no camera calibration.* This algorithm to extract the HVTC was applied to a set of twelve different texture patterns (of 3D scenes) from the Brodatz's album, where we observed a similar behavior for most of the textures.

Practical approach to extract the HVTC from images, described in this paper is based on experimental observations only. However, theoretically the HVTC can be extracted from images in several ways (for instance, by measuring the fourier transform of the images). Currently, we are investigating alternate approaches to extract the HVTC from images.

Though the extraction of the HVTC needs some texture in the environment, it does not depend upon the type of texture in the environment.. Results presented in this paper are for textured surfaces only. Extension of the approach to extract the HVTC for textureless environments is currently being investigated. The HVTC described in this paper is good only in the region beyond the minimum clearance R_0. Currently we are studying the nature of visual fields for points whose ranges are less than R_0. We are also working on the implementation of the HVTC as a sensory feedback signal in action-perception closed-loop control system of a vision-based autonomous mobile vehicle to accomplish local navigation tasks such as collision avoidance as well as well maintenance of clearance in *a-priori* unknown outdoor environments.

Appendix A: Variations in Scale and Scale-Space Images

In this appendix the relation between the scale space images and the corresponding variations in scale are presented.

Since Equation (2) involves a convolution operator, it will be convenient to work in the frequency domain. Equation (2) can be written in frequency domain as follows:

$$L(u,v;\sigma) = G(u,v;\sigma)F(u,v) \qquad \text{(A1)}$$

where $L(u,v;\sigma) = F\{L(x,y;\sigma)\}$, $G(u,v;\sigma) = F\{G(x,y;\sigma)\}$, $F(u,v) = F\{f(x,y)\}$, $F\{(.)\}$ = Fourier Transform of (.).

Taking the natural logarithm on both sides of Equation (A1), we obtain the following relation:

$$\ln[L(u,v;\sigma)] = \ln[G(u,v;\sigma)] + \ln[F(u,v)] \qquad \text{(A2)}$$

and also we have the following:

$$G(u,v;\sigma) = A\exp(-k\sigma^2(u^2+v^2))$$

where A and k are constants independent of the scale σ .

Let $\ln[L(u,v;\sigma)] = M(u,v;\sigma)$. Hence Equation (A2) can be written as follows:

$$M(u,v;\sigma) = ln[A] + ln[F(u,v)] - k\sigma^2(u^2+v^2) \qquad \text{(A3)}$$

When there is a relative fixated motion between an observer and the same scene we have the following relations:

at time t = t$_i$, scale $\sigma = \sigma_i$ and image M = M$_i$ (i = 1, 2, ...).

For a given frequency (u,v) we can write the following relation:

$$M_1 - M_2 = k(u^2 + v^2)[\sigma_2^2 - \sigma_1^2] \tag{A4a}$$

$$M_2 - M_3 = k(u^2 + v^2)[\sigma_3^2 - \sigma_2^2] \tag{A4b}$$

Hence from Equations (A4a) and (A4b) we can write the following relation:

$$\frac{(M_2 - M_3) - (M_1 - M_2)}{(M_1 - M_2)} = \frac{(\sigma_3^2 - \sigma_2^2) - (\sigma_2^2 - \sigma_1^2)}{(\sigma_2^2 - \sigma_1^2)} \tag{A5}$$

If the interframe time interval Δt is very small, dividing the numerator and the denominator on both sides of Equation (A5) by Δt, Equation (5) can be written as follows:

$$\lim_{\Delta t \to 0} \frac{\left[\frac{(M_2 - M_3)}{\Delta t} - \frac{(M_1 - M_2)}{\Delta t}\right]}{\left[\frac{(M_1 - M_2)}{\Delta t}\right]} = \lim_{\Delta t \to 0} \frac{\left[\frac{(\sigma_3^2 - \sigma_2^2)}{\Delta t} - \frac{(\sigma_2^2 - \sigma_1^2)}{\Delta t}\right]}{\left[\frac{\sigma_2^2 - \sigma_1^2}{\Delta t}\right]} \tag{A6a}$$

$$\Rightarrow \frac{\left.\frac{dM}{dt}\right|_{t=t2} - \left.\frac{dM}{dt}\right|_{t=t1}}{\left.\frac{dM}{dt}\right|_{t=t1}} = \frac{\left.\frac{d}{dt}(\sigma^2)\right|_{t=t2} - \left.\frac{d}{dt}(\sigma^2)\right|_{t=t1}}{\left.\frac{d}{dt}(\sigma^2)\right|_{t=t1}} \tag{A6b}$$

Dividing on both sides of Equation (A6b) by Δt, we obtain the following relation:

$$\Rightarrow \lim_{\Delta t \to 0} \frac{\left.\frac{dM}{dt}\right|_{t=t2} - \left.\frac{dM}{dt}\right|_{t=t1}}{(\Delta t)\left.\frac{dM}{dt}\right|_{t=t1}} = \lim_{\Delta t \to 0} \frac{\left.\frac{d}{dt}(\sigma^2)\right|_{t=t2} - \left.\frac{d}{dt}(\sigma^2)\right|_{t=t1}}{(\Delta t)\left.\frac{d}{dt}(\sigma^2)\right|_{t=t1}} \tag{A6c}$$

$$\frac{\frac{d^2}{dt^2}(M)}{\frac{d}{dt}(M)} = \frac{\frac{d^2}{dt^2}(\sigma^2)}{\frac{d}{dt}(\sigma^2)} \tag{A7}$$

Appendix B: Variations in Range and Scale-Space Images

The right side of Equation (A7) in Appendix A (i.e., $\dfrac{\frac{d^2}{dt^2}(\sigma^2)}{\frac{d}{dt}(\sigma^2)}$) can be written as follows:

$$\frac{\frac{d^2}{dt^2}(\sigma^2)}{\frac{d}{dt}(\sigma^2)} = \frac{\frac{d^2}{dt^2}(\sigma)}{\frac{d}{dt}(\sigma)} + \frac{\frac{d}{dt}(\sigma)}{\sigma} \tag{B1}$$

