Chain Code Probabilities and Optimal Length Estimators for Digitized Three Dimensional Curves

Nahum Kiryati * **and** **Olaf Kübler**

Image Science Group

Communication Technology Laboratory

Swiss Federal Institute of Technology

CH-8092 Zurich, Switzerland

Abstract

Length estimation of three dimensional digitized curves is considered. Three dimensional 26-directional chain code representation is assumed. Recent results on $3-D$ Euclidean Distance Maps are inapplicable to optimal $3-D$ length estimation. To support optimal design of three dimensional length estimators, chain code probabilities in three dimensional chain encoded digital straight lines are derived. Optimal three dimensional length estimators are determined and their performance is evaluated analytically and experimentally.

Key words: chain codes, digital geometry, length estimation, perimeter estimation, 3-D shape analysis

*Nahum Kiryati is currently with the Department of Electrical Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel.

1 Introduction

Three dimensional shape analysis has received increasing attention in recent years. MRI, CT and other medical imaging systems, remote range sensors including Radar and Sonar, and computer vision techniques such as structured light and shape from stereo provide three dimensional volume or surface data.

Length measurement is an important element of shape analysis tasks; it is therefore not surprising that perimeter estimation of two dimensional shapes has received considerable attention in the last two decades. In three dimensional shape analysis, length estimation is the key to various measurements on surfaces [1].

The principal difficulty in length estimation is that length along a continuous curve in the preimage should be estimated from the chain code representation of the digitized curve. Information is lost in the digitization process, and the length estimation problem would be hopeless without incorporation of apriori knowledge. It must be assumed that the digitization grid is fine enough such that the digitized curves are reasonable approximations of the original continuous curves. This implies that curves can roughly be approximated by straight lines within pixels or voxels (volume elements).

Reconstruction of the original curve by interpolation and subsequent measurement of its length is computationally complex. Useful two dimensional length estimation algorithms classify the chain code [2, 3] links according to simple properties, and obtain a length estimate as a linear combination of the number of links in each class. These are referred to in this paper as *link classification* estimators. Their main advantage is their computational simplicity that results from their local nature: the contribution of a chain code link to the total length estimate depends only on its class, which is designed to depend on a few nearby links at most. The design of such estimators consists of defining the classification rules and fixing the weight constant for each class. *Link counting* estimators are degenerate link classification estimators, in which the length estimate is proportional to the total number of links.

This paper is concerned with estimating the length of three dimensional curves. The design procedure focuses however on straight lines: if curves can be approximated by straight lines within small neighborhoods or even within single pixels[1], then length estimators that are optimal for straight lines perform very well on general curves. In particular, assuming a uniform distribution of curve tangent orientations, then if the estimator is *unbiased* for straight lines of uniformly distributed orientations, the estimation error for curves is very small. The error approaches zero as the spatial resolution increases. It is however important not to confuse the goal, i.e., estimating the length of general *curves*, with the design procedure that is based on straight lines. In this paper, unlike references [5, 6], it is not assumed that the curve to be measured is a straight line segment. Also notice that if sufficiently detailed information on the distribution of local curvatures in the input curves exists, then for a given finite spatial resolution it might be even better to optimize the design using circular arcs rather than straight lines.

The organization of this paper is as follows. Section 2 reviews 2-D perimeter estimators. 3-D length estimation basics are considered in section 3. Section 4 is concerned with optimal design of 3-D link counting estimators. In section 5 the relation between the length estimation problem and the design of local distances for distance transformations is discussed, based on results from reference [1]. Important properties of 3-D chain encoded digital straight lines are derived in section 6, and applied to the optimal design of simple link classification length estimators in section 7. Experimental results are reported in section 8 and conclusions are discussed in section 9.

Preliminary results were reported in reference [7].

2 Brief Review of Two Dimensional Perimeter Estimators

The two dimensional perimeter estimation problem is to estimate the perimeter of a shape from the chain code of its digitized contour. The digitization scheme and chain encoding algorithm are an important part of the problem definition and should always be specified.

Koplowitz and Bruckstein [8] define a general class of link classification perimeter estima-

[1] A criterion for selecting the density of digitization grids is studied in reference [4].

tors

$$\hat{L} = \sum_{all\,P_i} \Psi\,\{\,prop\,P_i\} \tag{1}$$

where P_i is a link in the chain code, $prop\,P_i$ is a list of properties associated with P_i, and Ψ is a function that classifies and assigns weights to the chain links.

Early perimeter estimators were derived from the length of the digitized contour. In the framework of equation 1 and assuming 4-connected chain code representation, the chain links were not classified and one simply substituted $\Psi = 1 \;\; \forall P_i$ to obtain

$$\hat{L} = N \tag{2}$$

where N is the total number of links in the chain code.

Equation 2 defines a link counting estimator that is accurate for horizontal or vertical straight lines. For other line orientations, the estimation errors are large, and *biased* in the sense that the average estimation error (assuming a uniform distribution of line orientations) is not zero. In fact, for straight lines the length is never underestimated. The estimator can be made unbiased by setting Ψ to be an appropriate constant.

If 8-connected chain code representation is assumed, not all chain links contribute similarly to the length of the digitized contour. In the framework of equation 1, Ψ can be set to 1 for vertical and horizontal (even) links, and to $\sqrt{2}$ for diagonal (odd) links. This leads to

$$\hat{L} = N_e + \sqrt{2}N_o \tag{3}$$

where N_e and N_o respectively denote the number of even and odd links [2]. This estimator is however also biased. Various estimators of the form

$$\hat{L} = \Psi_e N_e + \Psi_o N_o \tag{4}$$

have been suggested [9, 10], with Ψ_e and Ψ_o chosen to satisfy certain unbiasedness and optimality properties.

Focusing on points (rather than on links) in 4-directional chain codes, it has been suggested to classify points as "corner" or as "non-corner" points, and to use an estimator of the form

$$\hat{L} = \Psi_c N_c + \Psi_n N_n \tag{5}$$

where N_c and N_n respectively denote the number of corner and non-corner points, and Ψ_c and Ψ_n are appropriate weights [8, 11]. Accounting for the number of corners has also been suggested for perimeter estimators based on 8-directional chain code [12]. See reference [13] for comparative study on 2-D link classification perimeter estimators.

Designing length estimators to satisfy optimality criteria in the case of straight lines requires, for any straight line orientation, to predict the number of chain links in each class specified in the estimation procedures. This is not difficult in the two dimensional case; see references [3, 8, 12, 14]. For example, in the 4-connected chain code representing a long digitized straight line of length L that makes an angle θ with the positive x-axis (assume without loss of generality $\theta \in [0, \pi/4]$), the number of horizontal links is $L\cos\theta$ and the number of vertical links is $L\sin\theta$. Minority codes appear isolated, so the number of corner points is $2L\sin\theta$, and the number of non-corner points is $L\cos\theta - L\sin\theta$. The number of links satisfying other simple properties can be similarly derived, enabling prediction of $\hat{L}$ as a function of θ and optimal design of length estimators.

The properties of the chain links that the above estimators use for classification are local in the sense that the contribution of a chain link to the estimated length depends only on its absolute direction (equations 3 and 4) or on its relative direction with respect to the previous chain link (equation 5). One estimator suggested in reference [8] uses a larger finite neighborhood.

It is known that in 4-connected chain code representation of infinite straight lines with orientation $\theta \in [\arctan 1/(j+1), \arctan 1/j]$, there are only groups of j and $j+1$ consecutive horizontal links ("horizontal runs") alternating with solitary vertical links. This observation led Koplowitz and Bruckstein [8] to also suggest a perimeter estimator that classifies each point according to the length of the longest run it belongs to. Thus, noncorner points are classified according to the run to which they belong, and corner points according to the longer of the two runs. An important characteristic of this estimator is that the number of classes to which a chain point can be classified, as well as the size of the neighborhood on which the classification is based, are in principle infinite, and are bounded only by the finite dimensions of the image, hence this perimeter estimator is not local. In practice, the number of classes can be bounded by assigning all points that belong to a run longer than a certain limit to one class. If the contribution of a point that belongs to a run of length j to the length estimate is taken to be $\sqrt{1+j^2}$, the estimator actually measures the length of a corner-smoothing polygonal approximation of the shape, and is similar to Wechsler's algorithm [15]. By appropriate tuning, unbiasedness and other optimality properties can be obtained. This is a bridge between link-classification local estimators, and estimators that are based on interpolation of the continuous contour and measurement of its length.

3 The 3-D Length Estimation Problem

The precise definition of the curve digitization process is crucial to the design and analysis of length estimators. Vaguely defined digitization schemes have been an obstacle to the development of optimal length estimators in the two dimensional case. Yet, in two dimensions the properties of digital *straight lines* obtained with various digitization schemes are quite similar. In three dimensions the dependence of the digital contour on the specifics of the digitization scheme is greater than in 2-D, so special care must be taken to clearly define the digitization method and the procedure for obtaining the chain code.

