

Finding Line Segments in the Ulam Square with the Hough Transform

Leszek J Chmielewski and Arkadiusz Orłowski

Faculty of Applied Informatics and Mathematics (WZIM),
Warsaw University of Life Sciences (SGGW), Poland
ul. Nowoursynowska 159, 02-775 Warsaw, Poland
{leszek.chmielewski,arkadiusz.orlowski}@sggw.pl
<http://www.wzim.sggw.pl>

Abstract. The regularities present in the Ulam spiral provided an incentive for interesting observations in the number theory. Therefore, we have made the Ulam square an object of analysis from the image processing perspective. A version of the Hough transform designed specially for detecting sequences of pixels forming segments of straight lines with the slope defined by an irreducible fraction was used to find line segments in the Ulam spiral. Angles which described the slopes of the segments had tangents p/q expressed by integers p from 0 to 10 and q from -10 to 10 (0 excluded). Due to storage limitations the squares with the side of length up to 5001 points which correspond to the largest prime 25 009 991 were analyzed at present. In such a square the longest segment has 16 primes and its tangent is 3 (3 up and 1 to the right). Segments of length 14 and 15 were absent. The number of shorter segments varied strongly, from one for a 13-point segment to tens of thousands for shorter ones.

Keywords: Ulam spiral, line segments, long, contiguous, Hough transform, image processing

1 Introduction and motivation

The question of finding an efficient formula for generating prime number is still a challenge. The well known Euler formula $p = n^2 - n + 41$ generates primes for $n = 1, 2, \dots, 40$ which form a continuous sequence, and for numerous other values of n (cf. [1], last lines of the paper). When we look at the images of the Ulam spiral [2] with the primes generated by the Euler formula, one of the phenomena which strikes us is that some of the points which represent the primes compose themselves into regular sequences, like for example those for $n = 20, 21, \dots, 40$. This is shown in Fig. 1.

In the following the prime and the point corresponding to this prime in the Ulam square will be treated as equivalent. The points in the square will also be called pixels, due to that the square will be treated as an image. The sequences of points located so that the ratio of increments of the two coordinates between consecutive points is constant, for example, 1:1 or 3:1, like in the lines marked

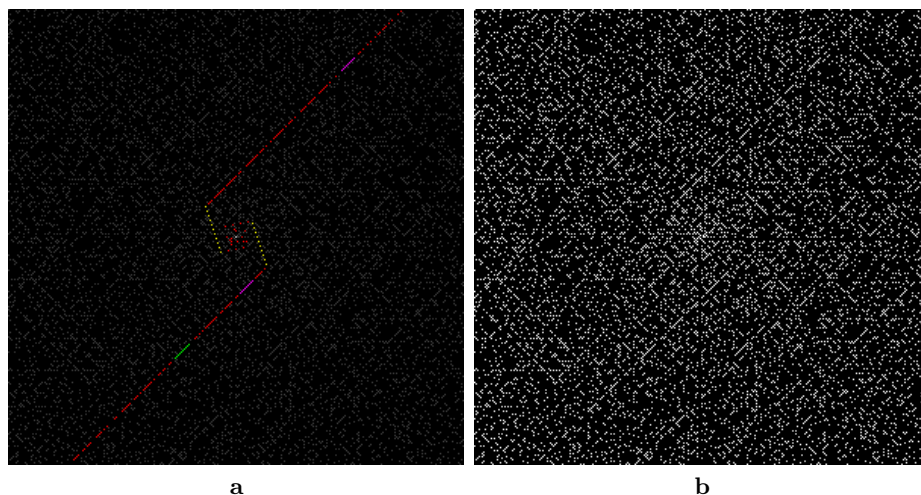


Fig. 1. (a) Central part of the Ulam square (301×301 points) with the points generated by the Euler formula. Yellow points (■): primes generated by $n = 20, 21, \dots, 40$; red, green and magenta points (■ ■ ■): remaining primes generated by the formula; grey points (■): remaining primes (grey level 63/255); white point (■): number 1, black (■): other non-primes. The left yellow line segment contains 11 points, the right one has 10 points. There is no other 11-point segment in this image but there is another 10-point one – marked with green (■), and two 9-point ones – marked with magenta (■). Shorter segments are not separately marked. (b) The same part of the Ulam square as in (a), with just primes marked with white (■), for comparison.

