cycle vectors when what we want is a single approximate solution? One approach we are actively considering involves the use of the cycle vectors to create new "pseudodata," which is then fed back and the algorithm restarted.

To be precise, the feedback method works as follows: assuming convergence of BI-SMART to a limit cycle of $N$ vectors, $z^1, z^2, \cdots, z^N$, we replace each old data value $y_i$ with the new value $(Pz^n)_i$, where $i$ is in the subset $S^{n+1}$, and we set $S^0 = S^{N+1}$. The new data vector is then used in exactly the same way the old $y$ was, beginning with the same initial vector and performing BI-SMART until convergence to a second limit cycle. The infinite sequence of data vectors calculated in this way can be shown to converge to a vector $y^\infty$, for which the system $y^\infty = Px$ is (nonnegatively) consistent. In simulations we have noted two interesting phenomena: first, this convergence is remarkably quick, with $y^\infty$ often reached within two or three repetitions of the feedback; and second, the solution $x$ of $y^\infty = Px$ is a strictly positive vector in most cases. This suggests that the repeated feedback is regularizing the problem; that is, effecting a filtering of the original data. The exact criterion implicitly used here is not yet understood. This feedback has been tried in connection with BI-EMML with some success [22].

## Acknowledgment

The author wishes to thank Prof. Y. Censor of the University of Haifa for helpful discussions on these matters.

## References

[1] A. Rockmore and A. Macovski, "A maximum likelihood approach to emission image reconstruction from projections," *IEEE Trans. Nucl. Sci.*, vol. NS-23, pp. 1428–1432, 1976.

[2] L. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. MI-1, no. 2, pp. 113–122, Oct. 1982.

[3] K. Lange and R. Carson, "EM reconstruction for emission and transmission tomography," *J. Comput. Assist. Tomog.*, vol. 8, pp. 306–312, 1984.

[4] Y. Vardi, L. Shepp, and L. Kaufman, "A statistical model for positron emission tomography," *J. Amer. Statist. Assoc.*, vol. 80, pp. 8–37, 1985.

[5] K. Lange, M. Bahn, and R. Little, "A theoretical study of some maximum likelihood algorithms for emission and transmission tomography," *IEEE Trans. Med. Imag.*, vol. MI-6, no. 2, pp. 106–114, June 1987.

[6] C. Byrne, "Iterative image reconstruction algorithms based on cross-entropy minimization," *IEEE Trans. Image Processing*, vol. 2, no. 1, pp. 96–103, Jan. 1993.

[7] S. Holte, P. Schmidlin, A. Linden, G. Rosenqvist, and L. Eriksson, "Iterative image reconstruction for positron emission tomography: a study of convergence and quantitation problems," *IEEE Trans. Nucl. Sci.*, vol. 37, no. 2, pp. 629–635, Apr. 1990.

[8] G. Herman and L. Meyer, "Algebraic reconstruction techniques can be made computationally efficient," *IEEE Trans. Med. Imag.*, vol. 12, no. 3, pp. 600–609, Sept. 1993.

[9] Y. Censor and J. Segman, "On block-iterative entropy maximization," *J. Inform. Optimizat. Sci.*, vol. 8, no. 3, pp. 275–291, 1987.

[10] P. Schmidlin, "Iterative separation of sections in tomographic scintigrams," *Nucl. Med.*, vol. 15, no. 1, 1972, Schatten Verlag, Stuttgart.

[11] J. N. Darroch and D. Ratcliff, "Generalized iterative scaling for log-linear models," *Annals Math. Statist.*, vol. 43, pp. 1470–1480, 1972.

[12] A. Lent, "A convergent algorithm for maximum entropy image restoration with a medical x-ray application," in *Image Analysis and Evaluation*, R. Shaw, Ed. Washington, DC: SPSE, 1977, pp. 238–243.

[13] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Statist. Soc.: Series B*, vol. 39, pp. 1–38, 1977.

[14] I. Csiszar and G. Tusnady, "Information geometry and alternating minimization procedures," *Statist. Decisions*, suppl. issue no. 1, pp. 205–237, 1984.

[15] I. Csiszar, "A geometric interpretation of Darroch and Ratcliff's generalized iterative scaling," *Annals Statist.*, vol. 17, no. 3, pp. 1409–1413, 1989.

[16] R. Gordon, R. Bender, and G. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography," *J. Theoret. Biol.*, vol. 29, pp. 471–481, 1970.

