

SRC-TR-87-94

On the Number of Digital Straight Line  
Segments

by

Carlos A. Berenstein  
David Lavine

On the Number of Digital Straight Line Segments

Carlos A. Berenstein\*  
Department of Mathematics and  
Systems Research Center  
University of Maryland  
College Park, Maryland 20742

and

David Lavine\*\*  
LNK Corporation, Inc.  
6811 Kenilworth Avenue, Suite 306  
Riverdale, Maryland 20737

ABSTRACT

Let  $L_N$  be the number of digital line segments of length  $N$  that correspond to lines of the form  $y = ax + \beta$ ,  $0 \leq a, \beta < 1$ . In a previous paper [4], a closed form expression for the quantity  $L_N$  was obtained. We prove an asymptotic estimate for  $L_N$  that might prove useful for many applications. Namely,

$$L_N = \frac{N^3}{\pi^2} + O(N^2 \log N).$$

An application to image registration questions is given.

\* Partly supported by the National Science Foundation.

\*\* Partly supported by a NASA grant.

## 1. INTRODUCTION.

Image processing problems in high accuracy matching, edge detection, and measurement can be approached using probabilistic methods in digital geometry. Using this approach we considered the problem of subpixel accuracy in feature based image registration and presented our work in a series of NASA sponsored symposia in 1983-85 [1,2,3]. The mathematical formulation of these ideas appear in detail in [4]. Related techniques have been developed by Dorst and Smeulders [5].

The digital nature of these problems gives rise to the need to study the properties of digital line segments, i.e. the digitization of real line segments. The original work characterizing which collections of pixels are digital lines was done by Rosenfeld and others ([6],[7]). In [7] it was observed that the digital line segments of length  $N$  through the origin, that is the digital line segments corresponding to lines of equation  $y = \alpha x$ ,  $0 \leq \alpha < 1$ , are in a one-to-one correspondence with the Farey series of length  $N$ ,  $\mathcal{F}_N$  (see next section for unexplained terms). It follows from well-known results in number theory that if  $L_N^0$  denotes the number of line segments of length  $N$  through the origin then

$$L_N^0 = \#\mathcal{F}_N = \frac{3N^2}{\pi^2} + O(N \log N).$$

We also note that  $L_N^0$  can be computed explicitly in terms of number theoretic functions but the asymptotic estimate given is more useful in many problems. In many situations we cannot know that the digital line segment we obtain, e.g. in edge detection,

passes through the origin. In [5] a useful method of representing arbitrary digital line segments of length  $N$  was introduced, and the number  $L_N$  of these segments corresponding to lines of the form  $y = ax + \beta$ ,  $0 \leq a, \beta < 1$  appeared then as a quantity that needed to be computed for the applications in [1,2,3]. In [4] we succeeded in giving an explicit expression for  $L_N$  in terms of arithmetical functions, again this expression (as  $L_N^0$ ) is very hard to compute and the corresponding asymptotic expansion would have been useful. In [4] we only succeeded in proving that

$$\frac{3}{4} \leq L_N / (N^3 / \pi^2) \leq \frac{10}{9},$$

if we disregard error terms of the order of magnitude  $O((\log N)/N)$ . These bounds are not trivial because a careless reading of [5] might have indicated that  $L_N \approx 2N^3 / \pi^2$ . We prove here in Theorem 4 that

$$L_N = \frac{N^3}{\pi^2} + O(N^2 \log N).$$

A recent note [10] considers the related problem of bit reduction for storage of the main code of line drawings, and an estimate of the form  $L_N = O(N^3)$  is rederived. In a very interesting paper [4], McIlroy considers the question of efficiently finding the digital line segment associated to a real line segment, his geometric approach is akin to the method used in [4]. In [4] we considered also a sort of converse problem to McIlroy's, given a digital line segment find a real line segment that had the least offset error. This arose in trying to obtain subpixel information from satellite data. We also obtained in [4] some upper and lower bounds for the average offset  $E_N$  for a digital

line segment of length  $N$ . We give in Theorem 6 an exact asymptotic formula for  $E_N$ . This quantity  $E_N$  represents how much subpixel accuracy you can expect in estimating line position from binary (black and white) digital data, it turns out to be about  $0.92/N$  for segments of length  $N$ .

It is our impression that both the results and the techniques used here will turn out to be useful in what we see as an emerging new area of research with applications to image processing, *digital integral geometry*, the theory from [8] carried in the context of digital lines, etc.

We would like to thank the referees for their many suggestions that led to improvements upon the first version of this manuscript, for pointing out references [10] and [11] to us, and specially for their suggestion that a more representative title was necessary.

## 2. BACKGROUND.

This section describes the parametrization [5] of digital lines mentioned in the introduction. We also recall the results obtained in [4] describing the set of all digital lines in terms of these parameters. We will be concerned with lines with slope in the range  $[0,1)$  and crossing the  $y$ -axis in the interval  $[0,1)$ . (The family of all lines can be reduced to this one modulo translation and relabeling of the axes.)

