PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Method for dating old handwritten manuscripts based on spectral photometry of ink in near infrared range

Igor Gurov
Mikhail Volkov
Konstantin Barsht
Svetlana Berezkina

SPIE.

Method for dating old handwritten manuscripts based on spectral photometry of ink in near infrared range

Igor Gurov*[a], Mikhail Volkov[a], Konstantin Barsht[b], Svetlana Berezkina[b]
[a]ITMO University, 49 Kronverksky ave., Saint Petersburg, 197101, Russia; [b]The Institute of Russian Literature (the Pushkin House) of the RAS, Makarov Embankment 4, Saint Petersburg, 195426, Russia

ABSTRACT

Several problems of spectral photometry analysis of handwritten documents have been studied like combined reflection from the both ink and paper material impregnated by ink, the illumination spatial non-uniformity with small light intensity, flexibility and instability of paper sheets. The images acquired at different wavelengths were spatially matched using developed computer algorithm and then regions of interests, i.e. ink lines were extracted. Then reflection spectra obtained at all pixels of ink lines were composed in vector representation, and comparison of the vectors was conducted. As the result, degree of correspondence between different fragments of text has been evaluated. The optical setup and its calibration procedures are considered in detail as well as results of image processing, matching, segmentation and spectral photometry evaluation of ink. Experimental results of different text fragments matching and estimates of text fragments spectral correspondence degree are presented and discussed.

Keywords: Dating of old manuscripts, F. Dostoevsky's handwritten manuscripts, near infrared radiation, image matching, ink reflectance, spectral photometry.

1. INTRODUCTION

Dating of old manuscripts is often important to evaluate accurate date of a document appearance. There exist several attributes allowing to provide dating like peculiarities of paper material and properties of ink. Properties of paper material relate to corresponding epoch, but accurate dating is problematic in this way. More accurate date estimate can be based on comparison of documents of known dates with the ones that have to be dated. To achieve this, it is possible to use the fact that ink chemical composition was essentially variable in the past due to handicraft manufacturing process with variable properties (see, e.g.,[1]). Assuming that both the dated and evaluated documents had been written using an ink from the same batch, it is possible to find their time correspondence. The problem consists in how to characterize ink chemical composition accurately. Additional difficulty consists in that often not allowed to damage a heritage object, and non-destructive methods only are applicable. Optical methods present such a possibility.

The considered approach had been suggested[2] to date more accurate the handwritten manuscripts by famous Russian writer F.M. Dostoevsky. There are known a few post letters written and accurately dated by the writer, but other handwritten documents were not dated. It is important to compare these documents with the letters to evaluate accurate dates of the former documents. For this purpose, spectral photometry method has been applied[2, 3]. The handwritten documents were illuminated by light in near infrared range at a few different wavelengths where spectral properties of ink are more noticeable, and reflected light images recorded by high resolving photo camera.

There have been found several problems of spectral photometry analysis of handwritten documents: combined reflection from the both ink and paper material impregnated by ink, the need in compensation of the rest illumination spatial non-uniformity with small light intensity, flexibility and instability of paper sheets (any mechanical contact like cover glass is not allowed) when switching light wavelengths. It is why the images acquired at different wavelengths were spatially matched using special computer software and then regions of interests, i.e. ink lines were extracted. Then reflection spectra obtained at all pixels of ink lines were composed in vector representation and comparison of the vectors was conducted.

In the paper, the optical setup and its calibration procedures are considered in detail as well as results of image processing, matching, segmentation and spectral photometry evaluation of ink. Results of different text fragments matching and degree of their correspondence are presented and discussed.

*gurov@mail.ifmo.ru; phone +7 812 571 6532; fax +7 812 315 7534

Optics for Arts, Architecture, and Archaeology VI, edited by Luca Pezzati,
Piotr Targowski, Proc. of SPIE Vol. 10331, 103310H · © 2017 SPIE
CCC code: 0277-786X/17/$18 · doi: 10.1117/12.2270089

2. MATERIAL AND METHODS

2.1 Optical setup

To evaluate reflectance spectra of ink, a sample under study is illuminated by radiation at several wavelengths in near infrared range. The images in reflected light recorded at each wavelength by high resolving digital video camera. Combined processing of the recorded set of video frames allows to extract useful information about local spectrum at each pixel of video camera thus giving a possibility to evaluate and compare spectra of different text fragments.

