NAT'L INST. OF STAND & TECH R.I.C.
A11104 938912

NIST
PUBLICATIONS

NISTIR 5791

Novel Active-Vision-Based Motion Cues for Local 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

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

QC
100
.U56
NO.5791
1996

NIST

ISTIR 5791

Novel Active-Vision-Based Motion Cues for Local 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

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

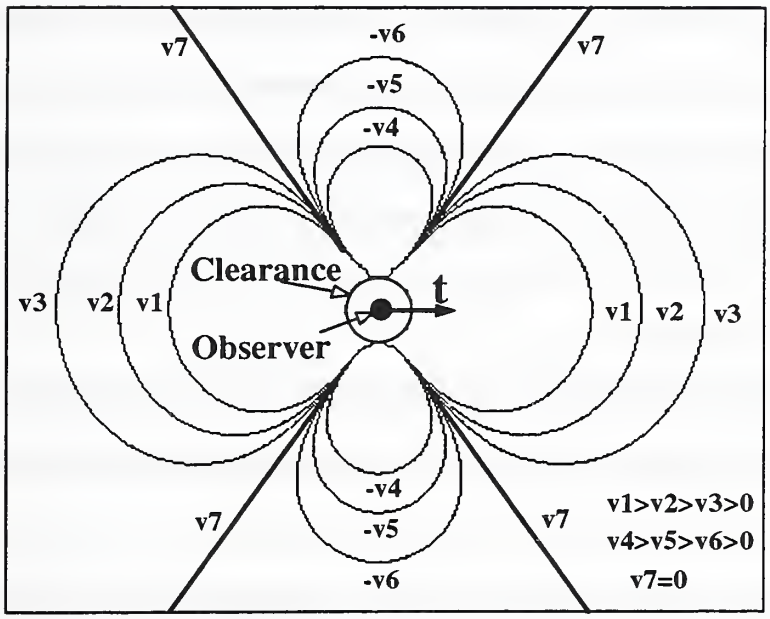
NATIONAL INSTITUTE OF STANDARDS
AND TECHNOLOGY
Arati Prabhakar, Director

Novel Active-Vision-Based Motion Cues for Local Navigation*

Sridhar R. Kundur[1] and Daniel Raviv[1,2]

[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

In the absence of *a-priori* information about the environment, an autonomous mobile robot relies on sensory information to make local judgments about its surrounding. Generation of a local collision-free path based on sensory data plays an important role in the control of the robot's interaction with the environment. This paper presents a novel approach based on *active-vision* paradigm, for generating local collision-free paths for mobile robot navigation, in indoor as well as outdoor environments. Two *measurable* visual motion cues that provide, some measure for a *relative change* in range as well *clearance* between a 3D surface and a *fixated* observer in motion are described. These visual cues are *independent* of the 3D environment and need no *a-priori* knowledge about it. For each visual motion cue, there is a *visual field* surrounding the moving observer. In other words, there are imaginary 3D surfaces attached to the observer that move with it, each of which correspond to a value of the cue. These visual fields can be used to *demarcate* regions around a moving observer into safe and danger zones of varying degree, which is suitable for making local decisions about the steering as well as speed commands to the mobile robot.

We describe a practical method to *extract* these cues from a sequence of images. This approach needs no feature tracking between images and almost no camera calibration.

Key Words: Active Vision, Visual Navigation, Collision Avoidance, Clearance

1 Introduction

Autonomous mobile robots may need several inputs such as the initial point, final destination, the trajectory to be traversed, model of the environment, etc., to accomplish navigation tasks. The problem of generating trajectories for mobile robot navigation in the presence of *a-priori* information about the environment has drawn the attention of several researchers (see for example [1]). If the robot has no (or partial) *a-priori* knowledge about the environment, autonomous navigation tasks become relatively difficult, as there may be obstacles in the robot's predetermined trajectory. To overcome this difficulty caused by the absence of *a-priori* information, additional on-board sensors such as vision, sonar, laser, etc., are necessary to obtain information about the robot's surrounding. In other words, the sensory data obtained from these on-board sensors is used to guide the robot. The task of generating paths around obstacles with minimum deviation from the global trajectory is often referred to as local navigation.

The area of generating collision-free local paths for mobile robot navigation has drawn the attention of several researchers in the recent past (see for example[2-14]).The idea of virtual forces acting on a robot usually referred to as potential field approach was suggested by Khatib [2]. According to this approach, all obstacles exert imaginary repulsive forces, while target exerts an imaginary attractive force on the robot. A resulting force vector F, which is the vector sum of the repulsive as well as the attractive forces on the robot for a given position. Using F, the next position of the robot in the environment is calculated, and the algorithm is repeated. Several other off-line path-planning approaches were developed based on this idea of virtual forces [3]. All these approaches are based on

the assumption that the model of the 3D world is known *a-priori* in which simple re-defined geometric shapes represent obstacles and the robot's path is generated off-line. Most approaches were not tested on real-sensory data.

A potential field approach of path planning was implemented on experimental mobile robots [4,5]. In this implementation each ultrasonic range reading is treated as a repulsive force vector. If the resultant of these repulsive forces exceeds a certain threshold value the robot stops and changes its direction of motion. Similar experimental procedure was reported in [6].

A virtual force field approach to motion planning in the presence of unknown static obstacles is presented in [7]. This method employs a 2D Cartesian histogram grid as a world model, that can be updated periodically with range data obtained by on-board sonar sensors. This 2D histogram generates a virtual system of forces acting on the vehicle. According to this approach all obstacles exert a repulsive force on the vehicle, whereas the target exerts an attractive force on it. Thus the vehicle is pulled towards the goal or pushed away from the obstacles.

One problem associated with the virtual force approach is that much of the valuable information about the local obstacles is lost due to the fact that it compresses all information into a single resultant force [8]. Problems arise when the clearance between the obstacles is small, in which case the robot sees the obstacles as a single obstacle and passage between them becomes impossible. For example, navigating through a doorway may become impossible as the repulsive forces from both the sides of the doorway result in a force that repels the robot from the doorway.

In order to overcome the difficulty of loss of information about local obstacles, an approach called the Vector Field Histogram is presented in [8]. On the basis of an occupancy grid, an obstacle density map is calculated for each of the polar sectors. Occupied cells contribute to this obstacle density map according to their proximity to the vehicle and also their degree of certainty. Using this map the robot navigates.

For mobile robot navigation tasks such as collision avoidance, maintenance of clearances, etc., there is no need to know the exact geometrical description of the obstacles in the environment. In other words, the robot does not need to know if the obstacle is a chair, table, tree, mountain, etc. This paper describes a new active-vision-based approach for local navigation decisions. The idea is based on *measurable* visual motion cues and their associated visual fields. We describe two such cues that provide a measure for a change in *range* as well as the *absolute clearance*. For each visual motion cue, there is a visual field associated with it, i.e., there is a family of imaginary surfaces attached to the observer and is moving with it. All the points that lie on a particular surface produce the same value of the cue. These visual motion cues can be used to demarcate the regions around a moving observer into safe and danger regions of varying degree which is very useful for local autonomous mobile robot navigation tasks such as collision avoidance, maintenance of clearance, etc.

This paper is organized as follows: section 2 describes the Visual Threat Cue (VTC), section 3 describes the Temporal Visual Threat Cue (TVTC), section 4 describes local path planning using the VTC as well as the TVTC. Section 5 the process of extraction of

the VTC and the TVTC from images, section 6 presents the experimental details. Section 7 describes the results and analysis.

2 The Visual Threat Cue (VTC)

2.1 Definition

Mathematically we define the VTC, for $R > R_0$ as follows [16]:

$$VTC = -R_0 \frac{\frac{d}{dt}(R)}{R(R - R_0)} \tag{1}$$

where R is the range between the observer and a point on the 3D surface, d(R)/dt is the differentiation of R with respect to time and R_0 is the desired minimum clearance and has the same units as R. The robot may be considered to be enclosed in this clearance region. Note that the units of the VTC are [time^{-1}].

Motivation for the definition of the VTC (Equation (1)) is presented in Appendix A

2.2 Visual Fields of the VTC

In this section we provide simulation results to show the location of points beyond the desired minimum clearance R_0 (i.e., $R > R_0$) in 3D space around an observer in motion, that have the same value of the VTC for a given motion of the camera. The VTC 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 VTC. The points that lie on a relatively smaller surface corresponds to a relatively larger value of VTC, indicating a relatively higher threat of collision. The VTC value near the clearance sphere is maximum. For the hemisphere in front of the observer

the VTC value approaches positive infinity as R approaches R_0. A positive value of the VTC corresponds to the region in front of the observer and a negative value corresponds to the region in back of the observer. A section of this visual field is shown in Figure (1). Note that this field is not a sphere in 3D. However it is symmetric about the instantaneous translational vector **t**.

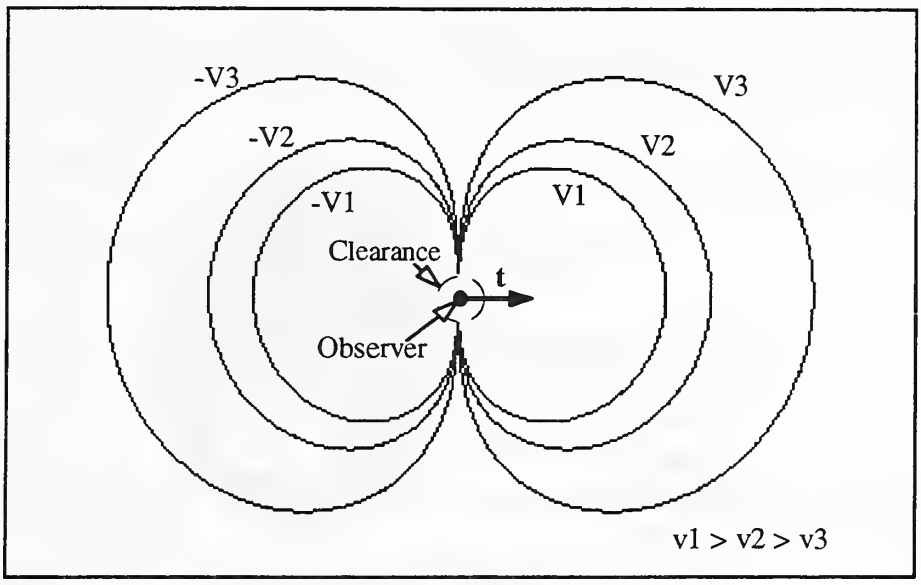


Figure (1): Cross section of the VTC, **t** is the instantaneous translational vector V_1, V_2, V_3 are positive values of the VTC

2.3 VTC as A Sensory Feedback Signal

The VTC divides the 3D space around the observer in motion into 3D regions as shown in Figure (1). Thus one can demarcate the region around an observer into safe and danger zones of varying degree (see Figure (2a)) A positive value indicates a decrease in relative range, and if this value exceeds the threshold for a safe zone then appropriate

control action has to be taken based on the actual value to accomplish the tasks such as collision avoidance, maintenance of clearance, etc.

The visual field associated with the VTC is analogous to the equipotential surface of an electric dipole [23]. This analogy might lead to a new direction in vision-based navigation systems by establishing a theoretical connection, between a well established scientific field such as Electromagnetics and the developing area of visual navigation.

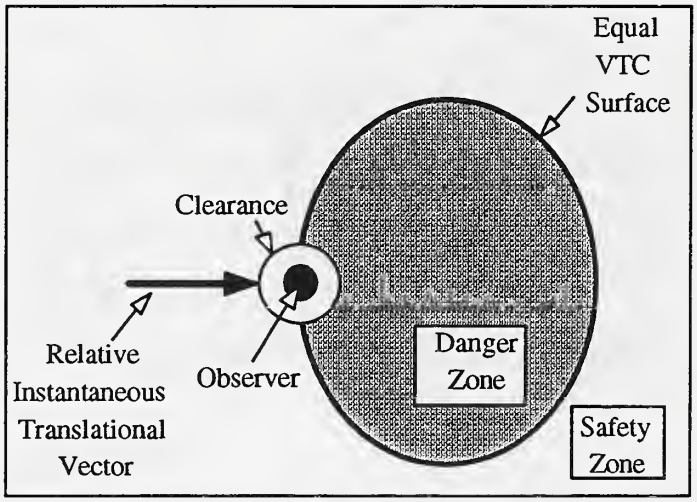


Figure 2(a): Qualitative VTC

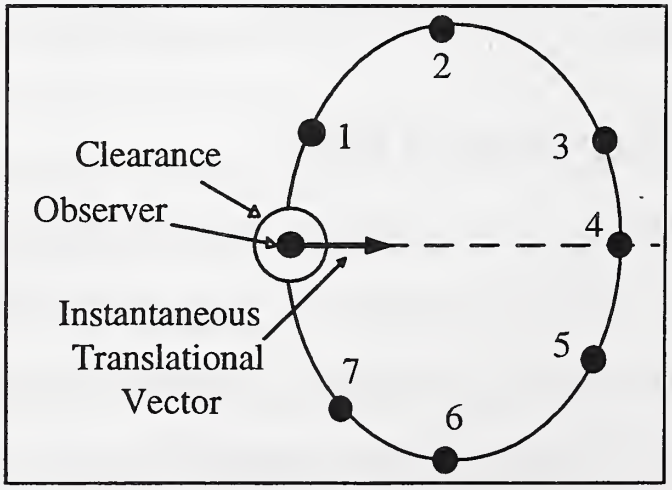


Figure (2b): Qualitative VTC

The problem in employing the VTC information alone for navigation tasks is explained as follows (refer to Figure (2b)): 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).

3 Temporal Variations of the VTC (TVTC)

3.1 Definition

Based on the definition of the VTC given in Equation (1), the temporal variations in the VTC can be expressed as follows (for $R > R_0$) as follows:

$$TVTC = \frac{d(VTC)}{dt} = \frac{2RR_0\dot{R}^2 + RR_0^2\ddot{R} - R_0R^2\ddot{R} - R_0^2\dot{R}^2}{R^2(R-R_0)^2} \qquad (2)$$

where R is the range between the observer and a point on the 3D surface, $\dot{R}$ is the differentiation of R with respect to time, $\ddot{R}$ is the differentiation of $\dot{R}$ with respect to time, and R_0 is the desired minimum clearance and has the same units as R. The robot may

be considered to be enclosed in this clearance region. Note that the units of the TVTC are [time^{-2}].

3.2 Equal TVTC Surfaces

In this section we provide simulation results to show the location of points beyond the desired minimum clearance R_0 in 3D space around an observer in motion, that have the same value of the TVTC for a motion of the camera. The TVTC 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 (3). There are regions in front and in the back of the observer that produce 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 degrees (see Appendix B).

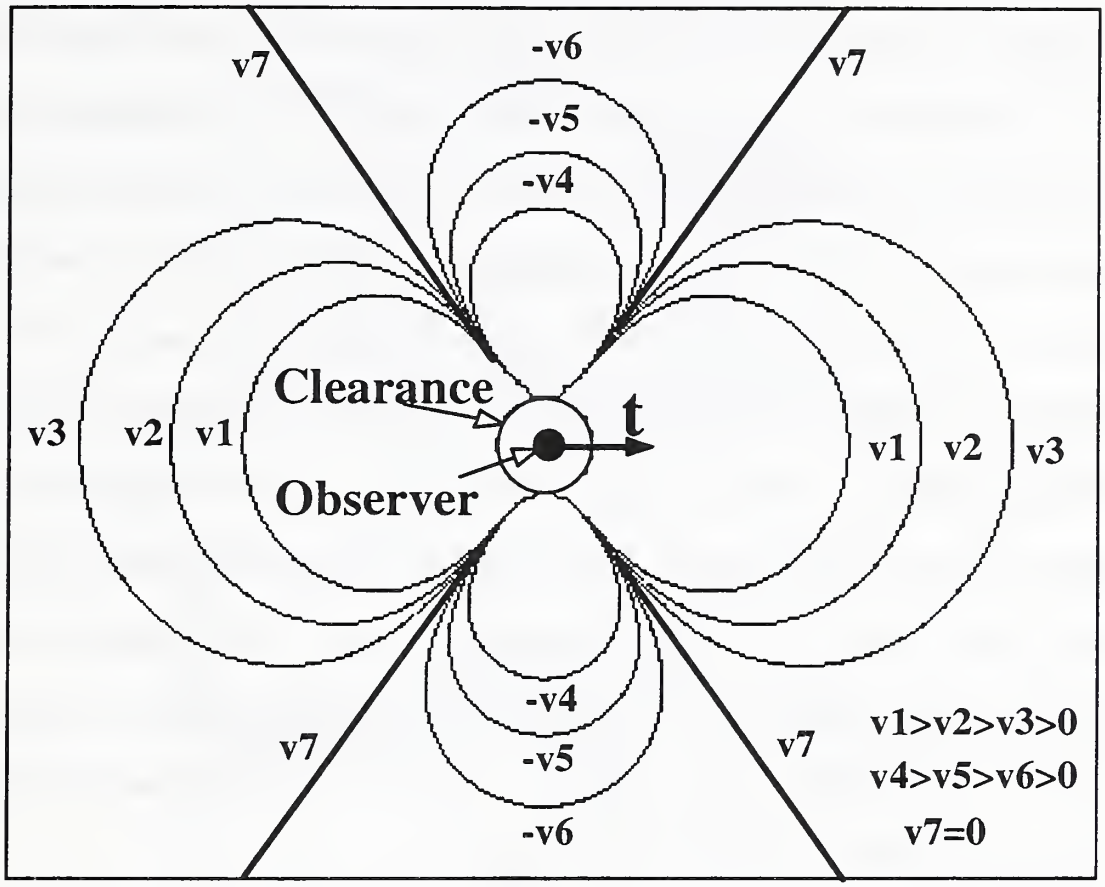


Figure (3): Cross section of the TVTC, **t** is the instantaneous translational vector

4 Local Path Planning Using the VTC and the TVTC

4.1 Demarcation of Space by the VTC and the TVTC

The VTC divides the 3D space around a moving observer into *two* regions as shown in Figure (4a). In Figure (4a): +VTC corresponds to a region where the VTC is positive and -VTC corresponds to the region where the VTC is negative. The TVTC divides the space around it into *two* regions as shown in Figure (4b) (positive in front and back and negative on the sides as shown in Figure (4b)). In Figure (4b): +TVTC corresponds to the region where the TVTC is positive and -TVTC corresponds to the region where the TVTC is negative. These visual motion cues if employed simultaneously could be used to demarcate the region around the observer into four different regions as shown in Figure (4c). In Figure (4c): +VTC, +TVTC corresponds to the region where the VTC and the TVTC are both positive; +VTC, -TVTC corresponds to the region where the VTC is positive and the TVTC is negative; -VTC, +TVTC corresponds to the region where the VTC is negative and the TVTC is positive; -VTC, -TVTC corresponds to the region where the VTC is negative and the TVTC is positive.