with green or yellow in Fig. 1, will be called segments, in the same sense as the segments of a straight line. A sequence will be called a segment when its subsequent pixels are the nearest possible at a given ratio of increments. So, in the case of the ratio 1:1 they will be immediate neighbors (with the Euclidean distance between their centers $\sqrt{2}$), while in the case of the ratio 1:3 they will not, and their centers will be at an Euclidean distance of $\sqrt{10}$. In the latter case, the segment will be considered terminated if the next pixel is farther than this distance.

It is apparent that the primes generated by the Euler formula do form line segments. It can be said, informally, that the longer the segment generated by the formula, the better candidate for the prime-generator the formula is. Therefore, considering the Ulam square from the image processing perspective, we shall be interested in looking for long segments. We expect that finding long segments in the Ulam square will be helpful in exploring some aspects of the number theory.

As it can be seen in Fig. 1b, long segments can not be easily found with the unassisted human eye. The longest segment belonging to the results of the Euler formula is the left yellow segment which has 11 points. The right yellow segment has 10 points. A number of questions can be asked: are these segments the only

segments 11 and 10 points long in the Ulam square? Are there longer segments in this image, and if so, how long, how many and in which locations?

The partial answers to some of these questions have been shown in Fig. 1a. In the Ulam square of the considered size, within the numbers generated by the Euler formula there are no more 11-point segments. There is one more 10-point segment besides the yellow one. Further, there are two 9-point segments, and so on. We shall see further in this paper that in this image there are more 10-point, 9-point, and shorter segments, but there is no other 11-point segment besides that generated by the Euler formula. We shall also see what is the longest segment in the square of 301×301 points. It will appear that there is no longer segment than that one in the Ulam square of the size up to 5001×5001 points.

A good method for finding lines and line segments in an image is the Hough transform (HT). After a broad review on HT [3] was published in 1993, there seemed to be a long break, intermitted only by the appearance of unpublished review reports, like for example [4]. This break was terminated by two recent review papers [5,6]. The first one is of interest for us because it reports on the papers in which differences between the Hough transforms for accurate and approximate objects were discussed. The comparison between the *digital* and *analog* Hough transforms was studied by Kiryati, Lindenbaum and Bruckstein [7]. They discuss the relation between the digital HT according to Cyganski, Noel and Orr [8,9], called there the Analytic Hough Transform, and the conventional HTs, which were analog in nature. Cyganski et al. analyzed digital arcs satisfying the chord property as well as those not satisfying the contiguity condition. They paid much attention to the question of representing all the possible digital straight lines being a digitization of a mathematical straight line segment.

In the Ulam square the points are not a representation of any real-world, analog lines. Neither they are the digital representations of any continuous digital lines passing the image. Therefore, the problem of their detection fits neither to the *analog*, nor to the *digital* version of the Hough transform in the sense defined in [7,8]. In this paper we shall use the version of the HT proposed by us previously in [10] and designed specially for the problem of our interest.

2 Method

The method was described in sufficient detail in [10]. Here we shall very briefly recall its main elements.

Let us locate the origin of the coordinate system O_pq in the central point of the Ulam square corresponding to the number 1. As written before, a line segment in the Ulam square is a sequence of pixels for which the increments $\Delta p, \Delta q$ of the coordinates p, q between the subsequent pixels fulfil the condition $\Delta p/\Delta q = n_1/n_2$, where n_1, n_2 are small integers and n_1/n_2 is an irreducible fraction. This fraction represents the slope of the line. The slope will be represented as an element of an array D_{ij} , with $i = n_1$ and $j = n_2$, called the *direction array*. It is assumed $i \geq 0$ for uniqueness. To relate (i, j) with the directions in the Ulam squares shown in the following, it should be said that i is pointing up

and j is pointing to the right. Therefore, for example, the yellow segments in Fig. 1a have the direction $i:j = 3:(-1)$ and the green one $1:1$. The dimensions of D are $[0, N] \times [-N, N]$. Therefore, a restricted number of angles can be represented, which is in conformity with the nature of the problem. Those elements of D which correspond to reducible and impossible fractions are not used. In each *used* element, the evidence for the existence of lines having the corresponding slope is accumulated.

