Diameter Distribution Approach to Circle Recognition in Binary Images

Agata Migalska[1]

Abstract: *In this paper, a diameter distribution approach to the circle recognition in the binary images is presented. A statistical test for verifying the circularity of an object in the image is proposed and an algorithm for calculating the most probable circle diameter is derived. Experimental results on images representing various shapes confirm the accuracy of the proposed method.*

Keywords: *digital circle, circle recognition, diameter estimation, statistical inference*

1. Introduction

Through the last few decades there has been a continuous interest in detecting and recognizing circles. This interest is due to the fact that circular objects with their coordinates in the real space are affected not only by the sampling error but also by the quantization error when coordinate values are rounded to the nearest integer.

There is a number of relevant applications of circle recognition and diameter estimation to industrial tasks, especially to automated visual inspection. Automated visual inspection is aimed at asserting that its geometrical (size or shape), structural (missing items or foreign objects) and superficial (surface quality, appearance) product characteristics are compliant with an established standard. These measurements serve to reject the defective products, adjust the manufacturing process as well as to gather the statistics, eg. on how the product diameters vary. Moreover, due to the natural variability in the products and different requirements on precision of the manufacturing process, it is reasonable to regard inspection as a process of making measurements that have to be checked statistically. The example applications of measuring the circular objects are shape and size of pizzas or biscuits which should be circular and fit into the box or package, size and roundness of the O-rings or piston heads [1]. In particular, controlling the diameter and circularity of an object is used in the diameter control system of crystal growth so that it grows into a perfect cylinder [2].

Traditionally, the circle detection in digital images is realized by means of Circular Hough Transform (CHT) [3, 4]. CHT is a two step procedure where in the first step an edge detector is used to deduce the center locations and radius values, while a second step involves peak detection by averaging, filtering and histogramming the transform space. The algorithm is robust under noise and partial occlusion, but its high computational and memory complexity makes it unsuitable for real-time applications. Moreover, each discrimination method has its drawbacks and none of them can be chosen for general purpose.

Another approach to the circle recognition is by its symmetric features [5, 6]. An object is symmetric if there is at least one symmetry that leaves the object unchanged. On the Euclidean plane a circle has two symmetries, rotational and reflectional. As a result, in order to detect whether an image represents a circle, symmetry detection methods can be used.

[1] Instytut Informatyki, Automatyki i Robotyki, Politechnika Wrocławska, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, `agata.migalska@pwr.wroc.pl`, http://agata.migalska.staff.iiar.pwr.wroc.pl

A method presented in this paper makes an attempt to detect whether the digital pattern $\mathbb{Q} \subset \mathbb{Z}^2$, where $\mathbb{Z}^2$ is a two-dimensional integer lattice on the Euclidean plane, is a representation of a circle in the real domain $\mathbb{P} \subset \mathbb{R}^2$ based on the cumulative distribution function of the circle diameters. When a circle center is of no interest, techniques based on diameter calculation are beneficial as only the points read from an image are considered and no additional erroneous point calculation is introduced. Moreover, the most probable circle diameter is determined and features of the distribution of the circle diameters are examined.

This paper is organized as follows. In section 2 the problem formulation followed by the discretization error estimation and the approximation of the sample distribution are given. In section 3 the proposed solution is explained. Obtained results are presented in section 4 and conclusion is made in section 5.

2. Problem Formulation

The binary digital images of various geometric shapes are considered. A shape represented in an image is a distorted, digitized version of an original analytical image.

It is assumed that the shape in the image is given by its boundary which is 1 pixel thick. Although the boundary tracking procedures generally do not guarantee the unit thickness of an edge, it can be achieved by virtue of thinning, a morphological operation used to reduce edges to unit thickness. It is also assumed that the diameter of the circle is at least 1 pixel long. Otherwise, a circle is represented by a single point.

The concept presented in this paper is based on the property of the circle that two arbitrarily chosen circle diameters are of the same length. The diameter is henceforth defined as the distance between two boundary pixels, an arbitrarily chosen pixel q_i and a pixel q_j, $i \neq j$, which is the most distant from q_i among all the pixels constituting the digital boundary.