Also we have the following relation between the scale and the range (see Equation (6)):

$$\sigma = k\left(\frac{1}{R_0} - \frac{1}{R}\right) \tag{B2}$$

$$\Rightarrow \frac{d}{dt}(\sigma) = k\frac{\frac{d}{dt}(R)}{R^2} \tag{B3}$$

$$\Rightarrow \frac{d^2}{dt^2}(\sigma) = k\frac{[R^2\frac{d^2}{dt^2}(R) - 2R\left(\frac{d}{dt}(R)\right)^2}{R^4} \tag{B4}$$

Dividing Equation (B4) by Equation (B3) we obtain the following relation:

$$\frac{\frac{d^2}{dt^2}(\sigma)}{\frac{d}{dt}(\sigma)} = \frac{\frac{d^2}{dt^2}(R)}{\frac{d}{dt}(R)} - 2\frac{\frac{d}{dt}(R)}{R} \qquad \text{(B5)}$$

$$\frac{\frac{d^2}{dt^2}(\sigma)}{\frac{d}{dt}(\sigma)} + \frac{\frac{d}{dt}((\sigma)}{\sigma} = \frac{\frac{d^2}{dt^2}(R)}{\frac{d}{dt}(R)} - 2\frac{\frac{d}{dt}(R)}{R} + \frac{R_0}{(R-R_0)}\frac{\frac{d}{dt}(R)}{R} \qquad \text{(B6)}$$

Appendix C: Fourier Transforms

In this section the Fourier Transforms of the 2D Gaussian function is derived.

$$G(x,y;\sigma)=\frac{1}{2\pi\sigma^2}\exp(-\frac{x^2+y^2}{2\sigma^2})$$

$$G(u,v;\sigma)=F\{G(x,y;\sigma)\}=\frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(-\frac{x^2+y^2}{2\sigma^2})\exp(-j(ux+vy))dxdy$$

$$G(u,v;\sigma)=F\{G(x,y;\sigma)\}=\frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty}\exp(-\frac{x^2}{2\sigma^2}-jux)dx\int\exp(-\frac{y^2}{2\sigma^2}-jvy)dy$$

From [63] we obtain the following results:

$$G(u,v;\sigma)=\frac{2\pi\sigma^2}{2\pi\sigma^2}\exp(-\sigma^2(u^2+v^2))=\exp(-\sigma^2(u^2+v^2))$$

Appendix D: Analysis of the HVTC

In this section each term in the HVTC expression is analyzed in this section.

Consider an observer centered co-ordinate system OXYZ as shown in Figure (D1). The origin of the co-ordinate system is attached to the observer and is moving along with it. Consider a point P in the stationary environment around the observer. Let **t** be the instantaneous translational vector, r be the range between the observer and the fixation point P.

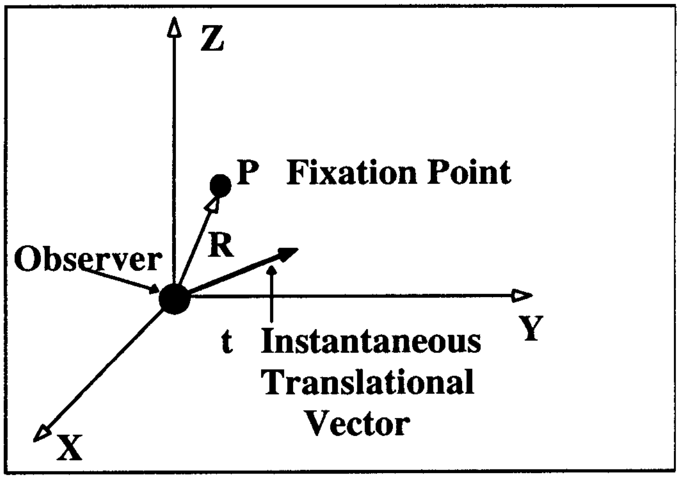


Figure (D1): Observer in motion, OXYZ is the observer centered co-ordinate system, **t** is the instantaneous translational vector, P is the fixation point, **r** is the range vector

The range between the observer and the fixation point r in the observer centered co-ordinate system can be written as follow:

$$R = \sqrt{x^2 + y^2 + z^2} \qquad \text{(D1)}$$

$$\frac{dR}{dt}=\frac{x\dot{x}+y\dot{y}+z\dot{z}}{\sqrt{x^2+y^2+z^2}} \tag{D2}$$

$$\frac{d^2R}{dt^2}=\frac{\sqrt{x^2+y^2+z^2}\,(x\ddot{x}+y\ddot{y}+z\ddot{z}+\dot{x}^2+\dot{y}^2+\dot{z}^2)}{x^2+y^2+z^2}-\frac{(x\dot{x}+y\dot{y}+z\dot{z})^2}{(\sqrt{x^2+y^2+z^2})^3} \tag{D3}$$

where (x, y, z) are the co-ordinates of the fixation point in OXYZ co-ordinate system, $(\dot{x}, \dot{y}, \dot{z})$ is the instantaneous translational vector **t** of the observer in OXYZ co-ordinate system. $(\ddot{x}, \ddot{y}, \ddot{z})$ is the instantaneous acceleration of the observer in OXYZ co-ordinate system, d(.)/dt, $d^2(.)/dt^2$ are the temporal derivatives of (.). For a uniform translational velocity the acceleration component $(\ddot{x}, \ddot{y}, \ddot{z})$ is zero. Hence Equation (D3) reduces to the following:

$$\frac{d^2R}{dt^2}=\frac{\sqrt{x^2+y^2+z^2}\,(\dot{x}^2+\dot{y}^2+\dot{z}^2)}{x^2+y^2+z^2}-\frac{(x\dot{x}+y\dot{y}+z\dot{z})^2}{(\sqrt{x^2+y^2+z^2})^3} \tag{D4}$$

The HVTC defined in Equation (10) is reproduced as follows:

$$\text{HVTC}=\frac{\frac{d^2}{dt^2}(R)}{\frac{d}{dt}(R)}-2\frac{\frac{d}{dt}(R)}{R}+\frac{R_0}{(R-R_0)}\frac{\frac{d}{dt}(R)}{R}$$

In the following sub-sections, representations of the entities on the right side of the above equation in terms of Equations [D1, D2, D4].