Fig. 1 shows schematic diagram of the custom made setup used in the research.

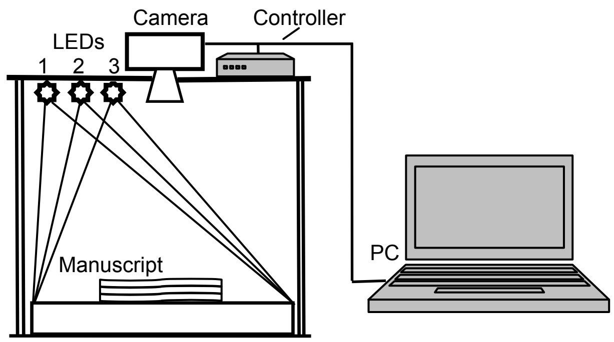


Figure 1. Experimental setup diagram. LEDs: NIR light emitting diodes.

The setup is based on a folding load-bearing framework equipped with digital video camera and light source including three light emitting diodes (LEDs) placed at upper part, whereas a manuscript under investigation is put at lower board. Digital video camera model NIKON D3100 (with removed infrared filter) has a CMOS light sensitive array with the sensitivity ISO 3200 and resolution 4608×3072 pixels. The LEDs illuminate a manuscript sequentially at the wavelengths 850, 880 и 950 nm being switched by a control unit.

The experimental setup satisfies all the requirements to work in archive rooms in autonomy, mobility and safety for documents. In particular, the radiation intensity is small and does not exceed in equivalent light power an ordinary thermal light source (illumination lamp) of 40W placed at the distance of 1 m from illuminated document.

2.2 Procedures of image processing

Image processing aimed to define parameters of the reflectance spectra with subsequent evaluation of ink properties correspondence within different manuscript fragments. To provide a possibility to compare different fragments, it was necessary first to correct spatial non-uniformity of illumination from each LED. Then the images recorded at different wavelengths were spatially matched to extract wavelength-dependent local reflectance at each video camera pixel. Finally, it was conducted the setup spectral calibration taking into account partial transparency of ink, such so obtained local spectrum depends also on reflection from a subsurface paper layer. Such local paper reflection component has to be removed. The above mentioned procedures are considered in detail below.

The non-uniformity of illumination can be determined using the fact that handwritten lines are much narrower in width with respect to the field of text. Besides of this, a non-uniformity of illumination varies slowly. This enables to apply local image modification under which the handwritten lines disappear, and one obtains the intensity distribution caused by a LED only. This estimate of illumination non-uniformity at each wavelength allows then to normalize image intensity that is necessary for subsequent accurate spatial matching of images.

There was not allowed to use any cover glass and mechanical contact with a manuscript sheet. It has been found mutual spatial shifts of images in the range about 2 ... 10 pixels caused by instability of a paper sheet position because of influence of external factors (like micro vibrations and small air flow). Due to short frame acquisition time the images are not blurred, but have a little bit different spatial positions as it is illustrated in Fig. 2.

It is seen that images in Figs. 2, a-c have different brightness and non-uniformity of illumination. The image in Fig. 2, c has small relative displacement in vertical and horizontal axes and a little bit rotated.

Full-frame image matching algorithm based on the criterion of minimum root mean square (RMS) intensity deviation was utilized to compensate image displacements.