Given the binary image, the coordinates of n points (boundary pixels) are read in Cartesian coordinate system. For each point, the most distant point is determined and the distance between them is calculated. The resulting set of distances constitute a real-valued sample $\mathbb{S} = (s_1, \ldots, s_n)$, $s_i \in \mathbb{R}$, of n diameters.

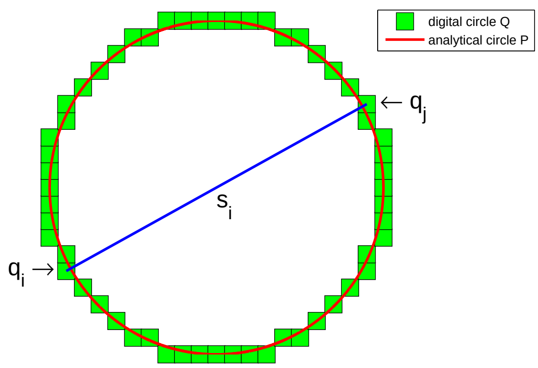

Fig. 1. Digital representation of the circle

2.1. Digitization Error

When the real-valued shape is mapped onto the integer lattice, its values are rounded to integer values, usually to the nearest integer. If the digital pattern in the binary image represents a circle and a diameter of this circle is equal s, it can be shown that each diameter of the digital pattern differs from s no more than $e_r = 2 + \sqrt{2}$. This finding form the basis for the construction of the circle detection statistical test further described in section 3.2.

2.2. Theoretic and Sample Distributions

The family of circle diameters distributions is a one-parameter location family of discrete distributions $\mathcal{P} = \{P(s), s \in \mathbb{R}\}$ with location parameter s. The cumulative distribution function (cdf) of $P(s) \in \mathcal{P}$ is a function $F_0 : \mathbb{R} \to \mathbb{R}$ given by

$$F_0(t) = \begin{cases} 0, & \text{if} \quad t \in (-\infty, s), \\ 1, & \text{if} \quad t \in [s, +\infty). \end{cases} \tag{1}$$

It is equivalent to stating that all the diameters of a circle are equal s with probability 1. The empirical cumulative distribution function (ecdf), i.e. the approximation of the cdf calculated from the sample, should converge to F_0 when the sample variance var approaches zero. However, in case of the digital images this convergence is distorted.

By means of statistical tests asserting normality of the distribution [7], it was discovered that the ecdf of the circle diameters is similar in shape to the cdf of the normal distribution with mean and variance equal to sample mean and sample variance. This is demonstrated in Figure 2. The plot was generated for the test image representing a circle of a diameter equal 200 drawn in GIMP 2.8. The sample $\mathbb{S}$ comprised of 1124 elements with sample mean $\bar{s} = 199.8457$ and sample variance $var = 0.28335$. Plot presents an empirical cumulative distribution function calculated from sample $\mathbb{S}$ and a cumulative distribution function of the normal distribution $N(\bar{s}, var) = N(199.8457, 0.28335)$ as well as a theoretical cdf for $s = 200$.

Therefore, it was concluded that the normal distribution can be accepted as a circle diameter sample distribution approximation.

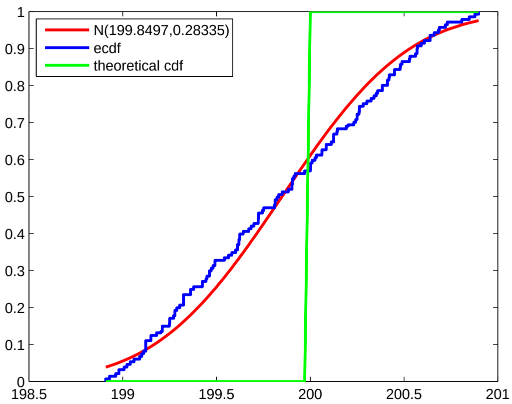

Fig. 2. Comparison of cumulative distribution functions of sample, normal and theoretical distributions.