D.1 Representation of $-2\frac{\dot{R}}{R}$ in terms of Equations [D1, D2]

Looming is defined as follows [10]:

$$-\frac{\dot{R}}{R} = L \tag{D5}$$

where L represents looming and $\dot{R} = \frac{dR}{dt}$

Rewriting Equation (D5) in terms of Equations (D1) and (D2) the following relation can be derived [10]:

$$(x+\frac{\dot{x}}{2L})^2 + (y+\frac{\dot{y}}{2L})^2 + z+\frac{\dot{z}}{2L})^2 = (\frac{\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}}{2L})^2 \tag{D6}$$

For a given instantaneous translational velocity ($\dot{x}$, $\dot{y}$, $\dot{z}$), Equation (12) represents a system of circles (spheres in 3D) as shown in Figure (5).

D.2 Representation of $\frac{\ddot{R}}{\dot{R}}$ in terms of Equation [D2,D4]

We can write ($\frac{\ddot{R}}{\dot{R}}$) in terms of Equations [D2, D4] as follows:

$$\frac{\ddot{R}}{\dot{R}} = \frac{(\dot{x}^2+\dot{y}^2+\dot{z}^2)}{(x\dot{x}+y\dot{y}+z\dot{z})} - \frac{(x\dot{x}+y\dot{y}+z\dot{z})}{(x^2+y^2+z^2)} \tag{D7}$$

Analysis of two entities on the right side of Equation (D7) is presented in the following subsections.

D.2.1 Analysis of $\dfrac{(\dot{x}^2+\dot{y}^2+\dot{z}^2)}{(x\dot{x}+y\dot{y}+z\dot{z})}$

Let:

$$\frac{(\dot{x}^2+\dot{y}^2+\dot{z}^2)}{(x\dot{x}+y\dot{y}+z\dot{z})}=T \qquad \text{(D8)}$$

Equation (D8) can be rewritten as follows:

$$x\left(\frac{\dot{x}}{\dfrac{\dot{x}^2+\dot{y}^2+\dot{z}^2}{T}}\right)+y\left(\frac{\dot{y}}{\dfrac{\dot{x}^2+\dot{y}^2+\dot{z}^2}{T}}\right)+z\left(\frac{\dot{z}}{\dfrac{\dot{x}^2+\dot{y}^2+\dot{z}^2}{T}}\right)=1 \qquad \text{(D9)}$$

Equation (D9) represents a systems perpendicular to the instantaneous translational velocity vector. This visual field associated with Equation (D9) is similar to the TTC concept in [11] and is measured in [time^{-1}] units. Any point that lies on a particular plane produces the same value of T and points that lie on a plane in front of the observer correspond to negative values and the points in the back of the observer correspond to positive values of T.

D.2.2 Analysis of $-\frac{(x\dot{x}+y\dot{y}+z\dot{z})}{(x^2+y^2+z^2)}$

The entity $-\frac{(x\dot{x}+y\dot{y}+z\dot{z})}{(x^2+y^2+z^2)}$ represents looming (see Equation (D5)) which is described in section D1.

D.3 Analysis of $\frac{R_0}{(R-R_0)}\frac{\dot{R}}{R}$

The entity $-\frac{R_0}{(R-R_0)}\frac{\dot{R}}{R}$ is defined as the Visual Threat Cue (VTC) [65], whose visual fields are shown in Figure (6).

The above manipulations show that the HVTC is a linear combination of the TTC, the Looming, the VTC.

Also a relation between the VTC and the HVTC can be derived as follows:

$$HVTC = \frac{\dot{VTC}}{VTC} - 2VTC \qquad \text{(D10)}$$

Appendix E: The Imaging System

The amount of blur in an image is characterized by the radius (refer to Figure (A1)) of the blur circle [24-26]. The expression for the radius of the blur circle, for a camera focused to a short distance u_0, can be derived as follows [26].