The elemental subset of pixels in the Ulam square which defines a line is a pair. Each pair of prime numbers is a voting set. If the slope corresponds to an element of D , the pair is stored in D_{ij} , where i/j is the reduced fraction $\Delta p/\Delta q$; otherwise it is neglected. In each element of D the votes can be stored in a vector or a list; in the current implementation the vectors were used to avoid keeping large numbers of pointers in memory. For each vote the line offset (defined as the intercept with the axis Op for horizontal lines and with Oq for the remaining ones) is stored. To make it possible to analyze the results of accumulation, the locations of voting points are also stored and the pairs and their primes are counted.

After the accumulation process is finished the accumulator can be analyzed with respect to such evidence like the existence of lines with large numbers of primes, the contiguous strings of points representing long segments in the square, the directional structure of the Ulam spiral, etc.

In general, the voting sets should be formed by considering each possible pair of primes. In the case of the search for contiguous strings of pixels this time-consuming procedure can be replaced by taking into account only such pairs for which the differences of coordinates correspond to the used elements of the direction array D .

In the implementation used for the results presented in this paper the number N which defines the dimensions of the direction array was assumed $N = 10$. It seemed reasonable not to consider pairs of pixels mutually farther than 10 as neighbors in the Ulam square.

At this stage we have used an implementation of the software which, due to the memory limit, restricted the size of the square to slightly more than 5001×5001 pixels. Therefore, the results for squares with side length up to 5001 will be presented. The number of primes in such a square is 1 566 540 and the largest prime is 25 009 991. The analysis of such a square with a typical PC took not more than several minutes. For smaller N , larger squares could be investigated, but the results were not increasingly interesting. Some of them will also be presented in the next Section. The memory limit can easily be overcome and in the future research we plan to analyze larger squares and larger prime numbers.

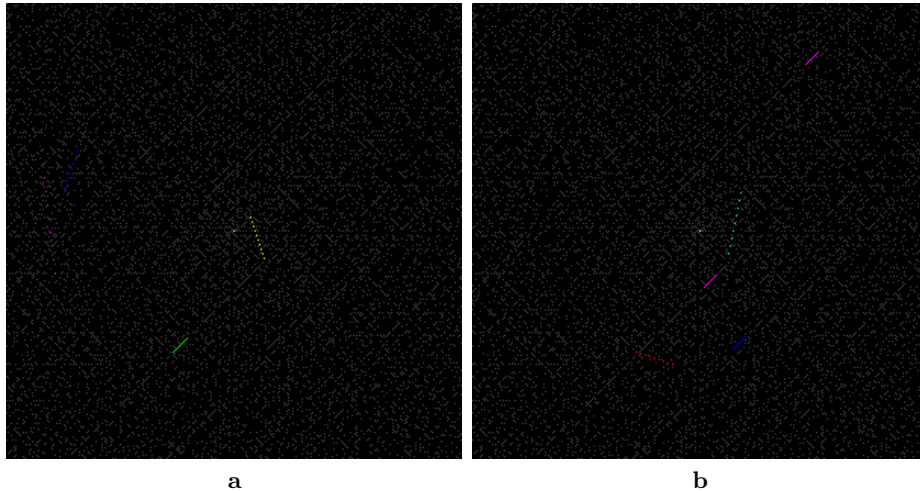


Fig. 2. (a) Segments 10 points long in the square of Fig. 1b. There are four such segments; two of them, marked with yellow (■) and green (■) belong to the set of points generated by the Euler formula, marked in Fig. 1a with the same colors, while the remaining two, marked blue (■) and magenta (■) do not belong to this set. (b) Segments 9 points long in the square of Fig. 1b. There are five such segments; two of them, marked magenta (■) belong to the set of points generated by the Euler formula, marked in Fig. 1a with the same colors, while the remaining three, marked red (■), blue (■) and cyan (■) do not belong to this set. The colors: white (), grey (■) and black (■) have the same meaning as in Fig. 1a.

