Removing the Noise from Cemetery Headstones

Cameron S. Christiansen
Brigham Young University
Department of Computer Science
Provo, Utah, USA
cam@byu.edu

William A. Barrett
Brigham Young University
Department of Computer Science
Provo, Utah, USA
barrett@cs.byu.edu

ABSTRACT

Many approaches exist for alphanumeric data acquisition from images, however very few make mention of engraved text and none to the knowledge of the author make mention of cemetery headstones. The ability to extract the textual information found on headstones is of great importance due to the valuable historical information that is, in most cases, authoritative and unique to that stone. Multiple groups continue to put forth great effort to index such data through a manual transcription process. Although the engraved characters are often of a common font (allowing for accurate recognition in an OCR engine), cemetery headstones present unique challenges to recognition that typical scene text images (street signs, billboards, etc.) do not encounter such as engraved or embossed characters (causing inner-character shadows), low contrast with the background, and significant noise due to weathering. We therefore propose a system in which the noisy background is removed through use of gradient orientation histograms, an ensemble of neural networks, and an automated implementation of graph cut, resulting in a clean image of the headstone's textual information.

Categories and Subject Descriptors

I.4.0 [**Computing Methodologies**]: Image Processing and Computer Vision—*General*; I.7.0 [**Computing Methodologies**]: Document and Text Processing—*General*

General Terms

Human Factors

Keywords

Image processing, OCR, Engraved Stone, Binarization

1. INTRODUCTION

Cemetery headstones contain a wealth of genealogical information that is largely untapped, unindexed and until recently electronically inaccessible. Such information could provide not only the vital birth and death information, but, in many cases, links to family relationships and clues as to where they lived. By acquiring headstone data, access is gained to historical information that is most often authoritative with no other source. Such information was engraved for the very purpose of lasting through time and passing on information to future generations, but the nature of the medium makes quick access to its information difficult.

The lack of availability of such records has led to significant efforts to capture information from headstones and make it searchable and available on-line [1][2][3]. For example, "Find A Grave"[2] has indexed over 50 million names worldwide, along with pictures of the headstone and/or cemetery, and sometimes the individual and their biographical information. However, the gathering of such information is a time-consuming process: capturing the photo, recording the cemetery and plot location, transcribing the information on the headstone, importing the photo and information onto a computer, then using a browser to upload the information to the web site. This is impractical and does not scale given the number of headstones that remain unindexed, the non-decreasing number of deaths ($\sim$ 2.5 million per year in the U.S.) and the time-consuming process described.

Given that the amount of headstone information that has been indexed is only a fraction of what could be available, it is impracticable to rely on manual transcription and incomplete to rely on cemetery records. Some estimates suggest that there are over 1 billion headstones in North America alone (personal communication with Billion Graves [1]). Furthermore, the data contained on these headstones continues to deteriorate each year and may be the only source of this information for 75% of the headstones. An automated process to transcribe this information on site would be of great benefit in time, effort and preservation of data.

The research field of textual information extraction from images and video has continued to grow in the last decade with the proliferation of mobile computing devices (i.e. smart phones) among the general population. A variety of approaches have been developed as discussed in previous surveys [7][10]. A common challenge in recognizing scene text is correcting the perspective skew of an image. However since complete control is assumed on the location of the camera in regard to the headstone rotational correction is a minor concern but will be discussed.

In the current literature only a couple have explored rec-

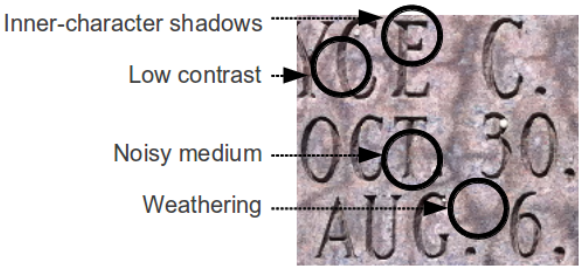

Figure 1: Sample of a noisy headstone

ognizing scene text in the terms of engraved text [5][12]. Challenges arise in recognizing headstone data for a number of reasons (Figure 1):

1. Headstone data is three-dimensional; characters are either recessed into the stone or embossed, both of which cause shadows and differences in color within the characters themselves.