To each such line we can associate a digital line by the following procedure. For each nonnegative integer  $a$ , if the line crosses the vertical line  $x = a$  at the point  $(a,b)$ , then

we mark the square pixel whose lower left hand corner is  $(a, [b])$ , where  $[b]$  denotes the integral part of  $b$ . The set of marked pixels obtained in this way is called the digital line associated to the original line. In practice, we are only interested in line segments, hence the integer  $a$  will be in the interval  $[0, N]$ .  $N$  will be called the length of the digital line segment. For this digital segment, one can assign a sequence  $\{c_j\}_1^N$  of zeros and ones as follows. Let  $b_0, b_1, \dots, b_N$  be the ordinates of the lower left hand corners of the pixels of that segment (Note that  $b_0 = 0$ ). Define

$$c_j = \begin{cases} 0 & \text{if } b_j = b_{j-1} \\ 1 & \text{otherwise} \end{cases} .$$

This sequence has  $N$  elements.

The period,  $q$ , of this sequence is defined to be the smallest integer such that there exists an infinite periodic extension of  $\{c_j\}_1^N, c_1, \dots, c_N, c_{N+1}, c_{N+2}, \dots$ , with period  $q$ . It is clear that  $1 \leq q \leq N$  and the case  $q = 1$  corresponds to a horizontal digital segment, i.e.  $c_j = 0$  for all  $j$ . Define  $p$  to be the number of ones in a period. If  $p$  is different from zero, then  $p$  and  $q$  are relatively prime.

The fourth parameter, called the shift  $s$ , can be defined by the property that  $0 \leq s \leq q - 1$  and that for every  $j, 1 \leq j \leq N$ , one has

$$(1) \quad c_j = [(j-s)(p/q)] - [(j-s-1)(p/q)].$$

If one reverses the above procedure for any quadruple of nonnegative integers  $(N, q, p, s)$  with  $1 \leq q \leq N$ ,  $p$  relatively prime to

$q$ ,  $0 \leq s \leq q - 1$ , the set of pixels whose lower left hand corner is  $(j, b_j)$ ,  $0 \leq j \leq N$  will form a digital line segment of length  $N$ . This quadruple  $(N, q, p, s)$  constitutes the parametrization of the family of digital line segments of length  $N$  proposed by Dorst and Smeulders [5]. To avoid trivial repetitions we only allow  $p = 0$  when  $q = 1$ . This case corresponds to the horizontal digital line segment.

The shift parameter  $s$  has the property that digital line segments through the origin (i.e. those arising from lines through the origin) are characterized by the value  $s = 0$ . In fact, as was already shown in [7], a digital line segment of length  $N$  through the origin always arises from the digitization of a line of equation  $y = p/q x$  with  $p$  and  $q$  relatively prime integers,  $1 \leq q \leq N$ ,  $0 \leq p < q$  (and  $1 \leq p < q$  if  $q \neq 1$ ). These parameters  $p$  and  $q$  are exactly the same as those appearing in  $(N, q, p, 0)$ . Furthermore, the correspondence between different digital line segments of length  $N$  through the origin and the above pairs  $p, q$  is a bijection. In number theory it is customary to call the set  $\mathcal{F}_N$  of rational numbers  $r$ ,  $0 \leq r < 1$ , with denominators less or equal to  $N$ , the Farey sequence of order  $N$  (cf. [9], to be precise in [9] the number 1 is included in  $\mathcal{F}_N$  but it is more convenient for us to modify the definition in this form). Note that  $\mathcal{F}_N$  coincides with the set of  $\{p/q: p \text{ and } q \text{ relatively prime, } 1 \leq q \leq N, 0 \leq p < q \text{ and } p \neq 0 \text{ if } q \neq 1\}$ . It is clear therefore that if we call  $L_N^0$  the number of digital line segments through the origin of length  $N$ , then  $L_N^0 = \#\mathcal{F}_N$ .

Following [9] one can give a close formula for  $L_N^0$ . Namely, let  $\varphi(n)$  = the number of integers  $m$ ,  $1 \leq m \leq n$ ,  $m \perp n$  (i.e.,  $m$  and  $n$  are relatively prime), e.g.  $\varphi(1) = 1$ ,  $\varphi(2) = 1$ ,  $\varphi(3) = 2$ ,  $\varphi(4) = 2$ , etc. Introduce an auxiliary function  $\Phi$  by

$$(2) \quad \Phi(x) := \sum_{1 \leq n \leq x} \varphi(n).$$

Theorem 330 from [9] states that

$$(3) \quad \Phi(x) = \frac{3x^2}{\pi^2} + O(x \log x),$$

where as usual we denote by  $O(f(x))$  a quantity bounded in absolute value by  $C|f(x)|$  for some constant  $C > 0$  and all sufficiently large values  $x$ . With this notation in mind we see that  $L_N^0$  has an explicit formula (Theorem 331, [9]):

$$(4) \quad L_N^0 = \Phi(N) = \varphi(1) + \dots + \varphi(N),$$

which is not very easy to compute for large values of  $N$ . For many applications the asymptotic expansion (3) suffices

$$(3') \quad L_N^0 = \frac{3N^2}{\pi^2} + O(N \log N).$$

Unfortunately, for arbitrary digital line segments the Dorst-Smeulders parametrization does not define a one-to-one correspondence. In [4] we gave a one-to-one correspondence between the family of all digital line segments and a subset of the quadruples  $(N, q, p, s)$ . This subset is determined by the single condition given in Proposition 1 below. In order to state this result, we must introduce an auxiliary integral parameter,  $\ell$ ,  $0 < \ell < q$ , given by the solution to the congruence equation,

$$(5) \quad \ell p \equiv -1 \pmod{q}.$$

Clearly the set of these  $\ell$  runs over the set of numbers relatively prime to  $q$  when  $p$  runs over the same set.