3 Results

3.1 Square of size 301 × 301 points

Segments shorter than 4 points do not indicate any stable tendency and are not interesting. Segments shorter than 6 points are very numerous and they will not be shown in this paper (see also Sect. 3.4). Segments of lengths from 6 to 16 points long will be presented here. We shall visualize them in the square 301×301 points due to that its size is small enough for presentation in the format of the paper. We briefly mention larger squares in the end of this Section.

We have already stated that the longest segment in the square of the considered size present in the set of primes generated by the Euler formula is 11 points long and there are no more segments of this length in this square. Now, we shall go on to segments 10 and 9 points long due to that they were already mentioned in relation with the Euler formula. There are four segments of length 10 and five segments of length 9. Their locations and slopes are shown in Fig. 2. Their relation to the primes generated by the Euler formula is outlined in the caption of this Figure. Shorter segments of the lengths 6, 7 and 8 points, can be seen in Fig. 3a, b and c.

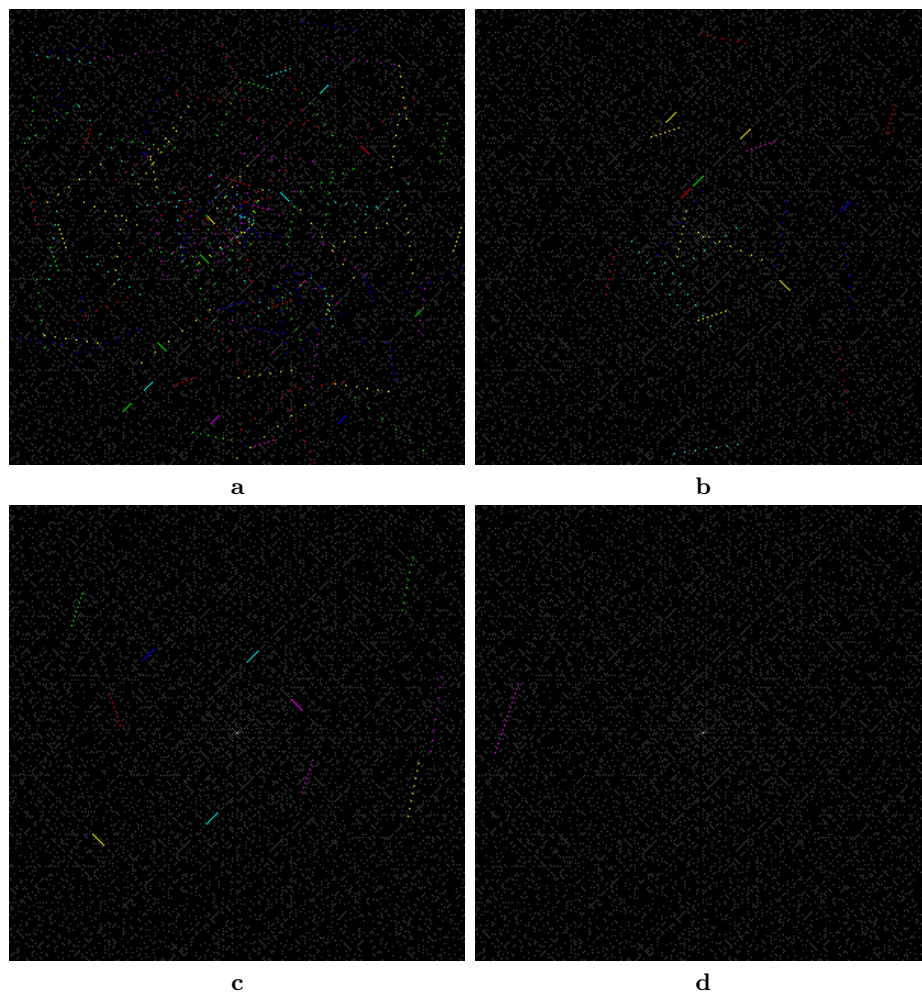


Fig. 3. Segments in the square of Fig. 1b: (a) 6 points long; (b) 7 points; (c) 8 points; (d) 16 points long – the longest segment in this square.