2. Much of engraved text has low contrast with its surrounding background.

3. The nature of the medium (i.e. granite, marble) causes a noisy and inconsistent background.

4. The great majority of engraved text suffers from weathering causing irregularities in character boundaries and increased noise in both the characters and the background.

To solve challenges of automatically extracting textual data from cemetery headstones we propose a system in which gradient orientation histograms are heavily used to accurately locate regions of text. The textual features are fed into a neural network whose output will then be used in the initialization of a graph to produce a min-cut. This process is described in more detail in section 3. Results, future work, and a conclusion are given in sections 4, 5, and 6, respectively.

2. PREVIOUS WORK

Research in the field of textual information extraction has been very active in the last decade due to the increasing availability of cameras on mobile computing devices, providing increased demand for technology and an increasing amount of user access to information. However, no previous research has been found that focuses on recognition of text found on cemetery headstones and very little [5][12] is found with respect to engraved characters. Related literature in text recognition is discussed.

2.1 Scene Text Recognition

The recognition of scene text has the challenge of large amounts noise and is similar to the described problem. Noise can come from inconsistent lighting, perspective skew, and inconsistencies that exist simply because the the text “lives” in the real world(weathering, blemishes). Therefore there is a wide spectrum of approaches used in recognizing scene text. Proposed systems have made use of connected components created by Sauvola binarization [6], and the Canny edge-detector [4]. Additional work has been done using the texture of the image through wavelets [9].

While both cemetery headstones and typical scene text are similar in nature, scene text recognition will often rely on a high contrast (such as that on street signs, billboards, etc.) with a plain background. Neither of the two can be expected within a headstone image and thus introduce a more difficult problem. Using Sauvola binarization will produce disjointed characters and added noise on a typical headstone (see Figure 5) due to low contrast or varied textual coloring. Additionally, purely texture based approaches will struggle due to the noise within the texture, causing a distortion of the features. Texture based features are valuable, however, and are used in the proposed system through gradient orientation histograms.

3. METHODS

In this section the feature extraction and use of these features for classification are described.

3.1 Feature Extraction

To recognize the text found within an image a set of discriminant features is desired. The image is first divided into a specified NxN window where each window overlaps neighboring windows by N/2 in both x and y. From each of these windows a set of features are extracted to form feature vectors.

Inspired by the use of a gradient orientation histogram in [11], the core set of features are derived from a gradient orientation histogram where the orientations are weighted by the gradient magnitude. It is found that in weighting the orientations, large peaks appear and allow for easy distinction between background and foreground. This is due to both consistent orientations on a true edge (such as a straight line) and the magnitude at which that edge is found. The histograms for background areas consist of various gradients, but of inconsistent directions causing a random and a greater distribution across the histogram (see Figures 2 and 3). Therefore, a balanced use of gradient magnitude and orientation is applied through a gradient orientation histogram to overcome the challenges of both low contrast and a noisy background.

To calculate the gradient histogram of a given area, 3x3 sobel kernels are first convolved with the image to find gradients in the x and y directions. Once the gradients are computed in each direction, the magnitude and orientation are computed with the following equations in their respective order:

$$\|\nabla I\| = \sqrt{\nabla I_x^2 + \nabla I_y^2} \tag{1}$$

$$\phi(\nabla I) = \tan^{-1} \frac{\nabla I_y}{\nabla I_x} \tag{2}$$

where ∇I_x and ∇I_y represent the gradients in the x and y

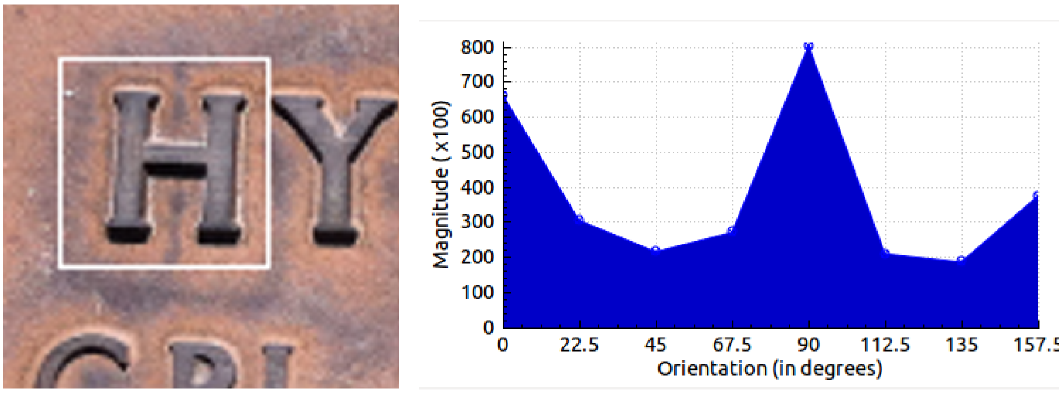

Figure 2: (Left) A 50x50 region (white box) of a headstone image from which high peaks are generated at 0 and 90 degrees in the gradient orientation histogram (right), distinguishing the area as text

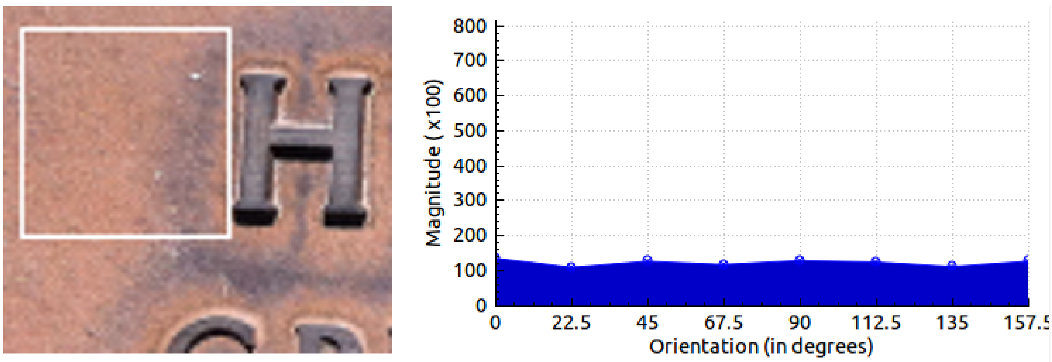

Figure 3: (Left) A 50x50 region (white box) of a headstone image from which no significant peaks are generated in the gradient orientation histogram (right), distinguishing the area as background

direction. The orientations are placed into one of 8 bins in the histogram; one bin for every $\frac{\pi}{8}$ radians from 0 to π.

3.2 Classification

The features found from the process described above are fed into a neural network, namely a multi-layer perceptron (MLP). Each feature acts as an individual input to the network.

The proposed system uses an ensemble of MLP's where each class (text, background, and artwork) correlates to a specific MLP. Each MLP in the ensemble is trained individually to classify one particular class versus all others. For example, the text MLP is trained on data in which all text classifications are labeled as 1 and all others as -1. This allows the particular MLP to learn that specific problem very well. The MLP's then each cast a vote in which the MLP which is the most confident classifies the data point.

3.3 Segmentation

Although the current system has not implemented more than what has been described, such classification is not the end goal of the system. The resulting confidence scores of the MLP's will be leveraged for use in the initialization of a graph for graph cut. We do not expect the classification to be perfect in capturing all of the text and therefore will use the globally optimal graph cut to fill in gaps that the classification may have left.

Additional features not used for classification will also give input to the n-weights within the graph such as proximity, aspect ratio, and area of connected components.

Initially, graph cut will be performed on a coarse window size to reduce the solution space. Subsequently, the coarse window size will be reduced to a smaller size upon which graph cut will again be applied. This process is repeated until the window size is no longer significant.

4. RESULTS

Text segmentation on headstones is a new and unique area of research such that there are no directly alternate systems to compare performance. Alternatively, the effectiveness of the proposed system in recognizing text regions is shown first through classification accuracy of the MLP ensemble and a light comparison to the respective Sauvola binarization[13] result.

4.1 Classification

The training set used consisted of 613 50x50 samples from various headstone images each classified as one of the three classes: text, background, or artwork. The gradient orientation histogram's 8 bins (normalized by the maximum histogram value) were used with the average histogram value to form 9 features for the neural network. According to typical setup of an MLP, 18 hidden nodes are used (number of inputs * 2). The test set consisted of 112 headstone image samples. The classification accuracy is shown below:

Data set	Text	Background	Art	Overall
Training	96.1039%	94.0594%	80.0000%	90.7015%
Test	90.6977%	54.7619%	66.6667%	71.4286%

The accuracy is shown for each class as well as the overall accuracy. The distinction between background and artwork is at times small and difficult to classify. For example, a user may have difficulty classifying a sample which the border around the headstone's information. Due to such ambiguity the two classes often work against each other and reduce the classification accuracy, thus the significant drop in accuracy for these classes on the test set. However, the class of concern is the text and we focus on its performance without heavy regard to the background and art classes.