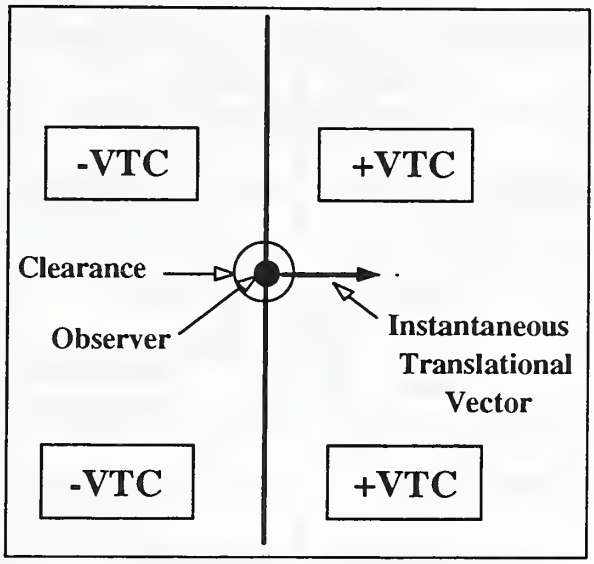


Figure (4a) : Qualitative Demarcation of Space by the VTC

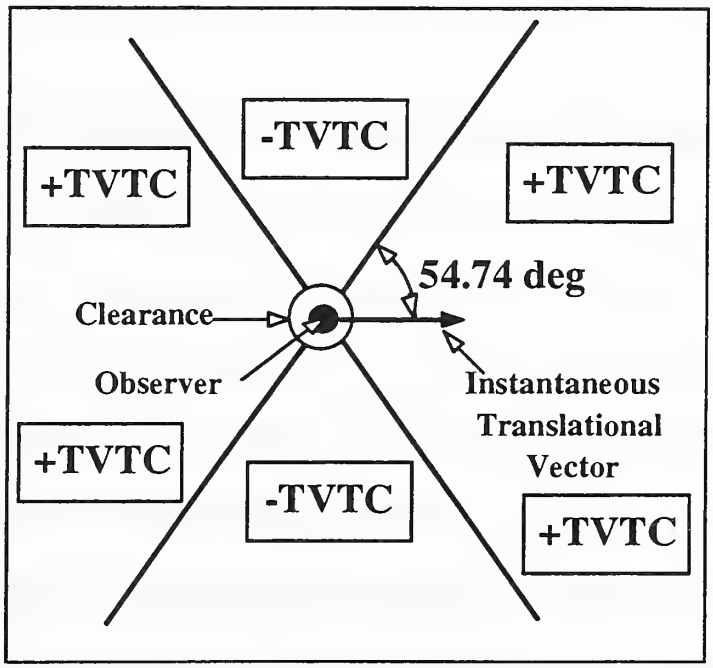


Figure (4b): Qualitative Demarcation of Space by the TVTC

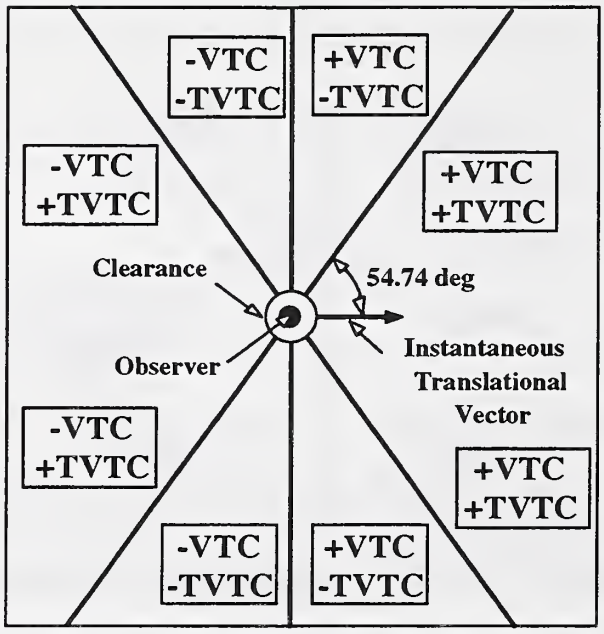


Figure (4c): Qualitative demarcation of the space around the observer

4.2 Local Navigation Decisions using the VTC and the TVTC

In this section we describe how the VTC and the TVTC can be used to make local navigation decisions such as collision avoidance, maintenance of clearance, etc. A flow chart of this process is shown in Figure (5). At each instance of time the observer scans its environment, saccading and fixating at many points, say for example N points, in its surrounding (for a good review of visual fixation and navigation refer to [18, 19]). For all the fixation points, the VTC is extracted and is compared against a threshold value. If the VTC value is less than the threshold value, then the fixation point corresponding to that VTC poses no threat as an obstacle. If the VTC value exceeds the threshold value, then it indicates that the corresponding fixation point is in the danger VTC zone (refer to Figure

(2a)). If the fixation point is in the danger zone, then we compute the TVTC value to identify, if the particular fixation point is a genuine threat, i.e., if the 3D point in space poses a threat of collision (refer to Figure (2b)). Thus at each instance of time the robot computes the TVTC only if the VTC exceeds the threshold value. The fixation points whose VTC values as well as the TVTC values are close to their corresponding threshold values produce a higher threat of collision. Using the values of the VTC and the TVTC it is possible to change the heading vector in the direction that produced lower values of the VTC. In [17] the VTC based control schemes have been designed and implemented to achieve the tasks of collision avoidance and maintenance of desired clearance from an *a-priori* unknown textured surface.

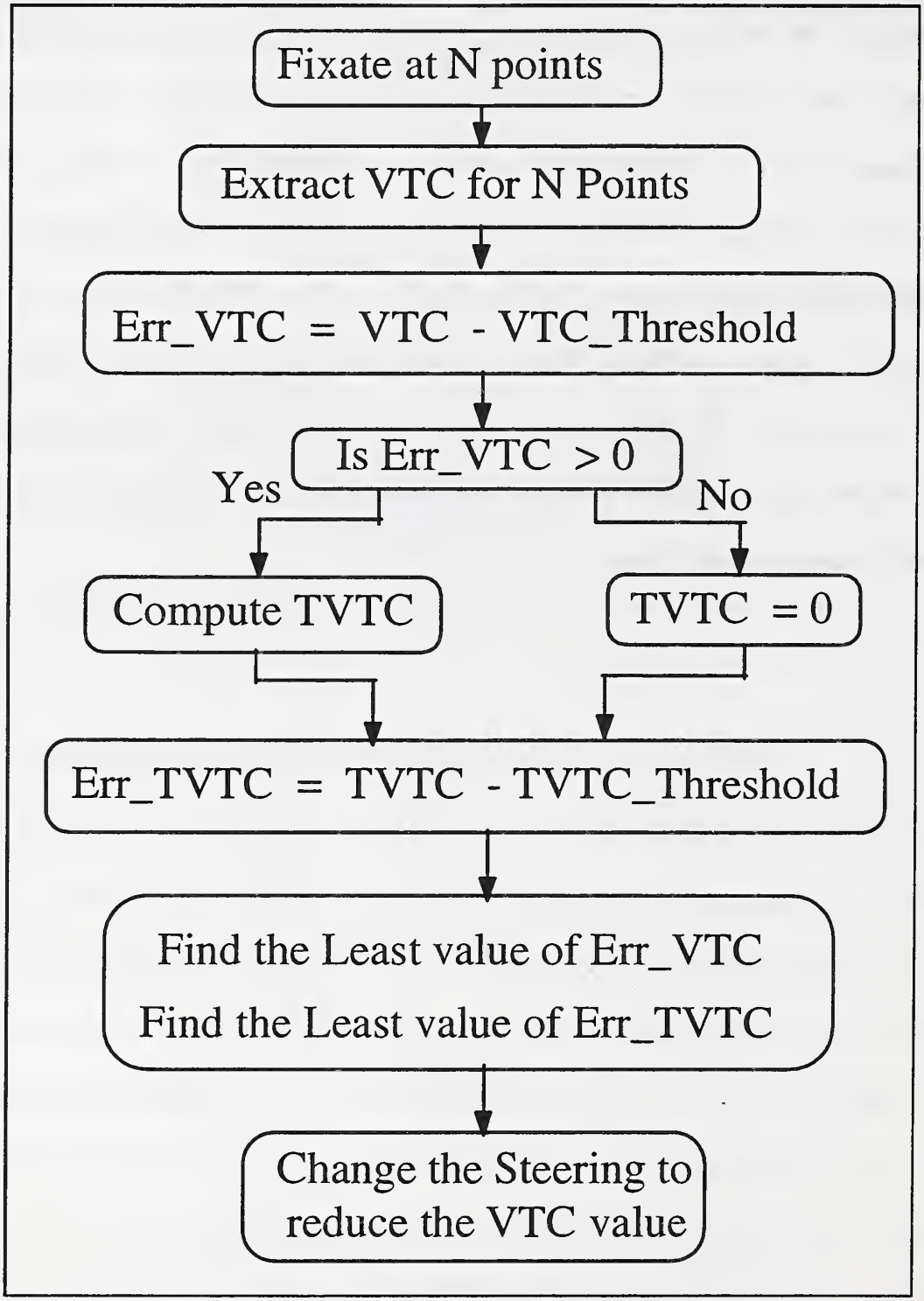


Figure (5): Qualitative local navigation

5 Extraction of the VTC and the TVTC

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

The VTC as well as the TVTC 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.

5.1 Related Work on 3D Surface Reconstruction using Defocused Images

Pentland [21] 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 [22] is another researcher who is actively involved in depth reconstruction approaches using defocused images. In [22] 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.

We present an alternate practical approach to extract the visual cues, by employing a global image variable known as the Image Quality Measurement (IQM). The IQM based approach to extract the VTCs needs no information about the radius of blur circle.

A practical and robust method to extract the VTC from a sequence of images of a 3D textured surface obtained by a *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. The camera is assumed to be of wide aperture.

5.2 Image Quality Measure (IQM)

Mathematically, the IQM is defined as follows [16]:

$$IQM = \frac{1}{|D|} \sum_{x=x_i}^{x_f} \sum_{y=y_i}^{y_f} \left(\sum_{p=-L_c}^{L_c} \sum_{q=-L_r}^{L_r} |I(x,y) - I(x+p, y+q)| \right) \tag{3}$$

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 (3) the IQM is a measure for the dissimilarity of gray level intensity in the image (in our experiments we arbitrarily chose Lc = 5 and Lr = 4).

The advantages of using this measure 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 spatio-temporal smoothing or segmentation, 3) It *does not need a model of the texture* and is suitable for many textures and 4) It is *simple* and can be *implemented in real time on parallel hardware.*

5.3 Extraction of the VTC and the TVTC from IQM

Based on experimental results (indoor as well as outdoor) [16], we observed that the relative temporal changes in the IQM behave in a very similar fashion to the VTC , i.e., $\frac{d(IQM)/dt}{IQM} \cong -R_0 \frac{d(R)/dt}{R(R-R_0)}$. This means that the VTC can be measured using the IQM. The VTC is *independent* of the magnitude of the IQM since it corresponds to relative temporal variations of the IQM. In other words, though the magnitude of the IQM is dependent upon the scene illumination, the VTC extracted from the relative temporal variations of the IQM is independent of the magnitude of the IQM. Experimental results to extract the VTC from the IQM values with out any temporal smoothing is described in [16]. The TVTC can be practically extracted from the VTC by calculating its temporal derivative. Since it involves the second derivative of the measured IQM value with respect to time, we have employed a temporal smoothing scheme for the IQM values, that uses the past measured IQM values to obtain the VTC as well as the TVTC. At any time instant all the IQM values available up to that time instant are used to find a sixth order polynomial that best fits the IQM values in a least squares sense (based on our simulation results of the mathematical model for the radius of the blur circle and the range, we found that the order of the polynomial has to be at least six in order to the minimize the error) (see Appendix C).

6 Experimental Details

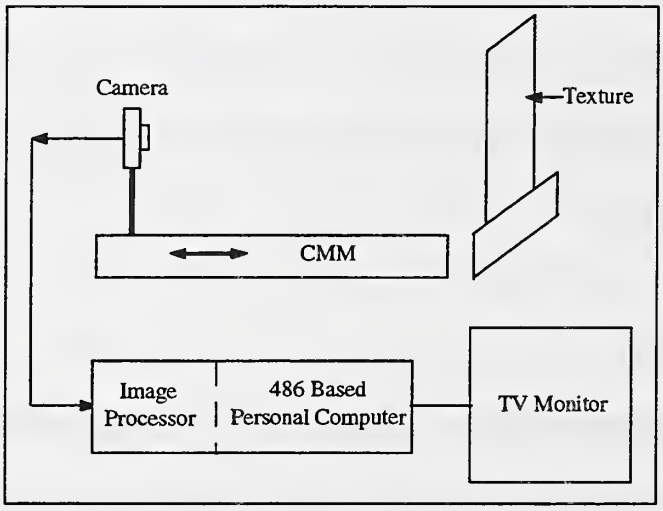
Several experiments were performed to study the variations in the IQM of image sequences in order to extract the VTC as well as the TVTC. The system used in the experiments include a Coordinate Measuring Machine (CMM), a CCD video camera, a 486-based personal computer, ITEX PC-VISION PLUS image processing system and several texture plates from Brodatz's album [15]. A block diagram of the connections is shown in Figure (6a).

6.1 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 (6b). 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 R_0 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 (6a) and Figure (6b) 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 and smooth the IQM, the VTC and the TVTC. For a given

texture, we computed these measures at 71 different distances and this was repeated for 12 different textures. The experimental results along with the textures used are presented in the following section.



Figure(6a): Block diagram of the Experimental setup

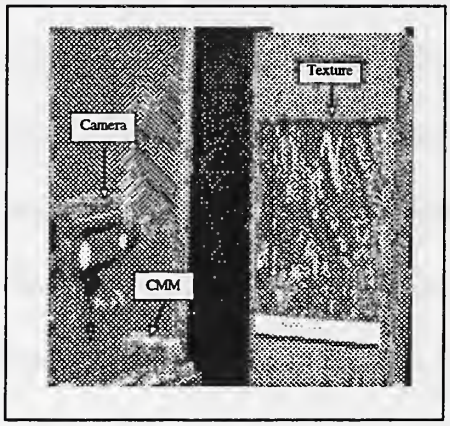


Figure (6b): CMM Set up

7 Results and Analysis

The IQM described in Equation (3) is extracted according to the procedure described in section 6. Since the extraction of the TVTC involves second derivative of IQM, we employed a curve fitting strategy described in Appendix [C]. Using at least six values of the measured IQM values in the immediate past, we fit a sixth order polynomial to the IQM values of the past. This sixth order polynomial is used to compute the VTC as well as the TVTC values.

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 7(a)-18(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 VTC as well as the TVTC is independent of the absolute magnitude of the IQM. Figures 7b-18b). The IQM plots shown in the Figures (7b-18b) are measured values and there is no temporal smoothing.

