UDK 681.3

A.V. Shehovtsov, A.N. Poletayeva

# IDENTIFICATION OF HANDWRITTEN MARKS AND DIGITS OF FONTS

*Summary. The main types of fonts using in up to date of operating system are considered. The knowledge in theory of pattern recognition are used. The problem of identification of figures of font by means of electric eye is noted. The conception of density of inking shape is determined. The method of identification of figures by given of density of inking shape is proposed. A priori data for different fonts and its size is briefly described. The recommendations of use fonts for identification of electric eye are given.*
*Keywords.pattern recognition, font identification, electric eye.*

## Introducing.

The modern operating systems (OS) can work with following fonts: raster and vectorial, scalable and built in the output device (printers). The raster fonts contain bit images of the definite sizes. If there is an inquired size there is its scaling, and at a large font size the contour of a character receives a serrate view. The vector fonts store the map in a vectorial view, that at scaling gives excellent outcome. They will be used basically on plotters. The scalable TrueType fonts have received the greatest distribution in different operating systems, being by the best compromise between raster and vector fonts. At a level of means of rendition of operating system the scaling of these fonts is stipulated at which one quality practically is not degraded [1].

The theory of pattern recognition includes set of the approaches and algorithms for the solution of a problem of identification. The analysis of density falls into to algorithms of identification. The activity is prepared for implementation of identification of the reports and questionnaire on referendum.

## Formulation.

The problem has arisen, at a building of a system of the automated workstation of the member of electoral commission, for the answer to a problem on a capability of identification of digits of a font and handwritten mark by means of photo cells and obtaining of outcome of identification with the greatest veracity.



 ISSN 1562-9945

**Purpose of activity.**

To define conditions, at which one it is possible to recognize digits and handwritten checkmarks with the help of the optical converter.

**Base material.**

Modern operating systems will use TrueType fonts thus more often most widespread are: Times New Roman, Arial, Courier New, as awakely used at clerical work On this cause we shall consider scaling of TrueType fonts. The altitude of digit in points depends on resolution of the output device and font size and sets as follows:

$$Height = -Size \cdot PxelPerInc / 72 \quad (1)$$

*Height* - altitude of a character in points;

*Size* - font size in points pt;

*PixelsPerInch* - resolution dpi (dot per inch- of points in inches).

The glow discretization of area identification of digit gives a main specification $d$, we shall call as its *density of plotting contour (DPC)*:

$$d = \frac{S_ц}{S_п + S_о - S_ц} \quad (2)$$

where Sц - area of the basic contour of digit;

Sп - area of a minimum orthogonal in which one will be placed digit of a number 0-9;

So - area around of digit;

The change of value of a font changes the size on (1) discretely and, therefore, even under ideal conditions of plotting parameter d will have some fluctuations on parameter of the size *size*.

Distribution DPC for a font Times New Roman 10pt, 600 dpi (Fig. 1), obtained by a programming way, demonstrates their manifestative distribution. Analyzing a volume range indispensable for difference most of composite segment, receive:

$$n = log_2 \left| \frac{d_{max} - d_{min}}{d_1 - d_2} \right| \quad (3)$$

where $n$ - digit capacity of a volume range in bits;

$d_{max}$ $u$ $d_{min}$ - absolute maximum and minimum density;

$d_1$ $u$ $d_2$ - nearest two adjacent points.

For the above mentioned case is received n = 5. Because of a discretization of key points of a font and therefore illconditioned plotting the given situation does not repeat for the adjacent sizes of fonts. For most often of used size 9pt we shall receive n=8, and for 11pt n=7.