Proposition 1. (see [4]). The family of digital line segments is in a one-to-one correspondence with the set of quadruples  $(N, q, p, s)$  such that the quantity

$$(6) \quad \lfloor (N-s)/q \rfloor q + \lfloor (s+\ell)/q \rfloor q - \ell$$

is positive.

For a fixed  $q$ ,  $1 \leq q \leq N$ , we want to compute the number  $L(N, q)$  of digital lines of length  $N$  and period  $q$ . Clearly,  $L(N, 1) = 1$ , so we can consider  $q > 1$ . Proposition 1 reduced the problem of counting the number of lines to the question of finding out for each  $\ell$  ( $1 \leq \ell < q$ ,  $\ell$  relatively prime to  $q$ ), the number of values of  $s$  for which the expression (6) is positive. It is clear that if  $N - s \geq q$ , then (6) is positive. The only time we must be careful is when  $N - s < q$ . This can only arise if  $N \leq q + s - 1 \leq 2q - 1$ , that is,  $(N+1)/2 \leq q$ . Hence, if  $q < (N+2)/2$ ,  $s$  can take arbitrary values and it follows that for  $2 \leq q < (N+2)/2$

$$(7) \quad L(N, q) = q\phi(q)$$

This formula is clearly also valid for  $q = 1$  since  $\phi(1) = 1$ . In the remaining range of  $q$ , one has to be more careful but a relatively simple argument which can be found in [4] leads to an explicit formula for the number  $L(N, q)$  in terms of arithmetic functions:

$$(8) \quad L(N, q) = (N - q + 2) \varphi(q) + \sum_{\ell} \min(2q - N - 2, q - \ell - 1, \ell - 1, N - q),$$

where the sum takes place over all values  $\ell$ ,  $0 < \ell < q$ , with  $\ell$  and  $q$  relatively prime.

We summarize the above as:

Proposition 2. (see [4]) Let  $L_N$  be number of digital lines of length  $N$  with both slope and  $y$ -intercept between 0 and 1.

Then

$$(9) \quad L_N = \sum_{q=1}^N L(N, q),$$

where  $L(N, q)$  is give by (7) if  $0 < q < (N+2)/2$  and by (8) otherwise.

### 3. Asymptotic formula for the number of lines.

The exact formula given in Proposition 2 is difficult to evaluate for large  $N$  and so an asymptotic formula becomes desirable. The derivation of such a formula is based on the heuristic fact that the numbers relatively prime to a given number  $q$  appear to be uniformly distributed in the interval  $[1, q]$ . In fact this heuristic can be made precise by the following auxiliary proposition:

Proposition 3. For fixed  $n$ , let the function  $F(x)$  be defined as the cardinality of the set of positive integers,  $p$ , such that  $p \leq x$  and  $p \perp n$ . Then

$$(10) \quad F(x) = \frac{\varphi(n)}{n} x + O(\log \log n).$$

The proof of Proposition 3 appears in the Appendix.

As a consequence of the definition of  $F$  we have for any  $a, b, 0 \leq a \leq b \leq n,$

$$(11) \quad \sum_{\substack{\ell \perp n \\ a \leq \ell \leq b}} 1 = F(b) - F(a) = (b-a) \frac{\varphi(n)}{n} + O(\log \log n).$$

Computing sums as Stieltjes integrals we also obtain from Proposition 3:

$$(12) \quad \sum_{\substack{\ell \perp n \\ a \leq \ell \leq b}} \ell = \frac{b^2 - a^2}{2} \frac{\varphi(n)}{n} + (b-a) O(\log \log n).$$

In effect,

$$\begin{aligned} \sum_{\substack{\ell \perp n \\ a \leq \ell \leq b}} \ell &= \int_a^b x dF(x) = xF(x) \Big|_a^b - \int_a^b F(x) dx \\ &= x^2 \frac{\varphi(n)}{n} \Big|_a^b + (b-a) O(\log \log n) - \frac{\varphi(n)}{n} \int_a^b x dx \\ &= \frac{1}{2} x^2 \frac{\varphi(n)}{n} \Big|_a^b + (b-a) O(\log \log n) \\ &= \frac{b^2 - a^2}{2} \frac{\varphi(n)}{n} + (b-a) O(\log \log n). \end{aligned}$$