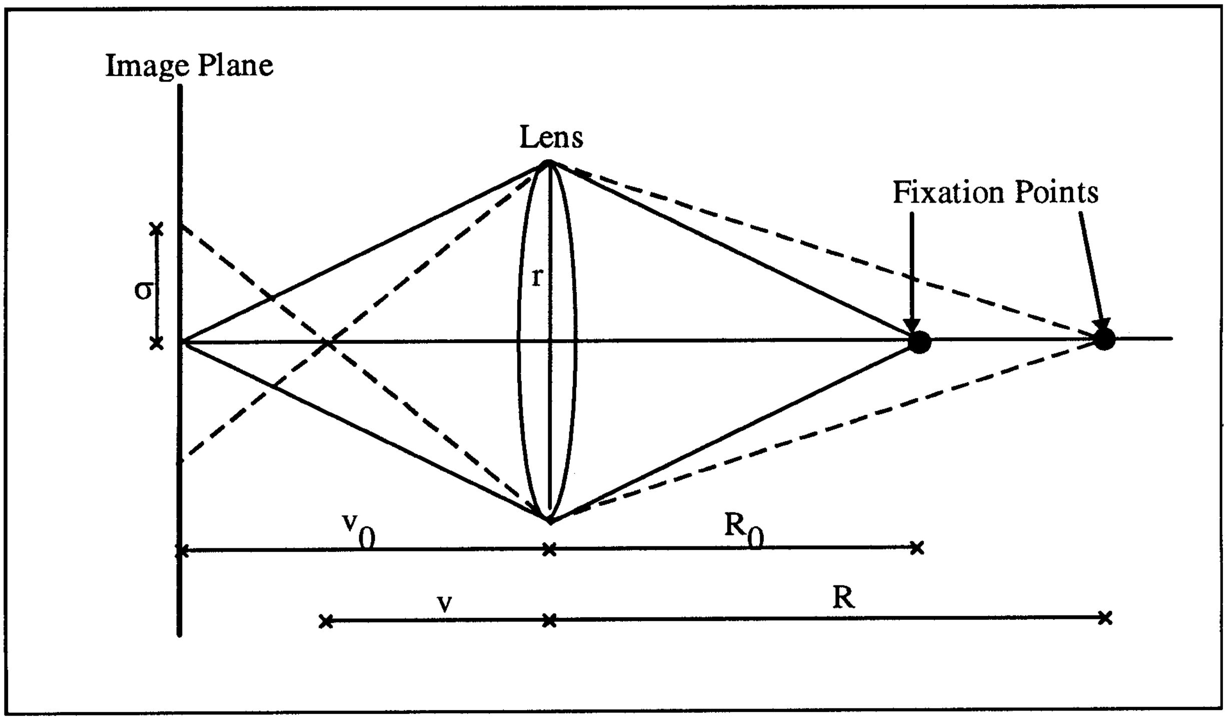


Figure E1: Imaging System

In Figure (E1) : v_0 = distance between the image plane and the lens, u_0 = distance between the lens and the scene for which the image is in focus, σ = radius of the blur circle (for u > v), r = radius of the lens, F = Focal length of the lens, f = focal number of the lens.

For objects in perfect focus the following Gaussian lens formula holds

$$\frac{1}{u_0}+\frac{1}{v_0}=\frac{1}{F} \tag{E1}$$

From Figure (E1) we also get the following relation

$$\frac{r}{v}=\frac{\sigma}{v_0-v} \tag{E2}$$

By combining Equation (E2) with Equation (E1) and replacing u by R we have the following relation:

$$R=\frac{Fv_0}{v_0-F-\sigma f} \tag{E3}$$

From which the following relation is obtained:

$$\sigma=\frac{Fv_0}{f}(\frac{1}{R_0}-\frac{1}{R}) \tag{E4}$$

where R is the distance between the object and the lens and R_0 is the distance to which the camera is focused to initially.

Appendix F: The Image Quality Measure (IQM)

Let (x,y) be the spatial coordinates of an arbitrary pixel in the image, where x, y are integers and I(x,y) be the intensity at (x, y). The inter-pixel distance is denoted by δ and is defined as follows.

$$\delta = (\Delta x, \Delta y) \qquad \text{(F1)}$$

where Δx = difference between the corresponding x coordinates of two pixels; Δy = difference between the corresponding y coordinates of two pixels.

The dissimilarity between the image intensities of pixels separated by the inter-pixel distance defined in Equation (B1) can be characterized by the City Block Metric (CBM), which is defined as [58, 59]:

$$CBM = \{|I(x,y) - I(x+\Delta x, y+\Delta y)|\} \qquad \text{(F2)}$$

where I(x,y) = intensity at pixel (x,y); I(x+Δx, y+Δy) = intensity at pixel (x+ΔX, y+Δy).

Several other dissimilarity measures may be used instead of the one used in Equation (F2). A detailed description of these measures is presented in [55].

Each pixel (x,y) in an image can be characterized by a matrix known dissimilarity matrix, which is basically a matrix of numbers that characterizes the dissimilarity of pixel intensities in the neighborhood of the pixel. For instance, if the texture is smooth, the dissimilarity is very low, hence the mean value of the dissimilarity matrix is low. This dissimilarity matrix is a (2Lr+1)X(2Lc+1), where Lc, Lr are positive integer constants. The (i,j) element of this matrix is the CBM defined in Equation (B2), i = -Lc, ...,-1,0,1,...Lc and j = -Lr,...,-1,0,1,...,Lr. The (0,0) element of the dissimilarity matrix is zero since $\{|I(x,y)-I(x+0,y+0)\}| = 0$. A matrix of numbers can be generated for any pixel.

This dissimilarity matrix can be used to generate a global image variable to indicate the smoothness of texture details in an image. Next we show how to generate a global measure which indicates the texture smoothness.

We select an arbitrary window in the image plane. Let x_i and x_f be the initial and final coordinates of the window along the x-direction respectively and y_i and y_f be the initial and final coordinates of the window along the y-direction respectively. For each pixel in the window selected we compute the sum of all elements in the dissimilarity matrix described above. Thus there are $(x_f-x_i)X(y_i-y_f)$ sums for the window selected, i.e., mathematically it can be described as follows.

$$DSM = \sum_{x=x_i}^{x_f} \sum_{y=y_i}^{y_f} \left(\sum_{p=-L_c}^{L_c} \sum_{q=-L_r}^{L_r} |I(x,y) - I(x+p, y+q)| \right) \tag{F3}$$

Where I(x,y) is the intensity at pixel (x,y) and x_i and x_f are the initial and final x-coordinates of the window respectively ; y_i and y_f are the initial and final y-coordinates of the window in the image respectively and Lc and Lr are positive integer constants, need not be equal.

Appendix G: Estimation of IQM using measured IQM

In this appendix, we describe the process of estimation of IQM using the measured values of the IQM.

The following Equation presents a relation between the radius of the blur circle and the range (see Equation (6)).

$$\sigma = k\left(\frac{1}{R_0} - \frac{1}{R}\right) \tag{G1}$$

In reality, only the left hand side of the above Equation is measurable (in the form of 1/IQM). In order to fit a curve for σ i.e., 1/IQM, we chose time as the independent variable, i.e., $\sigma_{approx} = f(t)$.

Let $f(t) = a_0+a_1t+a_2t^2+...+ a_nt^n$ (G2)

The minimum value of n that results in an estimate that minimizes the error between the σ_{approx} and $\sigma_{measured}$ is obtained on the basis of computer simulation as follows:

Using Equation (G1)a set of σ*'s* ($\sigma_{measured}$)is computed for various values of R. Using this set of σ's as our inputs, we tried to fit a polynomial of nth order in t as follows:

at $t = 0$, $\sigma = \sigma_1$; $t = \Delta t$, $\sigma = \sigma_2$; $t = 2\Delta t$, $\sigma = \sigma_3$; ...; $t = (n-1)\,\Delta t$, $\sigma = \sigma_n$;

Based on numerical simulation results, we found that the order of the polynomial has to be atleast six, in order to minimize the error between σ_{approx} and $\sigma_{measured}$. Hence, we employed a sixth order polynomial fit in computing the HVTC values using the IQM values measured from images.