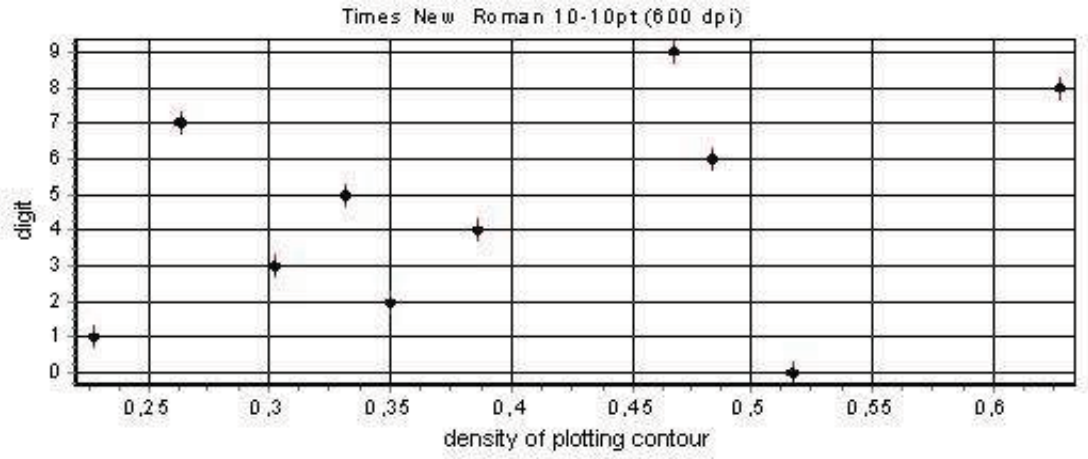


Figure 1 – Distribution DPC for a font Times New Roman 10pt (600 dpi)

Analyzing a segment with 9 up to 11pt (Fig. 2) can be noted, that at scaling of digits in DPC the fluctuations are watched. Augmenting a segment of the analysis up to 8-14pt (Fig. 4) there would be digits bigger of the size, which one, because of reduction error of plotting, have practically identical DPC and will derivate oil clotss of heightened density (for example, area of digit 1).

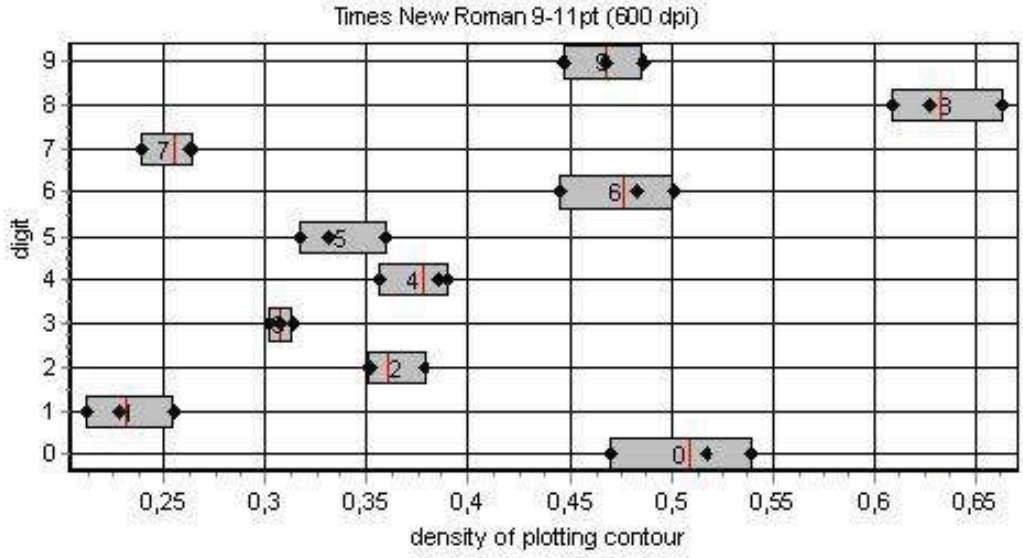


Figure 2 – Distribution DPC for a font Times New Roman 9-11pt (600 dpi)

Is apparent, that the unique definition of digits from range of the sizes 9-11pt on parsed density is impossible in connection with multiple

imposing of different digits. But from a Fig. 2 unique determinate digits 3 and 8, group of digits 9-6-0, 5-4-2 and 7-1.