In the proof of Theorem 4 below we will need another estimate of the same type, this one follows from (3). Let us define

$$(13) \quad G(x) := \sum_{1 \leq q < x} q \varphi(q)$$

Then we have

$$\begin{aligned} G(x) &= \int_1^x t d\zeta(t) = t\zeta(t) \Big|_1^x - \int_1^x \zeta(t) dt \\ &= x\zeta(x) - \frac{x^3}{\pi^2} + O\left(\int_1^x t \log t dt\right) = \frac{2x^3}{\pi^2} + O(x^2 \log x). \end{aligned}$$

Since this estimate will be used repeatedly below, we summarize it here

$$(14) \quad G(x) = \frac{2x^3}{\pi^2} + O(x^2 \log x).$$

Finally, let us introduce a function  $M(q)$  for  $(N+2)/2 \leq q \leq N$ ,

$$(15) \quad M(q) := \sum_{\substack{\ell \perp q \\ 1 \leq \ell \leq q-1}} \min(2q-N-2, q-\ell-1, \ell-1, N-q).$$

Actually the sum only uses  $2 \leq \ell \leq q-2$ , since the other two values of  $\ell$  contribute only zero to it.

These formulas and a modicum of computations will allow us to prove the asymptotic formula for  $L_N$  conjectured in [4].

Theorem 4. The following asymptotic formula holds

$$(16) \quad L_N = \frac{N^3}{\pi^2} + O(N^2 \log N).$$

Proof of Theorem 4. It is clear from Proposition 2 that computing  $L_N$  involves three kinds of terms. The first one can be handled using (13) and (14):

$$(17) \quad \sum_{1 \leq q < (N+2)/2} L(N, q) = \sum_{1 \leq q < (N+2)/2} q\varphi(q) = \frac{N^3}{4\pi^2} + O(N^2 \log N).$$

The second kind arises out of the first term in formula (8) (use first (2) and (13), then (3) and (14)):

$$(18) \quad \begin{aligned} \sum_{(N+2)/2 \leq q \leq N} (N+2-q)\varphi(q) &= (N+2) \left[ \Phi(N) - \Phi\left(\frac{N}{2}\right) \right] \\ &- \left[ G(N+1) - G\left(\frac{N+2}{2}\right) \right] = (N+2) \left[ \frac{3}{\pi^3} \left( N^2 - \left(\frac{N}{2}\right)^2 \right) + O(N \log N) \right] \\ &- \left[ \frac{2}{\pi^2} \left[ (N+1)^3 - \left(\frac{N+2}{2}\right)^2 \right] + O(N^2 \log N) \right] = \frac{N^3}{2\pi^2} + O(N^2 \log N). \end{aligned}$$

Finally we are left with the hardest term which requires repeated use of the estimates (11) and (12). This is the sum  $\sum_{(N+2)/2 \leq q \leq N} M(q)$ , where we have used the auxiliary function  $M$  defined by (15).

Note that if we graph the minimum as a function of  $\ell$ , we obtain a trapezoid, computing its "area" and correcting the "length of the base" to be  $\varphi(q)$  was the heuristic reasoning that led us to the statement of Theorem 4. To make this precise, we divide the range of  $q$  into two parts. Observe that  $2q-N-2 \leq N-q$  when  $q \leq \frac{2N+2}{3}$ . If we use first  $q$  in the range  $\frac{N+2}{2} \leq q \leq \frac{2N+2}{3}$  then

$$(19) \quad \text{minimum} = \begin{cases} \ell-1 & \text{if } 1 \leq \ell \leq 2q-N-1 \\ 2q-N-2 & \text{if } 2q-N \leq \ell \leq N-q \\ q-\ell-1 & \text{if } N-1+q \leq \ell \leq q-1 \end{cases}$$

Therefore, for a fixed  $q$ ,  $(N+2)/2 \leq q \leq (2N+2)/3$ ,  $M(q)$  can be estimated using (19), (11) and (12) by

$$\begin{aligned} M(q) &= \frac{(2q-N-1)^2-1}{2} \frac{\varphi(q)}{q} + (2q-N-2) O(\log \log q) \\ &\quad - (2q-N-2) \frac{\varphi(q)}{q} + O(\log \log q) \\ &\quad + (2q-N-2) \left[ (2N-3q) \frac{\varphi(q)}{q} + O(\log \log q) \right] \\ &\quad + (q-1) \left[ (2q-N-2) \frac{\varphi(q)}{q} + O(\log \log q) \right] \\ &\quad - \left[ \frac{(q-1)^2 - (N-q+1)^2}{2} \frac{\varphi(q)}{q} + (2q-N-2) O(\log \log q) \right] \\ &= (2q-N)(N-q) \frac{\varphi(q)}{q} + (7q-3N) O(\log \log q) \\ &= (2q-N)(N-q) \frac{\varphi(q)}{q} + O(q \log \log q). \end{aligned}$$

Where we have used  $N \leq 2q$  to simplify the last expression.