Appendix H: Depth of Field and Width of Aperture

There is a region in front of the camera in which the image remains almost in sharp focus. This region in front of the camera is usually referred to as the depth of field and depends upon several factors such as the width of the aperture of the camera, distance to which the camera is focused to initially, etc. This section presents a relationship between the depth of field of the camera and its parameters.

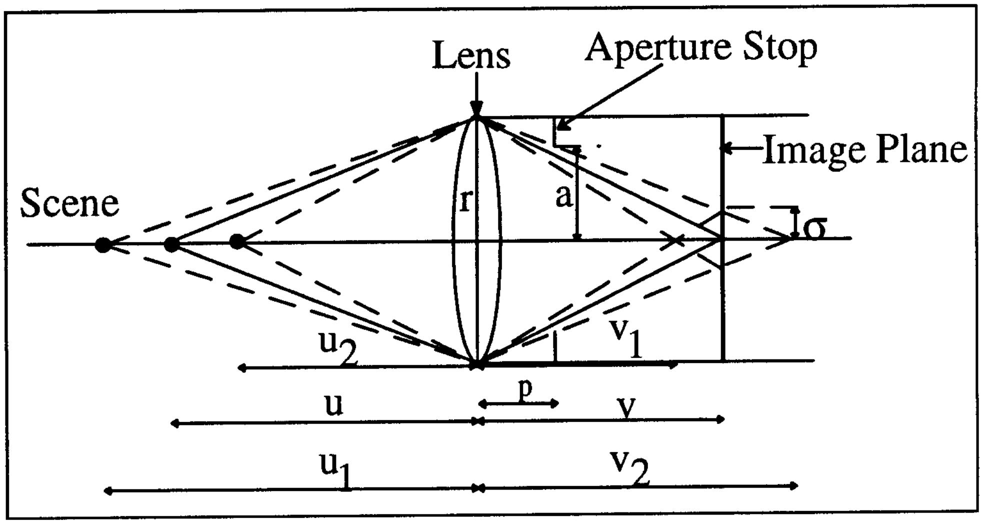


Figure (H1) : A Block Diagram of Lens and Shutter Arrangement

Let r be the radius of the lens, u be the distance to which the camera is focused to, u1, u2 be the lower and upper bounds on u for which the camera remains approximately focused, v be the distance between the image plane and the lens, v1, v2 be the corresponding images of u1 and u2, “a” be the width of the aperture, “p” be the distance between the aperture stop and the lens and σ be the upper bound on the radius of the blur circle so that the point is almost in perfect focus.

From Figure (H1) we can derive the following relations:

$$\frac{\sigma}{v - v_1} = \frac{a}{v_1 - p} \qquad \text{(H1)}$$

$$\frac{\sigma}{v_2 - v} = \frac{a}{v_2 - p} \qquad \text{(H2)}$$

Simplification of Equations (H1) and (H2) leads to the following relations:

$$v_1 = \frac{(p\sigma + av)}{(\sigma + a)} \qquad \text{(H3)}$$

$$v_2 = \frac{(p\sigma - av)}{(\sigma - a)} \qquad \text{(H4)}$$

By employing the Gaussian lens law we obtain the following expression for u_1 and u_2 from Equations (H3) and (H4):

$$u_1 = \frac{f(p\sigma + av)}{(p\sigma + av - f\sigma - fa)} \qquad \text{(H5)}$$

$$u_2 = \frac{f(p\sigma - av)}{(p\sigma - av - f\sigma + fa)} \qquad \text{(H6)}$$

where f = focal length of the camera, σ be the upper bound on the radius of the blur circle so that the point is almost in perfect focus (in otherwords, the point remains in almost perfect focus as long as its radius of blur circle is less than or equal to σ), a = width of aperture, v = distance between the lens and the image plane, p is a positive constant.

The depth of field of the camera is defined as follows (refer to Figure (H1)):

$$D = u_1 - u_2 \tag{H7}$$

If the aperture stop is very close to the lens, then $p \cong 0$, with this assumption an expression for the depth of field can be derived in terms of Equations [H5, H6] as follows:

$$D = u_1 - u_2 = \frac{fav}{(-f\sigma) + (-fa + av)} + \frac{fav}{(-f\sigma) - (-fa + av)} \tag{H8}$$

On simplification Equation (H8) leads to the following:

$$D = \frac{2f^2 av\sigma}{a^2 f^2 - 2a^2 vf + a^2 v^2 - f^2 \sigma^2} \tag{H9}$$

$$\Rightarrow \quad D = \frac{2f^2 v\sigma}{a(f^2 - 2vf + v^2 - f^2 \frac{\sigma^2}{a^2})} \tag{H10}$$

Width of the aperture *a* and the upper bound for radius of blur circle σ can be represented in terms of the width of the lens as follows:

$$a = \alpha r \tag{H11}$$

$$\sigma = \beta r \tag{H12}$$

where α, β are positive constants.

Combining Equations [H10-H12], expression for the depth of field can be written as follows:

$$D = \frac{2f^2 v\beta}{\alpha(f^2 - 2vf + v^2 - f^2(\frac{\beta^2}{\alpha^2}))} \qquad \text{(H13)}$$

When $\alpha \to 0$, it corresponds to a pin hole camera and when $\alpha \to 1$ it corresponds to wide aperture camera.

Qualitative plots depicting the behavior of the depth of field D as a function of the width of the aperture a and the initial distance to which the camera is focused to, are shown in Figures [H2,H3].