In this paper it is assumed that the continuous space is divided into cubic voxels. The digitization of a three dimensional curve consists of the voxels that the curve traverses. This set of voxels can be uniquely represented by a 6-directional chain code. It is also possible to represent a three dimensional curve by a 26-directional chain code, see Figure 1. In that case a voxel that is connected to two diagonally connected voxels may be omitted. The 26-directional chain code of a three dimensional digital contour is in general not unique. For example, a straight path between the two voxels (i,j,k) and $(i+2,j+1,k+1)$ can be described by either

- A direct link (incrementing i) and a major diagonal link (incremeneting i, j and k), or
- Two minor diagonal links, one that increments i and j and one that that increments i and k.

We have eliminated this ambiguity by specifying a starting voxel, a direction on the contour and by a "greedy" chain code generation scheme: if possible, add a major diagonal link; a minor diagonal link is the second priority and a direct link comes at the third priority. Other methods to obtain a 26-directional chain code representation can be defined; for a comparison of digital representations of curves in 3-D space see reference [16].

The 3-D length estimation problem has recently been addressed in reference [5]. One approach taken in that paper is direct extension to 3-D of the 2-D estimators defined by equations 2 and 3. Assuming a 26-directional chain code, these estimators are

$$\hat{L} = N \tag{6}$$

and

$$\hat{L} = N_1 + \sqrt{2}N_2 + \sqrt{3}N_3 \tag{7}$$

where N is the total number of links in the chain code, N_1 is the number of direct links, i.e., links that are parallel to one of the three main axes, and N_2 and N_3 are respectively the number of minor and major diagonal links in the chain code. Equation 6 describes a link counting estimator; equation 7 defines a basic link classification estimator. The 2-D length estimation experience suggests that these two 3-D length estimators are biased, and can be improved.

The next section focuses on the properties and optimal design of link counting 3-D length estimators. In particular, an unbiased estimator of the form

$$\hat{L} = \Psi N \tag{8}$$

(that reduces to equation 6 if Ψ is set to 1) is developed.

4 3-D Length Estimation from the Total Number of Links

Consider a long straight line segment of length L (in pixel side units). Placing one of its endpoints at the origin, the direction of the line is characterized by the angles θ and φ defined in Figure 2. The differences between the coordinates of the endpoints are

$$\Delta x = L\cos\theta\cos\varphi \tag{9}$$

$$\Delta y \; = \; L \cos\theta \sin\varphi \tag{10}$$

$$\Delta z \; = \; L \sin\theta. \tag{11}$$

Symmetry allows to assume without loss of generality that $0 \leq \varphi \leq \pi/4$ and $0 \leq \theta \leq \arctan(\sin\varphi)$; this implies that $\Delta x \geq \Delta y \geq \Delta z$.

With 26-directional chain coding, the total number of links in the representation of the line is

$$N \; = \; \Delta x \; = \; L \cos\theta \cos\varphi. \tag{12}$$

Combining equations 12 and 8,

$$\hat{L} \; = \; \Psi L \cos\theta \cos\varphi. \tag{13}$$

The estimation error is

$$\varepsilon(\theta,\, \varphi) \; = \; \frac{\hat{L}}{L} \; - \; 1 \; = \; \Psi \cos\theta \cos\varphi \; - \; 1, \tag{14}$$

bounded by

$$\frac{\sqrt{3}}{3}\Psi \; - \; 1 \; \leq \; \varepsilon(\theta,\, \varphi) \; \leq \; \Psi \; - \; 1. \tag{15}$$

In particular, by setting Ψ to 1, the error of the link counting estimator suggested in reference [5] (equation 6) is obtained, and is shown (as a function of φ and θ) in Figure 3; it is clearly biased.

To obtain an unbiased estimator assuming a uniform distribution of line orientations, Ψ should be chosen such that

$$\int_0^{\pi/4} \int_0^{\arctan\sin\varphi} \varepsilon(\theta,\, \varphi) \, \cos\theta \, d\theta d\varphi \; = \; 0. \tag{16}$$

$\cos\theta \, d\theta d\varphi$ is a differential solid angle element expressed in the coordinate system used in this paper.

Combining equations 14 and 16,

$$\Psi \int_0^{\pi/4} \int_0^{\arctan\sin\varphi} \cos^2\theta \cos\varphi \, d\theta d\varphi \; = \; \Omega, \tag{17}$$

where Ω, defined by

$$\Omega \equiv \int_0^{\pi/4} \int_0^{\arctan\sin\varphi} \cos\theta \, d\theta d\varphi \; = \; 0.261799, \tag{18}$$

is a solid angle (in *steradians*). Integration gives $\Psi = 1.2031$.

A simple unbiased link counting length estimator

$$\hat{L} = 1.2031N \tag{19}$$

has thus been obtained. The dependence of the estimation error in the case of straight lines on θ and φ is shown in Figure 4. The RMS error is

$$\varepsilon_{RMS} = \left[\frac{1}{\Omega}\int_0^{\pi/4}\int_0^{\arctan\sin\varphi}\varepsilon^2(\theta,\,\varphi)\,\cos\theta\,d\theta d\varphi\right]^{\frac{1}{2}} \tag{20}$$

and numerical integration yields $\varepsilon_{RMS} = 0.12042$, i.e., about 12 % .

The error of this unbiased estimator in estimating the length of straight lines is lower bounded by -30.5% and upper bounded by 20.3% . It is possible to modify Ψ such that the maximum absolute estimation error would be minimized. In particular, with $\Psi = 1.2679$ the maximum absolute error is minimized at 26.79%, but the estimator is biased.

An unbiased link counting estimator for the case of 6-directional chain code representation can be similarly designed. The total number of links in the 6-directional chain code of a straight line is

$$N = \Delta x + \Delta y + \Delta z = L(\cos\theta\cos\varphi + \cos\theta\sin\varphi + \sin\theta), \tag{21}$$

hence

$$\hat{L} = \Psi L(\cos\theta\cos\varphi + \cos\theta\sin\varphi + \sin\theta), \tag{22}$$

and the estimation error in this case is bounded by

$$\Psi - 1 \leq \varepsilon(\theta,\,\varphi) \leq \sqrt{3}\Psi - 1. \tag{23}$$

To achieve unbiasedness, Ψ must satisfy

$$\Psi\int_0^{\pi/4}\int_0^{\arctan\sin\varphi}(\cos\theta\cos\varphi + \cos\theta\sin\varphi + \sin\theta)\cos\theta\,d\theta d\varphi = \Omega. \tag{24}$$

Integration yields $\Psi = 0.6666$, so finally

$$\hat{L} = 0.6666N. \tag{25}$$

The dependence of the estimation error in the case of straight lines on θ and φ is shown in Figure 5. With this value of Ψ the RMS error is $\varepsilon = 10.1\%$, and the error bounds are -33.3% and 15.5%. It is again possible to modify Ψ such that the maximum absolute error would be minimized. This occurs at $\Psi = 0.7321$, and the resulting maximum error is again 26.79%.

To obtain more accurate length estimators, chain code links must be classified. To optimally design the weights, it is needed to predict how the links of chain encoded digital straight lines would be distributed among the classes. This is studied in the next sections.

5 Length Estimation and Distance Transformations

Simple link-classification 3-D length estimators are of the form

$$\hat{L} = \Psi_1 N_1 + \Psi_2 N_2 + \Psi_3 N_3 \tag{26}$$

where N_1, N_2 and N_3 are the number of the direct, minor diagonal and major diagonal links in the 26-directional chain code of the curve, and Ψ_1, Ψ_2 and Ψ_3 are weights to be designed. The estimation error is

$$\varepsilon(\theta, \varphi) = \frac{\hat{L}}{L} - 1 = \frac{\Psi_1 N_1 + \Psi_2 N_2 + \Psi_3 N_3}{L} - 1. \tag{27}$$

Ψ_1, Ψ_2 and Ψ_3 should be selected to obtain unbiased estimation and minimization of the RMS error over the set of straight lines in all possible orientations. Alternative optimality criteria can be defined.

Consider the design of local distances for 3-D distance transformations [17]. For the case of $3*3*3$ neighborhoods, the distance from the origin to a point on a large sphere around the origin is estimated to be

$$\hat{L} = \psi_1 n_1 + \psi_2 n_2 + \psi_3 n_3 \tag{28}$$

where n_1, n_2 and n_3 are the number of the direct, minor diagonal and major diagonal links in the 26-directional chain code that represents the shortest path, and ψ_1, ψ_2 and ψ_3 are referred to as local distances, and need to be designed. The estimation error is

$$\varepsilon(\theta, \varphi) = \frac{\hat{L}}{L} - 1 = \frac{\psi_1 n_1 + \psi_2 n_2 + \psi_3 n_3}{L} - 1. \tag{29}$$

ψ_1, ψ_2 and ψ_3 should be selected to obtain unbiased estimation and minimization of the RMS error over all possible orientations.

