Method of Extended Objects Identification on Low-Contrast Images

D. Bushenko and R. Sadykhov

Belarusian State University of Informatics and Radioelectronics
P.Brovka st., 6, Minsk, 220013, Belarus
d.bushenko@gmail.com

Abstract

Nowadays there is a demand of intelligent systems which are able to perform identification of color low-contrast extended objects such as textile fibers. Since this is a complex problem with no common way to solve it, detailed investigations are needed here. The paper presents a new way to extract and identify the extended objects such as textile fibers. The most complex problem here is the crossed objects splitting. The paper presents a new criteria for the selected object assessment which allows splitting the crossed objects using well-known object extraction techniques such as "Fuzzy Select Tool" algorithm. Finally, the paper discusses the hue histograms correlation as a tool to compare two extended objects. These two algorithms make together a methods of identification of extended objects on low-contrast images.

1 Introduction

Identification is a one of the main tasks of artificial intelligence. Since there is still no general method of identification which fits all the needs, each particular problem needs its own approach. Obviously, identification of such extended objects as textile fibers also needs special customization of more common identification techniques. The task of identification of textile fibers arises mostly in criminalistics. That is why a special identification method is needed to be developed which will succeed in the task of identification of textile fibers.

Basically, the textile fibers identification is a task of finding the resembling textile fibers taken from the victim's clothes and the clothes of the suspected person (*Sadykhov, 2009*). An expert is supposed to process tens of films with textile fibers using a microscope. A single experiment usually contains 20-30 films, and this large amount of material results in long time of manual processing. Sometimes it takes weeks to find an appropriate fiber on these films. Because of the long terms of manual films processing, a digital system of textile fibers identification was developed.

The system utilizes the following workflow.

1. The user scans all the films with a business-class scanner (the resolution is around 2400dpi).
2. The system preprocesses all the images.
3. The expert finds a particular textile fiber on the images and pinpoints it.
4. The system finds all the resembling textile fibers on the other images and highlights them.

Searching the resembling images involves a complex method with specially modified algorithms which allow (*Bushenko, 2009*):

1. Extracting of the pinpointed textile fiber.
2. Calculating the descriptors of the textile fiber.
3. Fast searching through the images to obtain the points where the resembling textile fibers are expected to be (*Bushenko, 2008*).
4. Extracting the objects around the found points.
5. Comparing the extracted objects using the pattern specified on the step 1.

The process of extracting of the pinpointed textile fiber (as well as the fiber around the found point) is the most complex problem here. The point of interest may be founded on the following types of objects:

1) paper background which is mostly white;
2) different types of alien objects such as particles of sand;
3) halo around an object which appears due to some physical effects while scanning the patterns under the thin film;
4) a single textile fiber;
5) a group of crossed objects.

Altogether, the task of object extraction here is the task of selection of the pixels which belong only to a single textile fiber and skipping the pixels of halo, background and crossed objects. The task of identification here means comparing two objects utilizing some distance function.

2 Extracted objects assessment

Let us introduce the following attributes of the objects which allow extracting them as single extended objects:

a) width of hue histogram;
b) minimal mean deviation of hue histogram;
c) length of the skeleton;
d) number of skeleton crossings.

A single monochrome textile fiber has a very narrow hue histogram. Ideally it should have just one peak on a particular color value, but due to different distortions while scanning the film, the histogram becomes wider. A typical hue histogram of a single textile fiber is shown on the Fig. 1. Here and further the hue values range from 0 to 359.

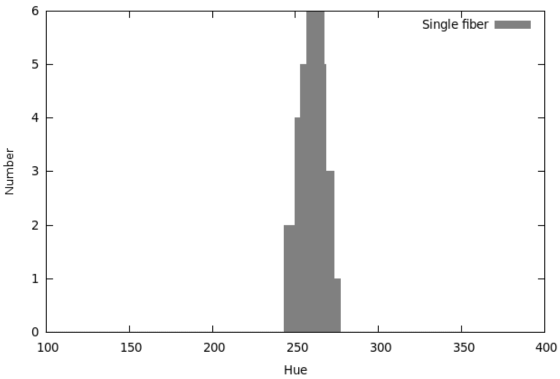

Figure 1: Hue histogram of a single textile fiber.