The longest segment has 16 points and has the slope $i:j = 3:1$ (Fig. 3d). It consists of the primes:

$$\begin{aligned}
 &75\,227(-13, -137), 74\,131(-10, -136), 73\,043(-7, -135), 71\,963(-4, -134), \\
 &70\,891(-1, -133), 69\,827(2, -132), 68\,771(5, -131), 67\,723(8, -130), \\
 &66\,683(11, -129), 65\,651(14, -128), 64\,627(17, -127), 63\,611(20, -126), \\
 &62\,603(23, -125), 61\,603(26, -124), 60\,611(29, -123), 59\,627(32, -122)
 \end{aligned}$$

where $a(b, c)$ denote: a – prime number, (b, c) – its coordinates in the Ulam square. It can be seen that the points are not far from the center of the square. It is interesting to note that there are no segments of lengths from 12 to 15 points in the considered square.

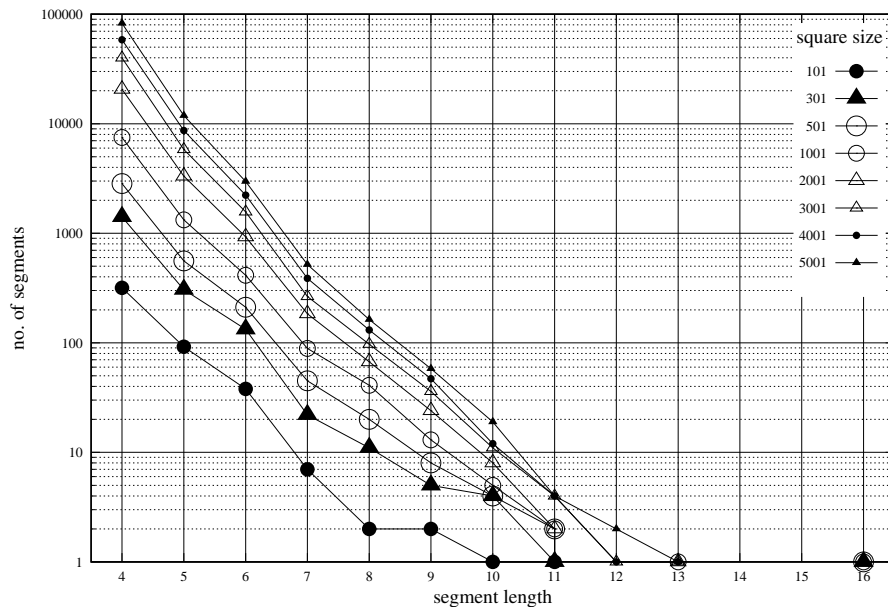


Fig. 4. Number of segments versus length of the segment for various sizes of the Ulam square. If the number of lines is zero the data point is absent.

3.2 Distribution of lengths of the segments

For the direction table of dimensions $[0, 10] \times [-10, 10]$ as specified before, the squares of dimensions up to 5001×5001 were analyzed. In the graph in Fig. 4 it is shown what was the distribution of the numbers of segments versus their lengths, for the squares of different sizes. It can be noticed that the numbers of short segments grow together with the size of the square, while the numbers of the longest segments change slightly or remain constant. As it can be expected, the longest segments are rare.

To show at least some of the results for longer segments and larger squares we have chosen the square of 1001×1001 points and the 9-point segments shown in Fig. 5. More images and results are available online (see Sect. 3.4).