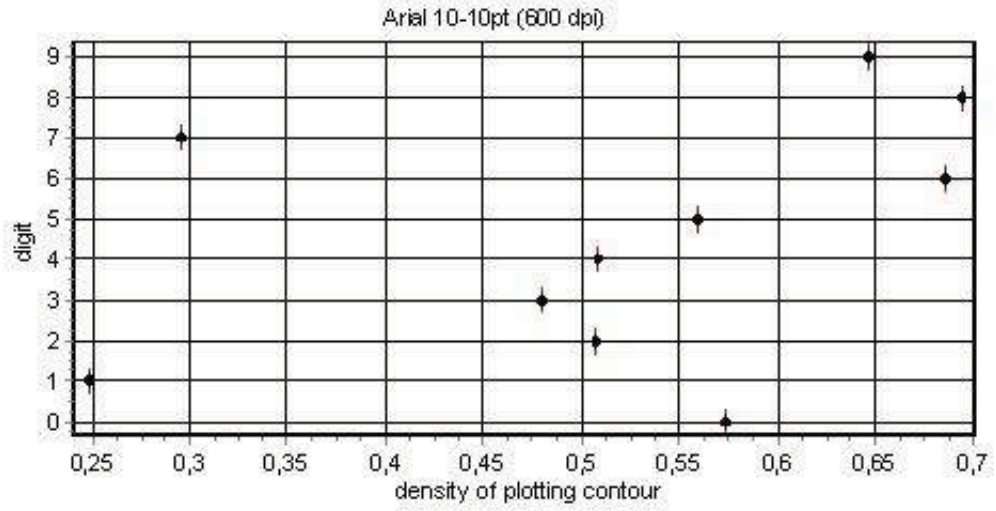


Figure 3 – Distribution DPC for a font Arial 10pt (600 dpi)

The analysis of distribution for a font Arial by the size 10pt (the Fig. 3) demonstrates a composite situation with identification of digits 2 and 4, that uprates a volume range n=10. Similarly for most often of used size 9pt we shall receive n=9, and for 11pt n=8. The obtained nonlinear dependences for fonts Times New Roman and Arial are connected with clout of a discretization of coordinates(co-ordinates) at plotting characters and as a consequent of a large dispersion of value DPC.

We shall put distribution DPC most often of used fonts Times New Roman (Fig. 4), Arial (Fig. 5) and Courier New (Fig. 6) for mesh sizes from 8 up to 14 at 600 dpi.

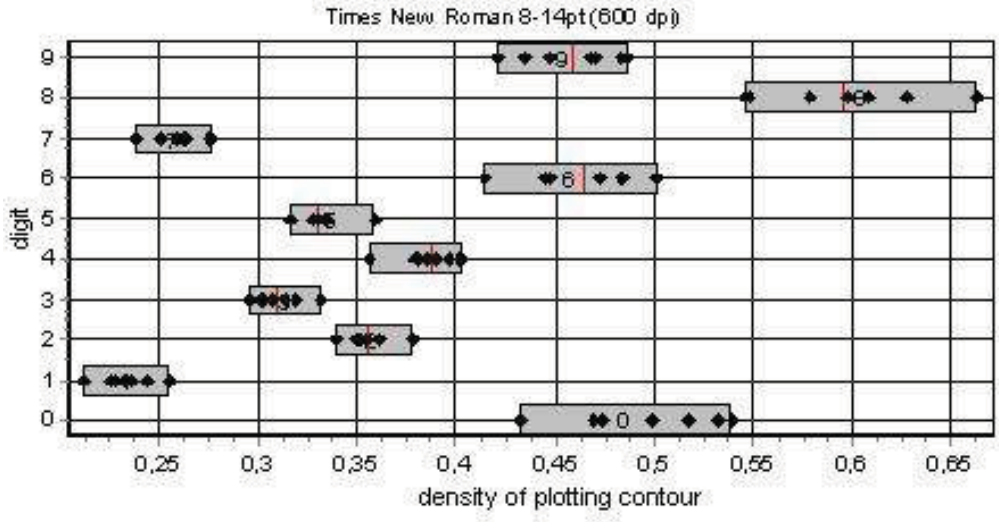


Figure 4 – Distribution DPC for a font Times New Roman 8-14pt (600 dpi)

For the propagate sizes of printing 8-14pt is received following groups 3-5-4-2, 9-6-0, 7-1 and only one unique area of digit 8.