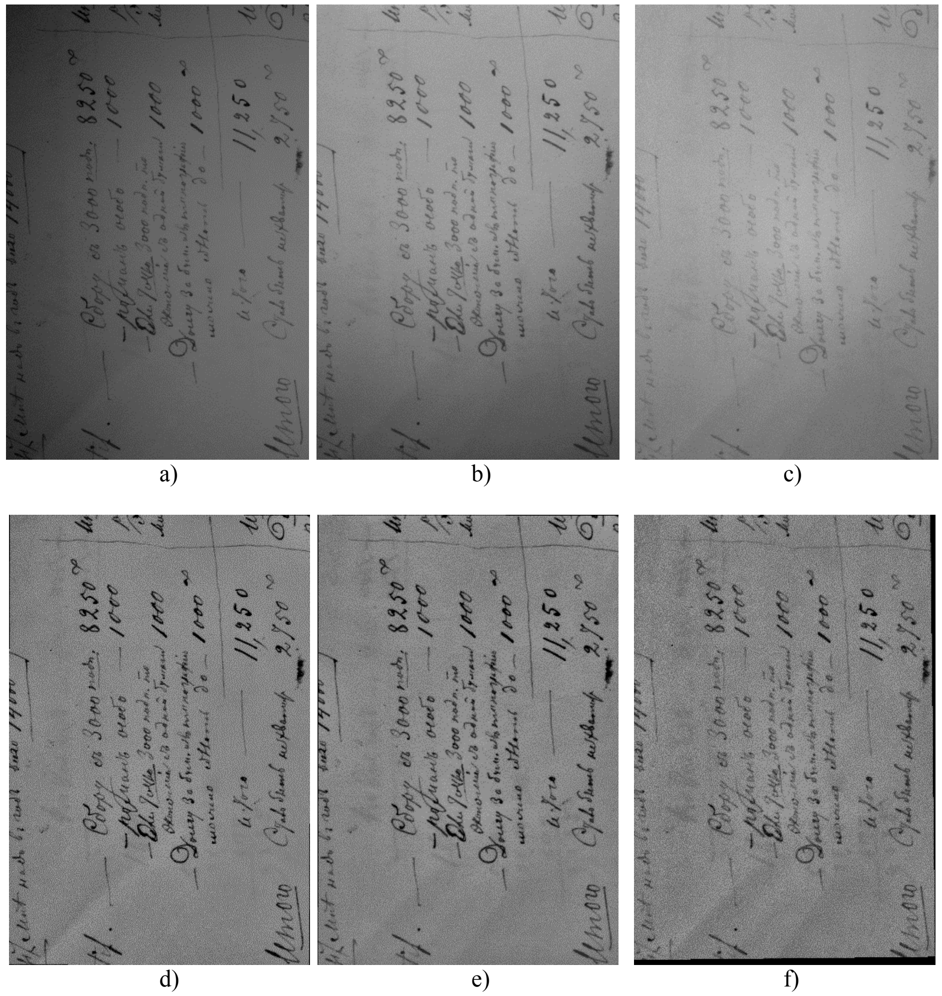


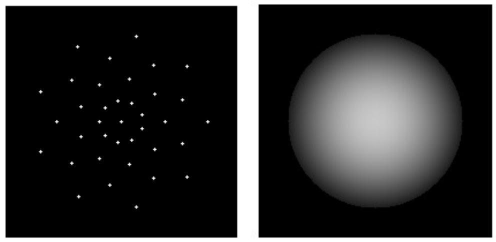
Figure 2. (a)–(c) Images of text fragments recorded at different wave lengths and (d)–(f) compensation of illumination non-uniformity and spatial displacements of images (displacements are clearly seen in figure (f)).

The conducted research shown partial transparence of ink lines in near infrared range. The evaluated reflectance degree depends on partial reflectance from paper material impregnated by ink varying in lateral directions of a field of view. The change is caused particularly by variable concentration of ink at different points. It is necessary to separate reflectance from ink and from paper material to extract information about pure ink reflectance. The problem is complicated by variable ink concentration and its variability over sheet area.

To solve this problem, it has been supposed that the reflection coefficients from ink and from paper are combined additively. Due to slow variation of paper reflectance in small area around ink line, it becomes possible to involve a paper small area near ink line to evaluate light reflection. However this procedure has to be applied locally because of variability of paper properties as well as ink lines within a paper sheet.