Similarly, if  $\frac{2N}{3} \leq q \leq N$ , then  $N-q \leq 2q-N-2$ , hence

$$(20) \quad \text{minimum} = \begin{cases} \ell-1 & 1 \leq \ell \leq N-q+1 \\ N-q & \text{if } N-q \leq \ell \leq 2q-N-2 \\ q-\ell-1 & 2q-N-1 \leq \ell \leq q-1 \end{cases}$$

The same computation leads to

$$(21) \quad M(q) = (2q-N)(N-q) \frac{\varphi(q)}{q} + O(q \log \log q),$$

which therefore holds in the whole range  $(N+2)/2 \leq q \leq N$ .

Hence, the last term we have to estimate to finish the proof of Theorem 4 is the following.

$$(22) \quad \begin{aligned} & \sum_{(N+2)/2 \leq q \leq N} M(q) \\ &= \sum_{(N+2)/2 \leq q \leq N} (2q-N)(N-q) \frac{\varphi(q)}{q} + O\left( \sum_{(N+2)/2 \leq q \leq N} q \log \log q \right) \\ &= 3N \sum \varphi(q) - 2 \sum q\varphi(q) - N^2 \sum \frac{\varphi(q)}{q} + O\left( \sum q \log \log q \right) \end{aligned}$$

where we have omitted the range of summation for simplicity. The first two terms can be estimated using (3) and (14). The remainder term can be estimated by comparison with the integral

$$\int^x t \log \log t \, dt = \frac{x^2}{2} \log \log x - \int^x \frac{t^2}{2} \frac{1}{\log t} \frac{1}{t} \, dt = O(x^2 \log \log x)$$

and hence it is smaller than  $O(N^2 \log N)$ . The only term in (22) we have not yet estimated is the one involving  $\frac{\varphi(q)}{q}$ . If we do it by integration by parts using the function  $\psi$  as we have done

before, we will end up with the correct leading term but a slightly bigger error term in Theorem 4, namely  $O(N^2 \log^2 N)$ . Since we will need a better estimate in Theorem 6, we state the following lemma which will be proved in the Appendix.

Lemma 5. Let  $H(x) = \sum_{1 \leq q \leq x} \frac{\varphi(q)}{q}$ , then

$$(23) \quad H(x) = \frac{6x}{\pi^2} + O(\log x)$$

Going back to the estimate of (22), we have now (using (3), (14) and (23)):

$$(24) \quad \sum_{(N+2)/2 \leq q \leq N} M(q) = 3N \left[ \bar{\Phi}(N) - \bar{\Phi}\left(\frac{N}{2}\right) \right] - 2 \left[ G(N+1) - G\left(\frac{N}{2}+1\right) \right] \\ - N^2 \left[ H(N) - H\left(\frac{N}{2}\right) \right] + O(N^2 \log N) = \frac{1}{4} \frac{N^3}{\pi^2} + O(N^2 \log N).$$

Collecting together (17), (18) and (24), we obtain

$$L_N = \frac{N^3}{\pi^2} + O(N^2 \log N)$$

which was the desired estimate. □

We have computed  $L_N$  using the exact formula (9) for  $N \leq 200$  and compared it with the leading term  $L'_N = \frac{N^3}{\pi^2}$  in (16). We found that  $L_N$  is systematically bigger than  $L'_N$ . The percent error steadily decreases from 5.5% for  $N = 100$  to 2.7% for  $N = 200$ . In fact, in this range of  $N$  we have

$$L_N = L'_N + c_N \frac{1}{\pi^2} N^2 \log N,$$

with  $c_N$  decreasing from 1.2 for  $N = 100$  to 1.08 for  $N = 200$ . It would seem natural to conjecture

$$L_N = \frac{N^3}{\pi^2} + \frac{N^2}{\pi^2} \log N + O(N \log N)$$

where the implied constant is of the order of magnitude of 0.1.

We do not know how to prove such a refined estimate.

#### 4. An application to subpixel registration accuracy.

As pointed out in the introduction, the origin of our research in this area was the problem of given a digital line segment, how to choose a line with that digitization and optimal in some sense. Let us denote  $\sigma$  the quadruple  $(N, q, p, s)$ . Denote by  $\lambda$  the affine function,  $\lambda(x) = \alpha x + \beta$ ,  $0 \leq \alpha, \beta < 1$ . We will write  $\lambda \in \sigma$  to indicate the digitization of the line segment of equation  $y = \lambda(x)$ ,  $0 \leq x \leq N$ , is exactly  $\sigma$ . We were interested in [1,2,3] in finding the line  $\lambda^* \in \sigma$  which minimized the (vertical) offset  $\varepsilon(\sigma)$  for a given  $\sigma$ :

$$(25) \quad \varepsilon(\sigma) := \min_{\lambda^* \in \sigma} \max_{\lambda \in \sigma} \max_{0 \leq x \leq N} |\lambda(x) - \lambda^*(x)|$$

In [4] we show that  $\lambda^*(x) = p/q x + \beta^*$ , where  $\beta^*$  can be found explicitly in terms of  $\sigma$  and

$$(26) \quad \varepsilon(\sigma) = \frac{1}{2q}.$$

The average offset  $E_N$  over all digital line segments of length  $N$  is then given by

$$(27) \quad E_N = \frac{1}{2} \sum_1^N \frac{L(N, q)}{q} / L_N$$

In [4] we obtained the (asymptotic) estimates

$$\frac{29}{30} \frac{1}{N} \leq E_N \leq \frac{59}{54} \frac{1}{N}$$

up to error terms of order  $O((\log N)/N^2)$ . Using Theorem 4 and the method of proof that led to it, we can obtain an exact asymptotic estimate.

