An Automatic System to Detect and Extract Text in Medical Images for De-identification

Yingxuan Zhu¹, PD Singh², Khan Siddiqui², Michael Gillam²
¹Dept. of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY
²Health Solutions Group, Microsoft Corp, Chevy Chase, MD

ABSTRACT

Recently, there is an increasing need to share medical images for research purpose. In order to respect and preserve patient privacy, most of the medical images are de-identified with protected health information (PHI) before research sharing. Since manual de-identification is time-consuming and tedious, so an automatic de-identification system is necessary and helpful for the doctors to remove text from medical images. A lot of papers have been written about algorithms of text detection and extraction, however, little has been applied to de-identification of medical images. Since the de-identification system is designed for end-users, it should be effective, accurate and fast. This paper proposes an automatic system to detect and extract text from medical images for de-identification purposes, while keeping the anatomic structures intact. First, considering the text have a remarkable contrast with the background, a region variance based algorithm is used to detect the text regions. In post processing, geometric constraints are applied to the detected text regions to eliminate over-segmentation, e.g., lines and anatomic structures. After that, a region based level set method is used to extract text from the detected text regions. A GUI for the prototype application of the text detection and extraction system is implemented, which shows that our method can detect most of the text in the images. Experimental results validate that our method can detect and extract text in medical images with a 99% recall rate. Future research of this system includes algorithm improvement, performance evaluation, and computation optimization.

1. INTRODUCTION

Medical imaging plays an important role in medical research and some of the medical images have burnt-in protected health information (PHI), such as ultrasound or secondary capture DICOM images. There is an increasing need to share medical images for research purposes, while the sharing must respect and preserve patient privacy, the identifiable information must be removed from the medical images before further research usage [1, 2]. Multiple solutions are available today that remove or modify DICOM header to de-identify DICOM images, but as there still are significant number of DICOM image types that contain burnt in PHI, true de-identification becomes a non-trivial task. Removing burnt in PHI is often performed manually, lack robust tools to perform the task, which is time-consuming and tedious, even for a trained expert. Although there is a large body of work on text detection and extraction in images and videos, little work has been done for de-identification of PHI. In this paper, we propose an automatic system to detect and extract burnt in text in medical images for de-identification purposes.

There has been a great deal of research on text detection and extraction [3-10] in the computer vision community. An automatic system is designed to detect text in images based on text spatial cohesion [3]. Edge features and edge based analysis have been used in text detection [5]-[6]. A histogram segmentation method was proposed to obtain the color threshold of characters by Liu et al. [8]. Moreover, clustering-based technique is widely applied to text extraction, e.g., it estimated globally matched wavelet filters using a collection of ground truth images, it extracted text using a K-means algorithm with coarse-to-fine method [9]-[10].
2. METHODS

In medical images, the burnt in PHI usually does not overlay anatomic structures or image pixels of interest for diagnosis and has high contrast to the background, as shown in Figure 1 (all the PHI used in this paper are fictitious). Figure 1 consists of four parts, background, anatomic structures, PHI, and other annotations and markups (e.g., measurements, labels, etc). We focus on separating the text, i.e. PHI and annotations, from background and anatomic structures in this paper. After the text is extracted from the image data, it is sent to optical character recognition (OCR), which can distinguish the PHI from other medical information [12]. As our intention is to create a clinically usable system, it is important that the system is robust, reliable, accurate and fast. We are aiming specifically for an ability to detect all the text (not just PHI) in these images with an accuracy \( \geq 99\% \).

Figure 2 shows the system diagram. First, the region variances of intensity values for each pixel are calculated and a threshold is applied to detect text regions in the image. Then, in order to eliminate the over-detected text regions from the first step, post processing with heuristic rules is applied to the detected texts after region grouping. In this paper, over-detection only indicates that non-text is detected as text, while under-detection means text has not been detected. Both over-detection and under-detection belong to misdetection. The post processing includes removing mis-detected anatomic structures, lines, etc. After that, we use level set method to extract the text from detected text regions for further OCR.

![Figure 1: An image with hypothetical PHI.](image1.png)

![Figure 2: Diagram of text detection and extraction system.](image2.png)
2.1 Text detection using intensity variances

Considering the remarkable contrast between text and background in medical images, we use intensity variances as a criterion to detect text. Variance is a first order statistic method; it can be used in images to detect the intensity differences, e.g., edges. Using intensity variance instead of just intensity value can reduce the misdetection caused by noise in image. Moreover, although the Gaussian or Laplace edge detection method can also be used to detect texts in the images, the detected edges of those methods usually include the edges between anatomic structures, which may result in over-detection.

Let $u$ be the input image, $(x, y)$ represent the location of a pixel, and $u(x, y)$ be the intensity value of that pixel.

Assume that the window’s size for calculating the region variance of one pixel is $N = (2n+1) \times (2n+1)$, which is the total number of pixels in that region. The region variance of a pixel can be defined as

$$v(x, y) = \frac{1}{N} \sum_{x'=-n}^{x+n} \sum_{y'=-n}^{y+n} (u(x', y') - \bar{u}(x, y))^2,$$

where $\bar{u}(x, y)$ is the intensity mean of that region, and $\bar{u}(x, y) = \frac{1}{N} \sum_{x'=-n}^{x+n} \sum_{y'=-n}^{y+n} u(x', y')$. If the threshold of variance for text is $thr$, then the texts in an image are the pixels which have $v(x, y) \geq thr$.