3.3 Larger Ulam squares and reduced direction table

To see whether a longer segment than 16 points can be found in larger squares, and to stay within the memory limit we have with the currently used software, some calculations were carried out with the limited set of directions. As it was written before, the dimensions of the direction table D are $[0, N] \times [-N, N]$ and up till now the value of N was 10. With N limited to 5 the arrays up to 9001×9001 points, and with N limited to 3 the arrays up to $10\,001 \times 10\,001$ points were analyzed. Some new segments of length 11 and 12 points were found. No

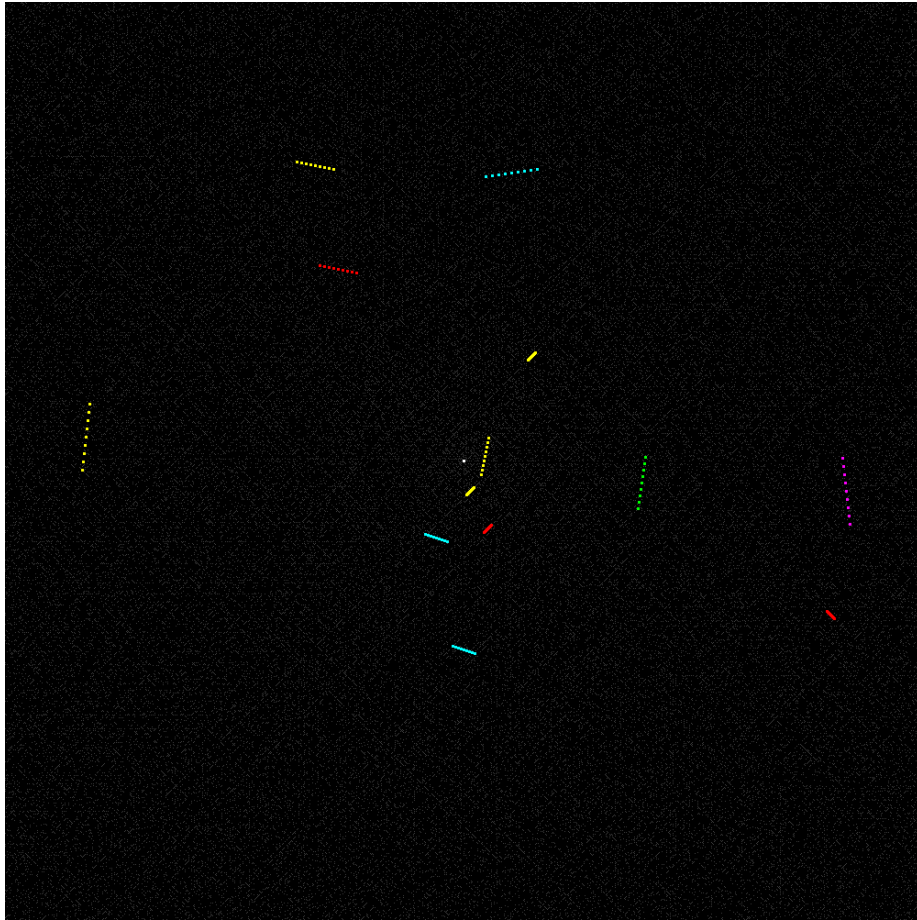


Fig. 5. Segments 9 points long in a 1001×1001 points Ulam square. For better visibility the white and color points were replaced with squares 3×3 pixels centered at respective points.

new segment of length 13 points and no segments of length 14 and 15 points were found. No other segment of length 16 was found besides that already seen in 301×301 square. No segment longer than 16 points was found. The square of $31\,623 \times 31\,623$ in which the largest prime would be over 10^9 could not be analyzed even for N limited to 1. Detailed results are available online (see next Section).

3.4 Online resource

The results are in a large part in the form of images which are too large to be presented in a page of a paper. Also the volume of numerical data is too large to

show them in tables. Therefore we have established a web page [11] in which the results of this one and our other papers on the prime numbers will be available.