Theorem 6. The average offset (27) is given by

$$(25) \quad E_N = 3(1-\log 2)\frac{1}{N} + O\left[\frac{\log N}{N^2}\right].$$

Proof. We have to estimate the numerator in the definition of  $E_N$ , i.e.,  $\sum_{q=1}^N (L(N,q)/q)$ . As we know from the proof of Theorem 4,

the hardest term in this sum has the form  $M(q)/q$ ,  $(N+2)/2 \leq q \leq N$ . From (21) we have

$$(26) \quad \frac{M(q)}{q} = -2\varphi(q) + 3N\frac{\varphi(q)}{q} - N^2\frac{\varphi(q)}{q^2} + O(\log\log q).$$

We have already worked with every term in (26) except for  $\varphi(q)/q^2$ . Let us handle this one first. We have, using Lemma 5, (recall  $H(1/2) = 0$ ):

$$\begin{aligned} R(x) &= \sum_{1 \leq q \leq x} \frac{\varphi(q)}{q^2} = \int_{1/2}^x \frac{1}{t} dH(t) = \frac{1}{t} H(t) \Big|_{1/2}^x + \int_{1/2}^x \frac{H(t)}{t^2} dt \\ &= \frac{6}{\pi^2} + \frac{6}{\pi^2} \int_{1/2}^x \frac{dt}{t} + O\left[\frac{\log x}{x}\right] \\ &= \frac{6}{\pi^2} \log x + \frac{6}{\pi^2}(1 + \log 2) + O\left[\frac{\log x}{x}\right]. \end{aligned}$$

Therefore

$$(27) \quad \sum_{(N+2)/2 \leq q \leq N} \frac{\varphi(q)}{q^2} = R(N) - R(N/2) = \frac{6}{\pi^2} \log 2 + O\left[\frac{\log N}{N}\right].$$

The same computations we did in Theorem 4 now lead to

$$(28) \quad \sum_{q=1}^N \frac{L(N,q)}{q} = \frac{6}{\pi^2} (1 - \log 2) N^2 + O(N \log N).$$

Finally,

$$E_N = \frac{1}{2} \sum_1^N \left[ L(N,q)/q \right] / L_N = \frac{3(1 - \log 2)}{N} + O\left(\frac{\log N}{N^2}\right). \quad \square$$

Remarks. 1. The estimate from [4] was

$$0.72/N \lesssim E_N \lesssim 1.09/N$$

while Theorem 6 says that

$$E_N \approx 0.92/N.$$

In other words, in a digital line segment of length  $N$ , on the average one cannot do better than about  $1/N$  of a pixel in the offset. Note that this depends on the data being binary, i.e., we are painting a pixel black if the line goes through it and leaving it white otherwise.

2. For the sake of completeness, we give here the formula from [4] for the choice of a minimizer  $\lambda^*$  of (25). As we said above, we have  $\lambda^*(x) = p/q x + \beta^*$ . To compute  $\beta^*$  in terms of  $\sigma$  let  $F(s) := s - [s/q]q$ , then

$$\beta^* := [F(s) p/q] - (F(s) p/q) + (1/2q).$$

## 5. Appendix.

We prove here the crucial Proposition 3 and also Lemma 5. The proof we give of Proposition 3 is elementary (i.e. does not use anything that cannot be found in an elementary textbook like [9]) and introduces some of the things needed in the proof of Lemma 5.

The Moebius function  $\mu$  is defined by

$$\mu(d) = \begin{cases} (-1)^r & \text{where } r \text{ is the number of distinct primes} \\ & \text{dividing } d \text{ if there exists no prime whose} \\ & \text{square divides } d \\ 0 & \text{otherwise.} \end{cases}$$

Hence  $|\mu(d)| = 1$  or  $0$  according to whether  $d$  is square-free or not.

Lemma 7. The following two estimates hold:

$$(29) \quad \sum_{d|n} \frac{|\mu(d)|}{d} = O(\log \log n)$$

$$(30) \quad \sum_{d|n} \frac{1}{d} = O(\log \log n)$$

Proof. The first sum is over the divisors of  $n$  that are square-free. It coincides with  $\prod_{p|n} (1 + \frac{1}{p})$  as it can be seen by computing this product and using the unique factorization of integers into products of primes. (The product runs over the prime divisors of  $n$ ). This is clearly as large as possible if  $n$  itself is the product of the first  $r$  primes,  $n = p_1 \dots p_r$ . We now estimate  $p_r$ . Theorem 414 from [9] says that for some constant  $C > 0$ ,  $\sum_{p \leq x} \log p \geq Cx$ , if we take  $x = p_r$  in this inequality we obtain

$$\log n = \sum_{i=1}^r \log p_i \geq Cp_r.$$

For  $x \geq 0$  we have  $\log(1 + x) \leq x$ , therefore

$$\log \sum_{d|n} \frac{|\mu(d)|}{d} = \log \prod_{p|n} (1 + \frac{1}{p}) \leq \sum_{p \leq p_r} \frac{1}{p} = \log \log p_r + B + o(1),$$

for some absolute constant  $B > 0$ . The last identity is Theorem 427 in [9]. Using the previous inequality this leads to

$$\prod_{p|n} \left(1 + \frac{1}{p}\right) = O(\log \log n),$$

which proves the estimate (29).