At first sight, the design of optimal weights for 3-D length estimators seems similar to the design of local distances for 3-D distance transforms. It has already been observed [1] that the similarity is misleading and follows from the false assumption that N_1, N_2 and N_3 are respectively equal to n_1, n_2 and n_3. The arguments prvided in reference [1] are as follows.

n_1, n_2 and n_3 are very easy to calculate. Assuming that $\psi_1 + \psi_3 < 2\psi_2$, the shortest path between two voxels on a 3-D grid always prefers a connection by a direct link and a major diagonal link to using two minor diagonal links. Suppose without loss of generality that the differences between the coordinates of the two end voxels of the path satisfy $\Delta x \geq \Delta y \geq \Delta z$. Then n_3 must be equal to Δz, n_2 is equal to $\Delta y - n_3$ and n_1 is equal to $\Delta x - n_2 - n_3$. This means that the minor diagonal link $x - z$ never occurs. Verwer [17] uses this approach to obtain the optimal local distances ψ_1, ψ_2 and ψ_3.

It may seem surprising that the situation is different in the design of length estimators, especially since in two dimensions the optimal design of weights for length estimators is mathematically identical to the design of local distances for distance transformations. This stems from the fact that all possible digital straight lines that connect two given pixels in 2-D have exactly the same number of direct and diagonal links in their 8-connected chain code. Moving to three dimensions, consider the infinite group of all possible continuous straight lines that pass through two distant end voxels. These continuous lines correspond to a set of 3-D digital straight lines and a respective set of 26-connected chain codes. The values of N_1, N_2 and N_3 are in general not identical for all the chain codes. Suppose for example that the differences between the coordinates of the end voxels satisfy $\Delta x \gg \Delta y \gg \Delta z$. Transitions in the y direction between voxels along the line are thus infrequent, and transitions in the z direction are extremely infrequent. In certain cases, transitions in the y and z directions will indeed take place near each other. A major diagonal link will then be added to the 26-connected chain code. But in most locations along the long line, transitions in the y and z directions will be separated by many transitions in the x direction. This leads to minor diagonal links of the $x - y$ type, as well as minor diagonal links in the $x - z$ direction!

Chain encoding of an actually digitized line cannot always "force" the replacement of a minor diagonal link in the $x - y$ direction and a minor diagonal link in the $x - z$ direction by a direct link and a major diagonal link.

The important and surprising occurance of $x - z$ links implies that N_1, N_2 and N_3 for straight lines are not in general equal to n_1, n_2 and n_3. While n_1, n_2 and n_3 are easy to compute, the calculation of N_1, N_2 and N_3 is more difficult. It is carried out in the next section and is thus an important contribution in itself. The chain code probabilities for $3 - D$ digital lines are then applied to the optimal design of weights for 3-D length estimators. The weights are certainly different than the local distances for 3-D distance transforms determined by Verwer [17].

In distance transformations the chain code of the path that corresponds to the *shortest estimated distance* between the source and destination voxels is constructed by the algorithm. For that case optimal results have been given by Verwer [17]. In length estimation problems, the original continuous curves dictate the structure of their chain codes: not all digital straight lines between voxels in $3 - D$ correspond to the shortest estimated distance. This is a different situation, and the results of [17] do not apply. Our research focuses on the length estimation problem.

Beckers and Smeulders have recently published another related paper [18]. Even though the paper does not make a clear distinction between length estimation and distance transformations, then following the above discussion, it should be categorized as dealing with the design of distance transformations. It is interesting to observe that the local distances found by Beckers and Smeulders [18] are different than those of Verwer [17]. The differences are partly due to the optimality criterion used in reference [18].

6 Chain Code Probabilities in the Representation of 3-D Digital Straight Lines

Consider again a straight line segment of length L (in pixel side units) as shown in Figure 2. Assume that L is large, and that (without loss of generality) $\Delta x \geq \Delta y \geq \Delta z$. Suppose this line is represented by a 6-directional chain code; then only three link directions appear in the chain code: x, y and z.

Let $P(x)$, $P(y)$ and $P(z)$ denote the probabilities that a chain link picked at random will respectively be in the x, y or z direction. Then

$$P(x) = \frac{\Delta x}{\Delta x + \Delta y + \Delta z}, \tag{30}$$

thus

$$P(x) = \frac{\cos\theta\cos\varphi}{\cos\theta\cos\varphi + \cos\theta\sin\varphi + \sin\theta}. \tag{31}$$

Similarly,

$$P(y) = \frac{\cos\theta\sin\varphi}{\cos\theta\cos\varphi + \cos\theta\sin\varphi + \sin\theta} \tag{32}$$

$$P(z) = \frac{\sin\theta}{\cos\theta\cos\varphi + \cos\theta\sin\varphi + \sin\theta}. \tag{33}$$

Suppose that one of the x links is picked at random; the probabilities that the next link is x, y and z are denoted $P(x|x)$, $P(y|x)$ and $P(z|x)$ respectively. In a similar way the probabilities $P(x|y)$, $P(y|y)$, $P(z|y)$, $P(x|z)$, $P(y|z)$ and $P(z|z)$ can be defined, but note that the assumption that $\Delta x \geq \Delta y \geq \Delta z$ and the properties of digital straight lines dictate that $P(y|y) = P(z|z) = 0$.

A parametric representation of the line is

$$x(t) = x_0 + t\cos\theta\cos\varphi \tag{34}$$

$$y(t) = y_0 + t\cos\theta\sin\varphi \tag{35}$$

$$z(t) = z_0 + t\sin\theta \tag{36}$$

where (x_0, y_0, z_0) is a point on the line, and t is the parameter. Assume without loss of generality that the coordinate system origin is translated to an appropriate corner of a voxel under consideration, such that (x_0, y_0, z_0) is the point of entry of the line to that voxel, $0 \leq x_0, y_0, z_0 \leq 1$, and within the voxel $t \geq 0$.

If the current link in the chain code is x, then obviously $x_0 = 0$. For the next link to be x, there must be $t > 0$ that satisfies the following conditions:

$$t\cos\theta\cos\varphi = 1 \tag{37}$$

$$y_0 + t\cos\theta\sin\varphi < 1 \tag{38}$$

$$z_0 + t\sin\theta < 1. \tag{39}$$

Elimination of t leads to

$$0 < y_0 < 1 - \tan\varphi \tag{40}$$

$$0 < z_0 < 1 - \frac{\tan\theta}{\cos\varphi}. \tag{41}$$

Assume that the ratios $\Delta x/\Delta y$, $\Delta x/\Delta z$ and $\Delta y/\Delta z$ are irrational numbers. Recalling that underlying straight lines are in the continuous domain, only a countable set of of pairs (θ, φ) does not satisfy this assumption, so "almost all" pairs satisfy it, and a pair (θ, φ) picked at random satisfies the assumption with probability 1. If L grows to infinity, and assuming that the preceding x link was selected at random, y_0 and z_0 can be regarded as independent random variables, uniformly distributed between 0 and 1. This implies that

$$P(x|x) = \left(1 - \frac{\tan\theta}{\cos\varphi}\right)(1 - \tan\varphi). \tag{42}$$

Under the same assumptions and using similar derivations, it can be shown that

$$P(y|x) = \tan\varphi\left(1 - \frac{\tan\theta}{2\cos\varphi}\right) \tag{43}$$

$$P(z|x) = \frac{\tan\theta}{\cos\varphi}\left(1 - \frac{1}{2}\tan\varphi\right) \tag{44}$$

$$P(x|y) = 1 - \frac{\tan\theta}{2\cos\varphi} \tag{45}$$

$$P(z|y) = \frac{\tan\theta}{2\cos\varphi} \tag{46}$$

$$P(x|z) = 1 - \frac{1}{2}\tan\varphi \tag{47}$$

$$P(y|z) = \frac{1}{2}\tan\varphi. \tag{48}$$

Properties of digital lines dictate that in the chain code of a line satisfying the above assumptions, an x link can be followed only by one of the following patterns:

- Another x link.

- A y link followed by x.
- A z link followed by x.
- A y link followed by z followed by x.
- A z link followed by y followed by x.

The probabilities that these patterns appear after a randomly selected x link are denoted $P(x|x)$, $P(yx|x)$, $P(zx|x)$, $P(yzx|x)$ and $P(zyx|x)$ respectively. $P(x|x)$ has already been determined (equation 42).