The developed algorithm is based on formation of a mask around ink lines and involves adaptive image modification applied iteratively thus providing extrapolation of paper reflectance under an ink line. Fig. 3 represents an example of local discrete averaging mask and weighting function to introduce local image modification. The mask size is changed iteratively and it is applied locally around each image pixel. The adaptation criterion is based on calculation of relative number of dark pixels with respect to the number of brighter pixels within the mask. When this ratio becomes lower than a threshold value, the mask size is fixed. Thus the degree of image modification is dependent on local image variability.

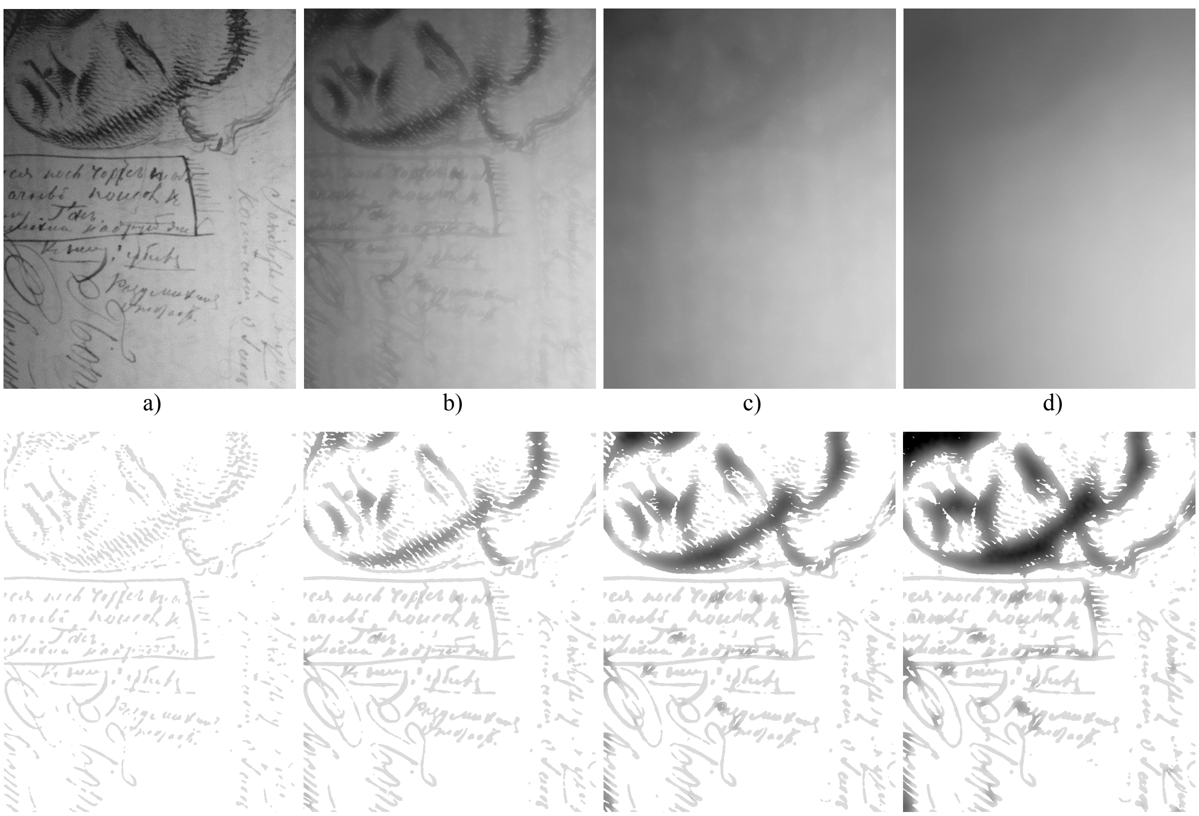
As the result, the algorithm removes sharp image lines, but does not change slowly varying component of an image intensity distribution determined by paper reflectance nearby a former place of ink line.



a) b)

Figure 3. (a) An example of the mask for selecting pixels within a local image area and (b) pixels weighting function represented as grey scale map.