To prove the second estimate one needs to show that

$$\left(\sum_{d|n} \frac{1}{d}\right) / \left(\sum_{d|n} \frac{|\mu(d)|}{d}\right) \leq c.$$

The same reasoning that leads to computation of  $\sum_{d|n} \frac{|\mu(d)|}{d}$  as

$$\prod_{p|n} \left(1 + \frac{1}{p}\right) \text{ leads to the inequality } \sum_{d|n} \frac{1}{d} \leq \prod_{p|n} \left(1 + \frac{1}{p} + \frac{1}{p^2} + \dots\right).$$

Therefore

$$\begin{aligned} \left(\sum_{d|n} \frac{1}{d}\right) / \left(\sum_{d|n} \frac{|\mu(d)|}{d}\right) &\leq \prod_{p|n} \frac{\left(1 + \frac{1}{p} + \frac{1}{p^2} + \dots\right)}{\left(1 + \frac{1}{p}\right)} \leq \prod_p \frac{\left(1 + \frac{1}{p} + \frac{1}{p^2} + \dots\right)}{\left(1 + \frac{1}{p}\right)} \\ &= \prod_p \frac{\left(1 - \frac{1}{p}\right)^{-1}}{\left(1 + \frac{1}{p}\right)} = \prod_p \left(1 - \frac{1}{p^2}\right)^{-1} = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} . \quad \square \end{aligned}$$

### Proof of Proposition 3:

First we note that Theorem 62 of [9] gives

$$\frac{\varphi(n)}{n} = \prod_{p|n} \left(1 - \frac{1}{p}\right) = 1 - \sum_{p|n} \frac{1}{p} + \sum_{\substack{p|n \\ q|n \\ p \neq q}} \frac{1}{pq} - \dots,$$

where, in these formulas,  $p, q, r, \dots$  represent primes.

Let

$$\begin{aligned}
 F(x) &= \sum_{\substack{\ell \perp n \\ 1 \leq \ell \leq x}} 1 = [x] - \sum_{p|n} \left\lfloor \frac{x}{p} \right\rfloor + \sum_{\substack{p|n \\ q|n \\ p \neq q}} \left\lfloor \frac{x}{pq} \right\rfloor - \sum \left\lfloor \frac{x}{pqr} \right\rfloor + \dots \\
 &= x \left[ 1 - \sum_{p|n} \frac{1}{p} + \sum_{\substack{p|n \\ q|n}} \frac{1}{pq} \dots \right] + \text{error} = x \frac{\varphi(n)}{n} + \text{error}.
 \end{aligned}$$

The error term arises out of replacing  $\left\lfloor \frac{x}{p} \right\rfloor$  by  $\frac{x}{p}$ , and so on.

Using Lemma 7 we can now estimate the error term:

$$\text{Error} \leq \sum_{p|n} \frac{1}{p} + \sum_{\substack{p|n \\ q|n \\ p \neq q}} \frac{1}{pq} + \dots = \sum_{d|n} \frac{|\mu(d)|}{d} = O(\log \log n)$$

Therefore we obtain the estimate (10)

$$F(x) = \frac{\varphi(n)}{n} x + O(\log \log n) \quad \square$$

Proof of Lemma 5. We first use formula 16.3.1 from [9]:

$$\frac{\varphi(n)}{n} = \sum_{d|n} \frac{\mu(d)}{d},$$

therefore

$$H(x) = \sum_{1 \leq n \leq x} \frac{\varphi(n)}{n} = \sum_{1 \leq n \leq x} \sum_{d|n} \frac{\mu(d)}{d} = \sum_{d, d'} \frac{\mu(d)}{d},$$

where the last sum is over every pair of positive integers  $d, d'$  such that  $dd' \leq x$ , as one sees by writing  $n = dd'$  in the intermediate expression. Hence we have

$$\begin{aligned}
H(x) &= \sum_{1 \leq d \leq x} \frac{\mu(d)}{d} \sum_{1 \leq d' \leq x/d} 1 = \sum_{1 \leq d \leq x} \frac{\mu(d)}{d} \left[ \frac{x}{d} \right] \\
&= x \sum_{1 \leq d \leq x} \frac{\mu(d)}{d^2} - \sum_{1 \leq d \leq x} \frac{\mu(d)}{d} \left\{ \frac{x}{d} \right\}
\end{aligned}$$

where  $\{a\}$  stands for the decimal part of number  $a$ . We need one more transformation of this expression

$$(31) \quad H(x) = x \sum_{d=1}^{\infty} \frac{\mu(d)}{d^2} - \left[ x \sum_{x < d < \infty} \frac{\mu(d)}{d^2} + \sum_{1 \leq d \leq x} \frac{\mu(d)}{d} \left\{ \frac{x}{d} \right\} \right].$$