Provided that a digital pattern represents a circle, the upper bound of the variability in the calculated diameters is equal e_r. Therefore, the circle diameter sample distribution can be approximated with the normal distribution with variance less than e_r.

3. The Proposed Solution

The proposed solution comprises of two steps: circle diameter calculation (section 3.1) and shape test for circularity (section 3.2).

3.1. Circle Diameter Calculation

Let us assume that the shape in the image represents a circle. In statistical inference the distance between empirical and theoretical cumulative distribution functions is expressed by means of Kolmogorov distance D_n. By Glivenko-Cantelli theorem [8], when sample size approaches infinity then Kolmogorov distance should converge to zero. Thus, the diameter s_0 is an integer value for which the Kolmogorov distance D_n between the theoretical cumulative distribution function F_0 and the empirical cumulative distribution function F_n is minimal. The formula for diameter calculation is given by equation 2, where $R_s = [\lfloor \min(\mathbb{S}\rfloor, \lceil \max(\mathbb{S})\rceil] \subset \mathbb{N}$ and $\mathbb{S}$ is a sample of diameters obtained from an image.

$$s_0 = \inf_{s_0 \in R_s} D_n = \inf_{s_0 \in R_s} \left[\sup_{s \in \mathbb{S}} |F_0(s) - F_n(s)| \right] \tag{2}$$

3.2. Shape Test

If the shape in the image is a circle, then the empirical cumulative distribution function is bounded above by the function F_u (3) and bounded below by a function F_l (4).

$$F_u(s) = \begin{cases} F(s_0 - 0.5, e_r) & s < s_0 - 0.5 \\ 1 & s \geq s_0 - 0.5 \end{cases} \tag{3}$$

$$F_l(s) = \begin{cases} 0 & s < s_0 + 0.5 \\ F(s_0 + 0.5, e_r) & s \geq s_0 + 0.5 \end{cases} \tag{4}$$

where F is the cdf of the normal distribution.

Let us define two sets of points, U and L, for which bounding functions are exceed by ecdf.

The test to determine whether the shape in the digital image represents a circle is defined as follows.

- *Hypotheses*

$$H_0: \text{ shape in the image represents a circle with a diameter equal } s_0 \tag{5}$$

$$H_1: \text{ shape in the image does not represent a circle} \tag{6}$$

- *Test statistic*

$$C = ||L|| + ||U|| \tag{7}$$

- *Critical region*

$$W = [1, +\infty) \tag{8}$$

The boundary cumulative distribution functions F_u and F_l defined in equations 3 and 4, respectively, express the biggest acceptable departure from the theoretical cumulative distribution function F_0 that could be justified by the coordinates values rounding error. Consequently, these functions define the maximal departure from the ideal circle by its representation in the binary digital image. If the empirical distribution function exceeds the boundary cdfs at any point, then the allowed rounding error e_r is exceeded and the shape in the image is unlikely to represent a circle. Therefore, the proposed critical region W (8) is defined so that the hypothesis, that a shape represents a circle, is rejected if a diameter empirical cumulative distribution function exceeds the boundary function F_u or F_l at any point.

The test allows for the shape to slightly departure from the ”ideal” digital circle representation. This tolerance is justified by the presence of sampling noise and other sources of variability in the digital images.

4. Obtained Results

In order to verify the correctness and robustness of the proposed solution, several tests were run. All the test images were generated either in MATLAB 7.12.0 (R2011a) or in GIMP 2 version 2.8, a free graphical tool used in computer graphics and web design. All the test images represented black shape boundaries on the white background.

Selected test results are presented in Table 1 and Figure 3. The image numbers in Figure 3 refer to the corresponding numbers in Table 1.