Iterative algorithm removes sharp ink lines by image modification adopted to local fragment of the text as it is illustrated in Fig. 4. This enables to evaluate partial reflectance of paper material extrapolated under ink line. An example of the extrapolation is illustrated in Fig. 5.



a) b) c) d)

e) f) g) h)

Figure 4. Extrapolation of reflected light intensity distribution under ink lines. (a)–(d) Adaptive local removal of ink lines at 1, 3, 7 and 20th iteration step, respectively; (e)–(h) corresponding local mask size represented by grey scale maps: the darker a point is, the larger is the mask size related to this point.

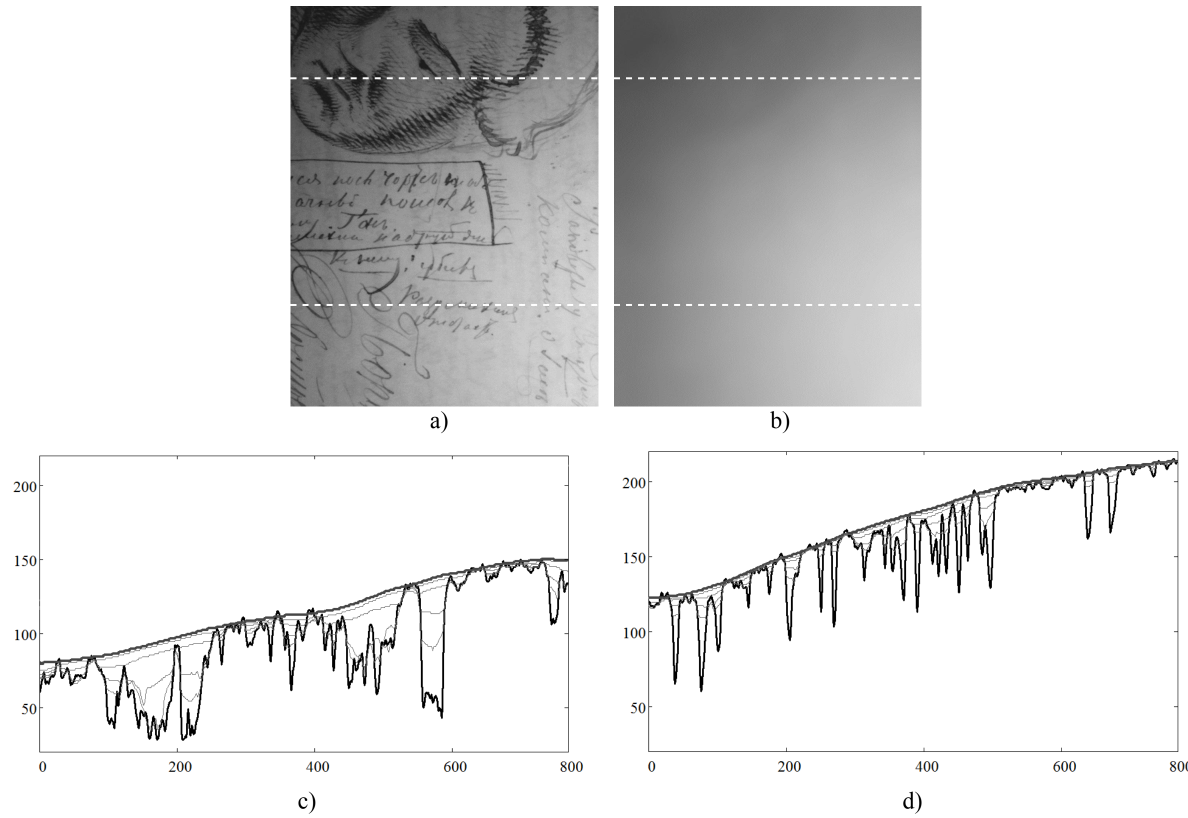


Figure 5. (a) Initial image, (b) the same image with removed ink lines and (c)-(d) estimates of ink reflection degree (expressed in 8-bit representation) along dotted lines marked in figure (a). In figures (c), (d), upper curves illustrate reflection from paper material. Thin grey lines relate to image processing iteration steps (from top to bottom).