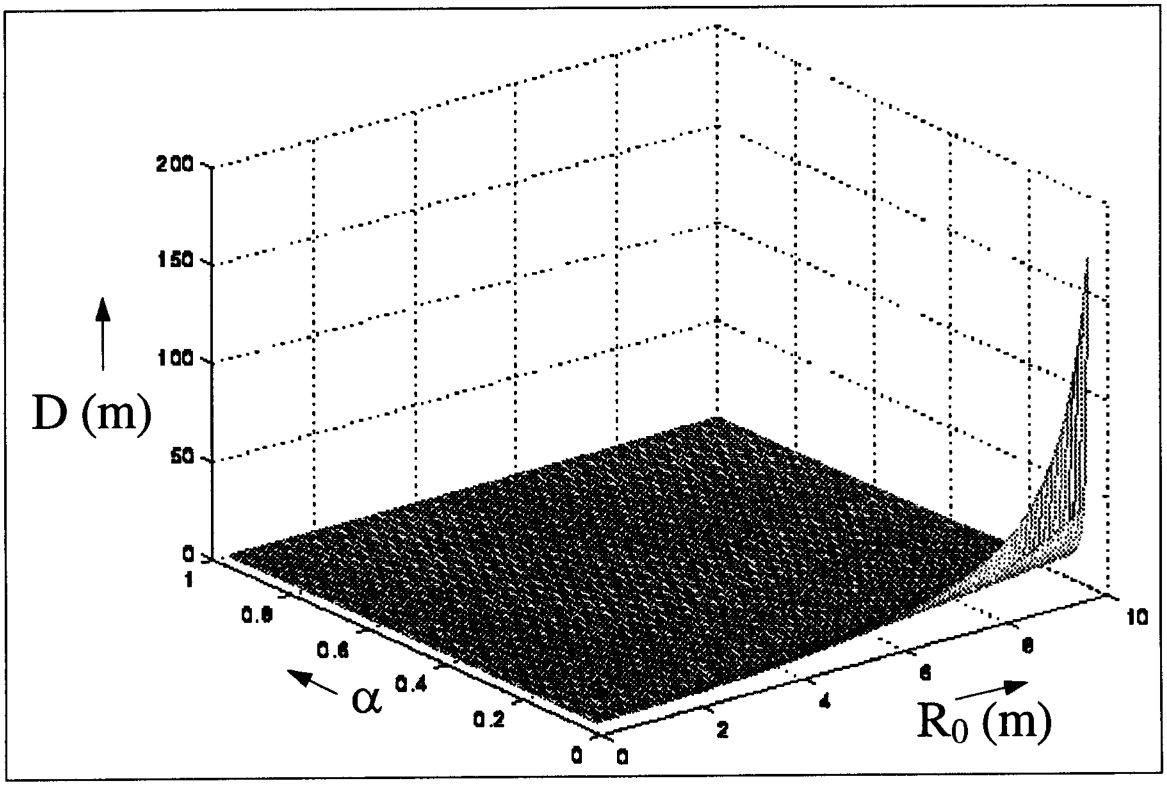


Figure (H2) : A qualitative plot of the depth of field (D) as function of the aperture width (α) and the initial focus adjustment (R_0): f = .105m, β = 0.0005

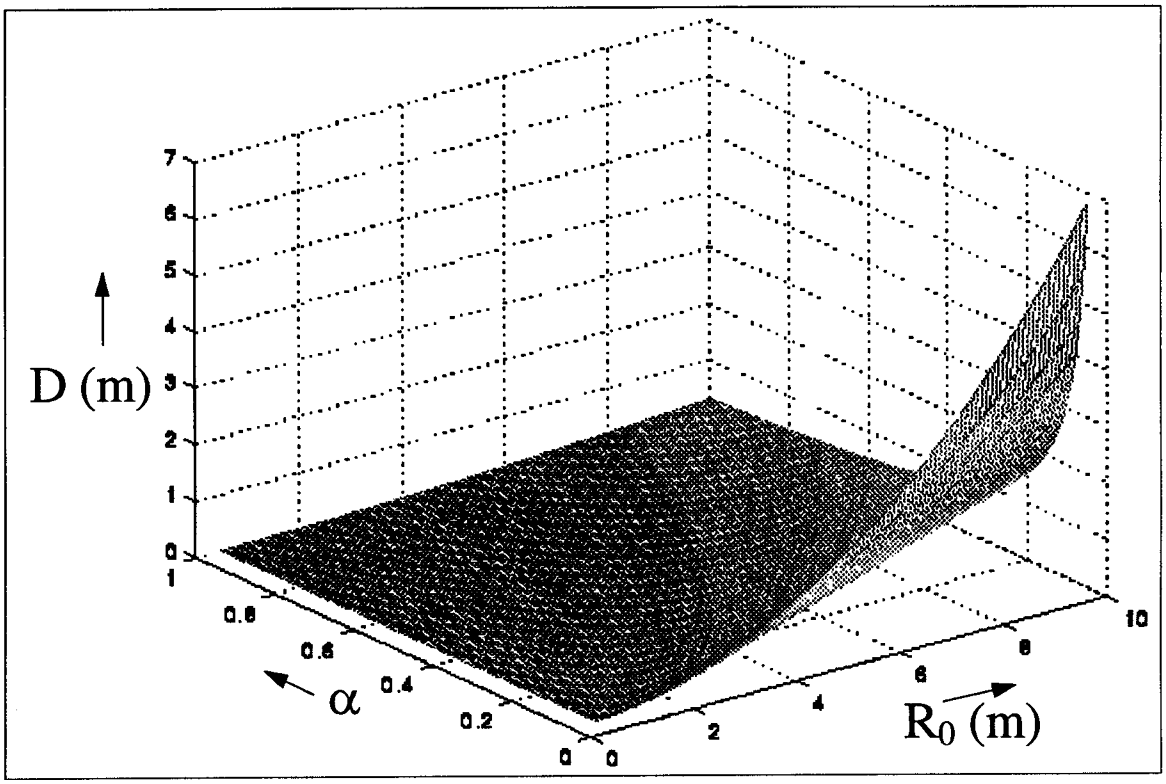


Figure (H3) : A qualitative plot of the depth of field (D) as function of the aperture width (α) and the initial focus adjustment (R_0): f = .315m, β = 0.0005

From Figure (H2, H3), it can be seen that the depth of field increases as the width of the aperture is decreased, also it increases with the initial focus adjustment namely R_0. The depth of field is minimum when the aperture is fully open.