[17] S. Zenios and Y. Censor, "Parallel computing with block-iterative image reconstruction algorithms," *Appl. Num. Math.*, vol. 7, pp. 399–415, 1991.

[18] R. Aharoni and Y. Censor, "Block-iterative projection methods for parallel computation of solutions to convex feasibility problems," *Linear Algebra Applic.*, vol. 120, pp. 165–175, 1989.

[19] H. Hudson, B. Hutton, and R. Larkin, "Accelerated EM reconstruction using ordered subsets," *J. Nucl. Med.*, vol. 33, p. 960, 1992.

[20] H. Hudson and R. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, no. 4, pp. 601–609, Dec. 1994.

[21] C. Byrne, "Erratum and addendum to 'Iterative image reconstruction algorithms based on cross-entropy minimization',," *IEEE Trans. Image Processing*, vol. 4, no. 2, 1995.

[22] W. Xuan, "Evaluating a feedback technique in image reconstruction from projections," preprint, Dept. of Statistics, MacQuarie University, 1995.

[23] C. Byrne, "Iterative reconstruction algorithms based on cross-entropy minimization," in *Image Models (And Their Speech Model Cousins)*, S. Levinson and L. Shepp, Eds. Institute for Mathematics and its Applications, vol. 80, Dec. 1995.

[24] ——, "Accelerating the multiplicative algebraic reconstruction technique (MART) and related block-iterative methods by equation rescaling," *IEEE Trans. Image Processing.*, to be published.

# On the Metric Properties of Discrete Space-Filling Curves

C. Gotsman and M. Lindenbaum

***Abstract*—A space-filling curve is a linear traversal of a discrete finite multidimensional space. In order for this traversal to be useful in many applications, the curve should preserve "locality."**

**We quantify "locality" and bound the locality of multidimensional space-filling curves. Classic Hilbert space-filling curves come close to achieving optimal locality.**

## I. Introduction

Denote $[N] = \{1, .., N\}$. A discrete $m$-dimensional space-filling curve of length $N^m$ is a bijective mapping $C : [N^m] \longrightarrow [N]^m$ such that $d(C(i).C(i+1)) = 1$ for all $i \in [N^m - 1]$, where $d()$ is the Euclidean metric. In other words, the curve $C$ of length $N^m$ traverses all $N^m$ points of the $m$-dimensional grid with side length $N$, making unit steps and turns only at right angles. For a historical account of classical space-filling curve constructions, see [8].

Space-filling curves are useful in applications where a traversal (scan) of a multidimensional grid is needed. Some algorithms perform

Manuscript received September 14, 1994; revised September 6, 1995. A preliminary version of this work was presented at the IEEE International Conference on Pattern Recognition, Jerusalem, 1994. The associate editor coordinating the review of this paper and approving it for publication was Prof. Charles A. Bouman.

The authors are with the Computer Science Department Technion—Israel Institute of Technology, Haifa, Israel.

Publisher Item Identifier S 1057-7149(96)03178-8.

local computations on neighborhoods, or exploit spatial correlation present in the data; therefore, the preservation of "locality" during the traversal is desirable. By "locality," we mean that the traversal reflects proximity between the points of $[N]^m$, namely, that points close in $[N]^m$ are also close in the traversal order, and vice versa. Sample applications are image halftoning (see [11] and references therein), data organization [4], data compression [5], and color quantization [9].

The little work on this topic to date has addressed only one direction of this question, namely, how to design space-filling curves such that points close in the multidimensional space are also close along the curve. In general, as we shall show later (Theorem 2), this is impossible—for every space-filling curve $C$, there will always be at least one pair of close points in $[N]^m$ that are very far apart along $C$. However, as these cases are rare, on the *average* , the situation will be much better. Perez *et al.* [7] quantify this using the average locality measure

$$L(C) = \sum_{i,j\in[N^m],i<j} \frac{|i-j|}{d(C(i),C(j))} \tag{1}$$

and describe a hierarchical construction for 2-D space-filling curves, which comes close to minimizing this measure.

Mitchison and Durbin [6] investigate similar measures of locality, taking into account only short (unit) Euclidean distances. This is because they regard the grid as an abstract lattice graph, ignoring its underlying geometry. They treat general 2-D mappings $C : [N^2] \longrightarrow [N]^2$ (not necessarily defining a curve). Their family of measures, parametrized by $q \in [0,1]$, is

$$\mathcal{L}_q(C) = \sum_{\{i,j\in[N^2]:i<j,d(C(i),C(j))=1\}} |i-j|^q. \tag{2}$$

Interestingly enough, for the case $q = 1$, which may be compared with the measure (1), the optimal mapping turns out to be quite different from that in [7] (it is not even a curve).