3. A plot depicting the theoretical VTC and the VTC extracted from the images (Figures 7c-18c).

4. A plot depicting the theoretical TVTC and the TVTC extracted from the images (Figures 7d-18d).

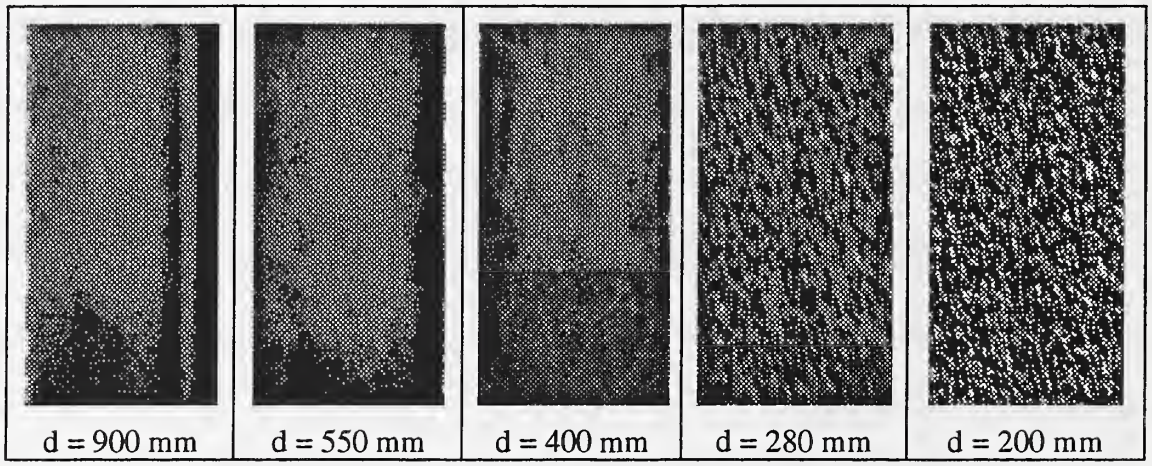


Figure (7a): Image Sequence for Texture D4, d is the relative distance

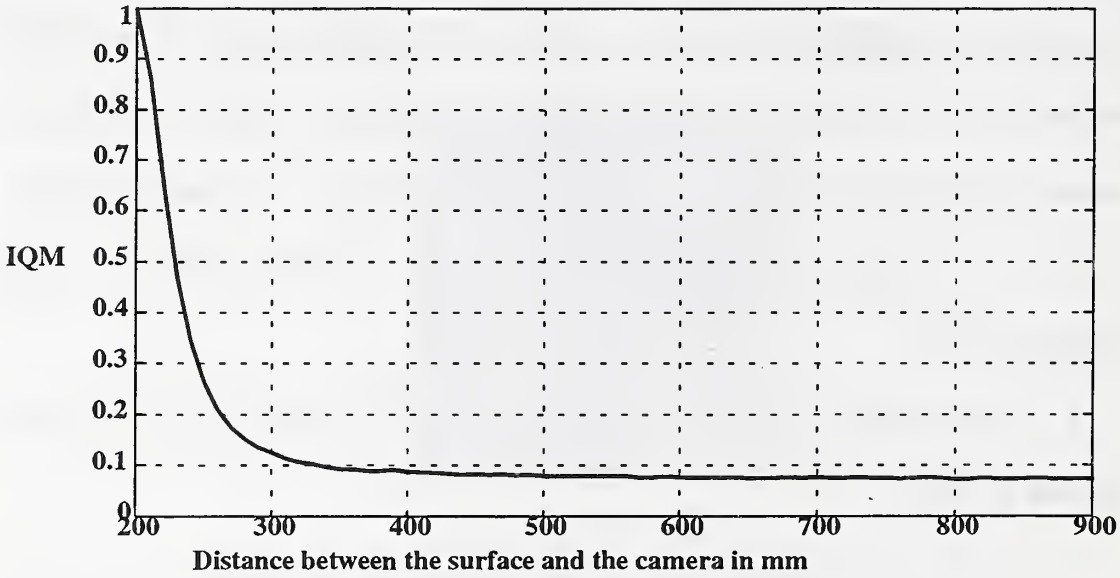


Figure (7b): Measured IQM vs. Distance between the camera and surface for D4

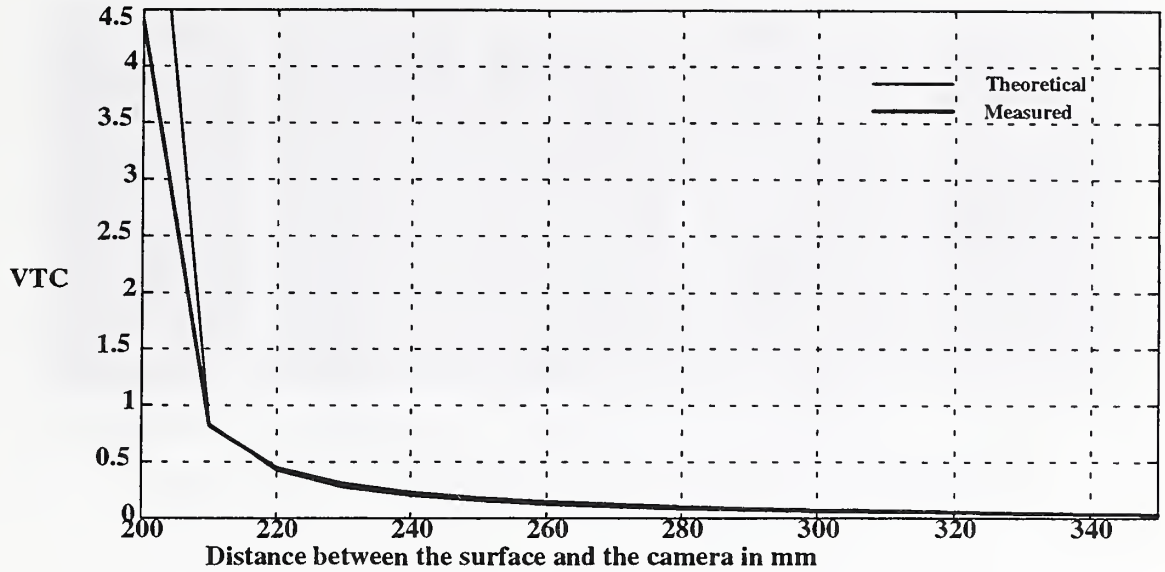


Figure (7c): VTC vs. Distance between the camera and surface for D4

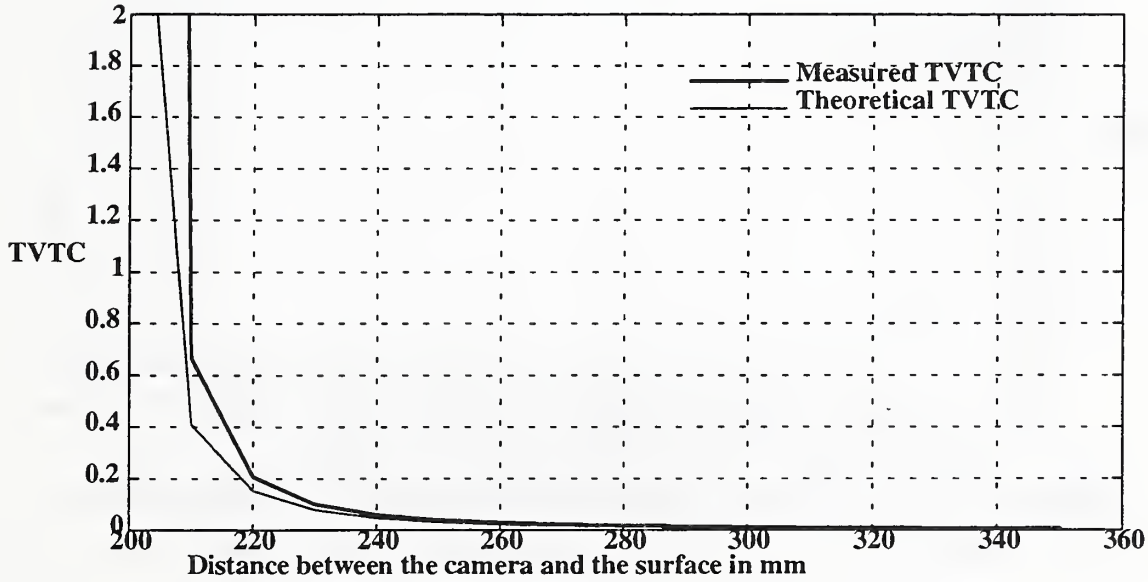


Figure (7d): TVTC vs. Distance between the camera and surface for D4

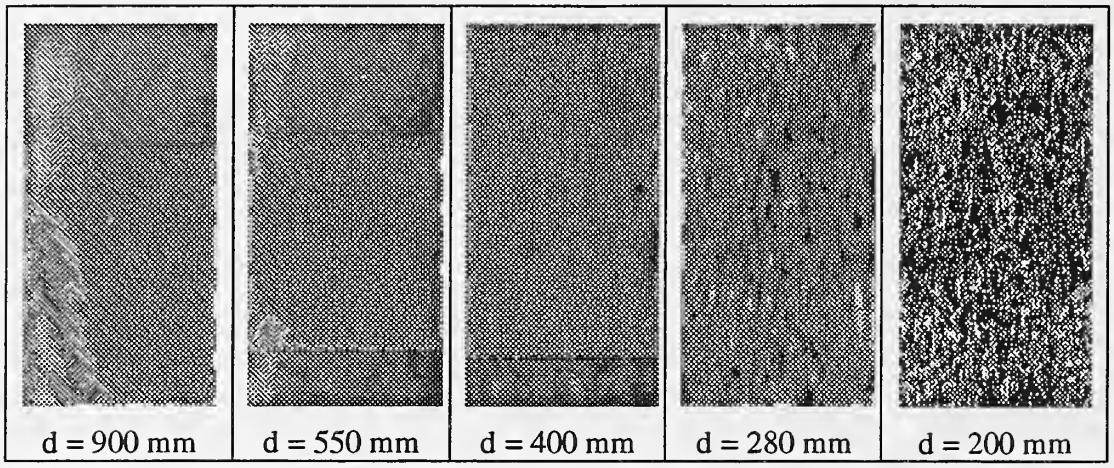


Figure (8a): Image sequence for Texture D9, d is the relative distance

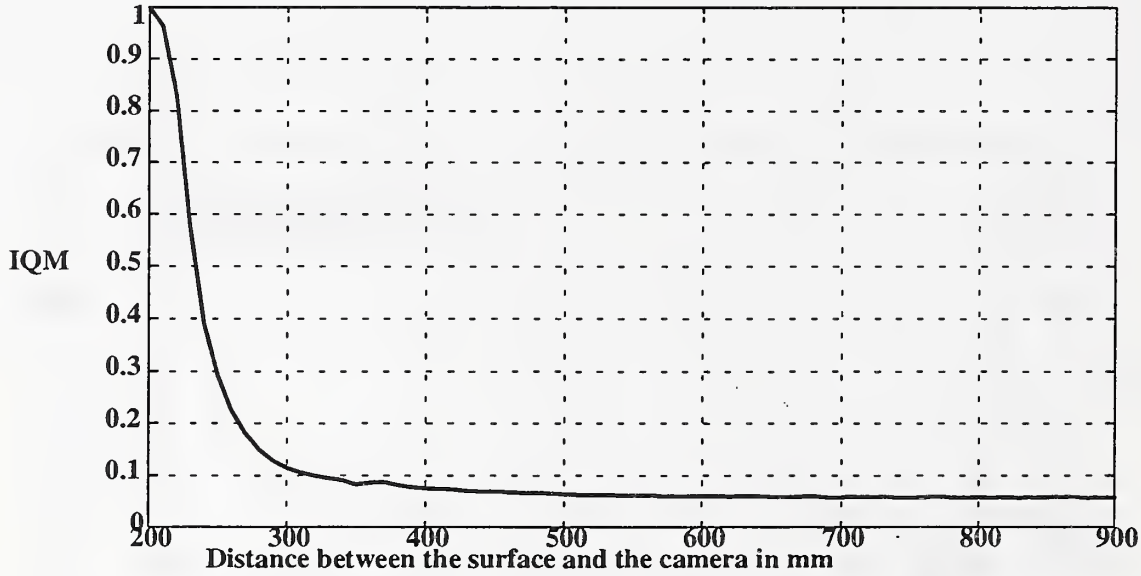


Figure (8b): Measured IQM vs. Distance between the camera and surface for D9

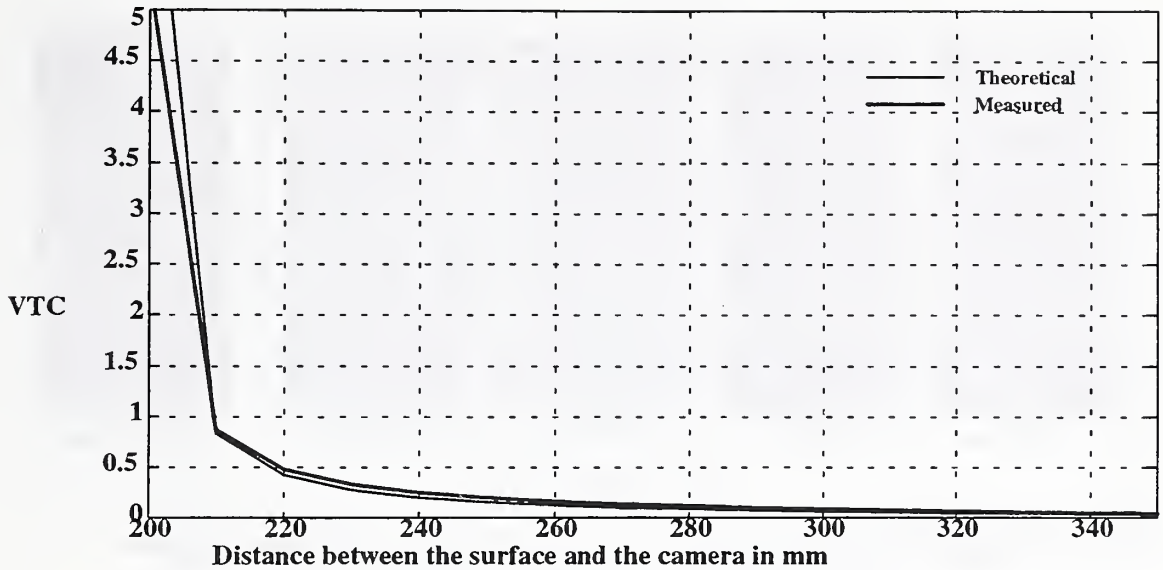


Figure (8c): VTC vs. Distance between the camera and surface for D9

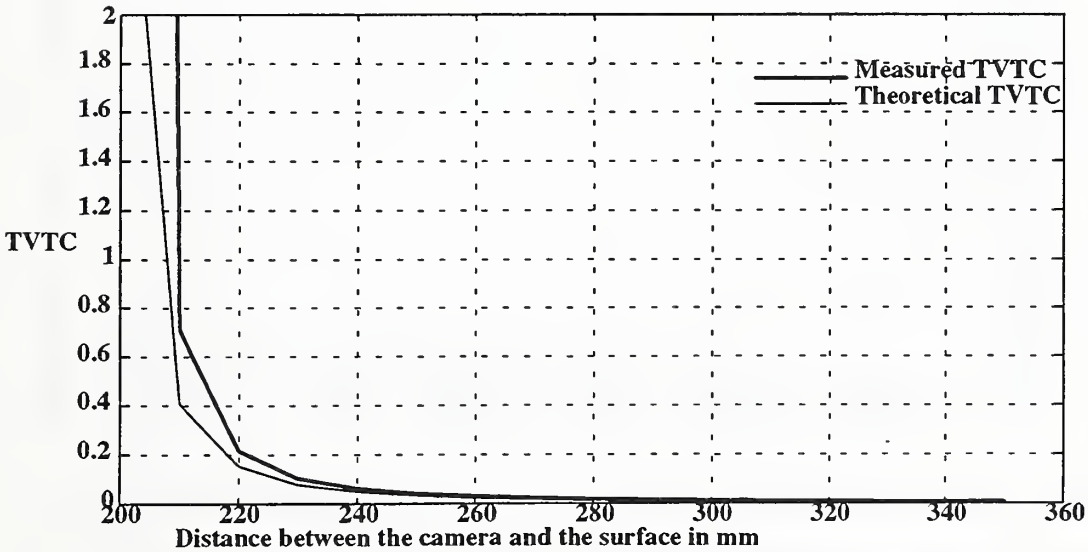


Figure (8d): TVTC vs. Distance between the camera and surface for D9

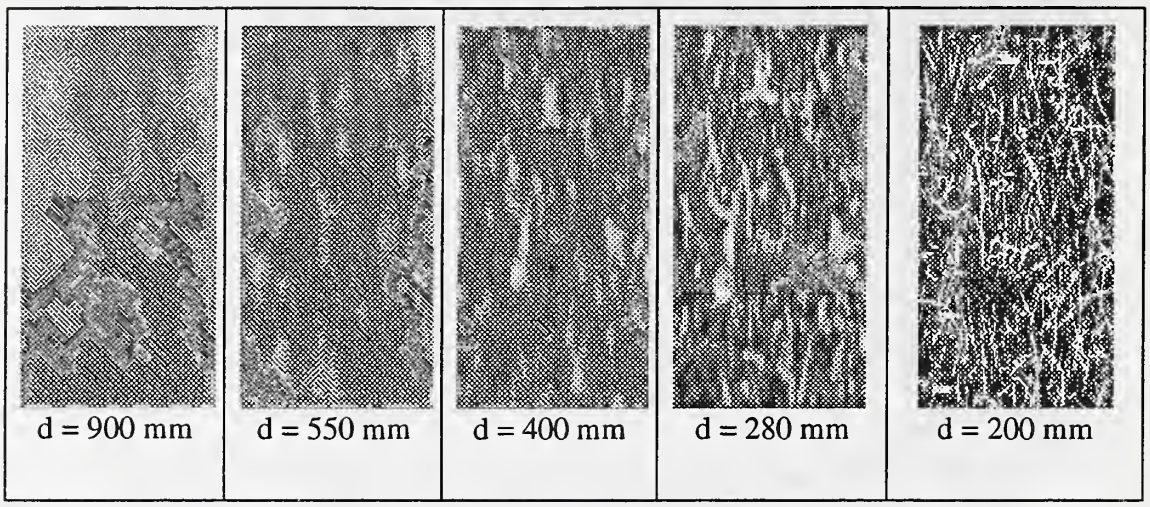


Figure (9a): Image sequence for texture D110, d is the relative distance

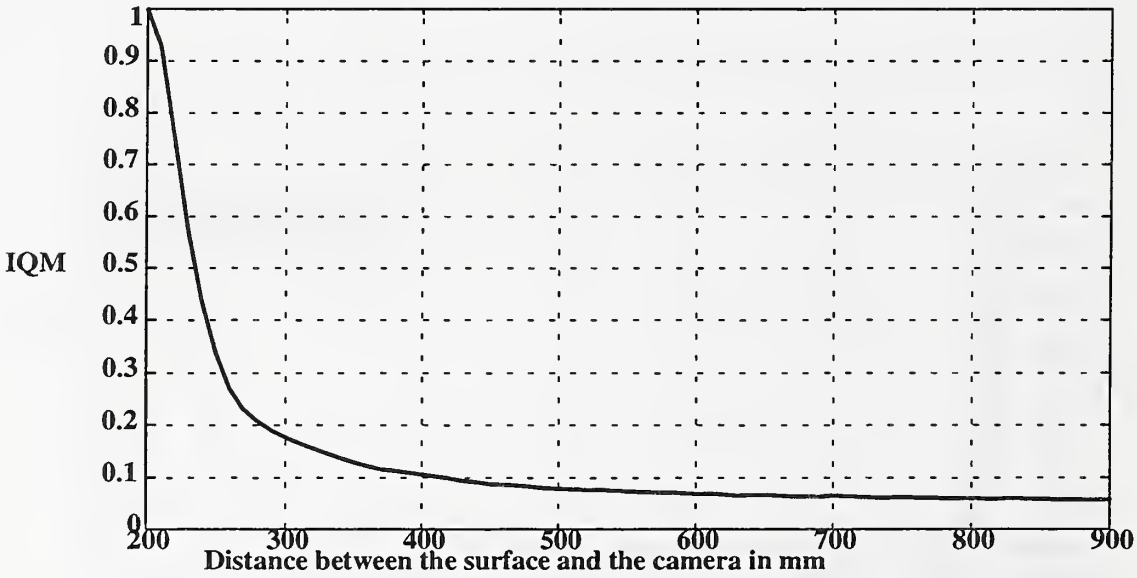


Figure (9b): Measured IQM vs. Distance between the camera and surface for D110

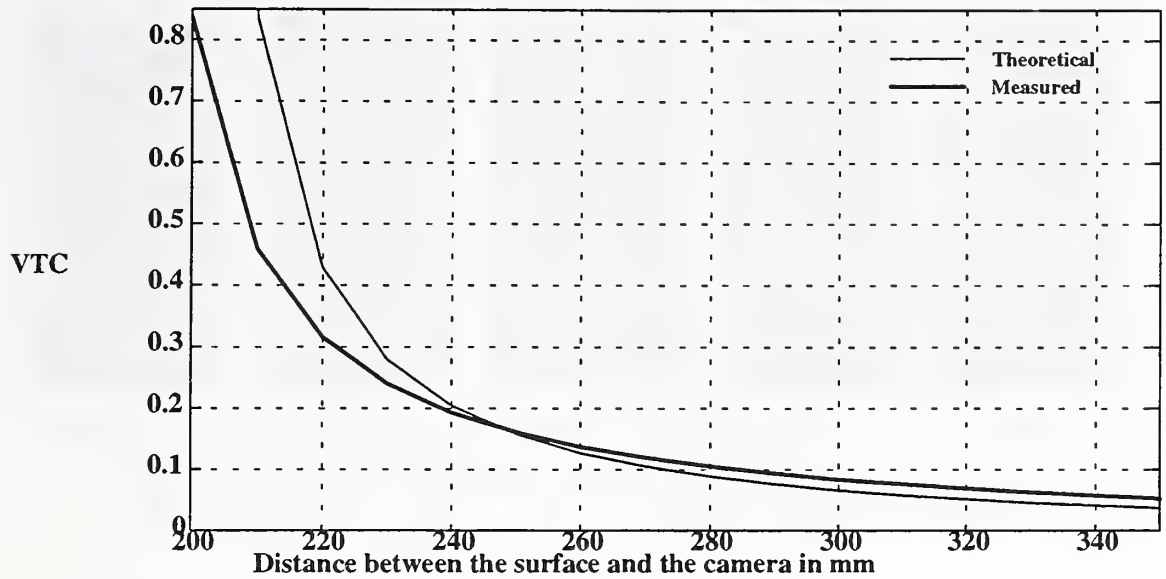


Figure (9c): VTC vs. Distance between the camera and surface for D110

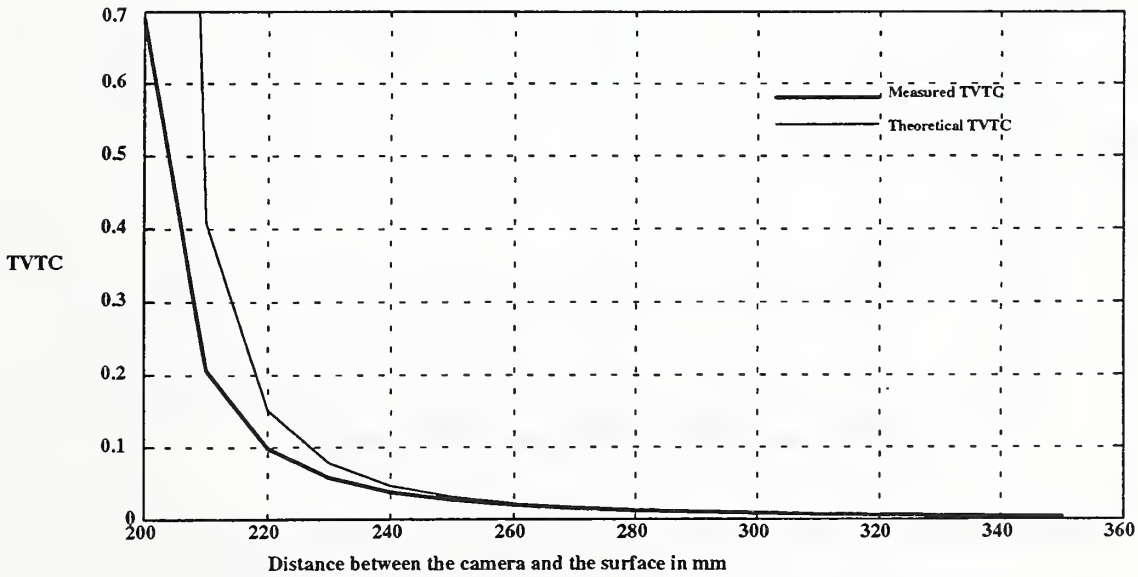


Figure (9d): TVTC vs. Distance between the camera and surface for D110

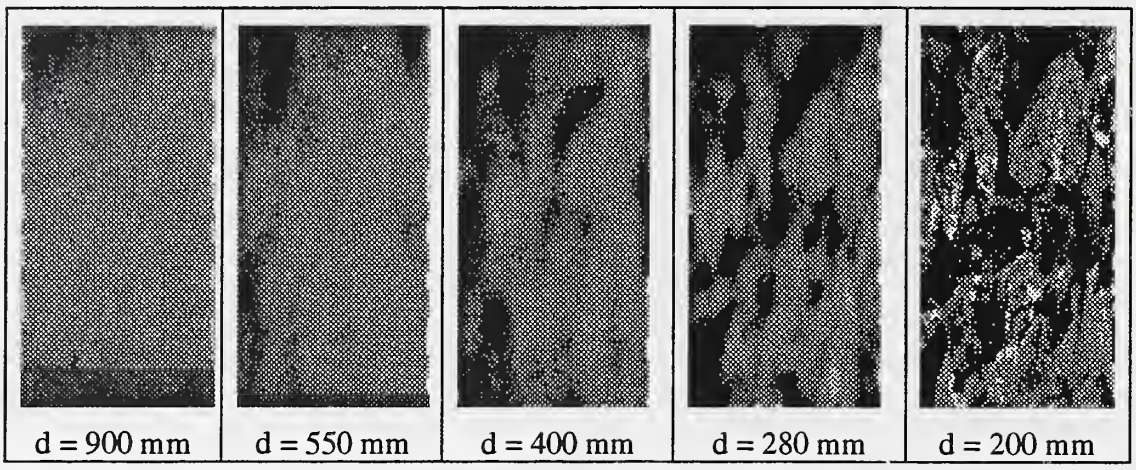