A single monochrome textile fiber extracted roughly has width about 10 pixels and includes the pixels which belong to the paper background, fiber's halo and the fiber itself. The typical histogram of such a fiber is shown on the Fig. 2.

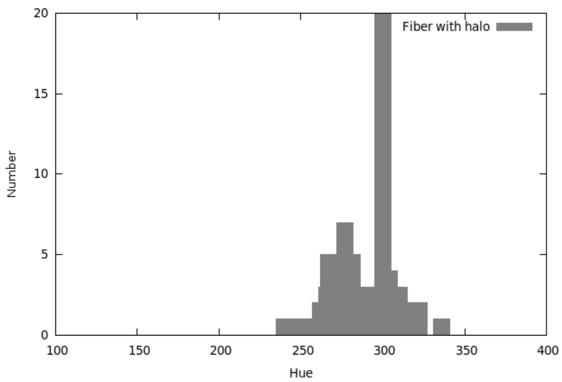

Figure 2: Hue histogram of a roughly extracted textile fiber.

The left area of the histogram between hue values (approximately) 240 and 290 belongs to the pixels of the background. The highest peak of the hue value 270 belongs to the fiber's halo, and the rest of the histogram belongs to the fiber itself.

While going ahead with the investigations of the hue histogram one can easily prove that the more non-uniform objects are shown on the image – the wider the histogram is. That is why the width of the hue histogram may be used as a part of the formal criteria for the object classification.

Let us agree that the width of the hue histogram (W(Hue)) is the quantity of hue values whose number value (on the N axis) is greater than zero.

$$W(Hue) = \sum_{i=0}^{359} \begin{cases} 1, if\ N(hue_i) > 0, \\ 0, otherwise \end{cases} \quad (1)$$

The next discussed attribute – hue value. It is a value in discrete finite cyclic circle of numbers without minimum and maximum values (*Blahut, 1989*). That is why it is impossible to calculate the mean value and thus – the variance of the hue histogram. But variance is a miningful characteristic which could be useful in the task of classification. As discussed in (*Sheldon, 2004*), it shows how a set of data values distributes around its mean. Instead of the variance and standard deviation it is proposed here to use a special parameter minimal mean deviation of hue histogram (d_{min}) as defined below.

$$d_j = \sum_{i=0}^{359} \left(\sqrt{(h_i - h_j)^2} \right) \cdot \frac{n_i}{N} \quad (2)$$

$$d_{min} = min\{d_0, d_1, \dots, d_{359}\} \quad (3)$$

Here h is the hue value in the interval [0; 359], n – its corresponding number value on the histogram.

When we are calculating the difference (h_i – h_j) one should bear in mind the cyclic nature of the hue set. Therefore consider expression:

$$(h_i - h_j) = \begin{cases} (h_i - h_j), if \sqrt{(h_i - h_j)^2} < 180 \\ 359 - (h_i - h_j), otherwise. \end{cases} \quad (4)$$

The most useful property of the minimal mean deviation is its ability to work correctly in discrete finite cyclic circle of numbers without minimum and maximum values.

Similarly to the parameter discussed above, it is possible to state here, that the lesser is the minimal mean deviation, the better the object was extracted.

The length of the skeleton is a parameter which cannot really help classifying the object because of wide variety of possible textile fiber sizes. But it is crucial for building the correct hue histogram which can describe the selected object. The common rule says: the longer the skeleton is the more accurate will be the calculated object's statistics.

The problem is in selecting minimal, optimum and maximum possible sizes of the textile fiber skeleton which will allow identifying the object. For instance, in (*Bendat*, *1986*) it is proposed to use selections of at least 25 objects; in (*Rumsey*, *2003*) the value 30 is mentioned but the proposed value is at least 100. While there are different possible lengths of textile fibers, usually the skeleton length around 50 is optimal because the longer value usually means undesired crossings with the other objects.

Altogether, let us assume that the optimum desired skeleton length is 50.

A single textile fiber usually does not cross itself, though it is possible in a general case. Basically, the textile fiber crosses other objects (another textile fibers or alien objects). Consequently, the fewer crossings the skeleton has, the better the object was extracted.

Having the discussed parameters one can develop a special rating function which will allow assessing the objects on how accurately they were extracted.