The conditions for the second pattern (yx) to follow an x link are:

$$t_1 \cos\theta \cos\varphi < 1 \quad (49)$$

$$y_0 + t_1 \cos\theta \sin\varphi = 1 \quad (50)$$

$$z_0 + t_1 \sin\theta < 1, \quad (51)$$

which are the conditions for the next link to be y, and

$$(t_1 + t_2) \cos\theta \cos\varphi = 1 \quad (52)$$

$$t_2 \cos\theta \sin\varphi < 1 \quad (53)$$

$$z_0 + (t_1 + t_2) \sin\theta < 1, \quad (54)$$

that ensure that the consecutive link is again x. t_1 and t_2 are two positive parameters. By eliminating t_1 and t_2 these conditions reduce to

$$1 > \quad y_0 \quad > 1 - \tan\varphi \quad (55)$$

$$1 - \frac{\tan\theta}{\cos\varphi} > \quad z_0 \quad > 0. \quad (56)$$

With y_0 and z_0 being again independent uniformly distributed random variables,

$$P(yx|x) = \tan\varphi \left(1 - \frac{\tan\theta}{\cos\varphi}\right). \quad (57)$$

Similarly,

$$P(zx|x) = (1 - \tan\varphi) \frac{\tan\theta}{\cos\varphi}, \quad (58)$$

and

$$P(yzx|x) = P(zyx|x) = \frac{\tan\theta\tan\varphi}{2\cos\varphi}. \tag{59}$$

It is stressed that these probabilities do not apply to the (countable) set of straight lines with θ and φ leading to rational ratios between Δx, Δy and Δz. For those lines y_0 and z_0 are periodic, and are not uniformly distributed.

If a 26-directional chain code is used, the links can be classified according to their length: direct links of length 1, minor diagonal links of length $\sqrt{2}$ and major diagonal links of length $\sqrt{3}$. Let P_1, P_2 and P_3 respectively denote the probability that a chain link picked at random belongs to one of these classes. The relation between 6-directional and 26-directional 3-D chain codes is simple for straight lines; it follows that

$$P_1 = P(x|x) \tag{60}$$

$$P_2 = P(yx|x) + P(zx|x) \tag{61}$$

$$P_3 = P(yzx|x) + P(zyx|x). \tag{62}$$

Explicitly, for $0 \leq \varphi \leq \pi/4$ and $0 \leq \theta \leq \arctan(\sin\varphi)$,

$$P_1 = \left(1 - \frac{\tan\theta}{\cos\varphi}\right)(1 - \tan\varphi) \tag{63}$$

$$P_2 = (1 - \tan\varphi)\frac{\tan\theta}{\cos\varphi} + \tan\varphi\left(1 - \frac{\tan\theta}{\cos\varphi}\right) \tag{64}$$

$$P_3 = \frac{\tan\theta\tan\varphi}{\cos\varphi}. \tag{65}$$

In the next section these results are applied to the optimal design of 3-D length estimators.

7 3-D Link Classification Length Estimators

This section focuses on optimal design of length estimators of the form

$$\hat{L} = \Psi_1 N_1 + \Psi_2 N_2 + \Psi_3 N_3 \tag{66}$$

where N_1, N_2 and N_3 are the number of the direct, minor diagonal and major diagonal links in the 26-directional chain code of the curve, and Ψ_1, Ψ_2 and Ψ_3 are the weights to be designed. The estimation error is

$$\varepsilon(\theta, \varphi) = \frac{\hat{L}}{L} - 1 = \frac{\Psi_1 N_1 + \Psi_2 N_2 + \Psi_3 N_3}{L} - 1. \tag{67}$$

The number of links N in the 26-directional chain code representation of a long straight line of length L is given by equation 12. Thus, for "almost all" angles in the domain $0 \leq \varphi \leq \pi/4$ and $0 \leq \theta \leq \arctan(\sin\varphi)$,

$$N_1 = NP_1 = L(\cos\theta\cos\varphi - \sin\theta)(1 - \tan\varphi) \tag{68}$$

$$N_2 = NP_2 = L[(1 - \tan\varphi)\sin\theta + \tan\varphi(\cos\theta\cos\varphi - \sin\theta)] \tag{69}$$

$$N_3 = NP_3 = L\sin\theta\tan\varphi. \tag{70}$$

With $\Psi_1 = 1$, $\Psi_2 = \sqrt{2}$ and $\Psi_3 = \sqrt{3}$, the second estimator suggested in [5] is obtained (equation 7). In the case of straight lines, the dependence of the estimation error of this estimator on φ and θ is shown in Figure 6. The estimation is clearly biased; in fact, the length of straight lines is never underestimated. In the sequel a length estimator that is unbiased for long straight lines (assuming uniform distribution of orientations) with the least RMS error possible for unbiased estimators is developed.

Unbiasedness implies that

$$C(\Psi_1, \Psi_2, \Psi_3) = \frac{1}{\Omega}\int_0^{\pi/4}\int_0^{\arctan\sin\varphi} \varepsilon(\theta, \varphi)\cos\theta d\theta d\varphi = 0. \tag{71}$$

Carrying out the integration yields:

$$C(\Psi_1, \Psi_2, \Psi_3) = 0.299738\Psi_1 + 0.394096\Psi_2 + 0.137358\Psi_3 - 1 = 0. \tag{72}$$

Ψ_1, Ψ_2 and Ψ_3 must be chosen to minimize the RMS error, or equivalently its square

$$\varepsilon_{MSE}(\Psi_1, \Psi_2, \Psi_3) = \frac{1}{\Omega}\int_0^{\pi/4}\int_0^{\arctan\sin\varphi} \varepsilon^2(\varphi, \theta)\cos\theta d\theta d\varphi \tag{73}$$

while satisfying the unbiasedness constraint (equation 72).

This constrained minimization problem can be solved by Lagrange optimization. The Lagrangian is defined as

$$T(\Psi_1, \Psi_2, \Psi_3, \lambda) = \varepsilon_{MSE}(\Psi_1, \Psi_2, \Psi_3) + \lambda C(\Psi_1, \Psi_2, \Psi_3) \tag{74}$$

where λ is a Lagrange multiplier. The optimal solution satisfies

$$\frac{\partial}{\partial \Psi_1} T(\Psi_1, \Psi_2, \Psi_3, \lambda) = 0 \tag{75}$$

$$\frac{\partial}{\partial \Psi_2} T(\Psi_1, \Psi_2, \Psi_3, \lambda) = 0 \tag{76}$$

$$\frac{\partial}{\partial \Psi_3} T(\Psi_1, \Psi_2, \Psi_3, \lambda) = 0 \tag{77}$$

$$\frac{\partial}{\partial \lambda} T(\Psi_1, \Psi_2, \Psi_3, \lambda) = 0. \tag{78}$$

These conditions lead to the following set of linear equations:

$$\alpha_{11}\Psi_1 + \alpha_{12}\Psi_2 + \alpha_{13}\Psi_3 + 0.2997\lambda = \beta_1 \tag{79}$$

$$\alpha_{21}\Psi_1 + \alpha_{22}\Psi_2 + \alpha_{33}\Psi_3 + 0.1374\lambda = \beta_2 \tag{80}$$

$$\alpha_{31}\Psi_1 + \alpha_{32}\Psi_2 + \alpha_{33}\Psi_3 + 0.3941\lambda = \beta_3 \tag{81}$$

and to equation 72, where

$$\alpha_{ij} \equiv \frac{2}{\Omega} \int_0^{\pi/4} \int_0^{\arctan \sin \varphi} \frac{N_i}{L} \frac{N_j}{L} \cos \theta d\theta d\varphi, \tag{82}$$

and

$$\beta_i \equiv \frac{2}{\Omega} \int_0^{\pi/4} \int_0^{\arctan \sin \varphi} \frac{N_i}{L} \cos \theta d\theta d\varphi. \tag{83}$$

This system of equations can be expressed in matrix form as follows, with $\{\alpha_{ij}\}$ and $\{\beta_i\}$ replaced by their numerical values:

$$\begin{bmatrix} 0.293884 & 0.211097 & 0.0389764 & 0.299738 \\ 0.211097 & 0.33871 & 0.100297 & 0.394096 \\ 0.0389764 & 0.100297 & 0.0684386 & 0.137358 \\ 0.299738 & 0.394096 & 0.137358 & 0 \end{bmatrix} \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} 0.599476 \\ 0.78819 \\ 0.274717 \\ 1 \end{bmatrix}. \tag{84}$$

The solution to the system is $\Psi_1 = 0.901588$, $\Psi_2 = 1.28876$, $\Psi_3 = 1.61524$, (and $\lambda = -0.00165172$). The unbiased length estimator with the least RMS error is therefore

$$\hat{L} = 0.9016 N_1 + 1.289 N_2 + 1.615 N_3. \tag{85}$$

The dependence of the estimation error in the case of straight lines on θ and φ is shown in Figure 7. It is lower bounded by -9.84% and upper bounded by 4.59% . The RMS error of this estimator is 2.88%, a very significant improvement with respect to the 10% – 12% RMS errors of the unbiased link counting estimators discussed in section 4, and with respect to the RMS error of the biased link classification estimator suggested in reference [5] (equation 7), which is also about 10%.