4.2 Image classification

Visualizing data has many advantages and can give insights as to how well the underlying system is performing. Although classification accuracy was given previously, a more visual approach to showing the system's performance is given. In figures 4, 5, and 6 the left image shows the Sauvola binarization of a headstone image and to the right the original image is given with exception to the areas classified as text by the described system highlighted. Sauvola binarization has shown insufficient for such a difficult medium as headstones as characters are left incomplete (the "P" in Figure 4) and the majority of the characters in Figure 5. As mentioned previously, the described system seeks to overcome such challenges through a global solution. The described system is able to find the regions of text, including those with which Sauvola struggled as seen in the images on the right.

5. FUTURE WORK

As mentioned in Section 3, the proposed system does not have all of the core features implemented. Such work is the immediate need in regard to future work and will be implemented as described previously.

Fine tuning of the neural network will likely result in improved classification results. During training only a specific number of epochs are run for the stopping criteria, engendering over-fit on the training set or alternatively pre-mature stopping. A more intelligent stopping criteria is to have a separate validation set upon which the stopping criteria is based: when the validation set stops improving, end the training.

Additionally, a thorough search for the best features, given the problem, will be beneficial.

6. CONCLUSIONS

Cemetery headstones contain valuable data that is difficult to access and to index. The current methods for indexing is difficult and time consuming, resulting in the need for a more automated system. Headstones are inherently noisy due to stone texture, shadows, low contrast, and weathering, rendering previous methods insufficient. A system is proposed in which gradient orientation histograms are used extensively in conjunction with a multi-layer perceptron and graph cut.

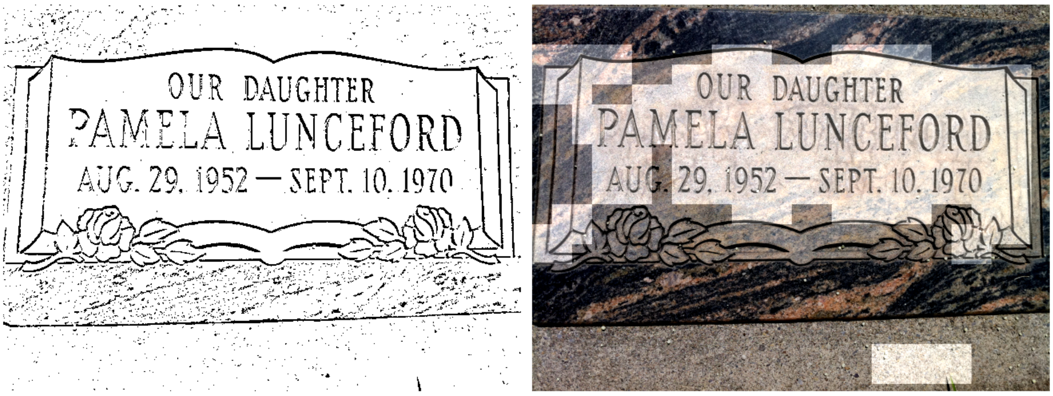

Figure 4: (Left) Sauvola binarization of a headstone image, (Right) Regions classified as text (lighter areas) with the proposed system