From Figure (H2,H3), it can be seen that as the desired clearance increases, the depth of field also increases. In other words for larger desired clearances R0, there will be larger error in maintaining desired clearances. This is not a very serious drawback for navigation tasks (especially maintaining clearances) as the tolerance in error is also large for large clearances.

References

[1] C. Thorpe, M. H. Hebert, T. Kanade and S. A. Shafer (1988), "Vision and Navigation for the Carnegie Mellon Navlab," *IEEE Transactions of Pattern Analysis and Machine Intelligence*, PAMI-10, No. 3, pp. 362-373.

[2] E. D. Dickmanns, B. Mysliwetz and T. Christians (1990), "An Integrated Spatio-Temporal Approach to Automatic Visual Guidance of Autonomous Vehicles," *IEEE Transactions on Systems, Man, and Cybernetics*", Vol. 20, No. 6, pp. 1273-1284.

[3] D. Coombs and K. Roberts (1992), ""Bee-Bot":Using Optical Flow to Avoid Obstacles", *Proc. SPIE Conf. on Intelligent Robots and Computer Vision XI: Algorithms, Techniques and Active Vision*, Boston, MA., pp. 714-721.

[4] A. M. Waxman, LeMoigne and B. Srinivasan (1987), "A Visual Navigation system for Autonomous Land Vehicles," *IEEE Journal of Robotics Automation*, RA-3, No. 2, pp. 124-141.

[5] G. S. Young, T. H. Hong, M. Herman and J. C. S. Yang (1992), "Obstacle Detection and Terrain Characterization using Optical Flow without 3D Reconstruction," Proc. *SPIE.* Vol. 1825, Intelligent Robots and Computer Vision XI, pp. 561-568.

[6] R. C. Nelson and J. Aloimonos (1989), "Obstacle Avoidance using Flow Field Divergence," IEEE *Transactions of Pattern Analysis and Machine Intelligence*, PAMI-11, No. 10., pp. 1102-1106.

[7] J. Santos-Victor, G. Sandini, F. Curotto, and Gabribaldi (1993), "Divergent Stereo for Robot Navigation: Learning from Bees," Proc. of *Computer Vision Pattern Recognition '93*, pp. 434-439.

[8] M. Tistarelli and G. Sandini (1990), "On Advantages of Polar and Log-Polar Mapping for Direct Estimation of Time-To-Contact from Optical Flow," Technical report *LIRA-TR3/90*, University of Genoa, Italy.

[9] D. Raviv (1992), "A Quantitative Approach to Looming," Technical Report *NISTIR 4808*, National Institute of Standards and Technology, Gaithersburg, Maryland.

[10] M. Brady and H. Wang (1992), "Vision for Mobile Robots," *Philosophical Transactions of Royal Society*, London, B337, pp. 341-350.

[11] D. N. Lee (1976), "A Theory of Visual Control of Braking Based on Information About time-to-collision,"*Perception*, Vol. 5, 437-459.

[12] R. Cipolla and A. Blake (1992), "Surface Orientation and Time-to-Contact from Image Divergence and Deformation," In Proc. of *European Conference on Computer Vision,* pp. 187-202.

[13] J. J. Gibson (1979), *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston.

[14] J. J. Gibson (1950a), *The Perception of Visual World,* Houghton Mifflin, Boston.

[15] D. N. Lee (1980), "The Optic Flow Field: The Foundation of Vision," *Philosophical Transactions of Royal Society of London*, B290, pp. 169-179.

[16] T. G. R.Bower, J. M. Broughton and M. K. Moore (1970), "The Coordination of visual and tactual input in infants," *Perception and Psychophysics*, Vol. 8, pp. 51-53.

[17] W. Schiff, J. A. Caviness and J. J. Gibson (1962), "Persistent Fear Responses in Rhesus Monkeys to the Optical Stimulus of "Looming"," *Science*, Vol. 136, pp. 982-983.

[18] H. H. Nagel (1987), "On the Estimation of Optical Flow: Relations between Different Approaches and Some New results," *Artificial Intelligence*, Vol. 33, pp. 299-324.

[19] A. Singh (1991), *"Optic Flow Computation: A Unified Perspective,"* IEEE Computer Society Press.

[20] J. L Barron, D. J. Fleet and S. S. Beauchemin (1992), "Performance of Optical Flow Techniques," *RPL-TR-9107*, Dept. of Computing and Information Science, Queens University, Canada.

[21] R. Bajcy and L. Liberman (1976), "Texture Gradients as Depth Cue," *Computer Graphics and Image Processing*, Vol. 5, pp. 52-67.

[22] J. Sato and R. Cipolla (1994), "Extracting the Affine Transform from Texture Moments," *TR-167*, Dept. of Engineering, Univ. of Cambridge, England.

[23] K. Joarder and D. Raviv (1994), "A New Method to Calculate Looming from Surface Texture," In Proc. of Proc. of *Computer Vision Pattern Recognition-94*, Seattle, WA., pp. 777-780.

[24] M. Subbarao and N. Gurumoorthy (1988), "Depth Recovery from Blurred Edges," In Proc. of Proc. of *Computer Vision Pattern Recognition-88*, pp. 498-503.

[25] S. K. Nayar (1992), "Shape from Focus System," In Proc. of Proc. of *Computer Vision Pattern Recognition-92*, pp. 302-308.

[26] A. Pentland (1987), "A New Sense for Depth of Field," IEEE *Transactions of Pattern Analysis and Machine Intelligence*, PAMI-9, pp. 523-531.

[27] J. J. Koenderink (1981), *Solid Shape*, MIT Press, Cambridge, MA.

[28] A. Rosenfeld and M. Thurston (1971), "Edge and Curve Detetction for Visual Scene Analysis," *IEEE Transactions on Computers*, C-20, pp. 562-569.

[29] A. Klinger (1971), Pattern and Search Statistics, in J. S. Rustagi (ed.*) Optimizing Methods in Statistics*, Academic press.

[30] S. Tanimoto and A. Klinger (1980) (eds.), *Structured Computer Vision*, Academic Press, NY.

[31] H. Samet (1990), *Applications of Spatial Data Structures*, Addison-Wesley, MA.