It is worth noting that the described above adaptive nonlinear algorithm differs from ordinary image blur due to in the former case the ink lines are removed without any influence to reflectance at adjacent points of paper material that takes place in the later case. This peculiarity is important from the viewpoint of system calibration for the subsequent spectral estimates of ink reflectivity.

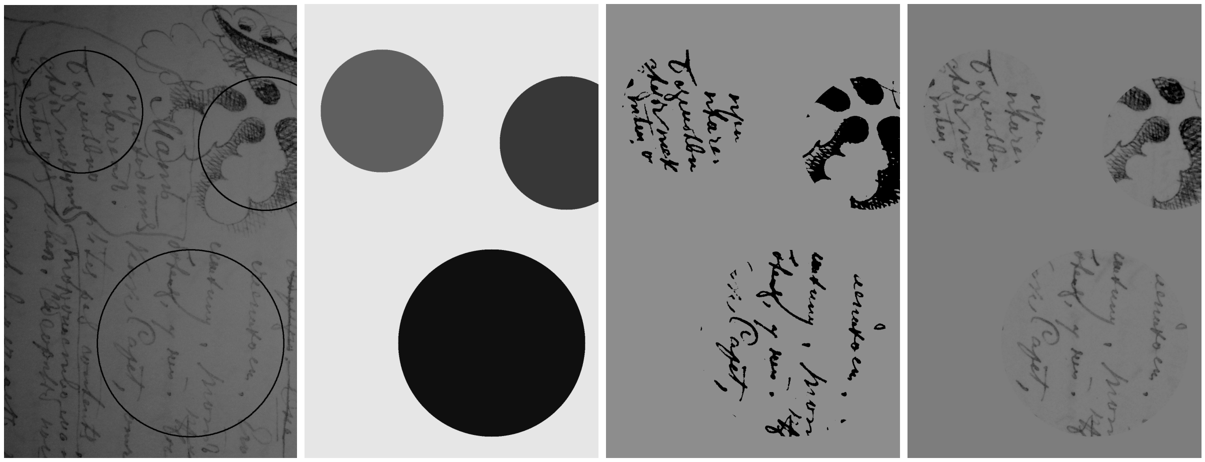
As it is seen in Fig. 5 the extrapolation yields correct estimates of the local reflection degree from paper material taking into account spatial non-uniformity of reflection. The extrapolation is applied to all the images recorded at different wave lengths. Removal of spatially extrapolated paper reflection from normalized image gives ink lines image. The procedures considered above provide local calibration to suppress influence of external disturbing factors and to obtain information related to pure ink reflection.

3. SPECTRAL PHOTOMETRY RESULTS

The following procedures are focused to compare ink spectral reflection originated at different text fragments and to present this information in the form convenient for subsequent textual criticism conducted by literary critic specialists.

Typical image of a sheet is comparably large and can include the parts where lines written by different ink. To extract regions of interest determined under assumption of uniform ink properties, an image is multiplied by spatial masks (see Fig. 6, a-b), within which the degree of ink reflection at different wavelengths is evaluated.

Fig. 6, a shows an image recorded under illumination at a particular wavelength. The masks in Fig. 6, b index areas by grey scale level to process them separately. Within a mask the image intensity is binarized to extract ink lines (see Fig. 6, c). The lines areas determine location of lines where reflected light is analyzed at different wave lengths. Fig. 6, d represents total light intensity as a sum of recorded lines at three different wave lengths, i.e. averaged intensity corresponding to the ink lines.



a) b) c) d)

Figure 6. (a) Initial text part, (b) regions of interest defined by masks and marked by grey scale level inside a mask, (c) extracted text parts and their binarization, (d) recovered ink lines intensity extracted by the binary masks (figure (c)) after removal of reflection by paper material.