2.2 Region grouping

Since the text information appears in series, grouping the connected text regions will be useful for further text localization and extraction. The connected text regions are grouped according to the 4-connected-objects rule which is illustrated in Figure 3. For each pixel as A in Figure 3, if A is detected as text, and if any one of A’s neighbors B is also detected as text, then that B and A will be grouped together.

2.3 Post processing

For a specific window size, the text region usually has the largest variance values, but in some cases, the detected text regions include non-text objects. Post processing is applied to the detected text regions to eliminate the over-detection. The post processing steps include:

1). Detect and remove the regions if their sizes are larger than $1/M$ of the image size.
2). Detect and remove the lines if there are any.
3). Remove the regions which are close to the center of the images, because those should be anatomic structures which are over-detected as text.

2.4 Level set based extraction
Level set method can combine different segmentation constraints to outline the text, but here, consider the computation time, we only use the \( C_V \) region based level set method, i.e., only use the two terms with parameters \( \lambda_1 \) and \( \lambda_2 \) in equation (9) of [11]. In this section, we use continue domain instead of discrete domain to represent an image.

Specifically, let \( f \) represent the boundary between text and non-text objects, using the notations defined above, we have the energy functional:

\[
F(f) = \lambda_1 \int_{\Omega} (u(x,y) - u_1)^2 H(f(x,y)) dxdy + \lambda_2 \int_{\Omega} (u(x,y) - u_2)^2 (1 - H(f(x,y))) dxdy
\]

where \( H(f) \) is the Heaviside function, and \( H(f(x,y)) = \begin{cases} 1, & \text{if } f(x,y) \geq 0 \\ 0, & \text{if } f(x,y) < 0 \end{cases} \)

and \( u_1 = \frac{\int_{\Omega} u(x,y)H(f(x,y)) dxdy}{\int_{\Omega} H(f(x,y)) dxdy}, \quad u_2 = \frac{\int_{\Omega} u(x,y)(1 - H(f(x,y))) dxdy}{\int_{\Omega} (1 - H(f(x,y))) dxdy}. \)

Since the images are separated into only two regions, text and non-text objects, and the text have remarkable contrast to the background, so both \( \lambda_1 \) and \( \lambda_2 \) are set to 1. After applying Euler-Lagrange equation to minimize \( F(f) \) with respect to \( f \), the text can be segmented out by

\[
\frac{\partial f}{\partial t} = \delta(f)[(u(x,y) - u_2)^2 - (u(x,y) - u_1)^2]
\]

where \( \delta(f) = \frac{d}{df} H(f) \).

2.5 Analysis

We calculate the accuracy by counting the number of detected text and comparing it to the number of ground truth text. The number of text is defined by the total number of letters and numbers in an image. The recall rate for one image is defined below.

\[
\text{recall rate} = \frac{\text{total number of correctly located texts}}{\text{total number of ground true texts}}.
\]

Our experiment is preformed on a server of 2.4 GHz Intel(R) Xeon(R) with 4 GB of RAM.

3. EXPERIMENTAL RESULTS

The following section presents the experimental results obtained by the method introduced in Section 2. In order to protect patients’ privacy, all PHI in the images is fictitious, while it has similar intensity values and contrast with the original ones. After text detection and extraction, the extracted text will be sent to OCR for text recognition to separate the PHI from the text of medical information, e.g., measurements.

3.1 Text detection and extraction
The detection result of Figure 1 is show in Figure 4, where all the text is blurred. Comparing the text regions in Figure 1 and Figure 4, we can see that all text is detected and located, while the anatomic structures are kept as good as the original image. Figures 5-6 show the detection and extraction results of medical images and a medical data image. The left column of Figures 5-6 represents the blurred text in each image, and the right column shows the extracted text by using level set method. Figure 6(a) and (b) show the result obtained from a medical image of data. Figures 4-6 validate that our method has a promising performance in text detection and extraction.
3.2 Prototype application

In our prototype, the text detection and extraction system is built as a standalone object which can be instantiated easily as an extension to any RIS/PAC system requiring de-identification capability. In order to illustrate this procedure and to test the proposed algorithm, we created a GUI of the prototype application for the whole system. The flowchart of this system is presented in Figure 7, and the screen shot of prototype application is shown in Figure 8. As shown in Figure 7, after the image data is loaded to the system, the user has options to either stay on the images of current patient, or step to the images of adjacent patients. Since some patients do not have any images to be read, the user can check the “Skip patient, if no PHI in image” to ignore the patients who do not have any images, and to speed up the procedure. The user can select to read or to blur the text information by checking “No PHI” or not.

![Flowchart of the system](image)

Figure 7: Flowchart of the system.
3.3 Accuracy and computation time

In our experiment, 128 images with PHI are tested, and their average recall rate is 0.9905. The average detection time for these images is 1.4456 seconds.

4. SUMMARY

In this paper, we propose an automatic text detection and extraction system for de-identification of medical images. This system is based on region intensity variances and level set method. Experimental experimental results show that our method can detect most of the text in medical images while keep the anatomic structures intact. Future work includes algorithm improvement, performance evaluation, computation optimization, and text recognition by OCR.

REFERENCES