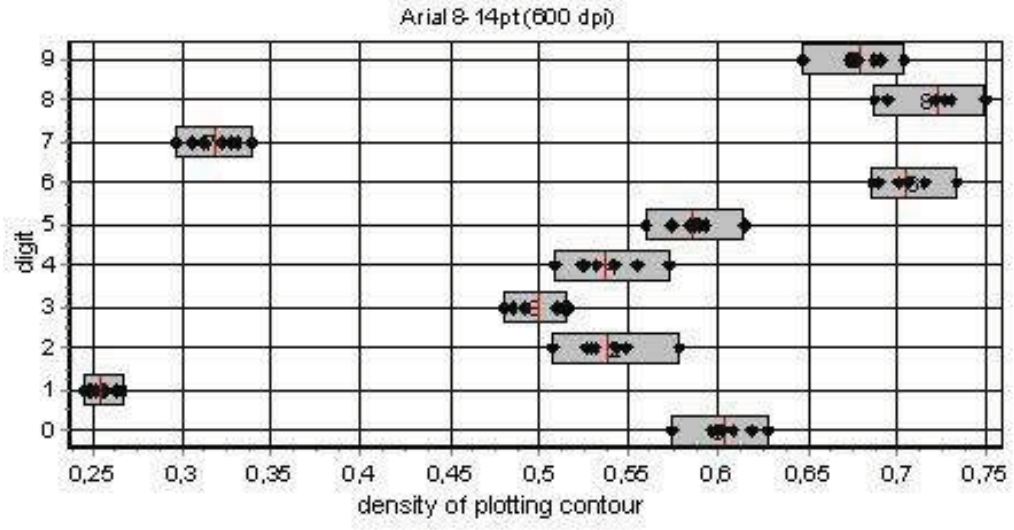


Figure 5 – Distribution DPC for a font Arial 8-14pt (600 dpi)

For a font Arial the similar distribution is obtained, except that now it is possible to define two digits 1 and 7, which one is far from groups 3-2-4-5-0 and 9-8-6.

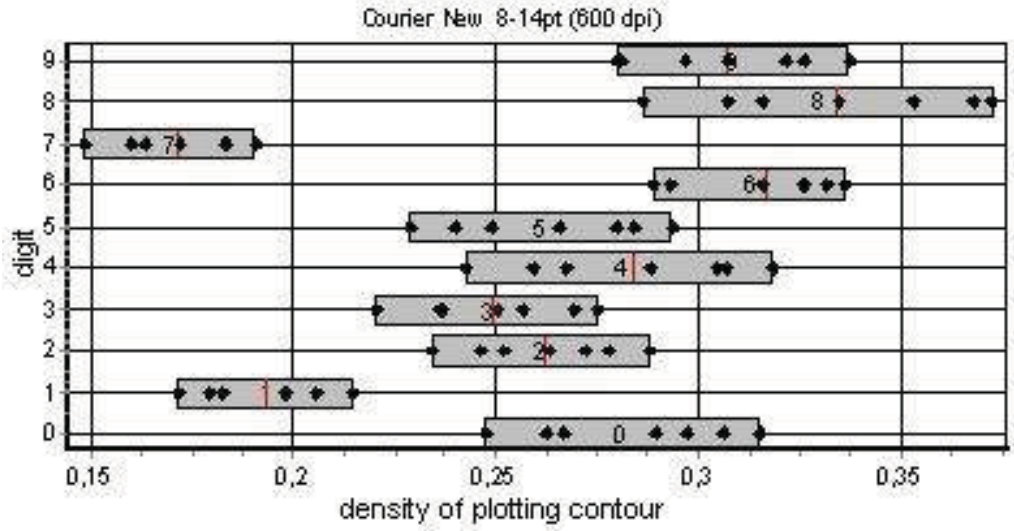


Figure 6 – Distribution DPC for a font Courier New 8-14pt (600 dpi)

The distribution of a font Courier New show on a Fig. 6 as against Times New Roman and Arial demonstrates multiple imposing DPC of digits. In this case it is impossible uniquely to determine digit on it density, there are two areas of overlap 7-1 and 9-8-6-5-4-2-0.

