ICDAR 2015 MultiSpectral Text Extraction Contest (MS-TEx 2015)

Rachid Hedjam∗†, Hossein Ziaei Nafchi ∗, Reza Farrahi Moghaddam∗, Margaret Kalacska† and Mohamed Cheriet∗

∗Department of Automated Manufacturing Engineering, ÉTS, University of Québec
1100, Notre-Dame Street West, Montreal, Quebec, Canada H3C 1K3
Email: rachid.hedjam@mcgill.ca, hossein.zi@synchromedia.ca, imriss@ieee.org, mohamed.cheriet@etsmtl.ca
†Department of Geography, McGill University,
805 Sherbrooke Street West, Montreal, QC, Canada H3A 2K6
Email: {rachid.hedjam, margaret.kalacska}@mcgill.ca

I. INTRODUCTION

In recent years, Multispectral (MS) imaging becomes a very important tool for historical document analysis. This technique is widely known as a non-invasive method of investigation thanks to its simultaneous use of ultraviolet, infrared and visible light. It enables conservators and art historians to obtain valuable information on ancient documents without causing any physical damages to the materials, and makes it possible to reveal the newly overwritten text, to distinguish and recognize the chemical material composing the ink, to detect signs of degradation in historical documents. It can also help to extract information from cultural heritage patterns which cannot be extracted using conventional color photography. Extracting (segmenting) the original text (old writing) from MS document image is a very important step for subsequent document image analysis and investigation. In order to facilitate comparison of the results of different algorithms and track their progress over time towards a level of human performance, it would be of great interest to obtain a standard benchmark and accurate ground-truth with the most representative information about the targeted samples. To this end, we have generated a dataset of 21 MS document images for training purposes 1 (see the URL bellow) [1], [2], and another dataset of ten MS document images for testing purposes.

II. DATA: ACQUISITION, CALIBRATION AND REGISTRATION

The datasets are generated from a set of historical manuscripts written between the 17th and 20th centuries, collected by Archives of Quebec (Canada). The documents are handwritten with iron gall-based (ferro-gallic) ink made from salts and tannic acid from vegetable sources [3], [4]. It was the standard writing and drawing ink from about the 12th century to the 19th century, and remained in use well into the 20th century. The MS document images have been recorded using the Synchromedia’s MS imaging system, which is composed of a CCD sensor [model Chroma X3 KAF 6303E (Kodak), with high quantum efficiency of 1100nm and resolution of 3072x2048 (6 Mega) pixels of 9x9 microns], and uses a set of eight (8) chromatic filters motorized and controlled by the software of the camera. Each chromatic filter acts as a band pass filter to produce a band image at a specific wavelength (see Table IV). The set of collected bands constitute the so-called MS information cube (see Fig. 2), which contains one spectral reflectance (or spectral signature) for each pixel. Briefly, the setup of our MS imaging system is shown schematically in Fig. 1, where the document under investigation is placed on a support facing the camera from above. Two tunable lamps are used to illuminate the document, which are usually positioned at a 45° angle, on the upper-left and upper-right sides of the camera, and close to the document, which allows to produce an ideal diffuse illumination where each area of the document is uniformly illuminated from all directions. The MS imaging is entirely enclosed in a light-proof cabinet, in order to prevent any stray light from external sources affecting the measurements.

Calibration the recorder spectral images must be done firstly before any subsequent processing to obtain the true spectral reflectance values for each pixel location on the document. This can be achieved by recording two frame images under the same recording conditions (i.e., same parameters): dark frame image (D) and white reference frame (S). The dark frame is measured by capturing a dark image by blocking the camera by using a cap; while the white reference frame is characterized simply by acquiring a light image of a homogeneous white surface. The spectral reflectance value at each location pixel s in a specific wavelength λ is given by [5], [6]:

\[ u(s, \lambda) = \mu(\lambda) S(s, \lambda) \]

where \( O \) is the recorded...
imperfect spectral image, \( t_o \) and \( t_s \) are the exposure times used respectively for the spectral image of the object (document) and the spectral image of the white reference frame. Experimentally, \( t_o = t_s \). \( \mu(\lambda) \) is the mean intensity value of the white reference frame. In addition, in order to compensate the problem of the geometric property difference between the different filters, the bands are aligned and registered using i2LAlign software, developed by DualAlign LLC.