8 Experiments

To complement the analytical results, an experimental evaluation of these length estimators in the case of circles in three dimensional space has been carried out. A generic experiment consisted of placing a circle of known radius in the XY plane, centered at the origin. The circle was rotated around the x and y axes at angles taken from a uniform distribution between 0 and π in each coordinate, then translated in all three coordinates by random values, taken from a uniform distribution between 0 to 1 in each coordinate. The circle was then digitized, its chain code was obtained, and its perimeter was estimated by each of the studied estimators and compared to the known radius of the original circle.

In practice, a set of 100 quintuples of randomly chosen transformation parameters has been repeatedly used per each tested radius and each of the studied estimators. This means that comparison of performance between estimators is always based on the same input circles. About 25 radii between 1 and 100 were used, with increments down to 0.5 in the lower range, up to 10 in the upper range.

The 100 outcomes per each radius and each tested estimator were processed to obtain the following quantities:

- The average (percent) estimation error. For unbiased estimators, the average can be expected to approach zero as the radius increases. For biased estimators the average is expected to approach the bias.

- The maximum and minimum (percent) estimation errors among the 100 trials. These indicate the variability of the estimates.

- The average of the *absolute* estimation errors (percent). This quantity is an important quality indicator.

Figure 8 shows the average error as a function of radius of the biased 26-directional link counting length estimator, suggested in reference [5] (equation 6); the minimum and maximum errors are also shown. Due to its large bias, this estimator is rather impractical for most applications.

Figure 9 shows the average, minimum and maximum errors as a function of radius for the unbiased version developed in this paper (equation 19) of the 26-directional link counting length estimator. The performance is dramatically improved, and the average error approaches zero as expected.

Figure 10 shows the average, minimum and maximum errors as a function of radius for the unbiased *6-directional* chain code link counting estimator (equation 25). The average error is nearly zero even at very small radii.

The average *absolute* errors of the above three link counting estimators are compared in Figure 11. It is clear that the performance of the biased estimator suggested in reference [5] (equation 6) is greatly inferior with respect to the two unbiased estimators suggested here. The unbiased estimator based on the 6-directional chain code is slightly but consistently superior to the estimator based on the 26-directional code.

Figure 12 shows the average, maximum and minimum errors as a function of radius for the (biased) link classification estimator (equation 7) suggested in reference [5]. Figure 13 shows the average, maximum and minimum errors as a function of radius for the link-classification estimator developed here (equation 85), which is the optimal unbiased estimator (for lines, in the RMS error sense) among all estimators of the general form of equation 66.

The average absolute errors of these two link classification estimators (equations 7 and 85) are compared in Figure 14. The good performance of the latter estimator is pleasing considering its simplicity.

Focusing on small radii, the average absolute errors of the three unbiased estimators developed in this paper are compared in Figure 15. The performance of the unbiased 26-directional link counting estimator (equation 19) is inferior with respect to the other two, especially at small radii. As expected, the link classification estimator (equation 85) is generally best, even though the simple 6-directional chain link counting estimator (equation 25) performs well at small radii.

9 Discussion

This paper deals with optimal design of length estimators for curves in three dimensional space from their chain codes. Only local estimators are considered, i.e., estimation algorithms in which the contribution of each chain link to the total length estimate does not depend on distant links. These include link counting estimators, with similar contribution for all links, and link classification estimators, in which the properties of a link determine its contribution; the latter have potential for better performance.

For reasons similar to those encountered in the design of 2-D perimeter estimators, 3-D length estimators should be designed to be unbiased for straight lines of uniform orientations. Unbiased length estimators can be expected to perform well on smooth curves with uniform distribution of tangent orientations.

The weight of a chain link is the only free parameter in the design of link counting estimators, so they are fully determined once the unbiasedness constraint is imposed. There are more degrees of freedom in the design of link classification estimators, hence further optimality criteria can be specified. The design goal of minimum RMS error for straight lines that was taken in this paper is one of many possible choices. In particular, it seems interesting to develop length estimators satisfying optimality criteria related to circular arcs.

The relation between the design of weights for length estimators and the design of local distances for Euclidean distance transformations has been considered in reference [1]. It has been demonstrated that while the two problems are mathematically similar in 2-D, they are inherently different in 3-D. In particular, the design of 3-D length estimators must be based on theoretical analysis of chain code probabilities in three dimensional chain encoded lines, and is thus more difficult than the design of 3-D distance transforms. Beyond the optimal estimators themselves, an important contribution of this paper is therefore the theoretical prediction of the number of direct, minor diagonal and major diagonal links (i.e., links of length 1, $\sqrt{2}$ and $\sqrt{3}$ respectively) in the chain code of an arbitrarily oriented line in 3-D. These results can be extended to enable prediction of the frequency of more complicated patterns in 3-D chain codes of lines, and provide a basis for the design of finer link classification length estimators. The lack of results on chain code probabilities made it impossible for the authors of [5] to optimize the design of their suggested estimators.

The experiments with perimeter estimation for circles in 3-D complement and validate the analytical results. It is pleasing that the average error of unbiased estimators is close to zero even at quite small radii, indicating the robustness of the suggested estimators. Indeed, finer link classification estimators can be expected to perform better on smooth curves, but due to the larger classification-support neighborhoods, their performance might rapidly deteriorate as curvature increases.

References

[1] N. Kiryati and G. Székely, Estimating Shortest Paths and Minimal Distances on Digitized Three-Dimensional Surfaces, *Pattern Recognition* **26**, 1623-1637 (1993).

[2] H. Freeman, Computer Processing of Line-Drawing Images, *Comput. Surveys* **6**, 57-97 (1974).

[3] J. Koplowitz, On the Performance of Chain Codes for Quantization of Line Drawings, *IEEE Trans. Pattern Analysis Machine Intelligence* **3**, 180-185 (1981).

[4] H. Freeman and J.M. Glass, On the Quantization of Line Drawing Data, *IEEE Trans. System Science and Cybernetics* **5**, 70-79 (1969).

[5] T.M. Amarunnishad and P.P. Das, Estimation of Length for Digitized Straight Lines in Three Dimensions, *Pattern Recognition Letters* **11**, 207-213 (1990).

[6] S. Chattopadhyay and P.P. Das, Estimation of the Original Length of a Straight Line Segment from its Digitization in Three Dimensions, *Pattern Recognition* **25**, 787-798 (1992).

[7] N. Kiryati and O. Kübler, On Chain Code Probabilities and Length Estimators for Digitized Three Dimensional Curves, Technical Report BIWI-TR-125, Institute for Com-

munication Technology, ETH Zurich (1991). Also in Proc. 11th ICPR, The Hague, The Netherlands, **A**, 259-262 (1992).

[8] J. Koplowitz and A.M. Bruckstein, Design of Perimeter Estimators for Digitized Planar Shapes, *IEEE Trans. Pattern Analysis Machine Intelligence* **11**, 611-622 (1989).

[9] F.C.A. Groen and P.W. Verbeek, Freeman-Code Probabilities of Object Boundary Quantized Contours, *Computer Graphics Image Processing* **7**, 391-402 (1978).

[10] Z. Kulpa, Area and Perimeter of Blobs in Discrete Binary Pictures, *Computer Graphics Image Processing* **6**, 434-451 (1977).

[11] D. Proffitt and D. Rosen, Metrication Errors and Coding Efficiency of Chain-Encoding Schemes For the Representation of Lines and Edges, *Computer Graphics Image Processing* **10**, 318-332 (1979).

[12] A.M. Vossepoel and A.W.M. Smeulders, Vector Code Probability and Metrication Error in the Representation of Straight Lines of Finite Length, *Computer Graphics Image Processing* **20**, 347-364 (1982).

[13] L. Dorst and A.W.M. Smeulders, Length Estimators for Digitized Contours, *Computer Vision Graphics Image Processing* **40**, 311-333 (1987).

[14] A.M. Bruckstein, Self-Similarity Properties of Digital Straight Lines, *Contemporary Methematics* **119** (1991).

[15] H. Wechsler, A New and Fast Algorithm for Estimating the Perimeter of Objects for Industrial Vision Tasks, *Computer Graphics Image Processing* **17**, 375-385 (1981).

[16] A. Jonas and N. Kiryati, Digital Representation of 3-D Curves: Survey and Analysis, Proceedings of the 2nd International Workshop on Visual Form, Capri, Italy (1994).

[17] B.J.H. Verwer, Local Distances for Distance Transformations in Two and Three Dimensions, *Pattern Recognition Letters* **12**, 671-682 (1991).

[18] A.L.D. Beckers and A.W.M. Smeulders, Optimization of Length Measurements for Isotropic Distance Transformations in Three Dimensions, *CVGIP: Image Understanding* **55**, 296-306 (1992).