Is interquartile, that for the analysis of the written character, the person recognizes set of yardsticks not only shape and edges, bending

and corners, and depth of a character and it DPC. If in last the imposing or close on density characters are watched, accordingly there can be a doubt in a regularity of identification. The distribution DPC on all area, as in font Times New Roman, will put to more unique identification of a character.

In a Fig. 7 the relation of overlap of areas DPC most often of used fonts is submitted: Arial, Courier New and Times New Roman for the large sizes from 150 up to 250pt (600 pdi).

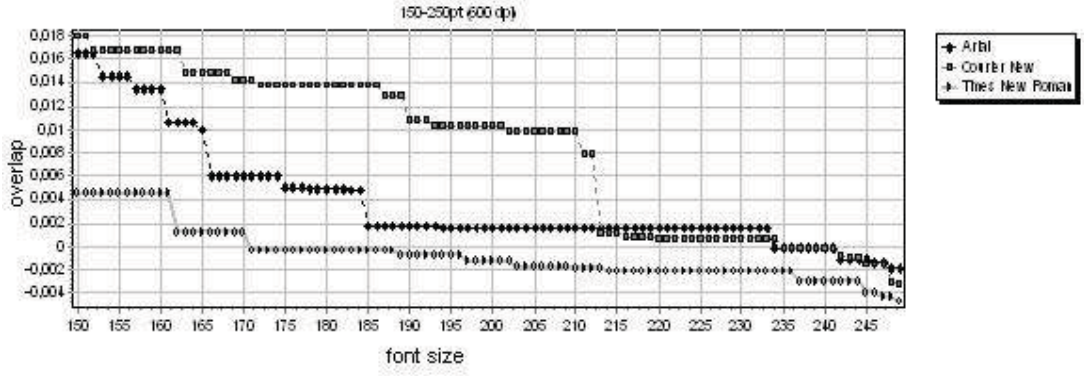


Figure 7 – Relation of overlap DPC from a font and its size.

The positive value of overlap of areas DPC will be derivate because of imposing of areas DPC of separate digits, negative testifies to existence of spacing interval between the nearest areas. The discontinuities of overlaps conditioned by chaotic nature of fluctuations DPC and indicate occurrence / petering of the next value, which one derivated limit of area; stable states on occurrence / petering of value located inside area.

The analysis is made disregarding errors - availability of discard of printing, interferences, color heterogeneity of a paper etc. At the same size notes on the paper carrier the defect results in ascending error, owing to increase of percent of the area of an introduced defect as contrasted to by general area of a font size.

The reviewed analysis has demonstrated: that there is a limit of a font size after which one overlap of density of plotting of digits does not descend. For Arial is received 234, Courier New - 235 and Times New Roman - 171. In figures 8, 9 and 10 the distributions DPC for the large sizes of fonts are figured, at which one there is no overlap.

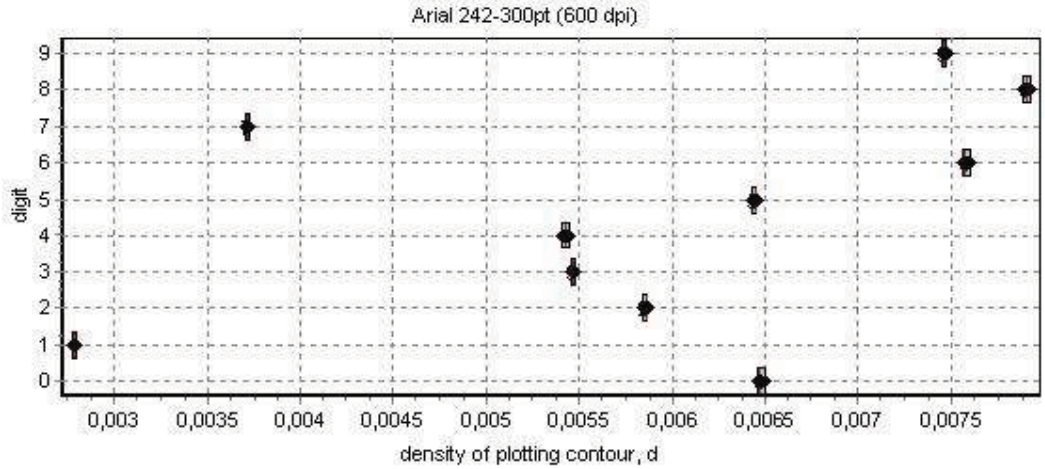


Figure 8 – Distribution DPC for a font Arial 242-300pt (600 dpi)