Figure (10a): Image Sequence for Texture D98, d is the relative distance

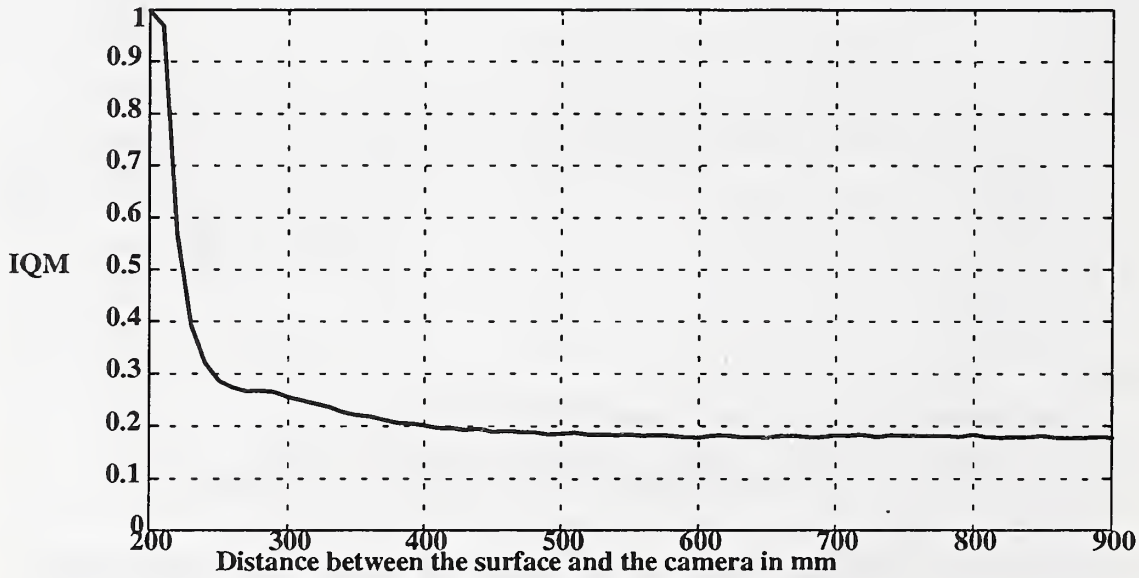


Figure (10b): Measured IQM vs. Distance between the camera and surface for D98

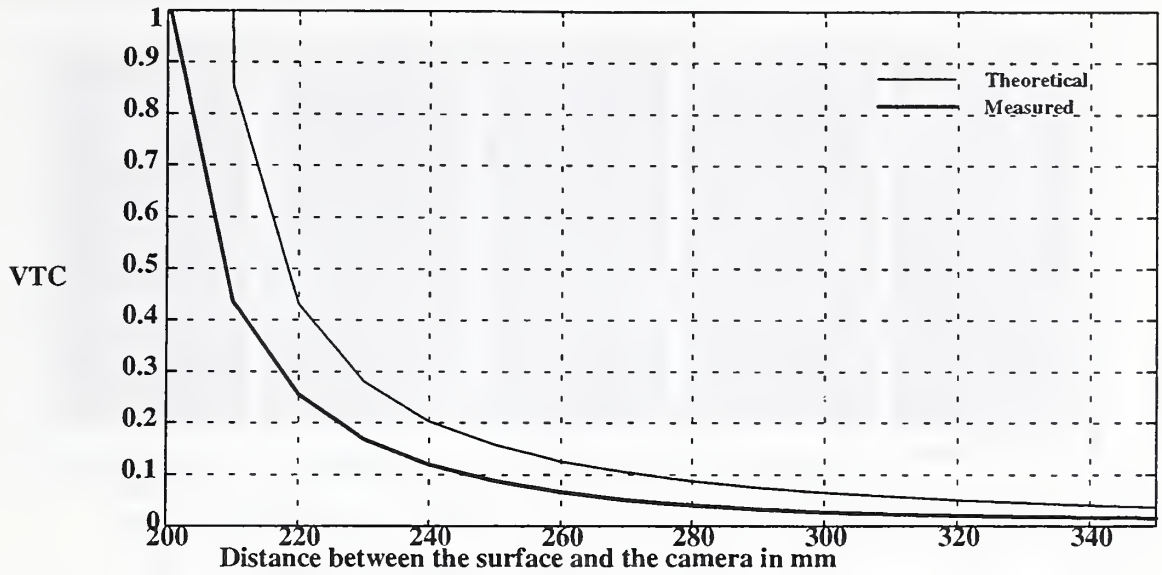


Figure (10c): VTC vs. Distance between the camera and surface for D98

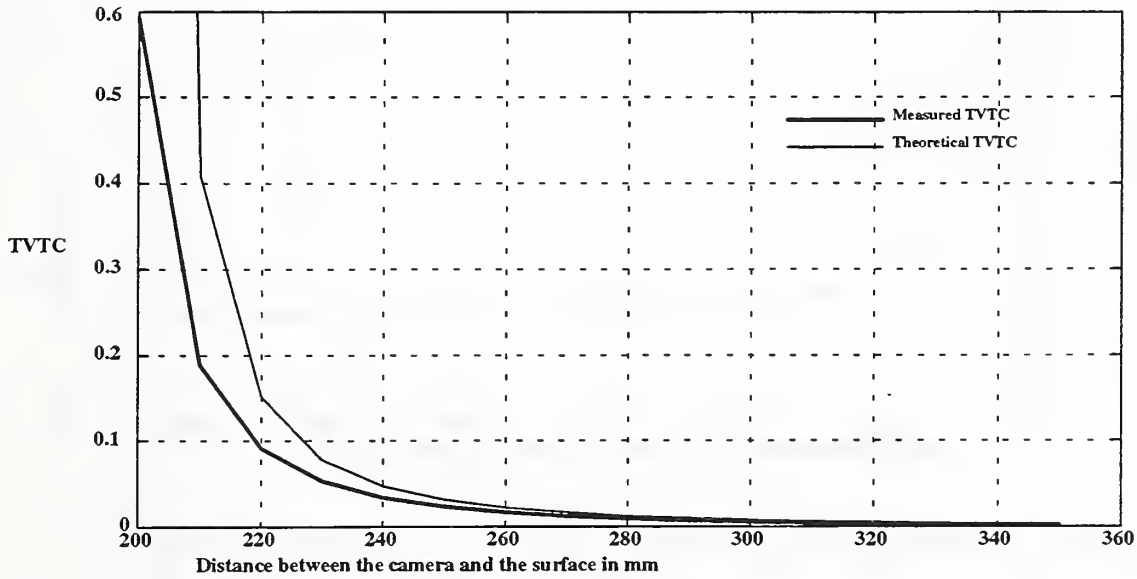


Figure (10d): Measured IQM vs. Distance between the camera and surface for D98

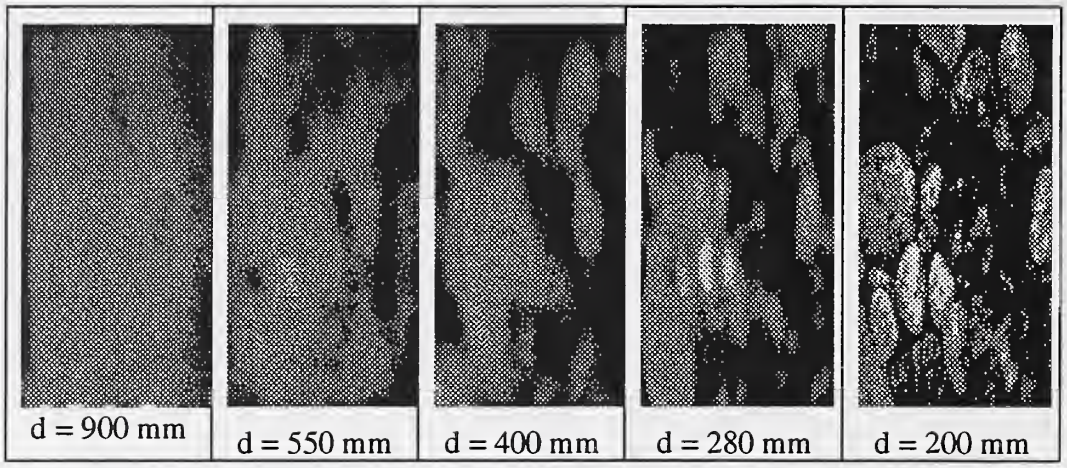


Figure (11a): Image Sequence for Texture D23, d is the relative range

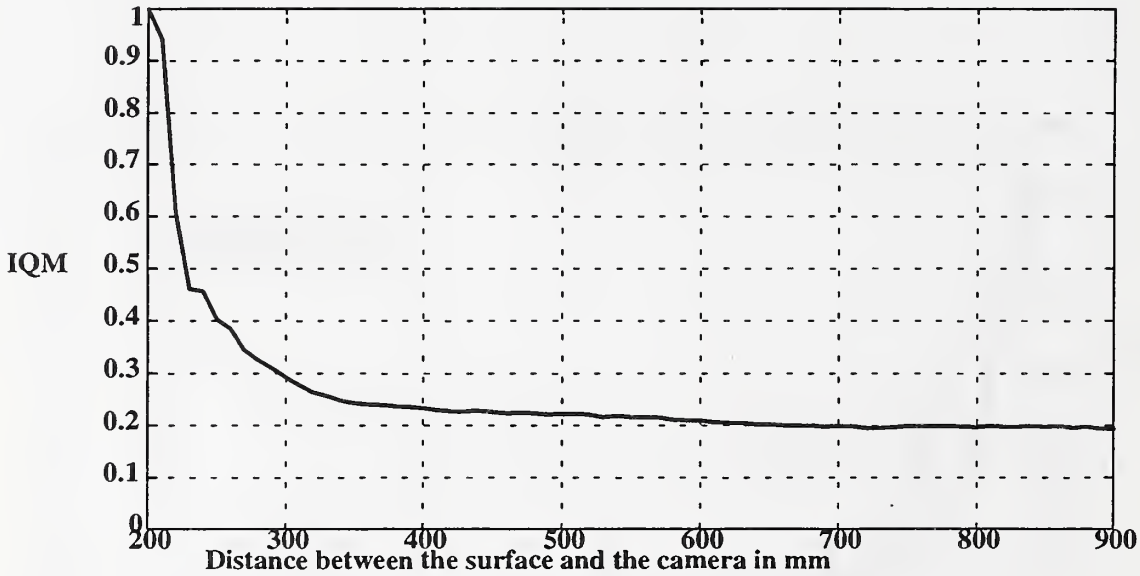


Figure (11b): Measured IQM vs. Distance between the camera and surface for D23

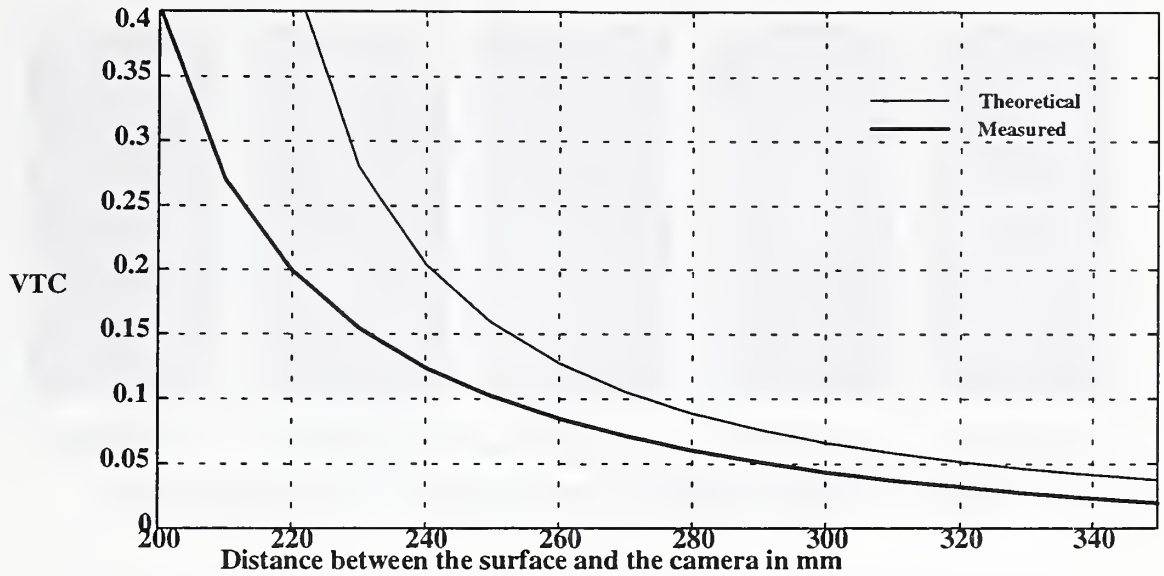


Figure (11c): VTC vs. Distance between the camera and surface for D23

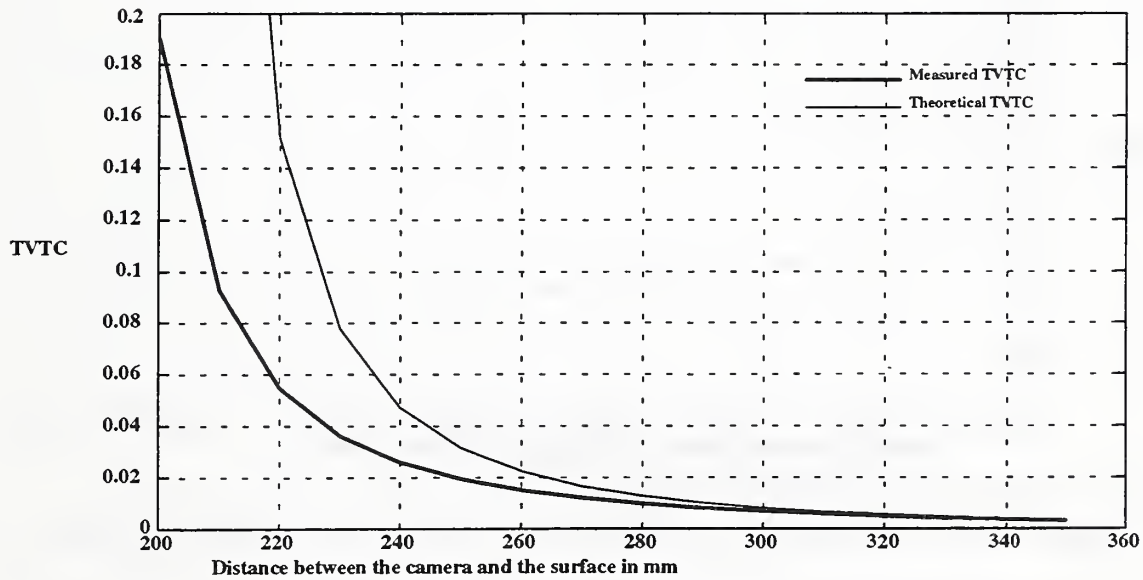


Figure (11d): Measured IQM vs. Distance between the camera and surface for D23

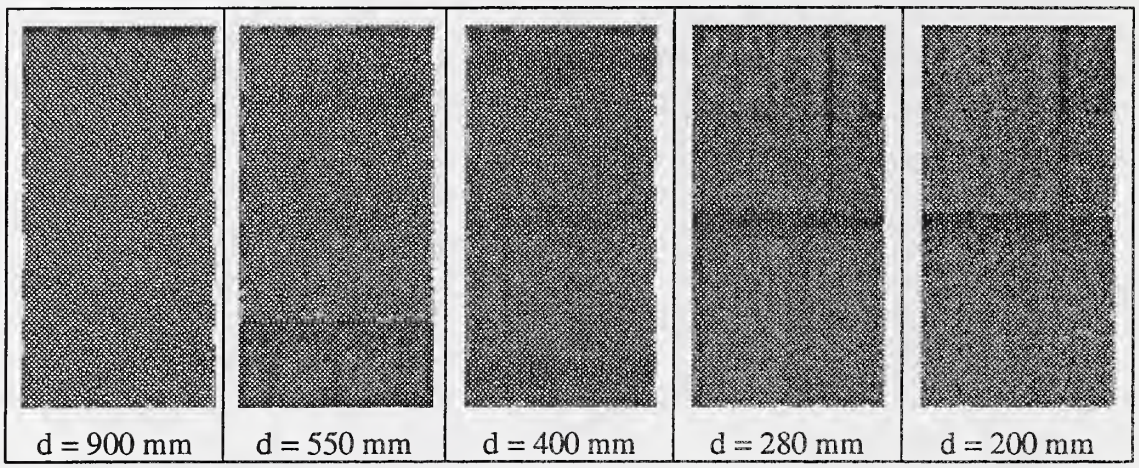


Figure (12a): Image Sequence for Texture D25, d is the relative range

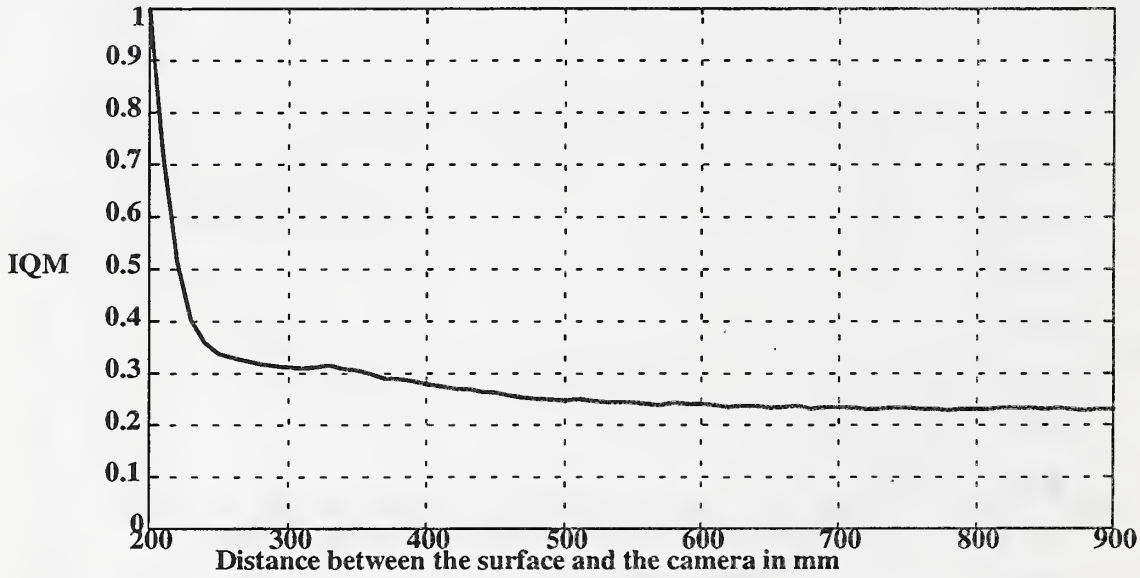


Figure (12b): Measured IQM vs. Distance between the camera and surface for D25

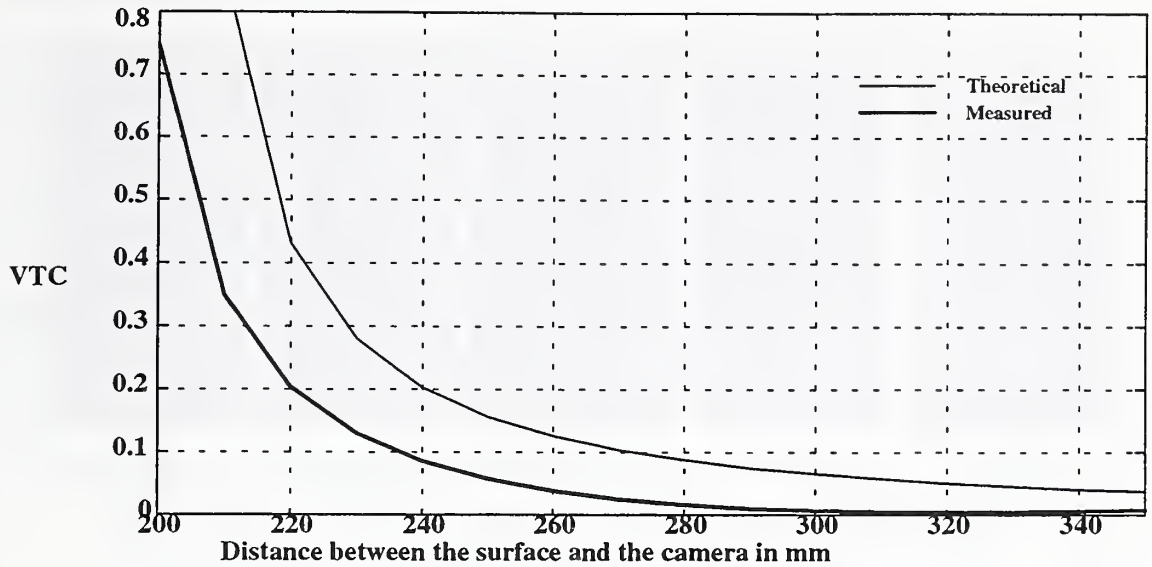


Figure (12c): VTC vs. Distance between the camera and surface for D25

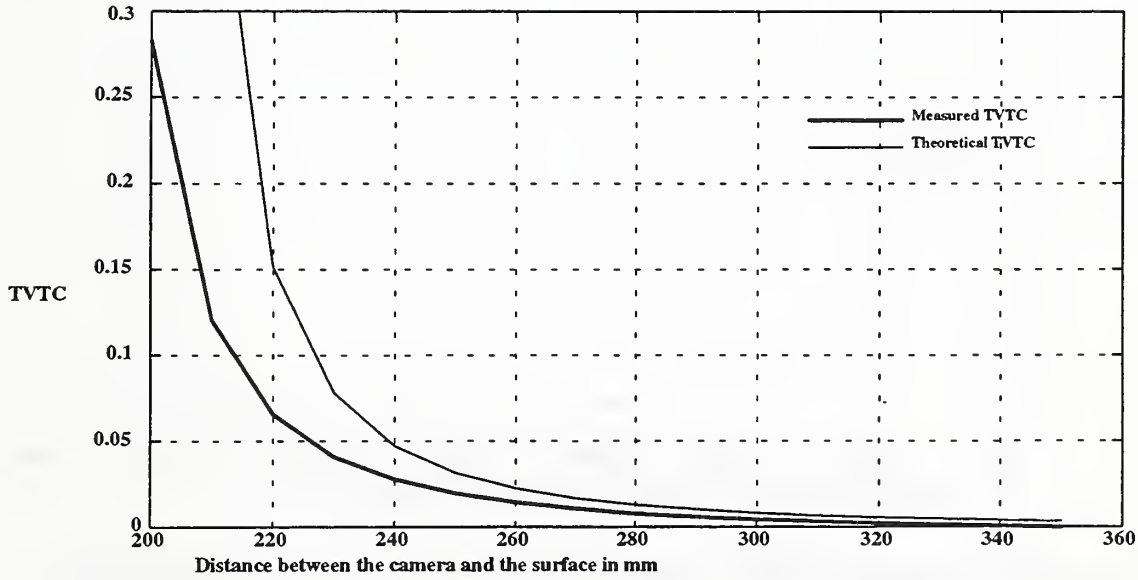


Figure (12d): TVTC vs. Distance between the camera and surface for D25

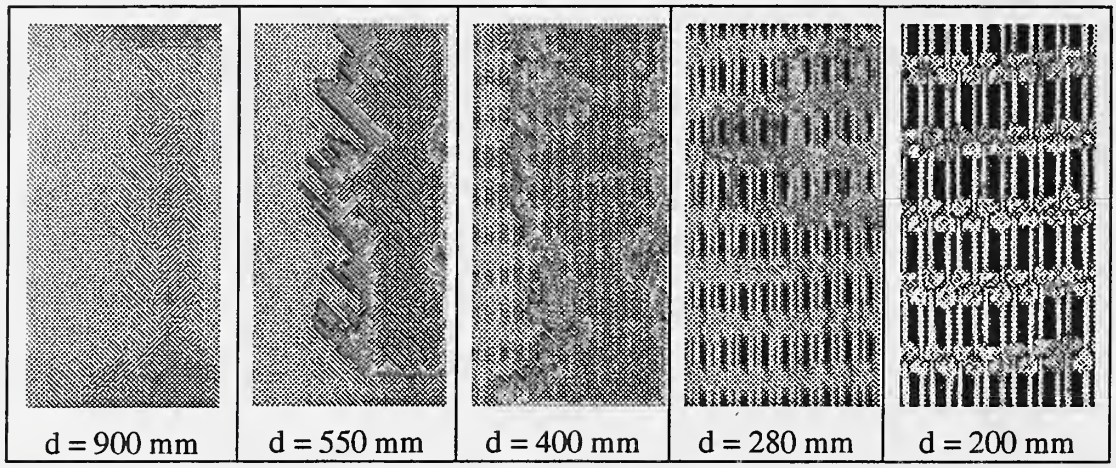


Figure (13a): Image Sequence for Texture D20, d is the relative range

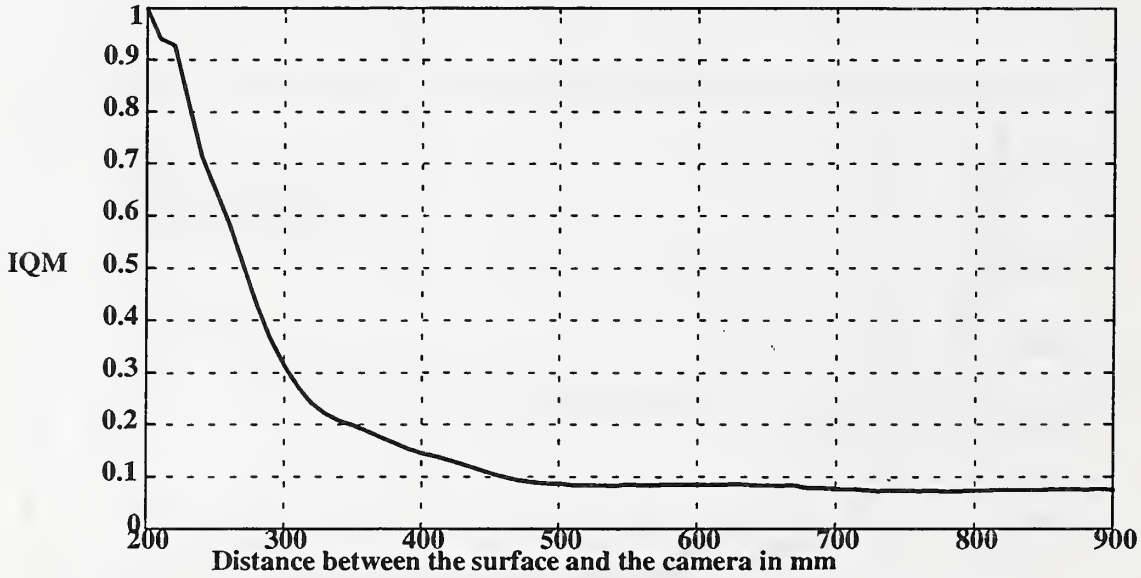


Figure (13b): Measured IQM vs. Distance between the camera and surface for D20