III. DATA FORMULATION AND ANALYSIS

Mathematically speaking, a MS image (as shown in Figure 2 (a)) is described as follows: \( u(s) : \Omega \subset \mathbb{R}^2 \to \mathbb{R}^d \), where \( s = [x, y]^T \in \Omega \), and \( \Omega \) is the domain of the MS image. Each pixel \( s \) is characterized by \( d \) independent spectral values (or spectral reflectance denoted by \( u(s) \)) represented by a vector of \( d \) components (for our MS imaging system \( d = 8 \)). The MS data set is then embedded in an \( d \)-dimensional vector space and the spectral signature of a pixel corresponds to a particular location in this space. The spectral signature (as shown in Figure 2 (b)) of a pixel \( s \) can be denoted: \( u(s) = (u_1(s), u_2(s), \ldots, u_d(s)) \) and then used as a feature vector in subsequent treatment procedures.

One interesting characteristic of the documents collected from the BAnQ (Bibliothèque et Archive National de Québec) is that the main (original) text is written with ink based on iron gall material and does not contain much carbon. Carbon absorbs IR radiation and reflects UV radiation, while iron transmits IR radiation, which is then reflected by the underlying constituents. This explains why the main text disappears from the spectral images when we move towards longer wavelengths (see Figure 3). In other words, the IR region makes iron gall ink transparent. In contrast, the annotations are written with pencil containing carbon, which allows them to be visible in the IR regions.

Now, we aim to explain briefly the spectral reflectance of the main element to be studied which is the original ink used for writing. The spectral analysis of the data provide the reader with important information about the different document image objects. Fig. 4 shows the intensity distribution of three samples of pixels picked manually from three different objects: original text, stamp, and background. The histogram of each sample at different wavelengths represents a tonal distribution, which can then be used as a feature for object separation.

Even though the text is clearly differentiable from the other objects as shown in Fig. 4 for example, this alone is not sufficient to make a definitive determination. One of the main reasons for the instability of the results is that the photometric properties of the original text are influenced by the volumetric concentration of the ink flows in each text zone [7]. Whenever the volumetric concentration is not uniform throughout the text, the distribution of its pixel intensities is multi-modal, and vice versa, as illustrated by Fig. 5 (2nd row).

IV. RAW DATA, ORGANIZATION AND SPECIFICATION

The training dataset is composed of 21 folders. The name of each folder is composed of the letter “z” and a
numeral number (e.g., “z97”). Each folder contains one MS document image with 8 spectral bands and one ground-truth image in the binary format with this name template: [folder-name][GT.png] (e.g., “z97GT.png”). The eight spectral bands are named as follows:

<table>
<thead>
<tr>
<th>Image</th>
<th>Wavelength(nm)</th>
<th>Illumination</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1s.png</td>
<td>340</td>
<td>UV</td>
</tr>
<tr>
<td>F2s.png</td>
<td>500</td>
<td>Visible 1</td>
</tr>
<tr>
<td>F3s.png</td>
<td>600</td>
<td>Visible 2</td>
</tr>
<tr>
<td>F4s.png</td>
<td>700</td>
<td>Visible 3</td>
</tr>
<tr>
<td>F5s.png</td>
<td>800</td>
<td>IR 1</td>
</tr>
<tr>
<td>F6s.png</td>
<td>900</td>
<td>IR 2</td>
</tr>
<tr>
<td>F7s.png</td>
<td>1000</td>
<td>IR 3</td>
</tr>
<tr>
<td>F8s.png</td>
<td>1100</td>
<td>IR 4</td>
</tr>
</tbody>
</table>

Generally speaking, a MS document image (in our dataset) may contain the following classes (see Fig. 6):

1) Original text (OT);
2) Annotations (AN);
3) Stamps (ST);
4) Degradation (DE);
5) Background (BG).

where OT and BG are present in all the document images. AN and ST are not always present. DE is always present in all the document images but with different degree of severity. It may contain several types of degradation that, virtually, cannot be enumerated.