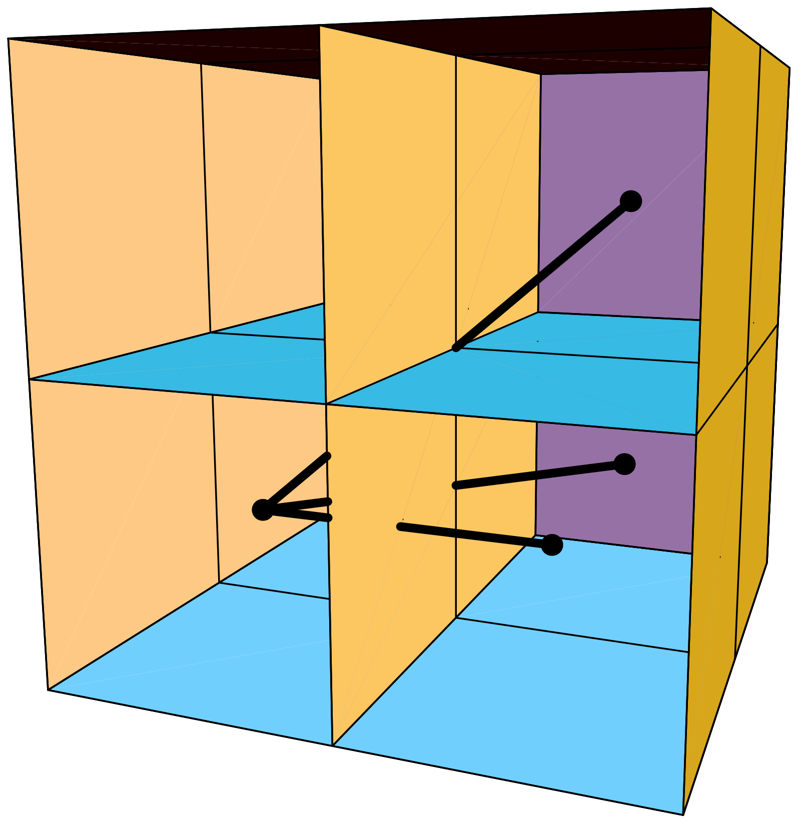

Figure 1: Link types in a 26-directional 3-D chain code: a direct link (parallel to one of the main axes), a minor diagonal link and a major diagonal link.

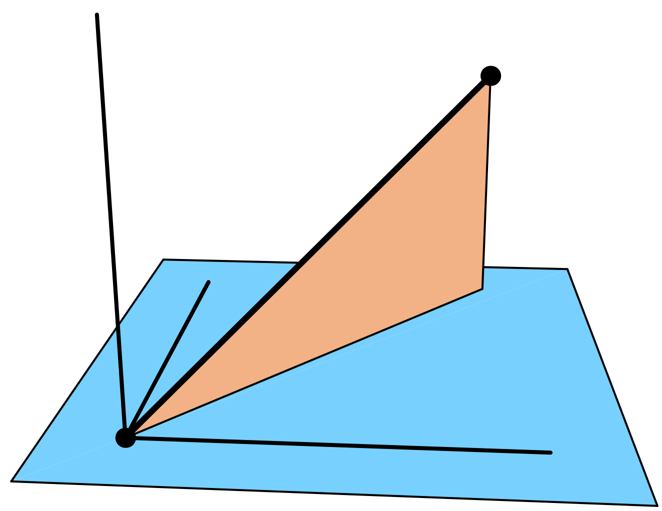

Figure 2: The definition of θ and φ.

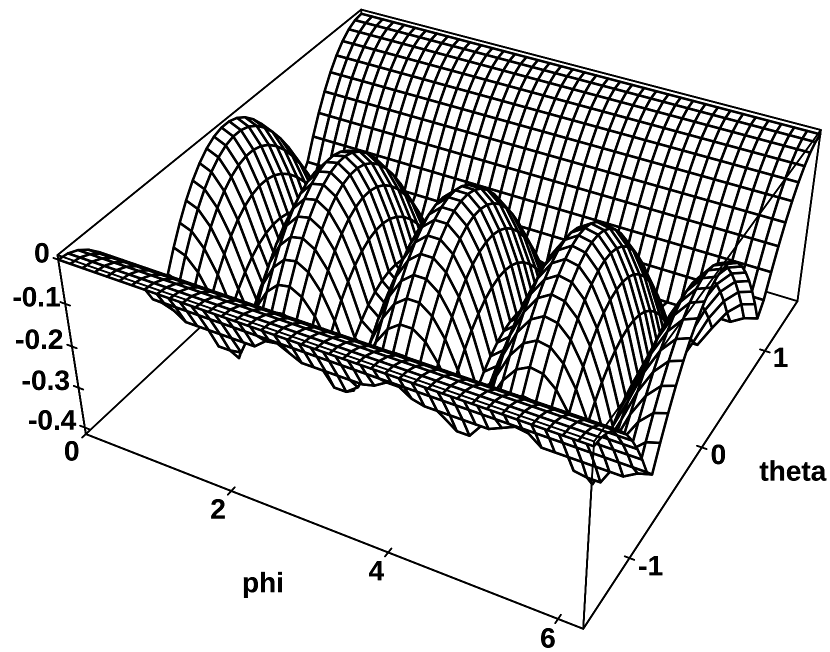

Figure 3: The dependence of the estimation error (in the case of straight lines) on φ and θ for the biased 26-directional link counting estimator. The error is non-positive, i.e., the length is never overestimated.

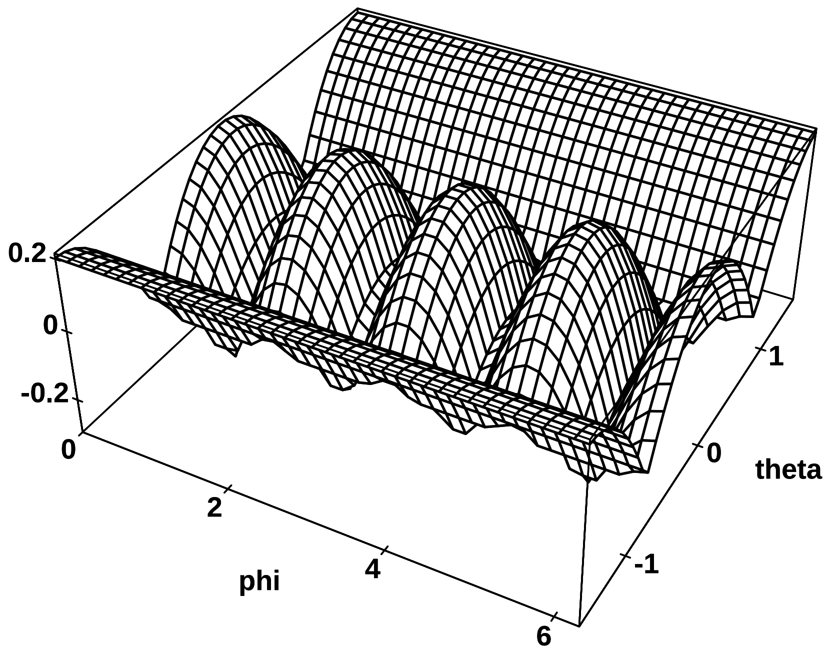

Figure 4: The dependence of the estimation error (in the case of straight lines) on φ and θ for the unbiased 26-directional link counting estimator.

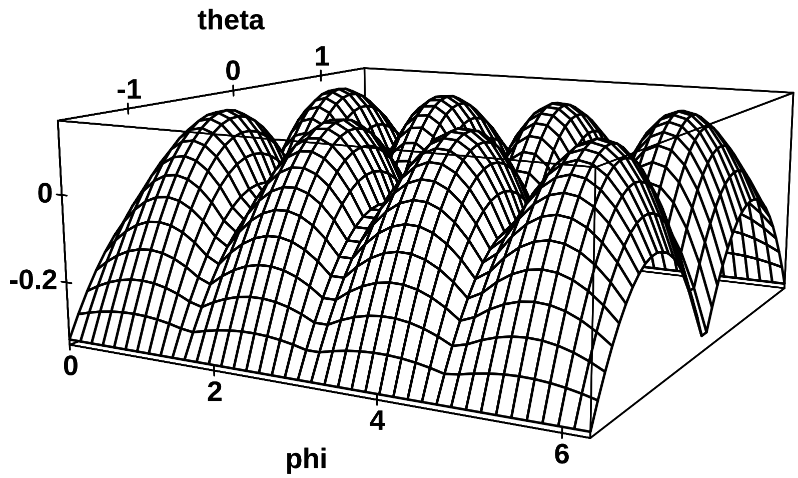

Figure 5: The dependence of the estimation error in the case of straight lines on φ and θ for the unbiased 6-directional link counting estimator.