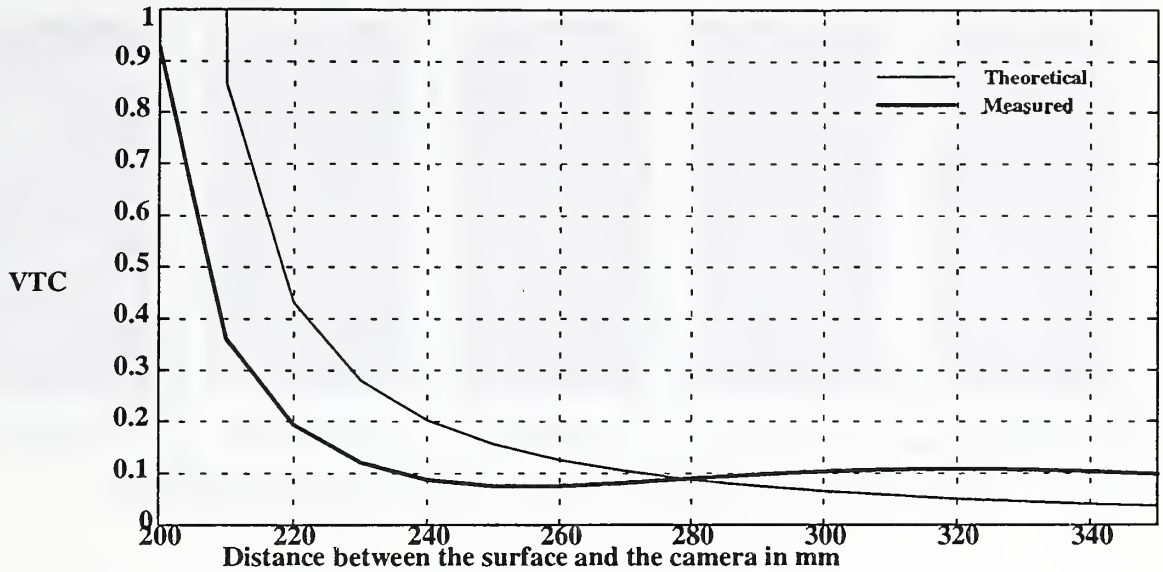


Figure (13c): VTC vs. Distance between the camera and surface for D20

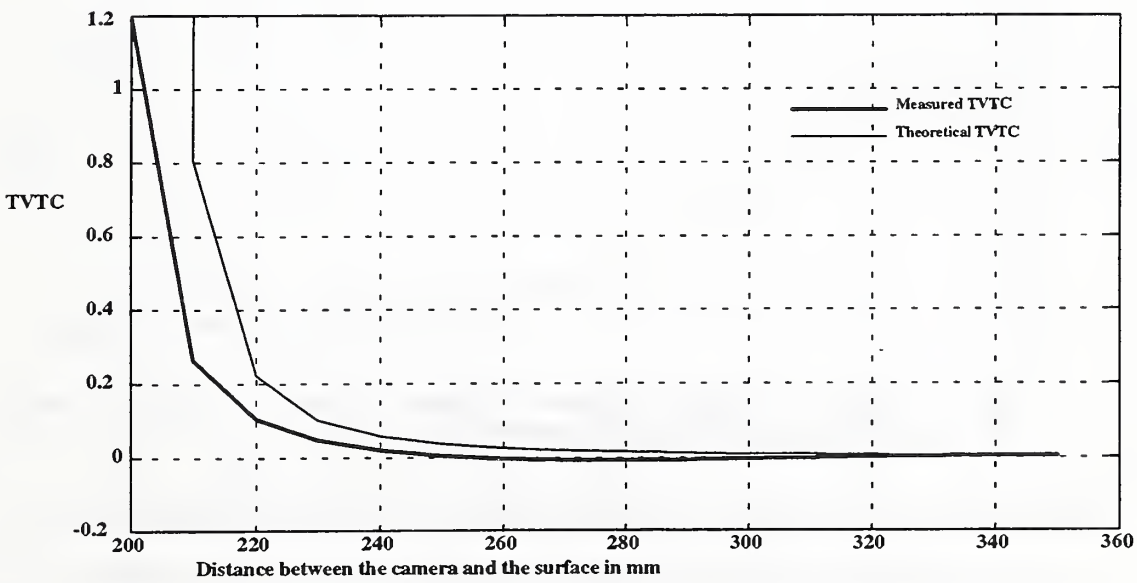


Figure (13d): TVTC vs. Distance between the camera and surface for D20

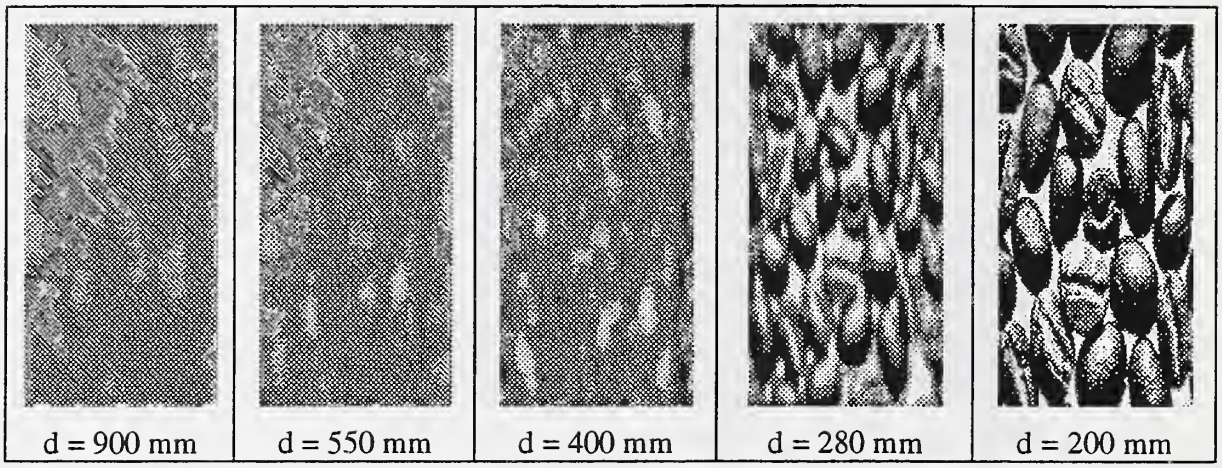


Figure (14a): Image Sequence for Texture D74, d is the relative range

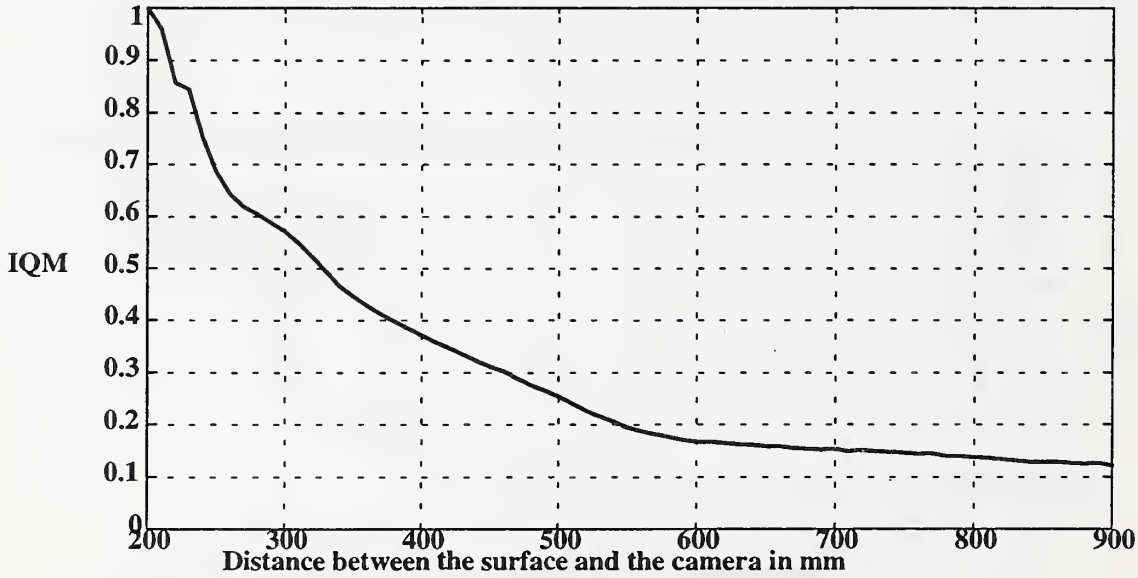


Figure (14b): Measured IQM vs. Distance between the camera and surface for D74

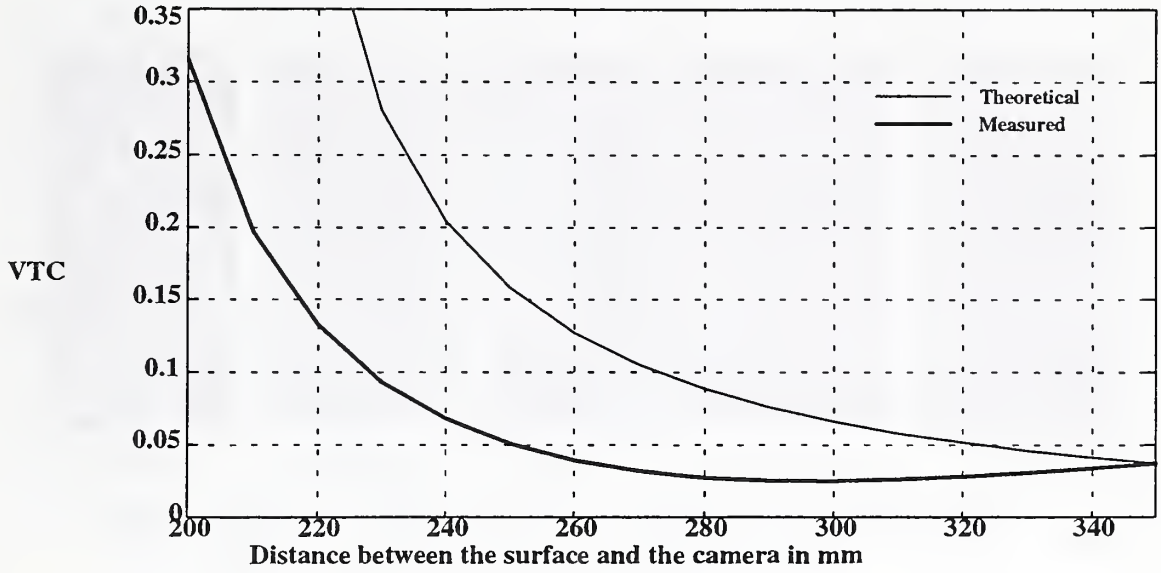


Figure (14c): VTC vs. Distance between the camera and surface for D74

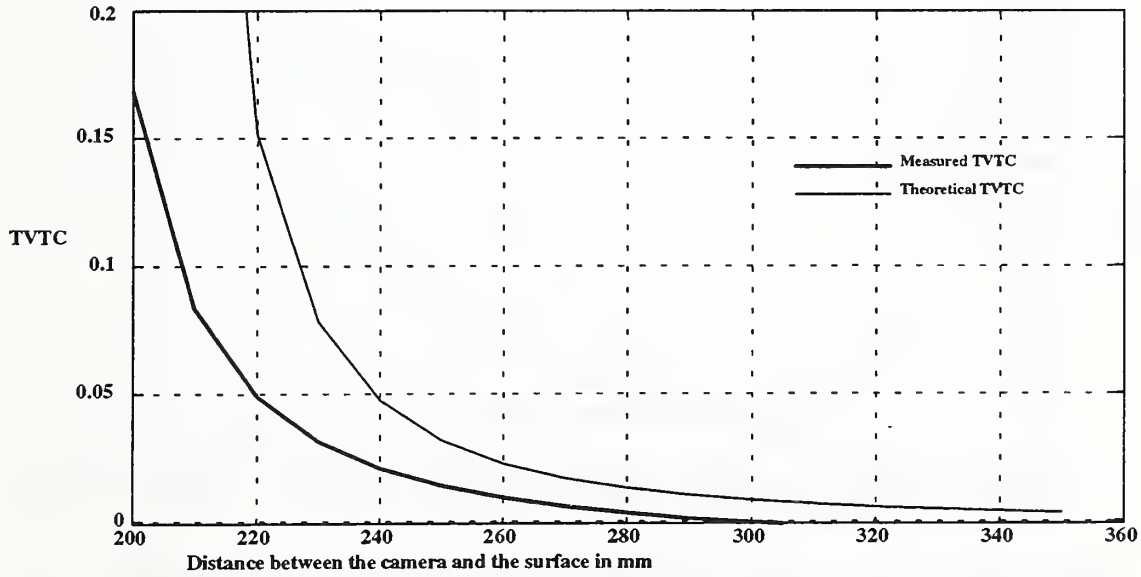


Figure (14d): TVTC vs. Distance between the camera and surface for D74

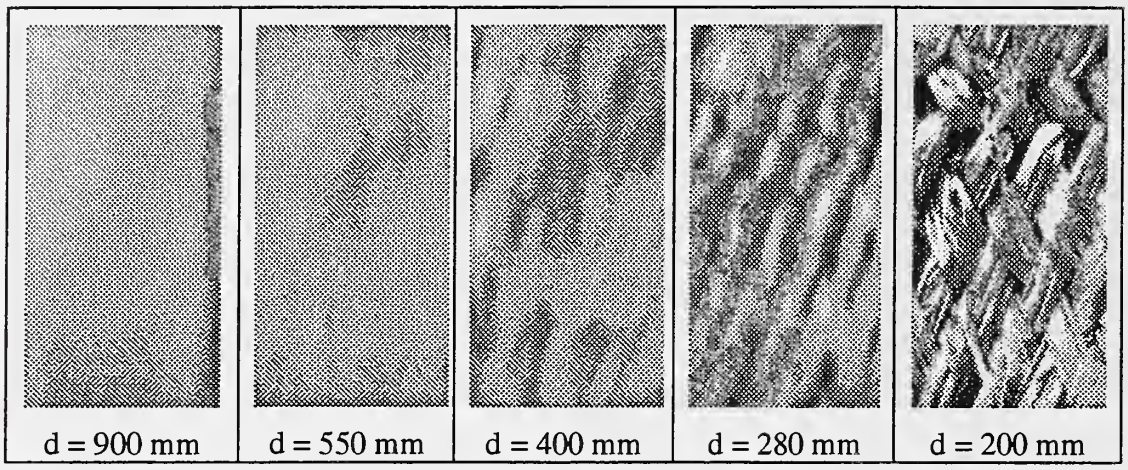


Figure (15a): Image Sequence for Texture D18, d is the relative distance

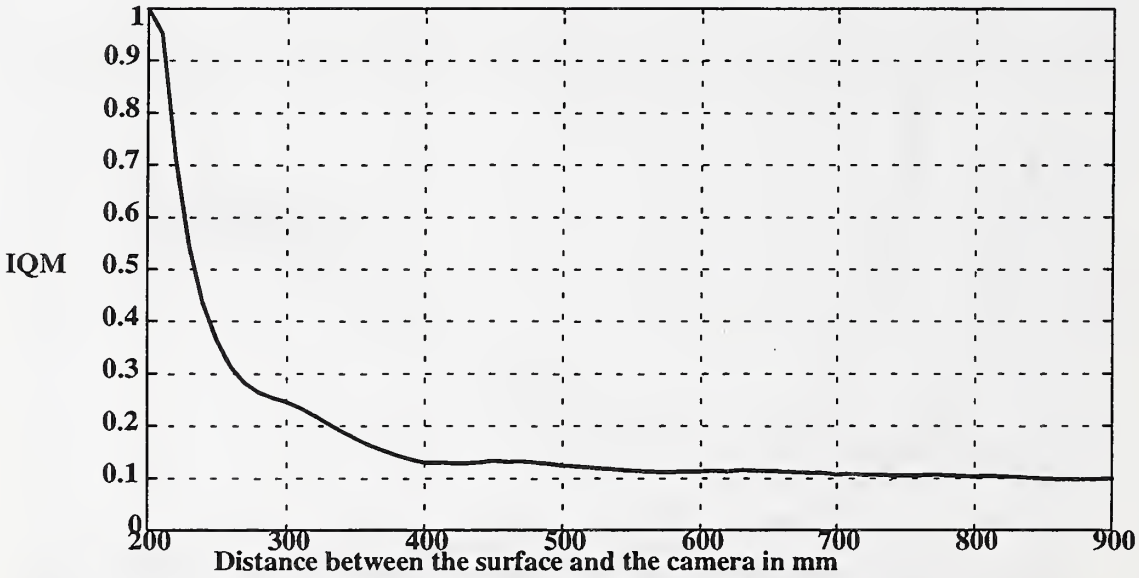


Figure (15b): Measured IQM vs. Distance between the camera and surface for D18

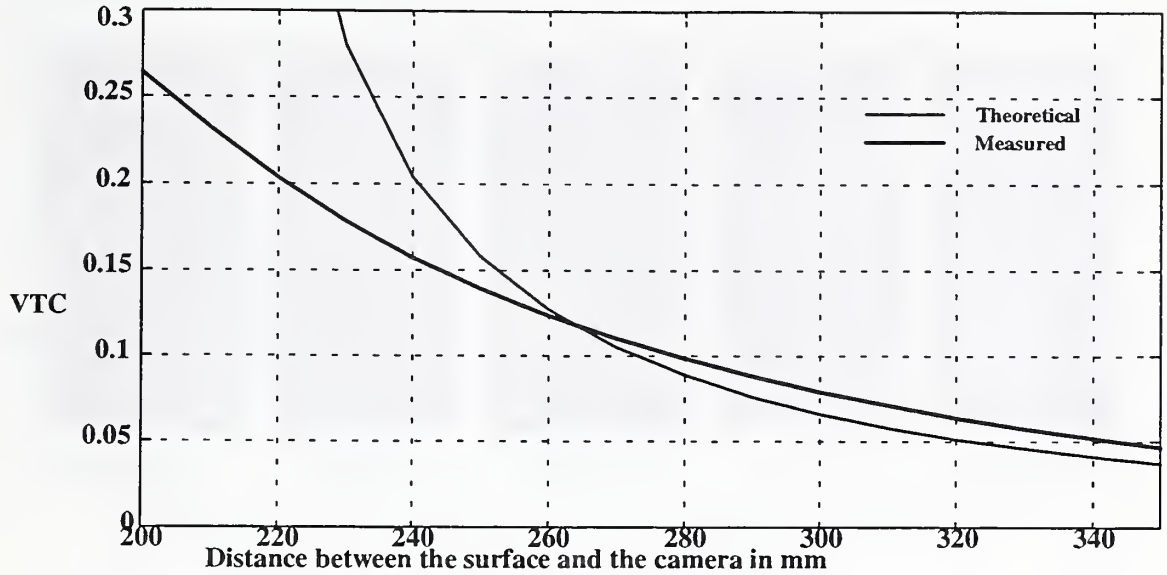


Figure (15c): VTC vs. Distance between the camera and surface for D18

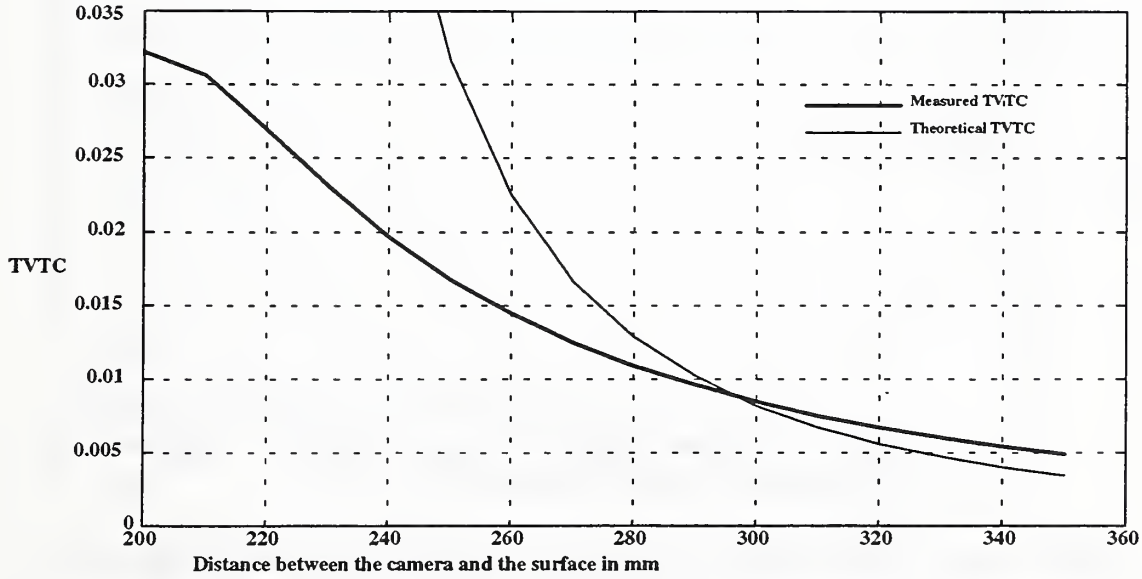


Figure (15d): TVTC vs. Distance between the camera and surface for D18

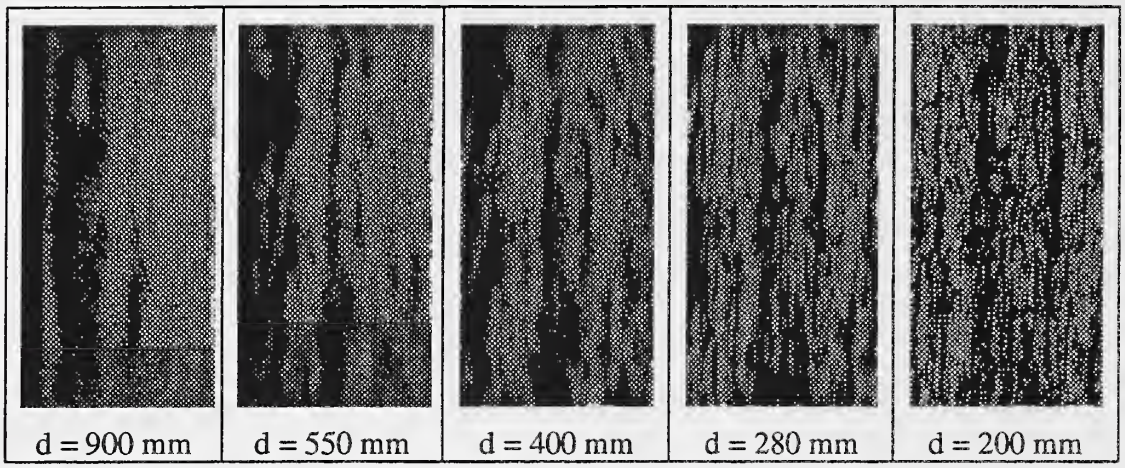


Figure (16a): Image sequence for Texture D12, d is the relative distance

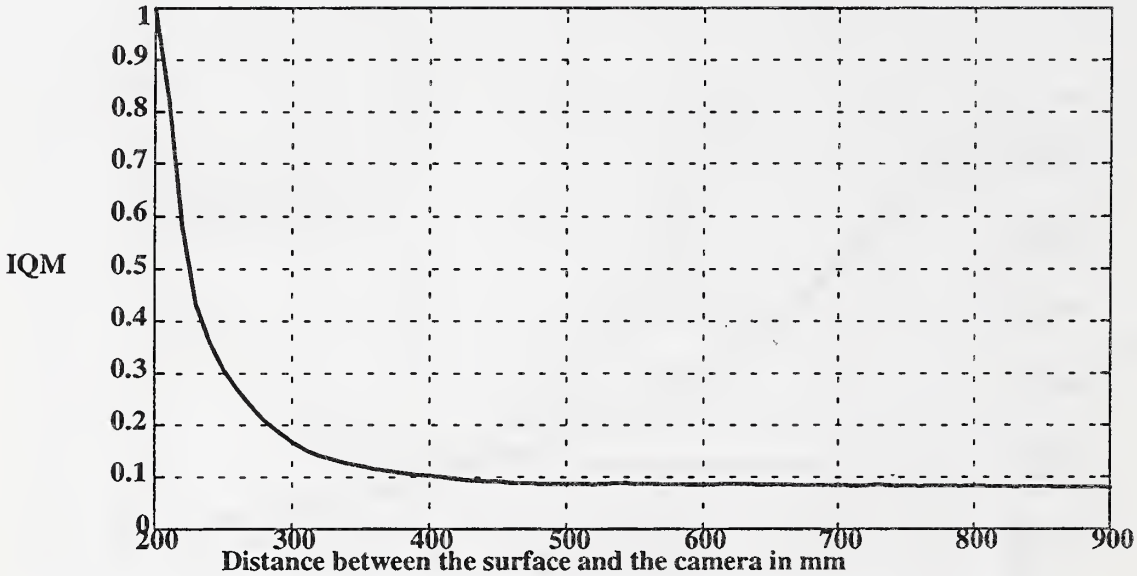


Figure (16b): Measured IQM vs. Distance between the camera and surface for D12

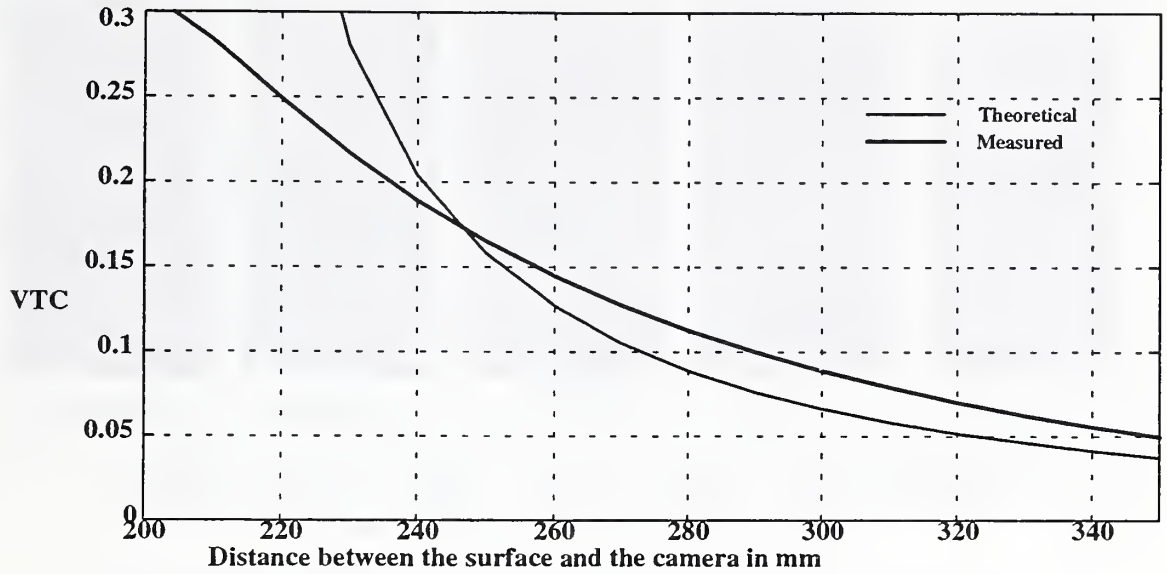


Figure (16c): VTC vs. Distance between the camera and surface for D12

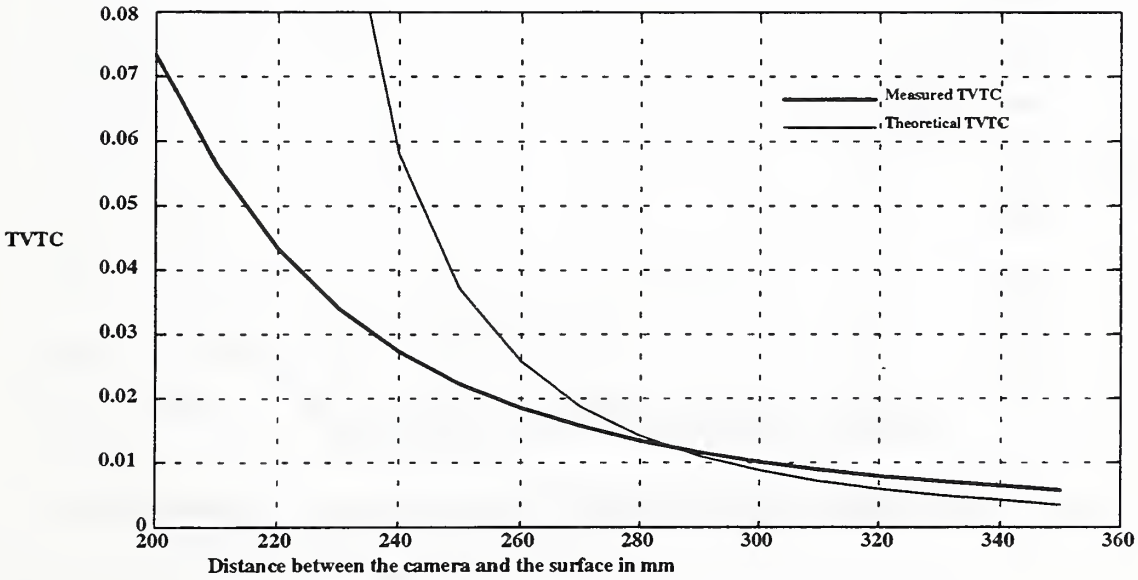


Figure (16d): Measured IQM vs. Distance between the camera and surface for D12