For the more interesting case $q < 1$, which de-emphasizes longer distances along the curve, Mitchison and Durbin prove the lower bound

$$\mathcal{L}_q(C) \geq \frac{1}{1+2q}N^{1+2q} + O(N^{2q})$$

and provide an explicit construction $C_N$ for any $N$ with good, albeit suboptimal, locality. They conjecture that the optimal mapping must define a curve with a "fractal" character.

Voorhies [10] defines a more heuristic measure of locality, related to computer graphics applications, and experimentally compares the measures obtained for a variety of space-filling curves. He concludes that the Hilbert curve [3] is superior to other curves in this respect.

In this correspondence, we mainly address the converse question, i.e., to which extent can two points, which are close in the traversal order along the curve, be far apart in the multidimensional Euclidean metric. To quantify this, we use the following measures:

$$L_1(C) = \max_{i,j\in[N^m],i<j} \frac{d(C(i),C(j))^m}{|i-j|} \tag{3}$$

$$L_2(C) = \min_{i,j\in[N^m],i<j} \frac{d(C(i),C(j))^m}{|i-j|}. \tag{4}$$

The use of the exponent $m$ in (3) and (4) is justified by the fact that the maximal distance between points of $[N]^m$ is $O(N)$ and between two points of $[N^m]$ is $O(N^m)$. This correspondence presents bounds on $L_1(C)$ and $L_2(C)$.

Certain curve designs may be used to produce a *family* of curves for increasing values of $N$: $\mathcal{C} = \{C_N : N = 1, 2, ...\}$ . In this case, it is interesting to investigate the limits (with a slight abuse of notation)

$$L_1(\mathcal{C}) = \lim_{N\to\infty} L_1(C_N)$$

$$L_2(\mathcal{C}) = \lim_{N\to\infty} L_2(C_N).$$

Essentially, we show that for any $m$-dimensional curve family $\mathcal{C}$, if these limits exist, then $L_1(\mathcal{C}) \geq 2^m - 1$, and $L_2(\mathcal{C}) = 0$.

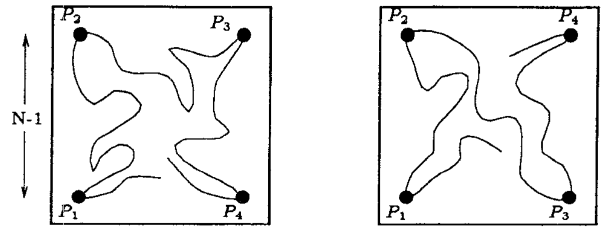



Fig. 1. Two possible distinct traversals through the four corner points of the 2-D case considered in the proof of Theorem 1. All other traversals are symmetric to these.

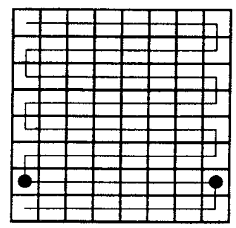


Fig. 2. Two-dimensional raster space-filling curve $R_N^2$ on $[N]^2$. For the two grid points at the two ends of any scan line, both the Euclidean distance and that along the curve, are $N-1$; therefore, $L_1(R_N^2) = N-1$.

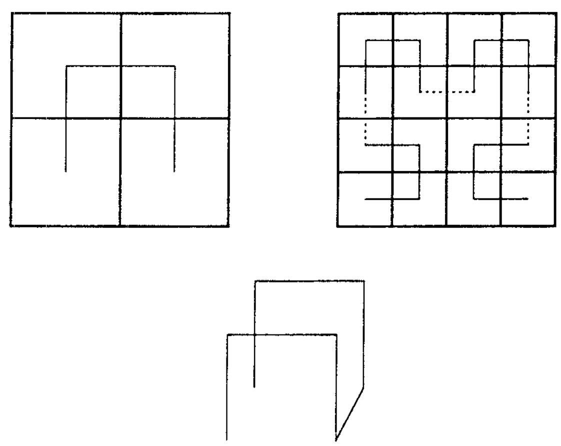
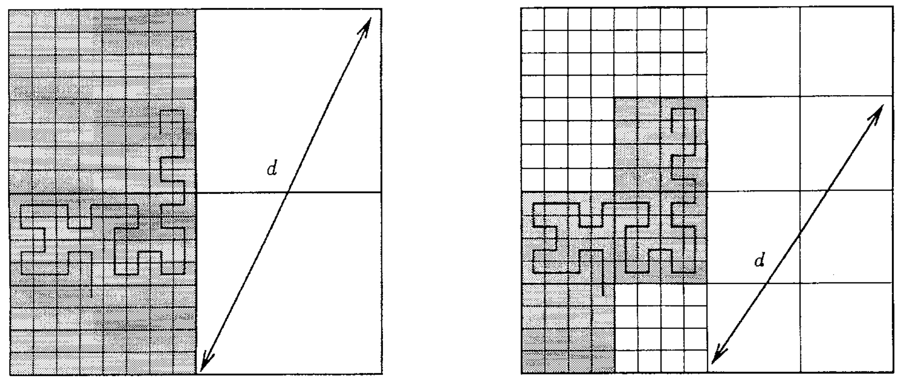