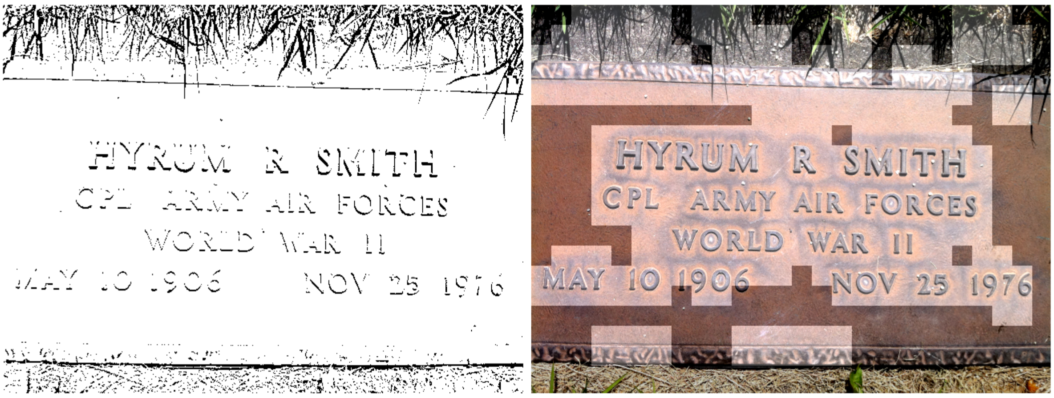

Figure 5: (Left) Sauvola binarization of a headstone image, (Right) Regions classified as text (lighter areas) with the proposed system

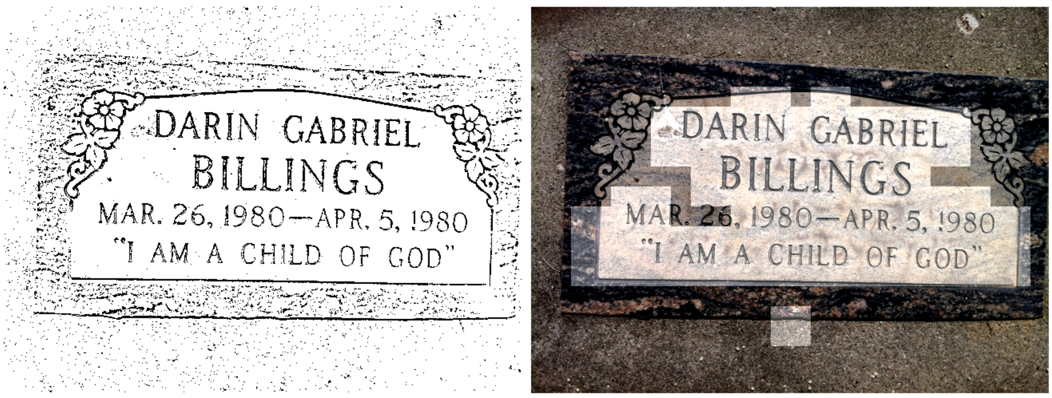

Figure 6: (Left) Sauvola binarization of a headstone image, (Right) Regions classified as text (lighter areas) with the proposed system

7. REFERENCES

[1] Billion graves @ONLINE, Apr. 2011.

[2] Find a grave @ONLINE, Apr. 2011.

[3] Names in stone @ONLINE, Apr. 2011.

[4] J. Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):679–698, 1986.

[5] U. Garain, A. Jain, A. Maity, and B. Chanda. Machine reading of camera-held low quality text images: An ICA-based image enhancement approach for improving OCR accuracy. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE.

[6] B. Gatos, I. Pratikakis, K. Kepene, and S. Perantonis. Text detection in indoor/outdoor scene images. In *Proc. First Workshop of Camera-based Document Analysis and Recognition*, pages 127–132, 2005.

[7] K. Jung, K. I. Kim, and A. K. Jain. Text information extraction in images and video: a survey. *Pattern Recognition*, 37(5):977–997, 2004.

[8] T. Kasar, J. Kumar, and A. Ramakrishnan. Font and background color independent text binarization. In *Proceedings of 2nd International Workshop on Camera Based Document Analysis and Recognition*, pages 3–9, 2007.

[9] H. Li, D. Doermann, and O. Kia. Automatic text detection and tracking in digital video. *Image Processing, IEEE Transactions on*, 9(1):147–156, jan 2000.

[10] J. Liang, D. Doermann, and H. Li. Camera-based analysis of text and documents: a survey. *International Journal on Document Analysis and Recognition*, 7(2):84–104, 2005.

[11] D. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[12] C. Mancas-Thillou and M. Mancas. Comparison between pen-scanner and digital camera acquisition for engraved character recognition. In *Proc. of the 2nd. International Workshop on Camera-Based Document Analysis and Recognition*, volume 130, page 137. Citeseer.

[13] J. Sauvola and M. Pietikäinen. Adaptive document image binarization. *Pattern Recognition*, 33(2):225–236, 2000.