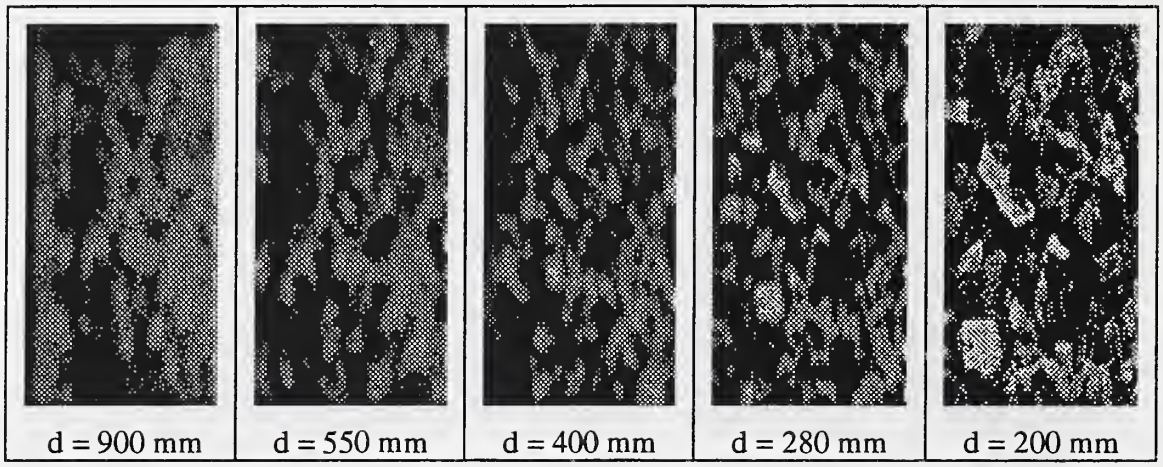


Figure (17a): Image Sequence for Texture D5, d is the relative distance

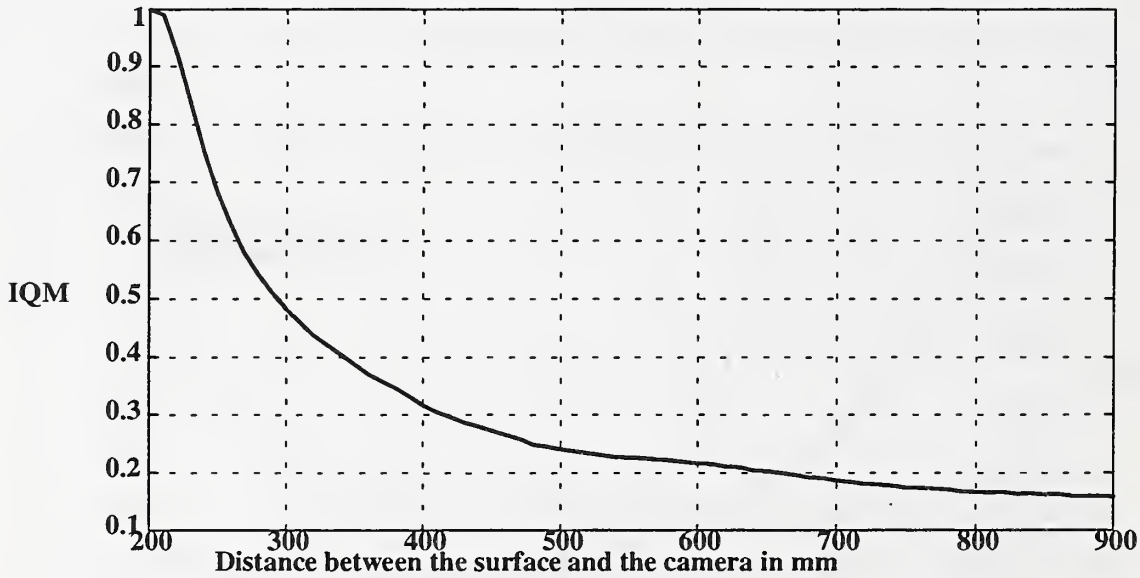


Figure (17b): Measured IQM vs. Distance between the camera and surface for D5

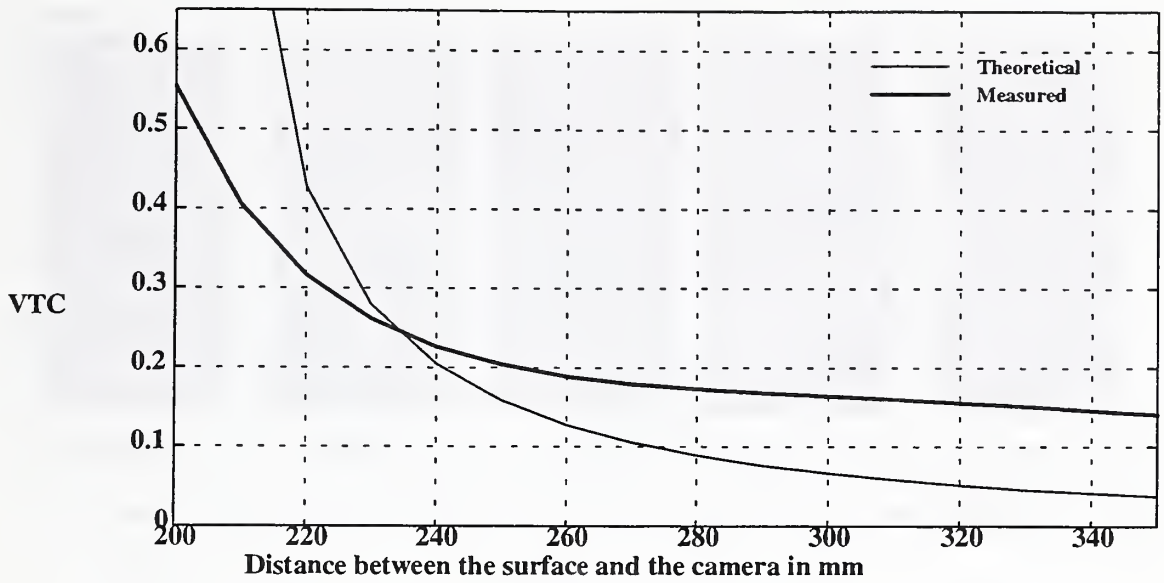


Figure (17c): VTC vs. Distance between the camera and surface for D5

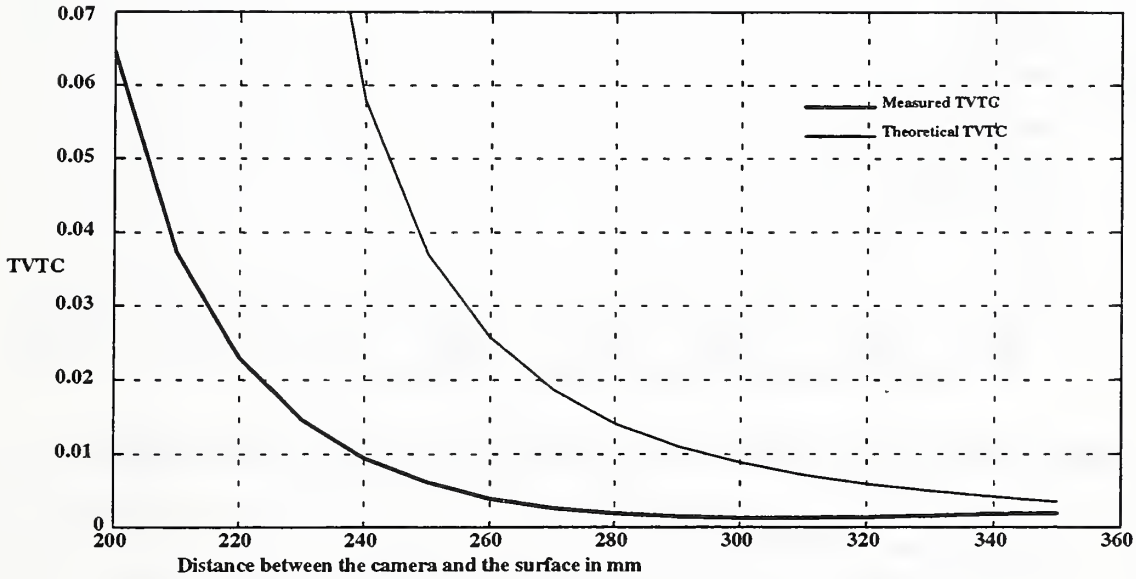


Figure (17d): TVTC vs. Distance between the camera and surface for D5

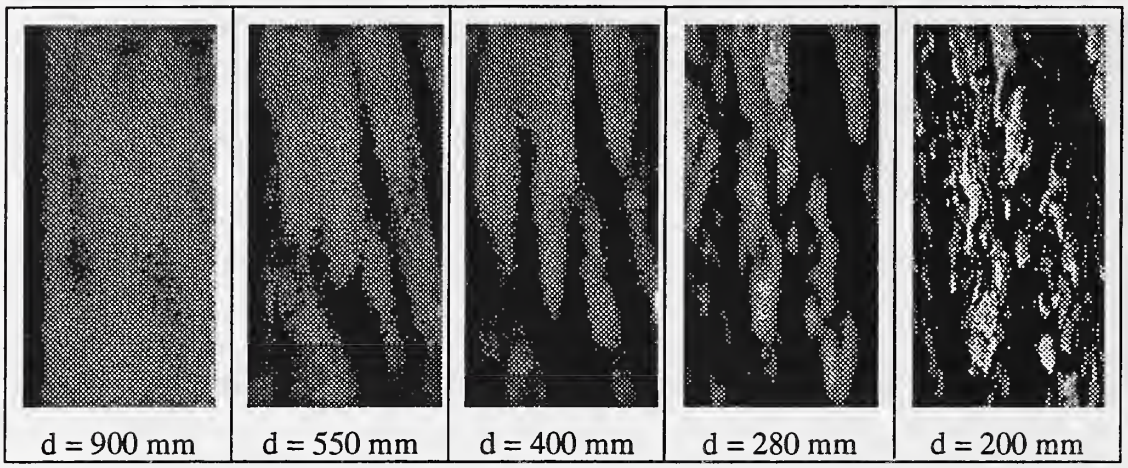


Figure (18a): Image Sequence for Texture D13, d is the relative distance

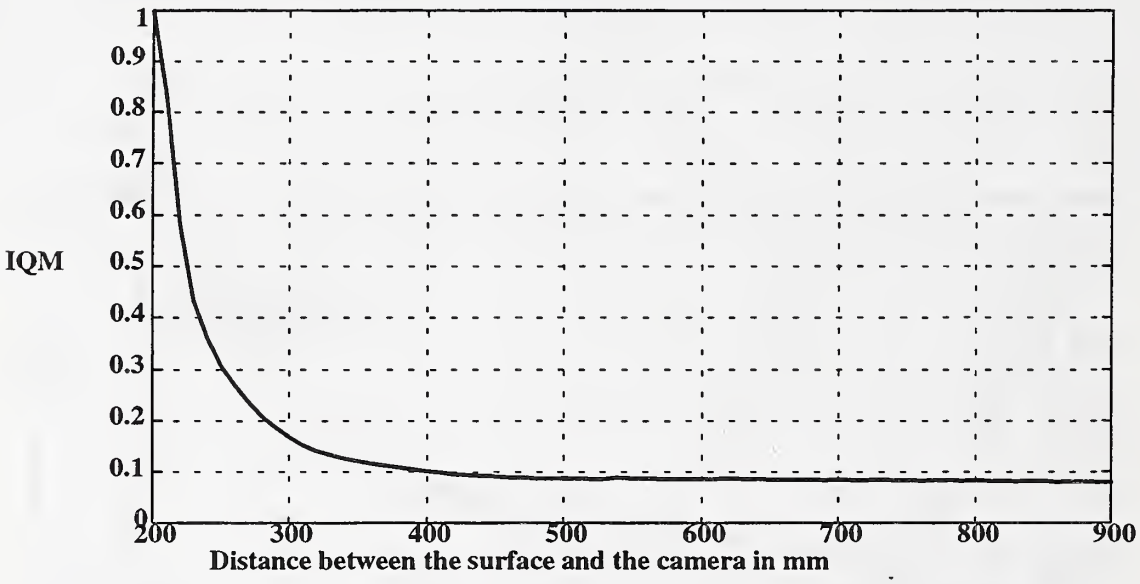


Figure (18b): Measured IQM vs. Distance between the camera and surface for D13

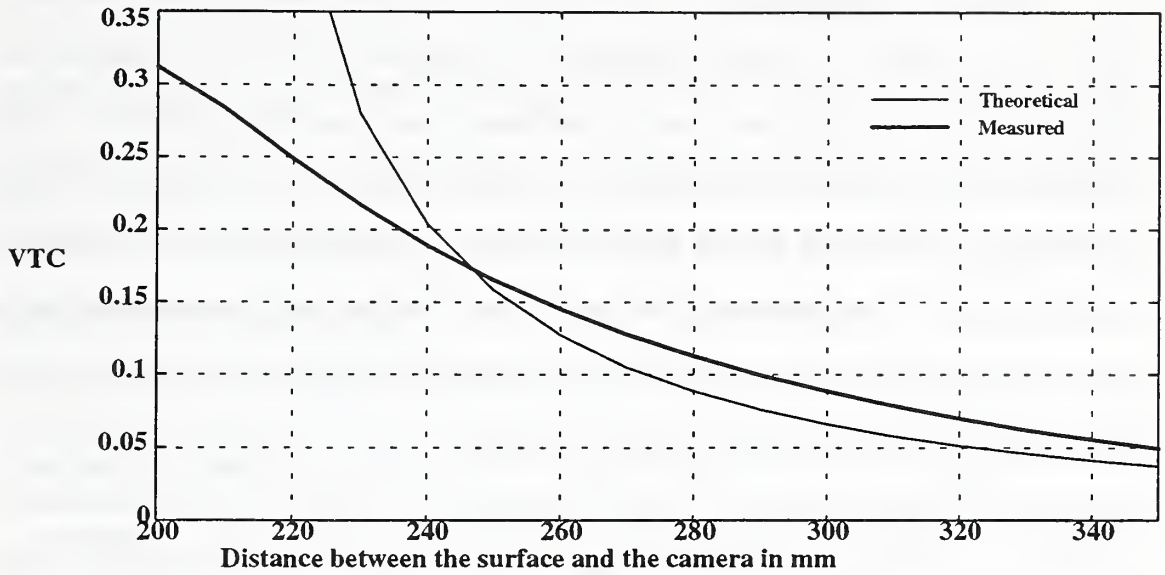


Figure (18c): VTC vs. Distance between the camera and surface for D13

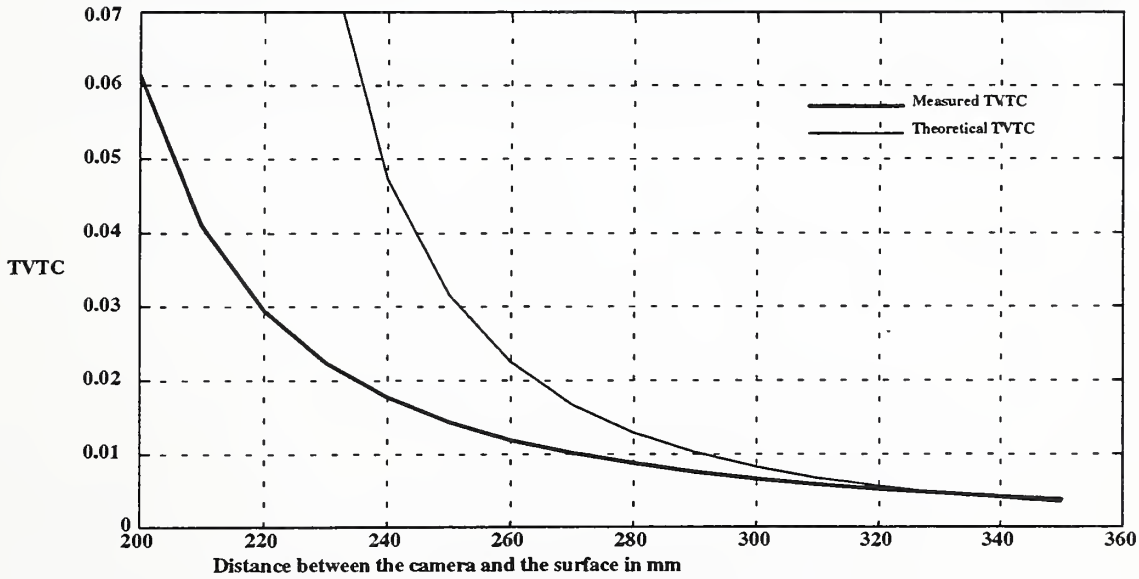


Figure (18d): TVTC vs. Distance between the camera and surface for D13

Conclusions and Future Work

This paper presents novel active-vision-based motion cues for local navigation tasks such as collision avoidance, maintenance of clearance, etc. The advantage of this approach is that it is based on observer coordinate system and the visual motion cues employed in the approach are measurable from image sequences. The approach is based on two measurable visual motion cues namely the VTC and the TVTC. Though the extraction of these cues need some texture in the environment, they *do not depend* on the type of the texture in the environment. We are currently working on the extension of the above mentioned cues for environments