The ink lines intensity distribution at each of the three wavelength is demonstrated in Fig. 7. Relation of intensities at the same fragment has to be constant characterizing a kind of ink. As one can see in Fig. 7 the algorithm works well due to taking into account the influence and then subsequent removal of local variations of ink impregnation depth into paper material with recovery of intensity in middle line parts as well as near the borders of ink lines, i.e. the intensity is almost identical for all the pixels within a fragment at each wave length. The final result is calculated as averaged value over ink points shown in each figure.

a) b) c)

Figure 7. Equalized ink lines intensities at three different wave lengths.

The set of ink intensities into a text fragment j can be expressed by the vector $\mathbf{r}_j = (r_{j1}\ r_{j2}\ r_{j3})^{\mathrm{T}}$, where numerical indices $k = 1, 2, 3$ relate to the wave lengths set. Evaluation of fragments spectral correspondence is based on calculation of Euclidian distance between vectors $\mathbf{r}_j$ and $\mathbf{r}_i$, namely $d_{ij} = \left[\sum_{k=1}^{3} (r_{ik} - r_{jk})^2 \right]^{1/2}$.

There were analyzed 52 text fragments, and their mutual spectral correspondence is expressed by the diagram shown in Fig. 8. The darker a matrix element is, the less is spectral difference between corresponding text fragments. It is seen in Fig. 8 that there variable correspondence is observed, and an additional separation rule has to be applied, for instance, based of threshold decision criterion.

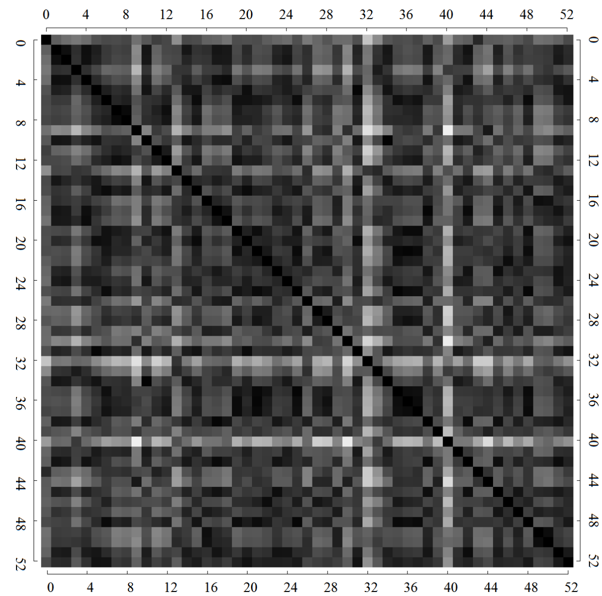
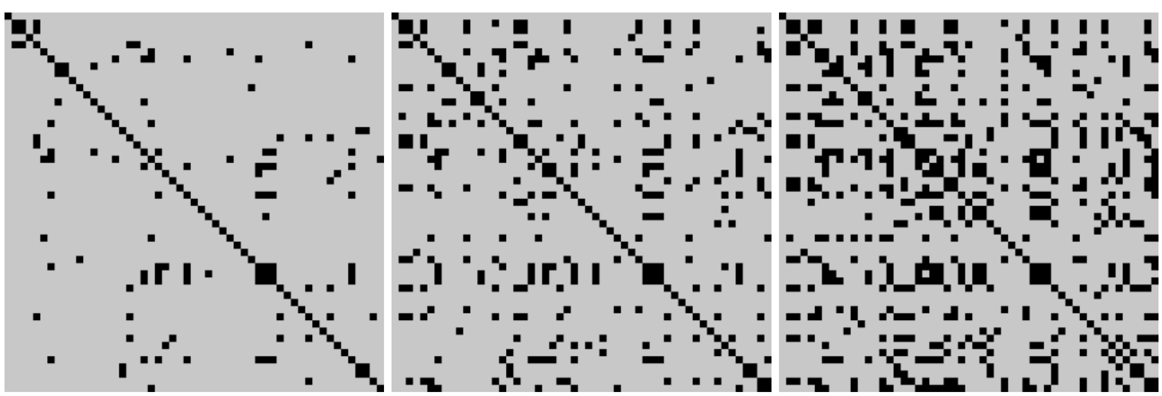


Figure 8. Spectral correspondence diagram obtained for 52 text fragments under evaluation. Degree of fragments difference shown by grey scale representation.