The desired rating function should:
1) raise when the skeleton length raises;
2) fall when the crossings number raises;
3) fall when the hue histogram width raises;
4) fall when the minimal mean deviation grows up.

There is no strict definition of such a rating function, each particular task may demand its own solution. In our experiments the most successful rating function looks as shown below.

$$R(obj_i) = \frac{a \cdot N_f}{\sqrt{b \cdot (W(Hue))^x + c \cdot (d_{min})^y + d \cdot (n_c)^z}}. \quad (5)$$

Here $R(obj_i)$ is the relative rating which shows how good is the object extracted; N_f is the fuzzy variable which depends on N – number of pixels forming the object skeleton; W(Hue) is the width of the hue histogram; dmin is the minimal mean deviation, n_c is the number of skeleton crossings; a, b, c, d, x, y, z – the coefficients which should be calibrated against each particular image type, scanner and its resolution. For example, in series of experiments with images of textile fibers acquired using the scanner Epson Perfection 3200 with the resolution 2400 dpi, it was found out that good quality is shown by the equation (5) with coefficients a = b = c = d = 1 and x = y = z = 3.

$$N_f = \begin{cases} N^2, if\ N < 50 \\ 50^2, otherwise, \end{cases} \quad (6)$$

$$n_{cf} = \begin{cases} \frac{10 n_c}{N}, if\ n_c > 10 \\ 0, otherwise, \end{cases} \quad (7)$$

$$R(obj_i) = \frac{N_f}{\sqrt{W(Hue)^3 + (d_{min})^3 + (n_{cf})^3}}. \quad (8)$$

3 Extended objects extraction

Correct object extraction is crucial for the quality of the objects identification. That is why existing object extraction algorithms are usually modified and optimized for special types of source data to achieve better quality of the extraction.

Let us discuss the algorithm for objects extraction named “Fuzzy Select Tool” in a well-known image manipulation program GIMP. As it is opensource software, anyone is allowed to download the source code and investigate it. The algorithm succeeds as follows.

1) The user sets up the allowed deviation (default value is 20).
2) The user pinpoints a pixel on the image.
3) The system acquires the specified pixel.
4) The system calculates the confidence interval using the value of the specified pixel and the allowed deviation:

$$\begin{aligned} R &\in [R_u - d; R_u + d] \\ G &\in [G_u - d; G_u + d] \\ B &\in [B_u - d; B_u + d], \end{aligned} \quad (9)$$

where (R_u, G_u, B_u) – red, green and blue values of the user specified pixel; d – allowed deviation.
5) The system selects all the pixels which are geometrically connected with the specified pixel and whose RGB-values fall into the confidence interval.

Consider the following input data (Fig. 3).

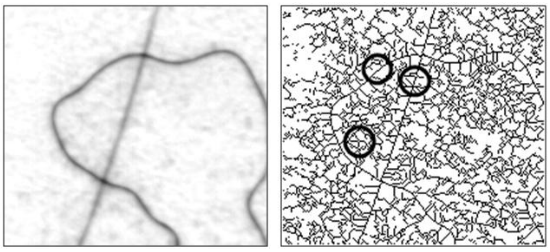

Figure 3: Source image with two crossed textile fibers of different colors (left) and its binarized halftone skeleton (right).

The left image of Fig. 3 contains two crossed textile fibers of blue and brown colors; the right image contains its binarized halftone skeleton. On the skeleton image there are three circles whose centers mark the points of interest. The bottom point of interest accidentally falls on the area between the fiber and the background. The other points of interest lay exactly on the fibers.

The results of the object extraction using the “Fuzzy Select Tool” algorithm with the allowed deviation set to 20 are shown on the Fig. 4.

The upper points result in the extracted skeleton branches which comparatively good correspond to the textile fibers. The only possible minus of the extracted objects could be some false branches of one or two pixels length which adjoin to the main branch. The bottom point of interest produces a complex fuzzy figure which is far from the textile fiber.