without texture. A control algorithm to accomplish collision avoidance as well as maintenance of clearance was implemented in real time on a six-degree-of-freedom flight simulator using the VTC information only. Currently we are working on the extension of this algorithm to real outdoor environments using a vision-based autonomous golf cart designed and built at Florida Atlantic University.

Appendix A : Motivation for the Definition of the VTC

Consider a luminous point P in a 3D scene located on the optical axis of the camera as shown in Figure (A1). Light emanating from point P scatters in all directions and a divergent bundle of rays emerges from the luminous point P as shown in Figure (A1) [20]. When this bundle of light passes through a convex lens, it converge to a point on the optical axis and again diverge as shown in Figure (A1). If the point of convergence lies exactly on the image plane (as shown in Figure (A1a)), then it would result in a sharp point on the image plane. On the other hand if the image plane lies in front of the point of convergence (as shown in Figure (A1b)) or beyond the point of convergence (as shown in Figure (A1c)), the bundle of rays is cut by the image plane. In such situations the retina will be illuminated not at one distinct point, but over a circular region corresponding to the bundle of rays [20]. The circular area on the image plane is usually referred as the blur circle, circle of confusion or circle of diffusion [20]. This circular patch on the image plane is characterized by its radius and is usually referred to as the radius of blur circle.