Tab. 1. Test statistic C, calculated diameter s_0, sample mean $\bar{s}$ and variance var for selected images

N°	**Image Description**	C	s_0	$\bar{s}$	var
1	Circle $d = 200$, GIMP 2	0	200	199.849663	0.283352
2	Ellipse, 200x195, GIMP 2	78	197	197.340328	3.558856
3	Square, diagonal $= 200$, GIMP 2	138	178	177.538622	147.286359
4	Hexagon, minimum bounding circle $d = 200$	195	190	190.149061	33.601873

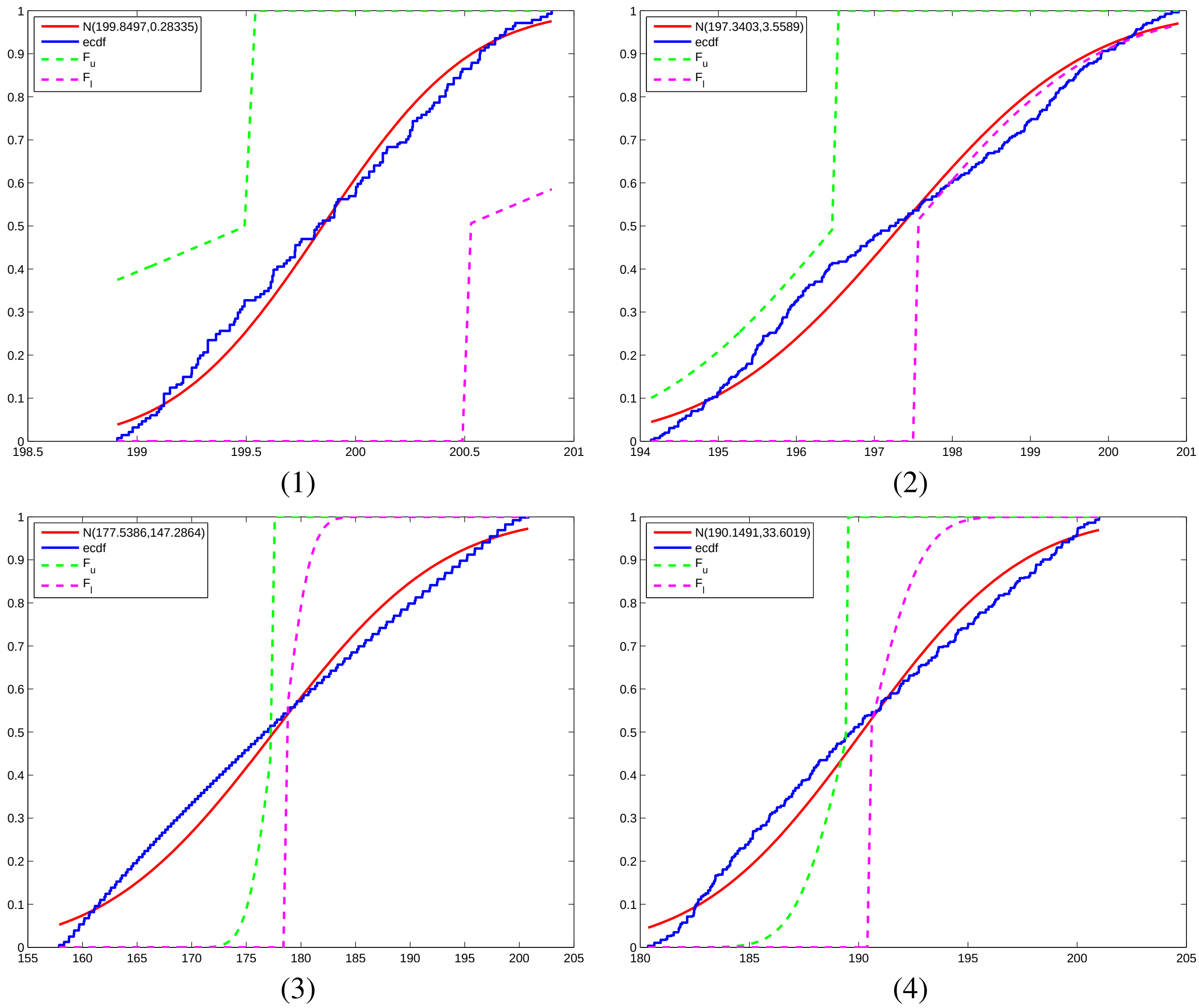

Fig. 3. Shape test results for several test images

Test results proved the accuracy of the proposed method of a circle recognition. Moreover, the calculated diameter was equal to expected value at all times. Due to the test tolerance, ellipses having shape close to circular and arcs which subtend an angle no lesser than 307° were also recognized as circles. On the other hand, for the shapes such as hexagon or square the null hypothesis was rejected.

5. Conclusion

In this paper, a new approach to circle recognition in the digital binary images has been presented. It has been observed that the sample distribution of the circle diameters can be approximated with the normal distribution having variance lesser than $e_r = 2 + \sqrt{2}$. The statistical test asserting the circularity of a shape has been presented and a formula for circle diameter caluclation derived. Moreover, through the experimental study the method has proved to be of high accuracy.

Should the work on the proposed method be continued, determining the distribution of the test statistic as well as resistance to noise evaluation will be addressed in the first place.

References

[1] E. Davies, Machine Vision: Theory, Algorithms, Practicalities, Morgan Kaufmann, 2005

[2] D. Hurle, Control of Diameter in Czochralski and Related Crystal Growth Techniques, J. Cryst. Growth, 42, 473–482, 1977

[3] R. Duda, P. Hart, Use of the Hough Transform to Detect Lines and Curves in Pictures, Commun. ACM, 15, 11–15, 1972