4 Summary and prospects

The Ulam square was analyzed from the image processing perspective. The straight line segments were looked for, in hope that this can be helpful in exploring the peculiarities of the number theory.

In the squares with the side of length up to 5001 points which correspond to the largest prime 25 009 991 the longest segment has 16 primes and its slope is 3:1 (3 up and 1 to the right). There were no segments of length 14 and 15. The number of shorter segments varied strongly, from one for 13-point segment to tens of thousands for shorter ones. Angles which described the slopes of the segments had tangents p/q expressed by integers $p \in [0, 10]$ and $q \in [-10, 10] \setminus \{0\}$. A version of the Hough transform designed specially for detecting such segments in the Ulam spiral was used. Detailed results can be seen in the web page [11].

In the future we shall develop the software with significantly weaker restrictions on memory, capable of analyzing a larger part of the Ulam square, and consequently, larger primes. Squares centered at other numbers than 1 will be possible to analyze.

In the data presented here it is not possible to observe to what extent the number of segments is going down together with the distance from the center of the square; however, such a phenomenon should be expected due to that the density of primes themselves is going down in this way. It would be interesting to see what is the relation of this decline to the decline in the density of primes according to the prime counting function.

References

1. Euler, L.: Extrait d'un lettre de M. Euler le Pere à M. Bernoulli concernant le mémoire imprimé parmi ceux de 1771. *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres* (1772) 35–36 eulerarchive.maa.org/pages/E461.html.
2. Stein, M.L., Ulam, S.M., Wells, M.B.: A visual display of some properties of the distribution of primes. *The American Mathematical Monthly* **71**(5) (1964) 516–520 [doi:10.2307/2312588](https://doi.org/10.2307/2312588).
3. Leavers, V.F.: Which Hough transform? *CVGIP: Image Understanding* **58** (1993) 250–264 [doi:10.1006/ciun.1993.1041](https://doi.org/10.1006/ciun.1993.1041).
4. Antolovic, D.: Review of the Hough transform method, with an implementation of the fast Hough variant for line detection. Department of Computer Science, Indiana University (2008) Technical report [TR663](#).
5. Mukhopadhyay, P., Chaudhuri, B.B.: A survey of Hough transform. *Pattern Recognition* **48**(3) (2015) 993–1010 [doi:10.1016/j.patcog.2014.08.027](https://doi.org/10.1016/j.patcog.2014.08.027).
6. Hassanein, A.S., Mohammad, S., Sameer, M., Ragab, M.E.: A survey on Hough transform, theory, techniques and applications. *CoRR* **abs/1502.02160** (2015) [arXiv:1502.02160](https://arxiv.org/abs/1502.02160).

7. Kiryati, N., Lindenbaum, M., Bruckstein, A.M.: Digital or analog Hough transform? *Pattern Recognition Letters* **12**(5) (1991) 291–297 [doi:10.1016/0167-8655\(91\)90412-F](https://doi.org/10.1016/0167-8655(91)90412-F).
8. Cyganski, D., Noel, W.F., Orr, J.A.: Analytic Hough transform. In: *Proc. SPIE. Volume 1260: Sensing and Reconstruction of Three-Dimensional Objects and Scenes*. (1990) 148–159 [doi:10.1117/12.20013](https://doi.org/10.1117/12.20013).
9. Liu, Y., Cyganski, D., Vaz, R.F.: Efficient implementation of the analytic Hough transform for exact linear feature extraction. In: *Proc. SPIE. Volume 1607: Intelligent Robots and Computer Vision X: Algorithms and Techniques*. (1992) 298–309 [doi:10.1117/12.57109](https://doi.org/10.1117/12.57109).
10. Chmielewski, L.J., Orłowski, A.: Hough transform for lines with slope defined by a pair of co-primes. *Machine Graphics & Vision* **22**(1/4) (2013) 17–25 [Open access: mgv.wzim.sggw.pl/MGV22.html].
11. Chmielewski, L.J., Orłowski, A.: Prime numbers in the Ulam square (2016) www.lchmiel.pl/primes [Online; accessed 14 July 2016].