Fig. 3. Recursive construction of the Hilbert curve: (top left) "Seed" $H_1^2$; (top right) $H_2^2$ constructed from four (rotated) versions of $H_1^2$. $H_k^2$ is constructed recursively in an analogous fashion; (bottom) "seed" $H_1^3$. $H_k^3$ is constructed recursively from the seed curve.

## II. Locality of Space-Filling Curves

In this section, we provide a lower bound on $L_1$ and an upper bound on $L_2$ for multidimensional space-filling curves.




$m = 2, n = 40, r = 3, d^2 = 8^2 + 16^2 = 320$

(a)

$m = 2, n = 40, r = 3, d^2 = 8^2 + 12^2 = 208$

(b)

Fig. 4. Upper bounds on $L_1(H_4^2)$: (a) Quantities of Theorem 3. Any Hilbert subpath of length $16 < n \leq 64$ must lie within two adjacent quadrants of size $8 \times 8$; (b) Case 5 of Theorem 4. The subpath of (a), as any Hilbert subpath of length $32 < n \leq 48$, is actually contained in four adjacent quadrants of size $4 \times 4$, resulting in a tighter bound on $d$.

*Theorem 1:* If $C$ is a discrete $m$-dimensional $(m > 1)$ space-filling curve on $[N]^m$, then

$$L_1(C) > (2^m - 1)(1 - 1/N)^m.$$

*Proof:* Consider the $2^m$ corner points $\{1, N\}^m$ of the $[N]^m$ grid. Any space-filling curve must start at some arbitrary grid point, pass through these corner points in some order, and then end at another arbitrary grid point. Consider the increasing sequence of indices $\{P_i\}_{i=1}^{2^m}$ of these corner points along any such ordering. Fig. 1 shows the two possible distinct orderings of these four indices for the case $m = 2$. The Euclidean distance between any two consecutive points is $d(C(P_i), C(P_{i+1})) \geq N - 1$. Since $P_{2^m} - P_1 < N^m$, there exists an $1 \leq i < 2^m$ such that $P_{i+1} - P_i < N^m/(2^m - 1)$. For those two points, we have $\frac{d(C(P_i), C(P_{i+1}))^m}{|P_i - P_{i+1}|} > (2^m - 1)(1 - 1/N)^m$. □

*Remark:* For the 2-D case, the limit constant given by Theorem 1 is 3. By a computerized exhaustive search, we have improved this to 3.25, implying that the bound of Theorem 1 is not tight. This was achieved by considering all possible paths through a specific configuration of nine points in the plane, analogously to the four corner points in the theorem proof. □

Whether $|i - j|$ can be considered a good estimate for $d(C(i), C(j))^m$ depends on the existence of a positive constant lower bound on $L_2(C)$. The answer to this is negative, relying on the following discrete analog of the classic topological theorem ([2], ch. 5, Theorem 2.3) that no mapping $f : [0,1] \longrightarrow [0,1]^m$ $(m > 1)$ can be continuous and possess a continuous inverse.

*Theorem 2:* If $C$ is a discrete $m$-dimensional space-filling curve on $[N]^m$, then

$$L_2(C) = O(N^{1-m}).$$

*Proof:* Choose a segment $S$ of $C$ of length at least $\frac{1}{4}N^m$ and at most $\frac{3}{4}N^m$. By the isoperimetric inequality on the multidimensional grid [1], the *boundary* of $S$, namely, the set of grid points in $S$ that have an immediate neighbor in $C - S$, denoted $\partial S$, satisfies $|\partial S| = \Omega(N^{m-1})$.[1] Form a list of the points of $\partial S$ sorted by their position along $C$, and choose the point $p$ at the middle of this list. The distance along $C$ from $p$ to the nearest endpoint of $S$ must now be at least $\lfloor \frac{1}{2}|\partial S| \rfloor$. This distance bounds from below the distance along $C$ between $p$ and any of its neighbors in $C - S$. The Euclidean distance between such a pair is 1, and the distance along the curve is $\Omega(N^{m-1})$, implying $L_2(C) = O(N^{1-m})$. □

[1] $\Omega(f(N))$ denotes a quantity that is not less than $cf(N)$ for some constant $c$ and sufficiently large $N$.