[4] C. Kimme, D. Ballard, J. Sklansky, Finding circles by an array of accumulators, Commun. ACM, 18, 120–122, 1975

[5] G. Marola, On the detection of the axes of symmetry of symmetric and almost symmetric planar images, IEEE Trans. Pattern Anal. Mach. Intell., 11(1), 104–108, 1989

[6] N. Bissantz, H. Holzmann, M. Pawlak, Testing for Image Symmetries - With Application to Confocal Microscopy, IEEE Trans. Inf. Theory, 55(4), 1841–1855, 2009

[7] H. Lilliefors, On the Kolmogorov-Smirnov test for normality with mean and variance unknown, J. Amer. Statist. Assoc., 62, 399–402, 1967

[8] R. Magiera, Modele i metody statystyki matematycznej. Tom II - Wnioskowanie statystyczne, Oficyna Wydawnicza GiS, Wrocław, 2007

ROZPOZNAWANIE OKRĘGÓW NA OBRAZACH BINARNYCH W OPARCIU O ROZKŁAD PRAWDOPODOBIEŃSTWA ŚREDNIC

W przeciągu ostatnich kilku dekad zaproponowano wiele metod rozpoznawania okręgów na obrazach cyfrowych, jednak prace nad znalezieniem precyzyjnego i szybkiego algorytmu nadal trwają.

W artykule przedstawiono metodę służącą do rozpoznawania okręgów na obrazach binarnych wykorzystującą własność, że dowolne dwie średnice okręgu są sobie równe. Wyznaczono górne ograniczenie błędu towarzyszącego wyznaczaniu średnicy okręgu, wynikającego z zaokrągleń rzeczywistych wartości współrzędnych punktów okręgu do wartości całkowitych. Na podstawie analizy rozkładu prawdopodobieństwa otrzymanych średnic wyznaczono maksymalną dopuszczalną rozbieżność pomiędzy rozkładem empirycznym a rozkładem teoretycznym. Zaproponowano test statystyczny do oceny kolistości figury na obrazie oraz algorytm służący do obliczania najbardziej prawdopodobnej średnicy okręgu. Poprawność proponowanej metody przetestowano na obrazach binarnych reprezentujących zarówno okręgi, jak i inne figury geometryczne.

Zaproponowana metoda może być szczególnie użyteczna w systemach automatycznej kontroli wyzyjnej okrągłych elementów wytwarzanych przemysłowo.