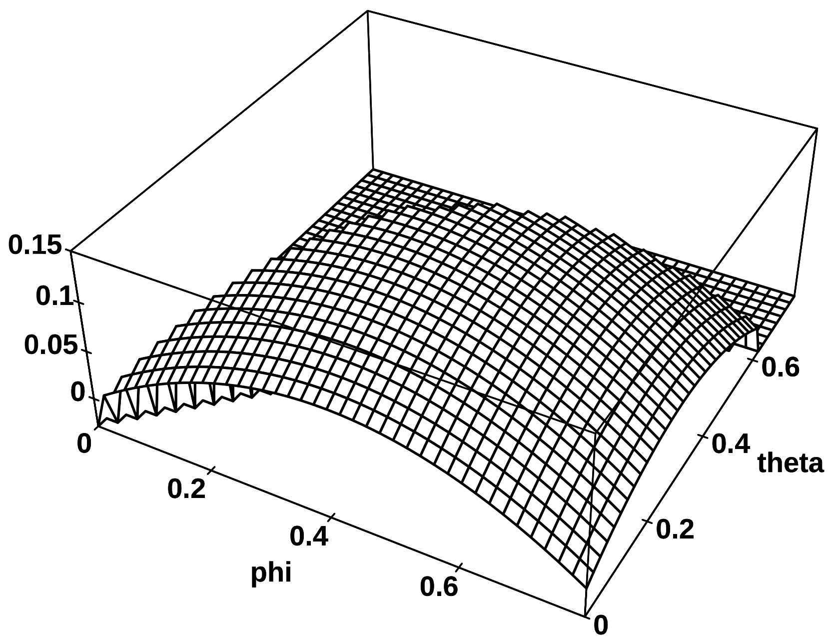

Figure 6: The dependence of the estimation error (in the case of straight lines) on φ and θ for the biased link classification estimator. The typical domain $0 \leq \varphi \leq \pi/4$, $0 \leq \theta \leq \arctan \sin \varphi$ is shown. The error is non-negative, i.e., the length is never underestimated.

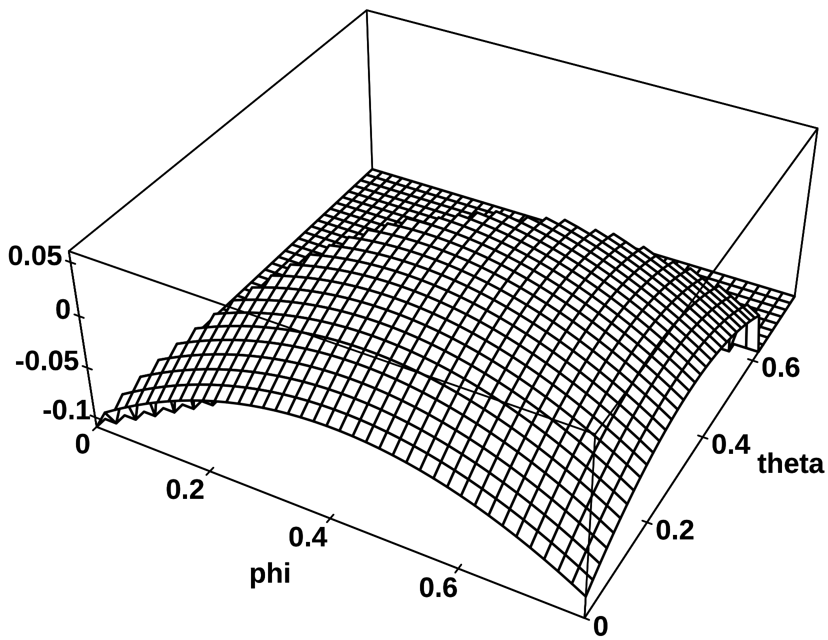

Figure 7: The dependence of the estimation error (in the case of straight lines) on φ and θ for the unbiased link classification estimator. The typical domain $0 \leq \varphi \leq \pi/4$, $0 \leq \theta \leq \arctan \sin \varphi$ is shown.

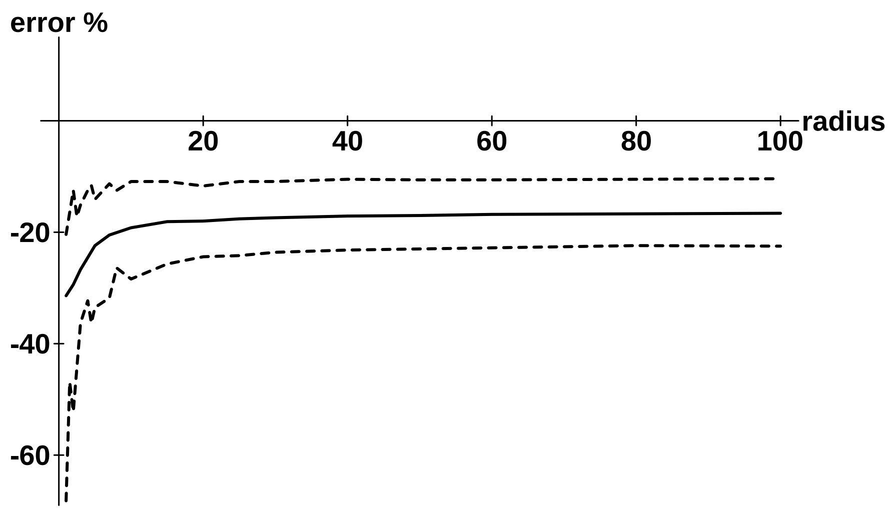

Figure 8: Perimeter estimation of circles in 3-D using the biased 26-directional link counting estimator. The solid curve shows the average error (of 100 experiments) as a function of radius. The dashed curves show the minimum and maximum errors encountered.

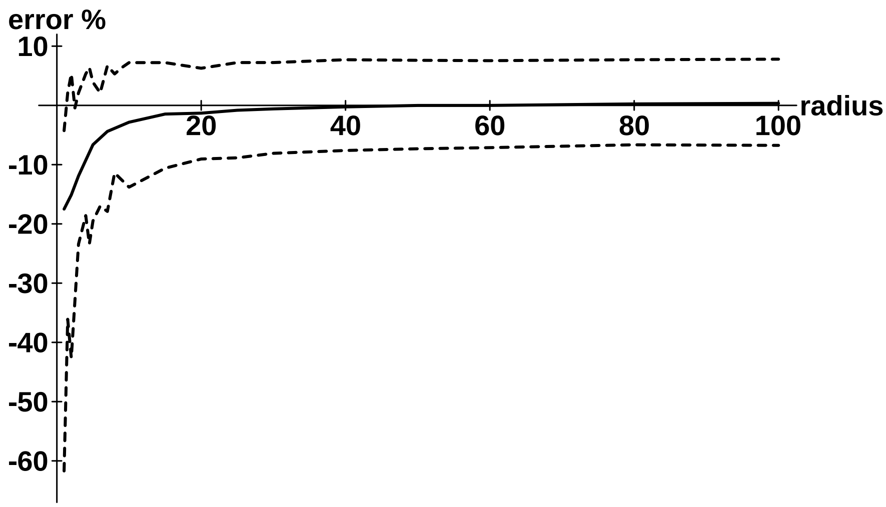

Figure 9: Perimeter estimation of circles in 3-D using the unbiased 26-directional link counting estimator developed in this paper. The solid curve shows the average error (of 100 experiments) as a function of radius. The dashed curves show the minimum and maximum errors encountered.

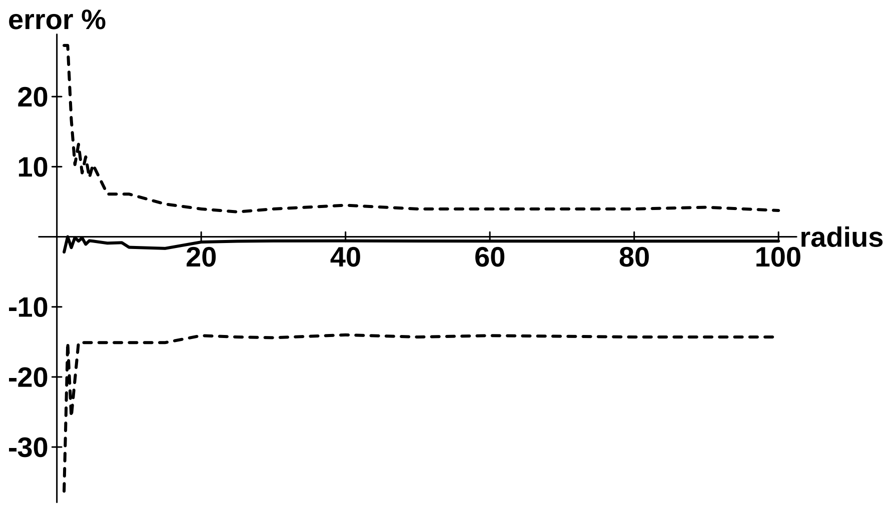

Figure 10: Perimeter estimation of circles in 3-D using the unbiased 6-directional link counting estimator developed in this paper. The solid curve shows the average error (of 100 experiments) as a function of radius. The dashed curves show the minimum and maximum errors encountered.