## III. Hilbert Curves

The standard $m$-dimensional raster space-filling curve on $[N]^m$, which is denoted $R_N^m$, does not have good locality properties, namely, $L_1(R_N^m) = \Omega(N^{m-1})$ (see Fig. 2). On the other hand, the Hilbert curve [3] is an excellent example of a locality-preserving space-filling curve. The $m$-dimensional Hilbert curve of order $k$, which is denoted $H_k^m$, may be constructed recursively for any $N = 2^k$, as described in Fig. 3. The locality of the Hilbert curve is demonstrated by the following theorem, showing that it is close to optimal (compare with Theorem 1).

*Theorem 3:* If $H_k^m$ is a $m$-dimensional Hilbert curve on $[N]^m = [2^k]^m$, then

$$L_1(H_k^m) \leq (m+3)^{m/2} 2^m.$$

*Proof:* Consider any subpath of length $n$ along $H_k^m$. There exists an integer $r$ such that $(2^m)^{r-1} < n \leq (2^m)^r$. The fact that once $H_k^m$ enters a grid quadrant of order $r$, it does not leave it until it has traversed all $(2^r)^m$ grid points in the quadrant implies that the subpath must lie in the union of two adjacent quadrants containing $(2^r)^m$ grid points each (see Fig. 4(a)). If this were not true, the length of the subpath would be greater than $(2^r)^m$. The diameter $d$ of the set of grid points traversed by the subpath satisfies $d^2 \leq (m - 1 + 4)(2^r)^2$ (by the Pythagorian theorem); therefore

$$\frac{d^m}{n} \leq (m+3)^{m/2} 2^m. \qquad \square$$

The upper bound of Theorem 3 is far from tight in all cases. For the 2-D case, we can improve the constant 20 obtained from Theorem 3 almost to its optimal value.

*Theorem 4:* If $H_k^2$ is a 2-D Hilbert curve on $[N]^2 = [2^k]^2$, then

$$6(1 - O(2^{-k})) \leq L_1(H_k^2) \leq 6\frac{2}{3}.$$

*Proof:* Using the terminology of the proof of Theorem 3, a more detailed analysis of subpath containment in quadrants of size $4^{r-1}$ (instead of quadrants of size $4^r$) shows that one of the following six possibilities must hold:

1) $\frac{4}{16}4^r < n \leq \frac{5}{16}4^r$: $d^2 < \frac{5}{4}4^r$, and hence, $d^2/n \leq 5$.

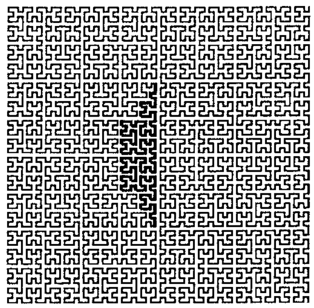


Fig. 5. "Worst" subpath of $H_6^2$ determining the value of $L_1(H_6^2)$. This path was found by exhaustive search, using a computer. An analogous structure is present in $H_k^2$ for any value of $k$ due to the recursive nature of the Hilbert curve.

2) $\frac{5}{16}4^r < n \leq \frac{6}{16}4^r$: $d^2 < \frac{29}{16}4^r$, and hence, $d^2/n \leq 5\frac{4}{5}$.
3) $\frac{6}{16}4^r < n \leq \frac{7}{16}4^r$: $d^2 < \frac{10}{4}4^r$, and hence, $d^2/n \leq 6\frac{2}{3}$.
4) $\frac{7}{16}4^r < n \leq \frac{8}{16}4^r$: $d^2 < \frac{10}{4}4^r$, and hence, $d^2/n \leq 5\frac{5}{7}$.
5) $\frac{8}{16}4^r < n \leq \frac{12}{16}4^r$: $d^2 < \frac{13}{4}4^r$, and hence, $d^2/n \leq 6\frac{1}{2}$.
6) $\frac{12}{16}4^r < n \leq 4^r$: $d^2 < 5 \cdot 4^r$, and hence, $d^2/n \leq 6\frac{2}{3}$.