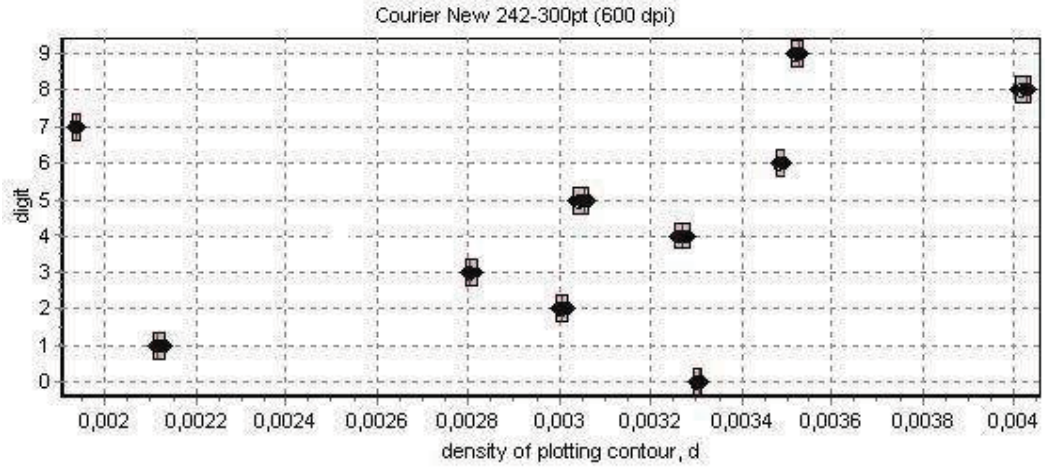


Figure 9 – Distribution DPC for a font Courier New 242-300pt (600 dpi)

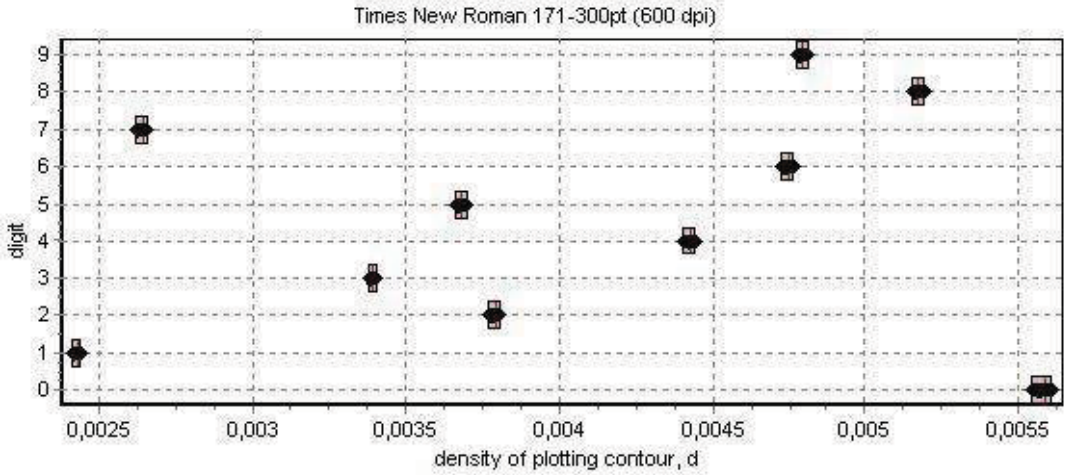


Figure 10 – Distribution DPC for a font Times New Roman 171-300pt (600 dpi)

 ISSN 1562-9945

In figure 11 the distribution DPC for a font Times New Roman 171-300pt is figured in the field of digits 6 and 9. The vertical indicates average value. The most interquartile values DPC collect about average value. The similar picture is watched in all areas of digits and does not depend on a type of a font. Therefore the supposition about existence of a normal distribution law DPC for digits will be logical depending on its of the size.