[32] S. G. Mallat (1989), "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," IEEE *Transactions of Pattern Analysis and Machine Intelligence*, PAMI-11, No. 7.

[33] G. Strang (1989), "Wavelets and Dilation Equations: A Brief Introduction," *SIAM Review*, Vol. 31, No. 4, pp. 614-627.

[34] T. Lindeberg (1993), "Detecting Salient Blob-Like Image Structures and their scales with a scale space primal sketch: A Method for Focus of Attention," *International Journal of Computer Vision*, Vol. 11, No. 3, pp. 283-318.

[35] V. Topkar, B. Kjell and A. Sood (1990), "Object Detection Using Scale Space,*" In Proc. SPIE, Vol. 1293, Applications of Artificial Intelligence VIII*, pp. 2-13.

[36] T. Lindeberg and J. O. Eklundh (1991), "Analysis of Aerosol Images Using the Scale Space Primal Sketch," *Machine Vision Applications*, Vol. 4, pp. 135-144.

[37] R. Hummel and R. Moniot (1989), "Reconstructions from Zero Crossings in Scale Space," IEEE *Transactions on Acoustics, Speech and Signal Processing*, Vol. 37, No. 12, pp. 2111-2130.

[38] T. Lindeberg and J. Garding (1993), "Shape from Texture from a Multiscale Perspective," Technical Report, Royal Institute of Technology, S-100 44, Stockholm, Sweden, Jan 1993.

[39] Y. Lu and R. C. Jain (1989), "Behavior of Edges in Scale Space," IEEE *Transactions of Pattern Analysis and Machine Intelligence*, PAMI-11, No. 4, pp. 337-356.

[40] A. P. Witkin (1983), "Scale-Space Filtering," *Proc. 8th. IJCAI* (Karlsruhe, Germany), pp. 1019-1022.

[41] J. Baubad, A. P. Witkin, M. Baudin and R. Duda (1986), "Uniqueness of the Gaussian Kernel for Scale-Space Filtering," IEEE *Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8, pp. 26-33.

[42] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink and M. A. Viergever (1991), "Scale and the Differential Structure of Images," *Image Vision and Computing*, Vol. 10, No. 6, pp. 376-388.

[43] J. J. Koenderink and A. J. van Doorn (1987), "Representation of Local Geometry in the Visual System," *Biological Cybernetics*, Vol. 55, pp. 367-375.

[44] J. J. Koenderink (1984), "The Structure of Images," *Biological Cybernetics*, Vol. , pp. 363-370.

[45] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink and M. A. Viergever (1989), "Scale Space: Its Natural Operators and Differential Invariants,"

[46] P. J. Burt (1981), "Fast Filter Algorithms for Image Processing," *Computer Vision Graphics and Image Processing*, Vol. 16, pp. 20-51.

[47] J. L. Crowley (1981), A Representation for Visual Information, Ph. D. Disseration Robotics Institute, Carnegie Mellon University, Pennsylvania, USA.

[48] C. A. Sher and A. Rosenfeld (1989), "Detecting and Extracting Compact Textured Regions using Pyramids," *Image and Vision Computing*, Vol. 7, No. 2, pp. 129-134.

[49] M. Subbarao (1987), "Direct Recovery of Depth-map I," *Proc. IEEE Computer Society Workshop on Computer Vision*, FL, pp. 58-65.

[50] S. R. Kundur and D. Raviv (1995), "An Image-Based Texture-Independent Visual Motion Cue for Autonomous Navigation," Technical Report *NISTIR 5567*, National Institute of Standards and Technology, Gaithersburg, Maryland.

[51] P. Brodatz (1966), *Textures: A Photographic Album for Artists and Designers*, Dover Publications, New York.

[52] R. M. Haralick, K. Shanmugam, and I. Dinstein (1973), "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-3, No. 6, pp. 610-621.

[53] R. W. Conners and C. A. Harlow (1980), "A Theoretical Comparison of Texture Algorithms," *IEEE Transactions on Pattern Recognition and Machine Analysis*, Vol. PAMI-2, No. 3, pp. 204-222.

[54] J. S. Weszka, Dyer C. R. and A. Rosenfeld (1976), "A Comparative Study of Texture Measures for Terrain Classification," *IEEE Transactions on Systems, Man and Cybernetics*, SMC-6, pp. 269-285.

[55] S. R. Kundur and D. Raviv (1995), "Change Detection Algorithm for Machine Vision Applications," Technical Report *NISTIR 5670*, National Institute of Standards and Technology, Gaithersburg, Maryland.

[56] A. Pentland, S. Scherock, T. Darrell and B. Girod (1994), "Simple Range Cameras Based on Focal Error," *Journal of Optical Society of America*, A, pp. 2925-2935.

[57] S. K. Nayar, M. Watanabe and M. Noguchi (1994), "Real-Time Focus Range Sensor," Technical Report, CUCS-028-94, Dept. of Computer Science, Columbia university, New York.

[58] L. Kaufman and P. J. Rousseeuw (1990), *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons Inc., New York.

[59] H. C. Romesburg (1990), *Clustering Analysis for Researchers*, Robert E. Krieger Publishing Company, Malabar, FL.

[60] S. R. Kundur and D. Raviv (1996), "Novel Active-Vision-Based Motion Cues for Local Navigation," Technical Report *NISTIR 5791*, National Institute of Standards and Technology, Gaithersburg, Maryland.

[61] A. Pentland (1987), "A New Sense of Depth of Field," *IEEE Transactions on Pattern Analysis and machine Intelligence*, PAMI-9, No. 4, pp. 523-531.

[62] M. Subbarao (1987), "Direct Recovery of Depth-Map I: Differential Methods," In *Proc. of the IEEE Computer Society Workshop on Computer Vision*, Miami Beach, FL., pp. 58-65.

[63] M. Abramowitz and I. A. Stegun (eds.) (1964), *Handbook of Mathematical Functions*, Applied Mathematics Series, National Bureau of Standards.

[64] S. R. Kundur and D. Raviv (1996), "Novel Active-Vision-Based Visual-Threat-Cue for Autonomous Navigation Tasks," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, San Francisco, California.