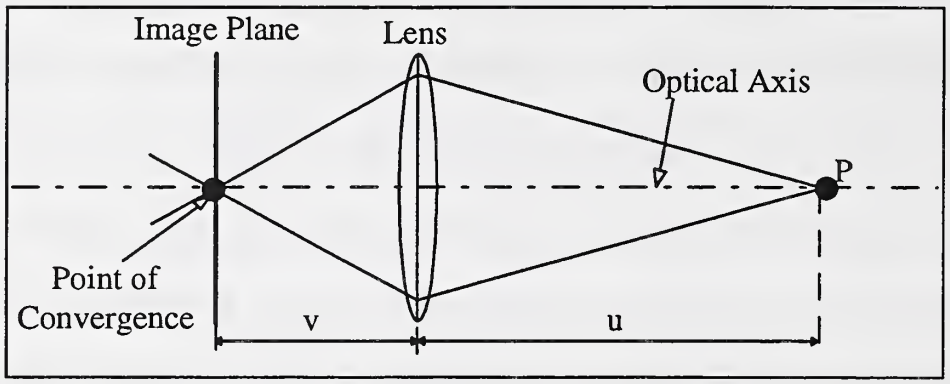


Figure (A1a): Lens in Perfect Focus; Image Forming on the Image Plane

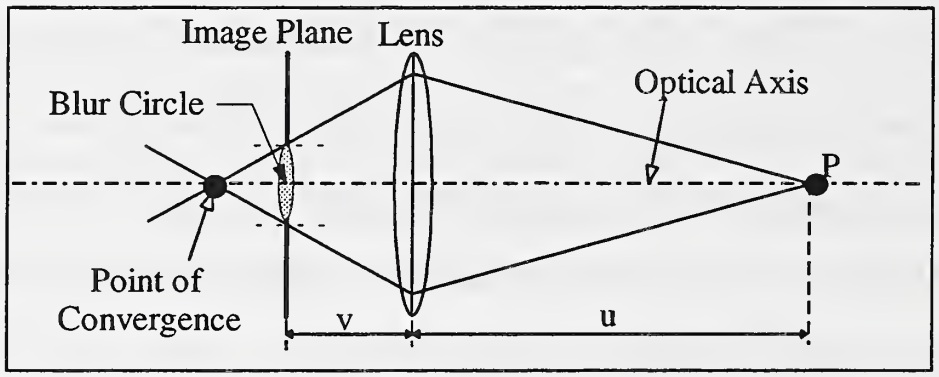


Figure (A1b): Lens Out of Focus; Image Forming in front of the Lens

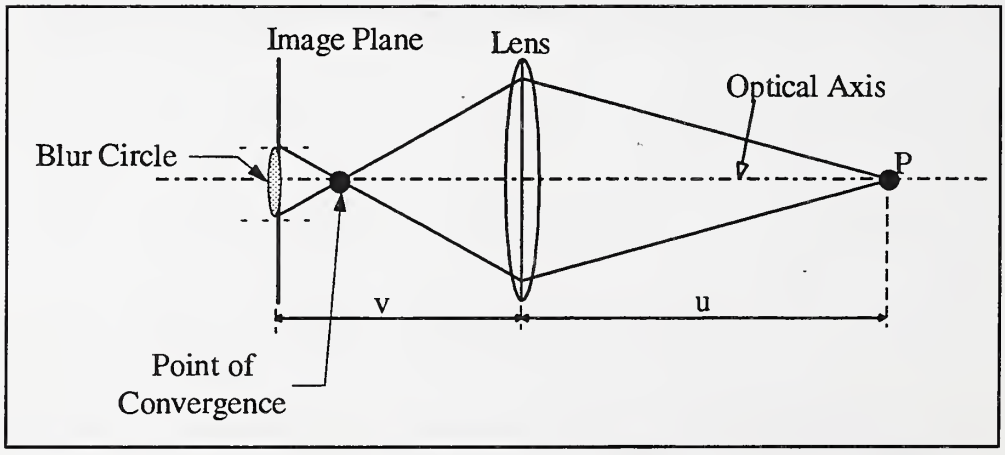


Figure (A1c): Lens Out of Focus; Image forming in front of the Image Plane

A.1 Expression for Radius of Blur Circle

Consider a fixed focus camera that is focused initially to some range R_0 as shown in Figure (A2). In such a camera the distance between the lens and the image plane v is fixed for a particular initial focus adjustment R_0. If this camera is used to fixate at points in 3D scene (see [18] and [19] for a good review on fixation), then a blur circle results whenever the range between the fixation point and the camera is not equal to R_0 the range to which the camera is focused to initially.

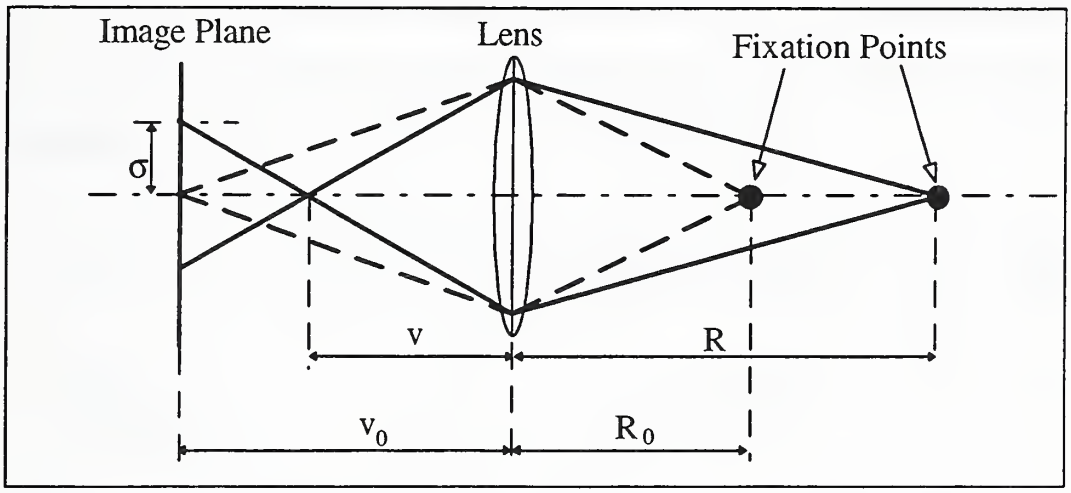


Figure (A2): Imaging System: v_0 = distance between the image plane and the lens, R_0 = distance between the lens and the scene for which the image is in focus, σ = radius of the blur circle, $R > R_0$

As mentioned earlier a blur circle results if the range R between the fixation point and the observer is greater than R_0 the distance to which the camera is focused to initially. In the analysis through out this dissertation it is assumed that $R > R_0$ (a similar analysis can be performed for case $R < R_0$). Based on this assumption namely $R > R_0$, an expression for the radius of blur circle is presented next.

From the geometry of Figure (A2) an expression for the radius of blur circle can be derived as follows [21]:

$$\frac{l}{v} = \frac{\sigma}{v_0 - v} \tag{A1}$$

where $2l$ is the width of the lens, v_0 is the distance between the image plane and the lens, v is the distance between the lens and the point of convergence of rays on the optical axis as shown in Figure (A2). Also we have the following relations based on Gaussian thin lens formula [21]:

$$\frac{1}{R_0} + \frac{1}{v_0} = \frac{1}{F} \tag{A2}$$

and

$$\frac{1}{R} + \frac{1}{v} = \frac{1}{F} \tag{A3}$$

The following relation can be derived by combining Equations (A1-A3):

$$\sigma = k\left(\frac{1}{R_0} - \frac{1}{R}\right) \tag{A4}$$

where $k = \frac{Fv_0}{f}$, F is the focal length of the camera, v_0 is the distance between the image plane and the lens, f is the f-number of the lens. For a fixed focus camera F, f and v_0 remain unchanged, hence k can be seen as a positive constant for a fixed focus camera.

A.2 Qualitative Analysis of the Radius of Blur Circle

Based on Equation (A4) graphical illustration of variations of the radius of blur circle as a function of the range (for ranges R greater than R_0) between the fixation point and the camera is shown in Figure (A3). From Figure (A3) as well as Equation (A4), it can be seen that for R greater than R_0, small variations in R around R_0 would lead to large variations in the radius of the blur circle σ and vice-versa. In other words, when the distance between the fixation point and the camera is very large compared to the distance R_0 to which the camera is focused to initially, then large changes in the ranges would lead to small variations in the radius of the blur circle. If $R \gg R_0$ then the radius of the blur circle changes very slowly.

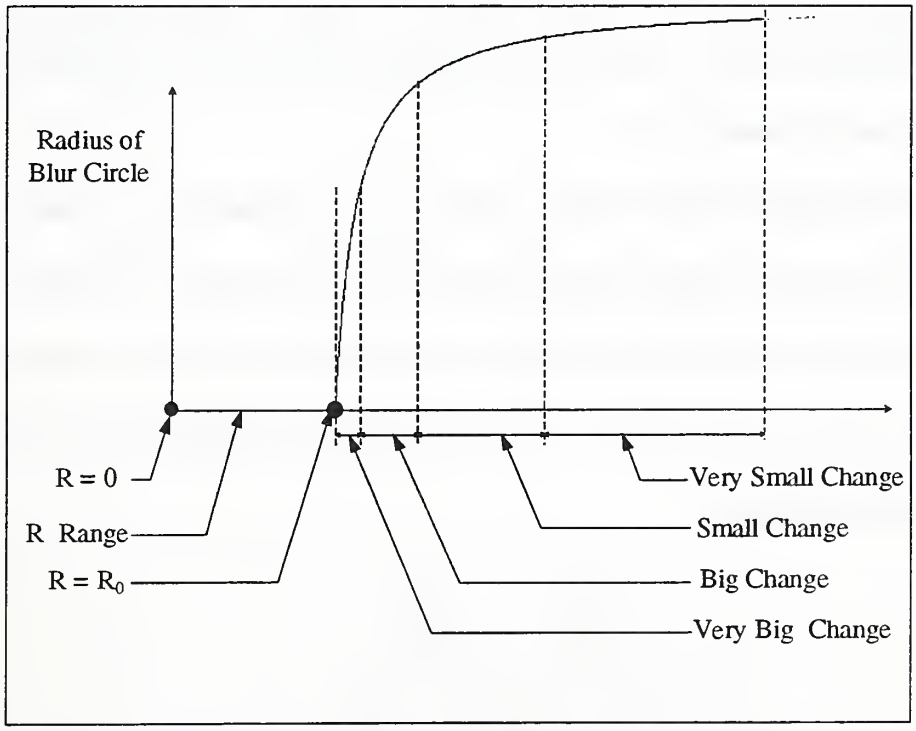


Figure (A3): A Qualitative plot of the radius of blur circle as a function of the range $R > R_0$

The relative variations in the radius of the blur circle provide good indication about the relative variations of range between the camera and the fixation point with respect to the initial focused distance R_0. This observation is the key behind the derivation presented in Equation (1).

An expression for temporal variations of the radius of blur circle can be derived by the differentiation of Equation (A4) with respect to time as follows:

$$\dot{\sigma} = k\frac{\dot{R}}{R^2} \tag{A5}$$

where $k = \frac{Fv_0}{f}$, F is the focal length of the camera, v_0 is the distance between the image plane and the lens, f is the f-number of the lens. For a fixed focus camera F, f and v_0 remain unchanged, hence k can be seen as a positive constant for a fixed focus camera, $\dot{R}$ is the temporal derivative of R.

From Equation (A5) it can be seen that the absolute temporal variations of the radius of blur circle depend upon several factors such as scene illumination, camera parameters such as focal length, width of the aperture, etc., which is expressed as k in Equations (A4, A5). The effect of k can be neutralized by taking the ratios of Equations (A4) and (A5) as follows:

$$\frac{\dot{\sigma}}{\sigma} = \frac{k\left(\frac{\dot{R}}{R^2}\right)}{k\left(\frac{1}{R_0} - \frac{1}{R}\right)} = \frac{\left(\frac{\dot{R}}{R^2}\right)}{\left(\frac{1}{R_0} - \frac{1}{R}\right)} \tag{A6}$$

From Equation (A6) it can be seen that the relative temporal variations of σ are independent of the scene illumination, camera parameters such as focal length, width of the aperture, etc. In other words, to extract relative temporal variations of the radius of the blur circle information about the camera parameters, scene illumination, etc., is not required. Also no camera calibration is necessary to extract it from images.

The right hand side of Equation (A6) is a function of R the range between the observer and the fixation point in the scene, $\dot{R}$ is the temporal derivative of R, R_0 is distance to which the camera is focused to initially. As a whole this function provides information about relative variations in range as well as clearance (R_0). In other words, information about variations in range can be obtained without measuring the absolute ranges. Since the entity on the right hand side of Equation (6) is measurable and also it provides information about relative variations in range, it is defined as a Visual Threat Cue.

Appendix B: Zero TVTC Surface

In this section we derive the Equation for the zero TVTC surface.

Consider an observer centered co-ordinate system OXYZ as shown in Figure (B1). 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.

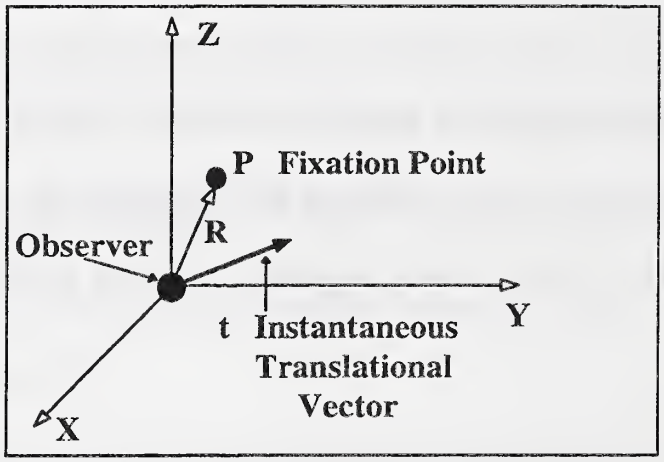


Figure (B1): Observer in motion, OXYZ is the observer centered co-ordinate system, **t** is the instantaneous translational vector, P is the fixation point, **R** is the range vector

The range **R** between the observer and the fixation point P in the observer centered co-ordinate system can be written as follow (for simplicity we chose only x-y plane, however the analysis can be easily extended to 3D space):

$$R = \sqrt{x^2 + y^2} \tag{B1}$$

$$\frac{dR}{dt} = \frac{x\dot{x} + y\dot{y}}{\sqrt{x^2 + y^2}} \tag{B2}$$

$$\frac{d^2R}{dt^2} = \frac{\sqrt{x^2 + y^2}\,(x\ddot{x} + y\ddot{y} + \dot{x}^2 + \dot{y}^2)}{x^2 + y^2} - \frac{(x\dot{x} + y\dot{y})^2}{(\sqrt{x^2 + y^2})^3} \tag{B3}$$

where (x, y, z) are the co-ordinates of the fixation point in OXYZ co-ordinate system, $(\dot{x}, \dot{y})$ is the instantaneous translational vector **t** of the observer in OXYZ coordinate system. $(\ddot{x}, \ddot{y})$ is the instantaneous acceleration of the observer in OXYZ coordinate system, d(.)/dt, $d^2(.)/dt^2$ are the temporal derivatives of (.). For a uniform translational velocity the acceleration component $(\ddot{x}, \ddot{y})$ is zero. Hence Equation (B3) reduces to the following:

$$\frac{d^2R}{dt^2} = \frac{\sqrt{x^2 + y^2}\,(\dot{x}^2 + \dot{y}^2)}{x^2 + y^2} - \frac{(x\dot{x} + y\dot{y})^2}{(\sqrt{x^2 + y^2})^3} \tag{B4}$$

By setting the TVTC in Equation (2) to zero leads to the following:

$$2RR_0\dot{R}^2 + RR_0^2\ddot{R} - R_0R^2\ddot{R} - R_0^2\dot{R}^2 = 0 \tag{B5}$$

By setting **t** = $(\dot{x}, 0)$ and substituting Equations [B1-B4] in Equation [B5] leads to the following:

$$R^3 - Rx^2 + R_0x^2 = R^2R_0 - R_0x^2 + 2Rx^2 \tag{B6}$$

Note that Equation (B6) is independent of the magnitude of observer's translational velocity.

By substituting $x = R\cos\theta$ in Equation (B6) and solving for $\cos(\theta)$ leads to the following relation:

$$\cos\theta = \pm\sqrt{\frac{1-k}{3-2k}} \tag{B7}$$

where $k = \frac{R_0}{R}$

for $R >> R_0$, $\cos\theta = \pm\frac{1}{\sqrt{3}}$

Appendix C: Smoothing of IQM using measured IQM Values

Since we need to compute the second derivative of the IQM, we need some temporal smoothing. The idea is based on using the past measured IQM values, a nth order polynomial is fit in a least squares sense, and then using this polynomial the derivatives of the IQM are computed. 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.

$$\sigma = k\left(\frac{1}{R_0} - \frac{1}{R}\right) \tag{C1}$$

In reality, only the left hand side of the above Equation is measurable (in the form of $(IQM)^{-1}$. In order to fit a curve to σ i.e., $(IQM)^{-1}$, we need an independent variable. An obvious choice would be the range. But in practice the range is not known *a-prori*. Hence we can not choose range as independent variable. Since *time* is a factor in the computation of the VTC as well as the TVTC from IQM values, we chose *time* as our independent variable, in other words $\sigma_{approx} = f(t)$.

Let $f(t) = a_0+a_1t+a_2t^2+...+ a_nt^n$ (C2)

In other words using the past measured IQM (i.e., σ^{-1}) values, we fit a nth order polynomial in a least squares sense.

We simulated the σ (i.e., IQM^{-1}) denoted as $\sigma_{measured}$ for various values of range R, using Equation (C1). Then employing these values as the input, we generated an nth order polynomial of the form given in Equation (C2). Then we computed the error between the

measured values and the smoothed values for various values of n. Based on our simulation results we observed that the order of the polynomial has to be at least six in order to minimize the error between the polynomial and the measured data.

References

[1] J. T. Schwartz and M. Sharir (1988), "A Survey of Motion Planning and Related Algorithms," *Artificial Intelligence*, Vol. 37, pp. 157-169.

[2] O. Khatib (1985), "Real-Time Obstacle Avoidance for Manipulators and Mobile Robots," *Proc. IEEE International Conf. on Robotics and Automation*, ST. Louis, MS, pp. 500-505.

[3] J. Barraquand, B. Langlois, and J-C. Latombe (1992), "Numerical Potential Field Techniques for Robot Path Planning," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 22, No. 2, pp. 224-241.

[4] R. A. Brooks (1989), "A robust layered control system for a mobile robot," *IEEE Journal of Robotics and Automation*, Vol. RA-2, #1, pp. 14-23.

[5] R. A. Brooks and J. H. Connell (1987), "Asynchronous distributed control system for a mobile robot," *Proc. SPIE, Mobile Robots*, Vol. 727, pp. 77-84.

[6] R. C. Arkin (1989), "Motor schema-based mobile robot navigation," *International Journal of Robotics Research*, August 1989, pp. 92-112.

[7] J. Borenstein and Y. Koren (1988), "Real time obstacle avoidance for mobile robots," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 19, #5, pp. 1179-1187.

[8] J. Borenstein and Y. Koren (1991), "Vector Field Histogram," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 19, #5, pp. 1179-1187.

[9] J. Barraquand, B. Langlois, J-C Latombe (1992), "Numerical Potential Field Techniques for Robotic Path Planning," *IEEE Transactions on Systems, Man and Cybernetics*, SMC-22, #2, pp. 224-241.

[10] Y. K. Hwang and N. Ahuja (1992), "A Potential Field Approach to Path Planning," *IEEE Transactions on Robotics and Automation*, Vol. 8, pp. 23-32.

[11] O. Khatib (1986), "Real-Time Obstacle Avoidance for Manipulators and Mobile Robots," *International Journal of Robotics Research*, Vol. 5, #1, pp. 90-98.

[12] P. Khosla and R. Volpe (1988), "Superquadratic Artificial Potentials for Obstacle Avoidance," in *Proc. of Conference on Decision and Control*, pp. 180-185.

[13] Y. Koren and J. Borenstein (1991), "Potential Field Methods and Their Inherent Limitations for Mobile Robot Navigation," *Proc. of the International Conference on Robotics and Automation*, Sacramento, CA, pp. 1398-1404.

[14] R. B. Tilove (1989), "Local Obstacle Avoidance for Mobile Robots Based on the Method of Artificial Potentials," *General Motors Research Laboratories*, Research Publication GMR-6650.

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

[16] S. R. Kundur and D. Raviv (1994), "An Image-Based Texture-Independent Visual Motion Cue for Autonomous Navigation," Technical Report *NISTIR 5567*, NIST, Gaithersburg, Maryland.

[17] S. R. Kundur and D. Raviv (1995), "Vision-Based Fuzzy Controllers for Autonomous Navigation Tasks," in *Proc. IEEE International Computer Vision Symposium, Miami.*

[18] D. Raviv and M. Herman (1994), "A Unified Approach to Camera Fixation and Vision-Based Road Following," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 24, #8, pp. 1125-1141

[19] J. Aloimonos, I. Weiss and A. Bandhopadhay (1987), "Active Vision," *Proc. First International Conference on Computer Vision*, pp. 35-54.

[20] H. V. Helmholtz (1925), *Treatise on Physiological Optics*, (Ed. J. P. C. Southall), Dover, NY.

[21] 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.

[22] 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.

[23] 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.