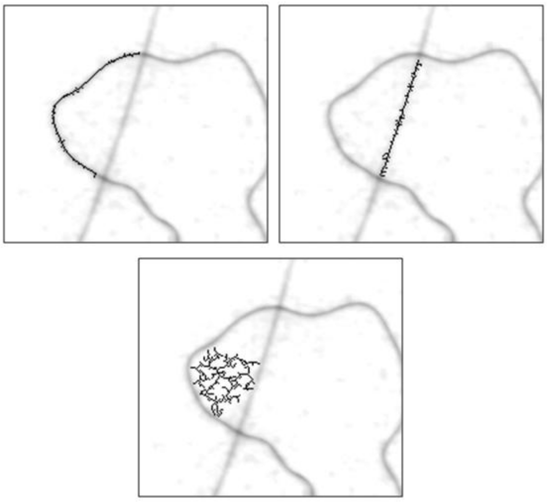

Figure 4: Object extraction using the “Fuzzy Select Tool” algorithm.

As it is discussed in the introduction, the case when the point of interest falls on the fiber’s halo or the background is a common one and occurs rather often. That is why, though the first two objects are extracted mostly correctly, the last one results in a false object which cannot be considered even as an extended object. For that reason the mentioned “Fuzzy Select Tool” algorithm should be enhanced.

The main drawback of the discussed “Fuzzy Select Tool” algorithm is in its unchangeable deviation. The idea of the algorithm modification is to find the optimum deviation value when the resulting object is extracted best. It succeeds as follows.

1. For each deviation value from the interval (default is from 2 to 25) select the particular deviation value.
a. Extract the object using the “Fuzzy Select Tool” algorithm.
b. Asses the extracted object using the criteria R(obj) discussed above.
2. From the set of acquired objects select the one with the maximum R(obj).

The results of the objects extraction using the “Adaptive Fuzzy Select Tool” algorithm are shown on the Fig. 5. Some parameters of the objects are shown in the Table 1.

Due to the quality rating R(obj) which shows how good the object was extracted, the deviation is selected for each object separately.

As it is shown in the Table 1, the last object has comparatively low rating (R(obj) = 100,21), showing that this object is extracted inaccurately. Such objects may be cleaned from the image using some threshold. The exact algorithm of how this threshold should be estimated is being investigated.

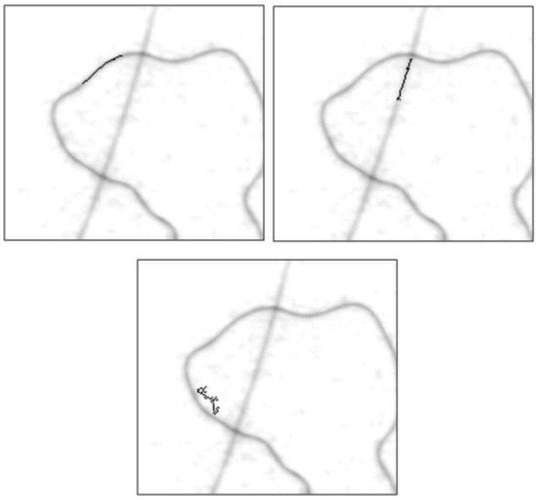

Figure 5: Object extraction using the “Adaptive Fuzzy Select Tool” algorithm.

Table 1: Parameters of the objects extracted with the “Adaptive Fuzzy Select Tool” algorithm.

Object	Deviation	R(obj)
1	13	286,00
2	6	402,95
3	9	100,21

In other words, the proposed “Adaptive Fuzzy Select Tool” algorithm is more agile than its ancestor. It is able to select the deviation for each particular object trying to extract it as a single extended object. If the extracted object has too low rating R(obj), it may be skipped using threshold on the R(obj) value.

4 Extended objects comparing

There are numerous ways to compare two objects. But the identification of textile fibers demands a distance function which could accurately show how close the two objects are.

It is proposed here to use correlation of the hue histograms as the distance function between two textile fibers (*Franchi, 2006*).

$$r(X,Y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad (10)$$

Fig. 6 – 8 compare the hue histograms of different textile fibers and table 2 contains their parameters and the correlation value of each object with the object 1.