The identification of digits is determined by area in which one is concluded given DPC and the more so is interquartile than closer is she to average value.

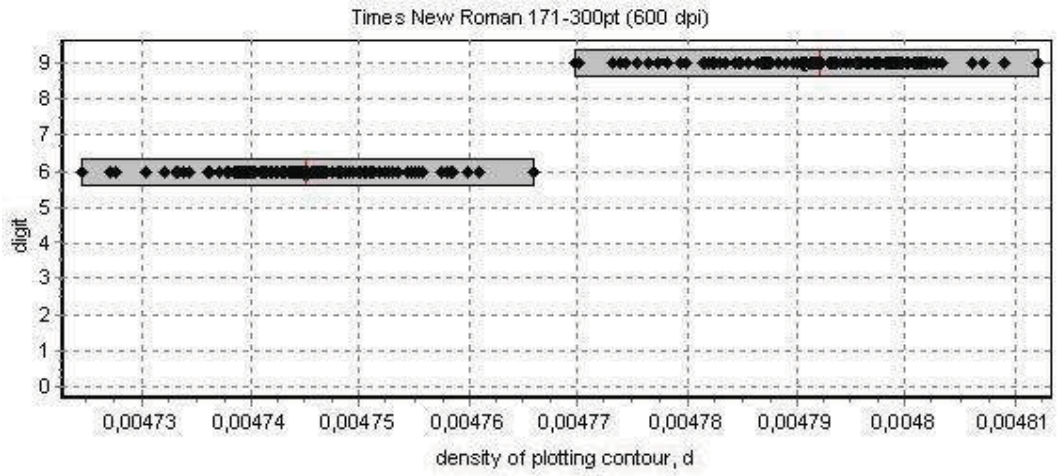


Figure 11 – Distribution DPC for a font Times New Roman 171-300pt (600 dpi), area of digits 6-9

**Conclusions**

1. In the reviewed fonts there is an area, behind which one the overlap of density of digits does not descend. Due to this the error-free identification of any digit is watched, but provided that it is known, what font will be used. Last descends that the relations between digits of a font on the miscellaneous sizes of its size are not formalized.

2. As a result of the analysis have received histograms separately on to each font, for all types of digits. It allows uniquely to identify unknowns fonts on availability as a minimum of 2 printed digits.

3. At the analysis of standard fonts such as, Arial, Courier New, Times New Roman the relations demonstrating distribution of density of digits of a font from the size of a size are obtained. For the sizes from 8 up to 14 points the least imposing of areas DPC was exhibited with a

font Times New Roman. For such class of fonts the unique recognition of digit from its density starts with the size more than 250 points.

4. At mesh sizes of a font size the identification of digit becomes complicated, since density of digits approaches, and it demands increase of sensitivity of the optical converter, besides, the errors of optical instruments today are great enough (up to 10 %, and it happens and above) because of existing technology of their manufacturing. Here it is necessary to remember possible defects (discard) of the map of digits at usage of standard printed means. At a static defect of printing and decreasing of the size of digits percent of an error DPC is augmented.

**The concluding.** The given analysis of fonts allows to work out the guidelines at the choice of fonts for recognition(identification) of the definite class of problems or documents. In particular for the reports for voting it is necessary to elect a font Times New Roman.

**LITERATURE**


1. Бардачов Ю.М., Ходаков В.Є., Шеховцев А.В., Шаганян С.М., Бараненко Р.В. «Вибори» - автоматизоване робоче місце працівника дільничної виборчої комісії. Вестник ХНТУ №1(24) 2006. стр. 10-14.
2. Фролов А., Фролов Г. Графический интерфейс GDI в Microsoft Windows. Том 14, М.: Диалог-МИФИ, 1994.
3. Калверт Чарльз Delphi 4. Энциклопедия пользователя: пер. с англ./ – К.: Издательство «ДиаСофт», 1998.
4. Вентцель Е.С., Овчаров Л.А. Прикладные задачи теории вероятностей. – М.: Радио и связь. 1983.
5 Роджерс Д., Адамс Дж. Математические основы компьютерной графики: Пер. с англ. – М.: Мир, 2001.