Now, one knows that by Theorem 287 from [9]

$$(32) \quad \sum_{d=1}^{\infty} \frac{\mu(d)}{d^2} = \left[ \sum_{n=1}^{\infty} \frac{1}{n^2} \right]^{-1} = \frac{6}{\pi^2}.$$

It is also clear that

$$(33) \quad \left| \sum_{x < d} \frac{\mu(d)}{d^2} \right| \leq \sum_{x < d} \frac{1}{d^2} \leq \frac{1}{x} + \frac{1}{x^2}.$$

We need only to estimate the last term of (31)

$$\left| \sum_{1 \leq d \leq x} \frac{\mu(d)}{d} \left\{ \frac{x}{d} \right\} \right| \leq \sum_{1 \leq d \leq x} \frac{|\mu(d)|}{d}.$$

Recall that we pointed out above that  $|\mu(d)| = 1$  or  $0$  according to whether  $d$  is square free or not. Let  $Q(y) =$  number of integers less or equal to  $y$  which are square free, that is,

$$Q(y) = \sum_{1 \leq d \leq y} |\mu(d)|$$

Theorem 333 from [9] gives us an estimate of  $Q$ , namely

$$Q(y) = \frac{6y}{\pi^2} + o(y^{1/2}).$$

We can now proceed as we have done several times before, writing sums as Stieltjes integrals

$$\begin{aligned}
 \sum_{1 \leq d \leq x} \frac{|\mu(d)|}{d} &= \int_1^x \frac{1}{y} dQ(y) = \frac{1}{y} Q(y) \Big|_1^x + \int_1^x \frac{Q(y)}{y^2} dy = \\
 &= \frac{6}{\pi^2} + \frac{6}{\pi^2} \int_1^x \frac{dy}{y} + O(x^{-1/2}) \\
 &= \frac{6}{\pi^2} (1 + \log x) + O(x^{-1/2}).
 \end{aligned}$$

Hence,

$$(34) \quad \left| \sum_{1 \leq d \leq x} \frac{\mu(d)}{d} \left\{ \frac{x}{d} \right\} \right| = O(\log x).$$

The three estimates (32), (33), (34) when used in (31) yield

$$H(x) = \frac{6}{\pi^2} x + O(\log x)$$

which is what we needed to prove. □

## 6. Conclusion:

We have shown that the number of digital line segments of length  $N$  that lie in the first octant can be estimated to be  $\frac{N^3}{\pi^2}$ . An application to subpixel accuracy was also given. We expect these theorems and methods will prove to be useful in other problems of digital integral geometry.

7. References:

- [1] D. Lavine, L. N. Kanal, C. A. Berenstein, E. Slud, C. Herman, "Analysis of Subpixel Registration Accuracy," Proceedings of the NASA Symposium on Mathematical Pattern Recognition and Analysis, Houston, TX, June 1983.
- [2] C. A. Berenstein, L. Kanal, D. Lavine, E. Olson, and E. Slud, "Analysis of Subpixel Registration," Presented at the NASA/Mathematical Pattern Recognition and Image Analysis Symposium, Houston, TX, June 1984.
- [3] D. Lavine, E. Olson, B. Lambird, C. Berenstein, and L. Kanal, "Analysis of Subpixel Accuracy," presented at the NASA/Mathematical Pattern Recognition and Image Analysis Symposium, Houston, TX, June 1985.
- [4] C. Berenstein, L. N. Kanal, D. Lavine, E. Olson, "A Geometric Approach to Subpixel Registration Accuracy," to appear in Computer Vision, Graphics, and Image Processing, 1987.
- [5] L. Dorst and A. W. M. Smeulders, "Discrete Representation of Straight Lines," IEEE Transactions of Pattern Analysis and Machine Intelligence 6 (1984), 450-462.
- [6] E. C. Kim and A. Rosenfeld, "Digital Straight Lines and Convexity of Digital Regions," IEEE Transactions on Pattern Analysis and Machine Intelligence 4 (1982), 149-153.
- [7] J. Rothstein and C. Weiman, "Parallel and Sequential Specification of a Context Sensitive Language for Straight Lines of Grids," Computer Graphics and Image Processing 5 (1976), 106-124.
- [8] L. A. Santaló, Integral Geometry and Geometrical Probability, Addison-Wesley, 1976.
- [9] G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford Univ. Press, 1986.
- [10] P. A. Sundar Raj and J. Koplowitz, "On Bit Reduction of Chain Coded Line Drawings," Pattern Recognition Letters 4 (1986), 99-102.
- [11] M. D. McIlroy, "A Note On Discrete Representation of Lines," AT&T Technical J. 64 (1984), 481-490.