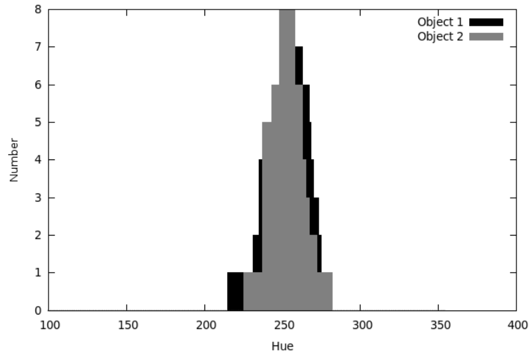

Figure 6: Hue histograms of the textile fibers 1(black) and 2 (grey).

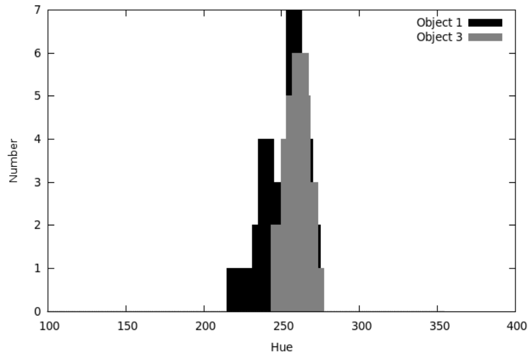

Figure 7: Hue histograms of the textile fibers 1(black) and 3 (grey).

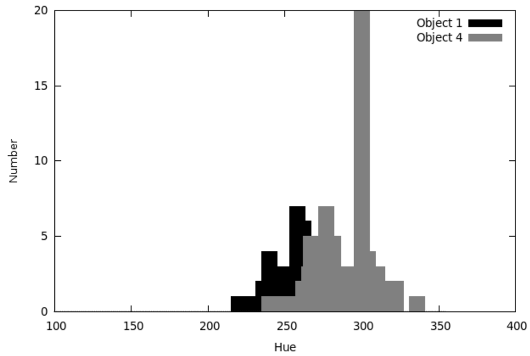

Figure 8: Hue histograms of the textile fibers 1(black) and 4 (grey)

Table 2: Hue histograms correlation.

Object	Mean	Deviation	Correlation
1 (etalon)	254	11	1,00
2	260	5	0,84
3	252	8	0,56
4	286	17	0,09

The hue histogram correlation fits our needs as it is invariant to any geometric deformations of the extended objects; also it is invariant to the objects' skeletons lengths. It shows the difference between two objects really accurately and responds only to the actual colors of the objects.

4 Conclusion

There are different areas where the automatic identification of the extended objects could be applied. One of these areas is criminalistics which deals with the identification of textile fibers. When automating this complex process one should cope with two problems: extended objects extraction and the objects identification. As there are no common algorithms which fit all the needs, special modifications were developed. The paper presents a new algorithm "Adaptive Fuzzy Select Tool" which can extract the textile fiber from the low-contrast non-uniform background and split it from the other fibers. Finally, the paper presents the way of comparing two extended objects using their hue histograms correlation. The discussed algorithms are implemented in the digital system of textile fibers identification (*Sadykhov, 2009*).

Reference

Blahut R. (1989), *Fast Algorithms for Digital Signal Processing*, Addison-Wesley, Wokingham.

Bendat J., Piersol A. (1986), *Random Data Analysis and Measurement Procedures*, John Willey & Sons, New York.

Bushenko D., Sadykhov R. (2008), "Segmentation of Extensive Objects on Low-Contrast Images", ICNNAI'2008, Minsk, Belarus, 2008, p.182-185.Sadykhov R., Bushenko D. (2009), "Digital System of Textile Fibers Identification", In proc. of MIPRO'2009, Opatija, Croatia, pp. 297-302.

Bushenko D., Sadykhov R. (2009), "Towards Near-Realtime Identification of Extended Objects in Low-contrast Images", In proc. of IMCSIT'2009, Mragowo, Poland, pp. 25-32.

Fanchi J. (2006), *Math Refresher for Scientists and Engineers*, 3d Edn., Willey-Interscience, Hoboken.

Rumsey D. (2003), *Statistics for Dummies*, Willey Publishing, Indianapolis.

Sadykhov R., Bushenko D. (2009), "Digital System of Textile Fibers Identification", MIPRO'2009, Opatija, Croatia, pp. 297-302.

Sheldon R. (2004), *Introduction to Probability and Statistics for Engineers and Scientists*, 3d edn., Elsevier Academic Press, Burlington.