Examples of correspondence diagrams obtained under different threshold value applied to the matrix shown in Fig. 8 are illustrated in Fig. 9. If one sets low threshold for admissible difference between fragments, only small number of fragments can be classified as similar to each other (see Fig. 9, a). When increasing the threshold, more number of fragments appear interpreted as similar (Figs. 9, b–c). The decision, whether fragments really correspond to each other is inferred under additional textual analysis conducted by literary critic specialists.



a) b) c)

Figure 9. Spectral correspondence diagrams extracted from the diagram in Fig. 8 under different threshold value.

4. DISCUSSION AND CONCLUSIONS

Old handwritten manuscripts are very complicated objects by their nature. Written lines vary significantly in style, width and local ink concentration. Ink impregnation depth in paper material is changeable as well as local properties of paper. The problem is that reference samples of original ink and paper materials are not available, and all reference and useful information has to be extracted from original sample of a manuscript. Spectral photometry method applied in this research developed under strong limits. There was impossible to employ high power light sources to provide high signal-to-noise ratio as well as any contact mechanical means to stabilize the object position during observation. It is why our approach was focused to accurate calibration of the system by applying original image processing algorithms to decrease influence of non-informative factors and to extract useful information at sequential processing steps.

The handwritten documents were illuminated by light in near infrared range at a few different wavelengths where spectral properties of ink are more noticeable. The first procedure of matching images recorded at different wave lengths required accurate calibration of each light source (LED) spatial intensity distribution without which accurate matching with sub-pixel accuracy is not possible. The principal point is to separate pure ink reflection and local reflection of paper material. The developed iterative nonlinear algorithm adopted to local properties of ink line and paper reflectance allowed obtaining information about pure ink reflection that provided necessary conditions to utilize subsequent spectral photometry method.

The procedures mentioned above need in processing of huge amount of video information because of recording many images with high resolution at a few wave lengths. The most computational power consuming procedures are image matching and separation of ink and subsurface paper material reflection. The problem was solved by application of parallel computational algorithms and software.

The main result of the research is expressed by equalized images of written ink lines related to different wave lengths (Fig. 7). These results present the basis for spectral photometry analysis and correct comparison of text fragments. As a result, there have been evaluated probabilities of correspondence between different fragments of text that allowed to enhance dating of several fragments[3] related to the famous book “Crime and punishment” written by F.M. Dostoevsky (1866).

ACKNOWLEDGEMENT

The research was partially supported by Russian Foundation for Basic Research, grant 14-06-00183.

REFERENCES

[1] Davids, T., [The History of Ink, Including its Etymology, Chemistry, and Bibliography], T. Davids & Co, New York (1860). Indiana: Repressed Publishing LLC (2012).

[2] Barsht, K.A., Rajhel, B.S., Sokolova, T.S., “A method of digital spectrophotometry in the study of a writer's manuscripts (on the example of “Siberian Notebook” by Fyodor Dostoyevsky)”, Bulletin of the Russian Academy of Sciences: Studies in Language and Literature 71 (4), 20–43 (2012).

[3] Barsht, K.A., Beryozkina, S.V., Volkov, M.V., Gurov, I.P., “Dating F.M. Dostoevsky's manuscripts by use of spectrophotometric method of ink analysis in infrared light”, Bulletin of the Russian Academy of Sciences: Studies in Language and Literature 75 (5), 48–61 (2016).

[4] Karimov, K.A., Volkov, M.V., “The phase correlation algorithm for stabilization of capillary blood flow video frames”, Proc. SPIE 9528, 952810 (2015).