For example, the subpath of Fig. 4(a) falls into category 5, as Fig. 4(b) illustrates. Taking the largest of the locality measures among these cases establishes the upper bound on $L_1(H_k^2)$.

For the lower bound, a subpath analogous to that illustrated in Fig. 5 (for $k = 6$, which was found by computer search) exists in $H_k^2$ for all $k > 1$ due to the recursive nature of the Hilbert curve. This subpath gives a locality measure of $6(1 - O(2^{-k}))$. Indeed, for $H_k^2$, it fills two adjacent quadrants of size $2^{k-3} \times 2^{k-3}$, two quadrants of size $2^{k-4} \times 2^{k-4}$ on either side of these two (aligned to a fixed direction), and so on until there are two quadrants of size 1. The Euclidean distance between the two endpoints is

$$d = 2\sum_{i=0}^{k-3} 2^i + 1 = 2^{k-1} - 1$$

and the distance along the curve is

$$n = 2\sum_{i=0}^{k-3} 4^i + 1 = \frac{2}{3}4^{k-2} + \frac{1}{3}.$$

Therefore, $d^2/n = 6(1 - O(2^{-k}))$. □

*Remark:* For the 3-D case, Theorem 3 yields $L_1(H_k^3) \leq 117.56$. By computer simulation, we have found that $L_1(H_k^3) \leq 23$. □

## IV. Discussion

There remains a considerable gap between the lower bound on $L_1$ for general space-filling curves (Theorem 1) and the upper bound on $L_1$ for the Hilbert curve family (Theorem 3). This leaves open the question of whether there exist families of space-filling curves with locality properties better than those of the Hilbert curves for all sizes.

It seems plausible that the Hilbert curves should also yield good results with respect to other measures of locality, such as that of Mitchison and Durbin [6]. These authors conjecture that the space-filling curve with optimal locality properties, measured by (2) with $q < 1$, must have a "fractal" character. Simulations performed by us show that, in agreement with this prediction, that in some cases, the (fractal) Hilbert curves indeed outperform the (nonfractal) curve constructed in [6].

In conclusion, we emphasize the practical implications of our results. Theorem 4 guarantees that if spatial correlation exists among the values of a discrete 2-D data array, a 1-D algorithm (such as that compressing a 1-D data stream) may scan the the array along a Hilbert curve, and the loss in data correlation along the scan will be bounded.

## Acknowledgment

Thanks to G. Agranov, A. Bruckstein and L. Kaplan for helpful discussions on this subject.

## References

[1] B. Bollobas and I. Leader, "Edge-isoperimetric inequalities in the grid," *Combinatorica*, vol. 11, no. 4, pp. 299–314, 1991.
[2] J. Dugundji, *Topology*. Boston, MA: Allyn and Bacon, 1966.
[3] D. Hilbert, "Uber die stetige abbildung einer linie auf ein flachenstuck." *Math. Ann.*, vol. 38, pp. 459–460, 1891.
[4] R. Laurini, "Graphical databases built on Peano space-filling curves," in *Proc. Eurographics '85*, pp. 327–338.
[5] A. Lempel and J. Ziv, "Compression of two dimensional data," *IEEE Trans. Inform. Theory*, vol. IT-32, no. 1, pp. 2–8, 1986.
[6] G. Mitchison and R. Durbin, "Optimal numberings of an $N \times N$ array," *SIAM J. Discr. Alg. Meth.*, vol. 7, no. 4, pp. 571–582, 1986.
[7] A. Perez, S. Kamata, and E. Kawaguchi, "Peano scanning of arbitrary size images," in *Proc. Int. Conf. Patt. Recogn.*, 1992, pp. 565–568.
[8] H. Sagan, *Space-Filling Curves*. New York: Springer-Verlag, 1994.
[9] R. J. Stevens, A. F. Lehar, and F. H. Preston, "Manipulation and presentation of multidimensional image data using the Peano scan," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-5, no. 5, pp. 520–527, 1983.
[10] D. Voorhies, "Space-filling curves and a measure of coherence," in *Graphics Gems II*, J. Arvo, Ed. New York: Academic, 1991, pp. 26–30.
[11] Y. Zhang and R. E. Webber, "Space diffusion: An improved parallel halftoning technique using space-filling curves," in *Proc. ACM Comput. Graphics Ann. Conf. Series*, 1993, pp. 305–132.