V. GROUND-TRUTH

Each ground-truth is a binary image. It is defined by black pixels, which correspond to the class OT, and white pixels, which cover the class BG and all other possible classes (AN, ST, DE, BG; see Fig. 7). The protocol of generating the ground-truth follows the method introduced in [8], [9] and consists of two main steps. First, images in different bands are processed and a rough binary image is produced. This rough binary image is then manually modified to generate final ground-truth.

VI. OBJECTIVE OF THE CONTEST

The objective is to extract only the class OT (original text) from the input MS document image. In our context, the term extracting can refer to segmentation, isolation, separation or retrieval. Therefore, there is no restriction on the proposed methods used for the original text extraction. Any kind of classification, binarization, image fusion, source separation, segmentation methods, etc. can be used. Accordingly, the classes BG, DE, ST and AN must be combined as one class, BG+, at the output of the submitted algorithm. The full description of the contest and the call for participation is available at http://www.synchromedia.ca/competition/ICDAR/mstexicdar2015.html.

VII. SUBMITTED EXE FILE

Each participant is required to submit one exe file (named MSTEX for Multispectral Text Extraction) of his (her) proposed method, which accepts as input a MS document image (a “README” text file is attached to the dataset explaining how to read the MS images), and produces as output a binary image (see Fig. 8.). The output must show the class OT in black color and all the rest (BG+) in white color.

MSTEX input output

VIII. OBJECTIVE EVALUATION

The following objective evaluation measures are used:

- F-measure [10]
• pF-measure: pseudo F-measure [10]
• PSNR: Peak Signal to Noise Ration [10]
• NRM: Negative Rate Metric [11]
• DRD: Distance Reciprocal Distortion [12]
• Kappa (κ) coefficient [13], which is well known in the domain of remotely sensed hyperspectral image classification. Kappa is intended to give the reader a quantitative measure of the magnitude of agreement between observers. The calculation is based on the difference between how much agreement is actually present ("observed" agreement $P_o$) compared to how much agreement would be expected to be present by chance alone ("expected" agreement $P_e$).

\[
\kappa = \frac{P_o - P_e}{1 - P_e}
\]


ew{ORGANIZERS} 

• Rachid Hedjam. Rachid received his M.Sc. and PhD. degrees in 2009 and 2013 from the university of Montreal and ÉTS (University of Quebec). R. Hedjam is now a postdoctoral fellow in the Department of Geography at McGill University, where he is working on multispectral image processing and pattern recognition for cultural heritage preservation, forensic applications and remote sensing.

• Hossein Ziasei Nafchi. Hossein received his B.Sc. and M.Sc. degrees in computer engineering from Islamic Azad University, Tehran, Iran. He is currently pursuing his Ph.D. studies with the Synchromedia Laboratory for Multimedia Communication in Telepresence, cole de technologie supérieure, University of Quebec, Montreal, QC, Canada. He has previously organized an image processing competition, the Persian Heritage Image Binarization Competition (PHIBC 2012). His research interests include low-level vision and document image processing.

• Reza Farrahi Moghaddam. Reza received his B.Sc. degree in Electrical Engineering and his Ph.D. degree in Physics from the Shahid Bahonar University of Kerman, Iran, in 1995 and 2003, respectively. He has been a Postdoctoral Research Fellow and a Research Associate with the Synchromedia Laboratory for Multimedia Communication in Telepresence, cole de technologie supérieure (University of Quebec) in Montreal (QC), Canada since 2007 and 2012, respectively. Dr. Farrahi has published more than 50 technical papers. His research interests include sustainability, behaviour analysis, green ICT, green economy, image processing, visual perception, and optimization. He is a member of the IEEE.

• Margaret Kalaeska. Dr. Kalaeska is an Associate Professor in the Department of Geography at McGill University. Her specialization is in remote sensing with a main research focus on the use of data collected from aircraft and satellites to address a range of ecological and forensic questions.

• Mohamed Cheriet. Mohamed received M.Sc. and Ph.D. degrees in Computer Science from the University of Pierre et Marie Curie (Paris VI) in 1985 and 1988 respectively. Since 1992, he has been a professor in the Automation Engineering department at the cole de Technologie Supérieure (University of Quebec), Montreal, and was appointed full Professor there in 1998. Prof. Cheriet is the founder and director of Synchromedia which targets multimedia communication in telepresence applications.

\new{REFERENCES} 