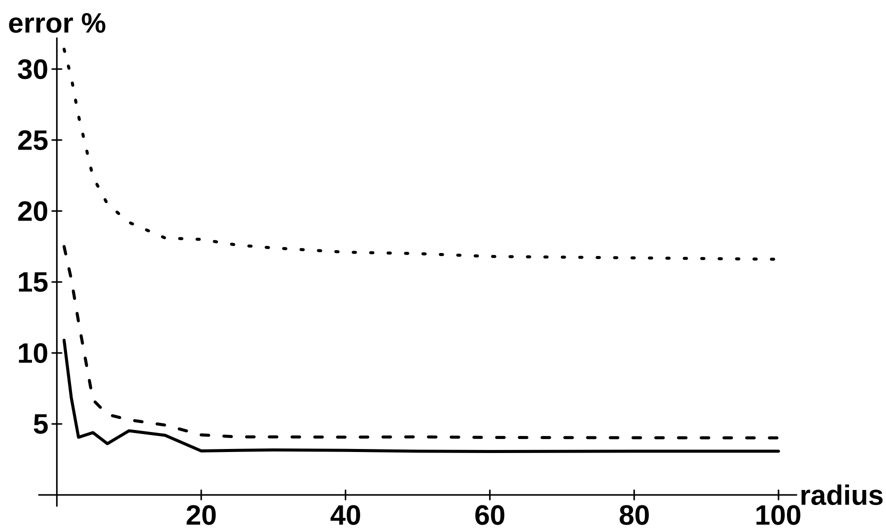

Figure 11: Perimeter estimation of circles in 3-D using link counting estimators. Average absolute errors (of 100 experiments) as a function of radius are compared. The dotted curve refers to the biased 26-directional link counting estimator. The dashed and solid curves respectively refer to the unbiased 26-directional and 6-directional link counting estimators, both developed in this paper.

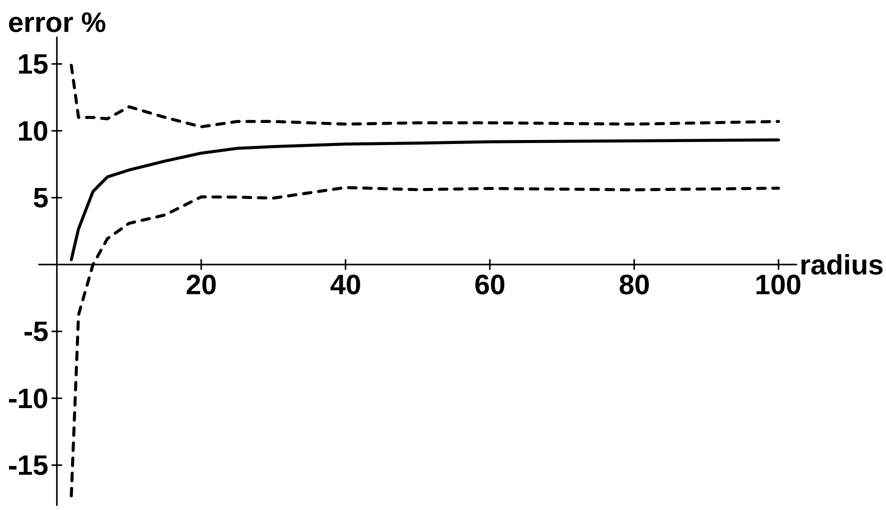

Figure 12: Perimeter estimation of circles in 3-D using the biased 26-directional link classification estimator. The solid curve shows the average error (of 100 experiments) as a function of radius. The dashed curves show the minimum and maximum errors encountered.

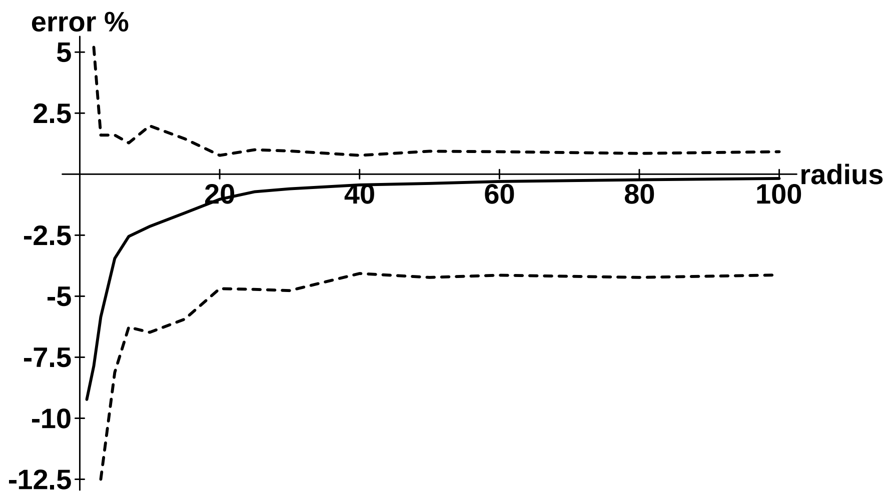

Figure 13: Perimeter estimation of circles in 3-D using the unbiased, minimum RMS error, 26-directional link classification estimator developed in this paper. The solid curve shows the average error (of 100 experiments) as a function of radius. The dashed curves show the minimum and maximum errors encountered.

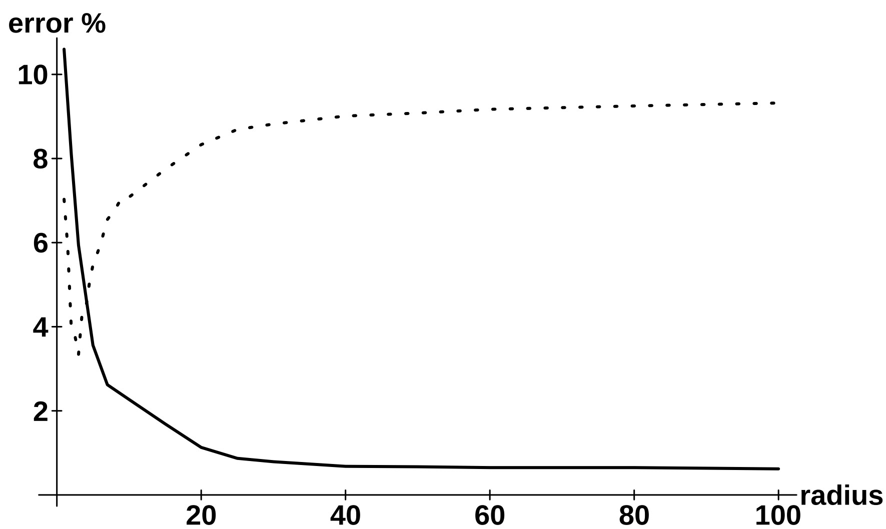

Figure 14: Perimeter estimation of circles in 3-D using link classification estimators. Average absolute errors (of 100 experiments) as a function of radius are compared. The dotted curve refers to the biased 26-directional link classification estimator. The solid curve refers to the unbiased, minimum RMS error, 26-directional link classification estimator developed in this paper.

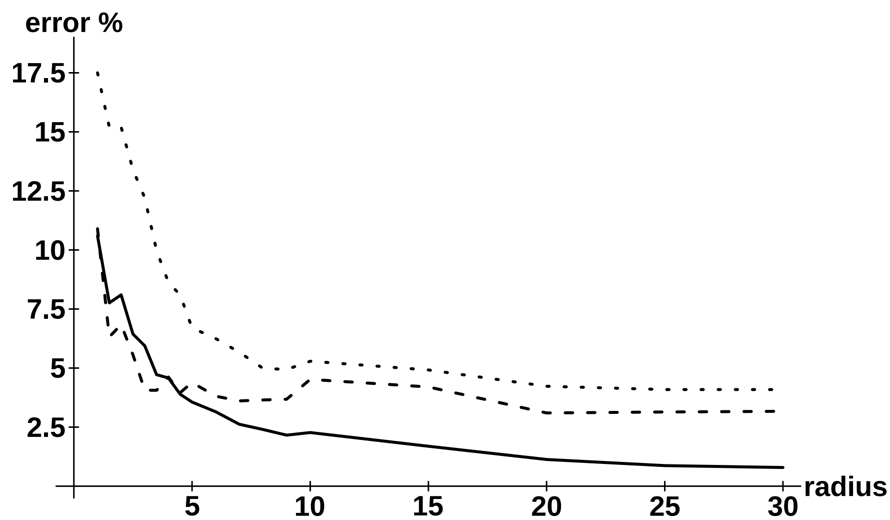

Figure 15: Perimeter estimation of circles in 3-D using the unbiased estimators developed in this paper. Average absolute errors (of 100 experiments) as a function of radius are compared. The dotted and dashed curves respectively refer to the unbiased 26-directional and 6-directional link counting estimators. The solid curve refers to the unbiased, minimum RMS error, 26-directional link classification estimator.