2012-11-27

Data Acquisition from Cemetery Headstones

Cameron Smith Christiansen
Brigham Young University - Provo

Data Acquisition from Cemetery Headstones

Cameron S. Christiansen

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

William A. Barrett, Chair
David W. Embley
Robert P. Burton

Department of Computer Science

Brigham Young University

November 2012

ABSTRACT

Data Acquisition from Cemetery Headstones

Cameron S. Christiansen
Department of Computer Science, BYU
Master of Science

Data extraction from engraved text is discussed rarely, and nothing in the open literature discusses data extraction from cemetery headstones. Headstone images present unique challenges such as engraved or embossed characters (causing inner-character shadows), low contrast with the background, and significant noise due to inconsistent stone texture and weathering. Current systems for extracting text from outdoor environments (billboards, signs, etc.) make assumptions (i.e. clean and/or consistently-textured background and text) that fail when applied to the domain of engraved text. Additionally, the ability to extract the data found on headstones is of great historical value. This thesis describes a novel and efficient feature-based text zoning and segmentation method for the extraction of noisy text from a highly textured engraved medium. Additionally, the usefulness of constraining a problem to a specific domain is demonstrated. The transcriptions of images zoned and segmented through the proposed system result in a precision of 55% compared to 1% precision without zoning, a 62% recall compared to 39%, an F-measure of 58% compared to 2%, and an error rate of 77% compared to 8303%.

Keywords: Text recognition, engraved characters, zoning, segmentation, noisy medium, headstones

ACKNOWLEDGMENTS

Great things are done when men and mountains meet.

William Blake

This is to those who helped me climb the many mountains I faced in the journey.

Thank you, Janell, for giving me the wings to fly and a journey to remember.

Thank you, kids, for making every evening home a party.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

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 the individual lived. By acquiring headstone data, access is gained to historical information that is authoritative and most often 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 physical nature of the medium is vulnerable to external forces (weather, earthquakes, etc.) and makes quick access to its content 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 (BillionGraves [2011], Find A Grave [2011], Names In Stone [2011]). For example, Find A Grave [2011] 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 image, recording the cemetery and plot location, transcribing the information on the headstone, importing the photo and information into 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 and the non-decreasing number of deaths (approximately 2.5 million per year in the U.S., Murphy et al. [2012]) and the time-consuming process described.

An alternative approach is taken by Names In Stone [2011] where cemetery records are digitized and made available. Cemetery records, however, are often out of date. Therefore, Names In Stone [2011] enlists participating cemeteries to provide up-to-date information. This, however, relies on the diligence of the cemetery sextons. Additionally, only plot number, names, and dates are recorded. Contextual information, such as relationships, is lost such as relationships and no image of the headstone is included.

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 BillionGraves [2011]). Furthermore, the text engraved 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 benefit in time, effort and preservation of data.

The research field of textual information extraction from images and video has continued to grow among the general population in the last decade with the proliferation of mobile computing devices such as smart phones. A common challenge in recognizing scene text is correcting the perspective skew of an image. However, complete control is assumed on the location of the camera in regard to the headstone, therefore rotational correction (in the 2-D plane) is a minor, but valid concern and will be briefly discussed. Additional skew (azimuth rotation and shear) is expected to be minimal and insignificant in terms of the OCR engine.

In the current literature only a couple of publications (Garain et al. [2008], Thillou and Gosselin [2006]) have explored recognizing engraved text. Challenges to recognition occur in such cases for a number of reasons (Figure 1.1):

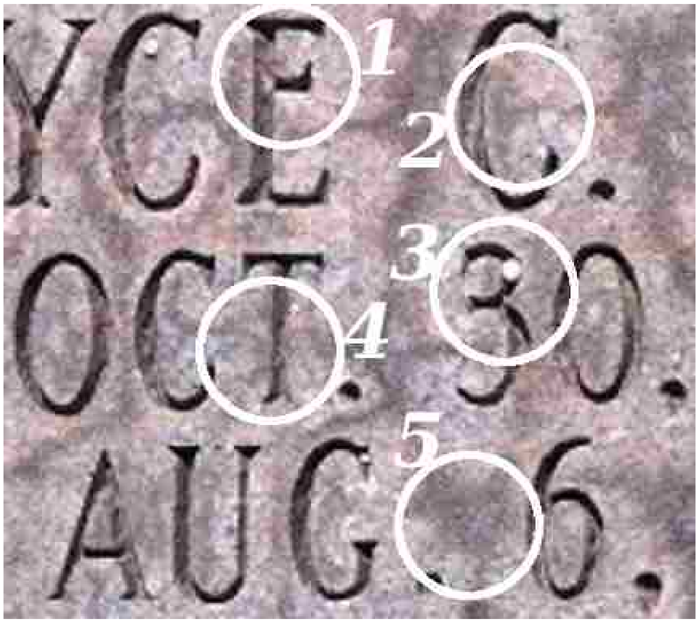


Figure 1.1: Sample of a noisy headstone. Challenges to recognizing text found on headstone images are circled and are as follows: **(1)** Three-dimensional characters causing inconsistent contrast and texture within the character, **(2)** inconsistent stone texture **(3)** high potential for additional noise (e.g. white circle within the character), **(4)** low contrast, and **(5)** Added noise due to weathering.

1. Engraved text 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. The stone texture causes a noisy and inconsistent background.

3. There is a high potential for additional noise in the real world. In-ground headstones incur noise through objects such as cut grass, dirt, leaves, and other miscellaneous objects that accumulate at ground level while upright headstones may have objects such as flowers or wreaths obstructing the text from view.

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

5. Engraved text suffers from weathering causing irregularities in character boundaries and increased noise in both the characters and the background.

The state of the art and arguably the best OCR engine (Mendelson [2009]), ABBYY FineReader (ABBYY [2012]), is unable to overcome the challenges that headstone images present. In Figure 1.2, an analysis is performed on ABBY FineReader's ability to transcribe the textual data found on a headstone image. The images shown in Figures 1.2a, 1.2c, and 1.2e are passed into the OCR engine and are shown with their corresponding transcriptions in Figures 1.2b, 1.2d, and 1.2f, respectively.

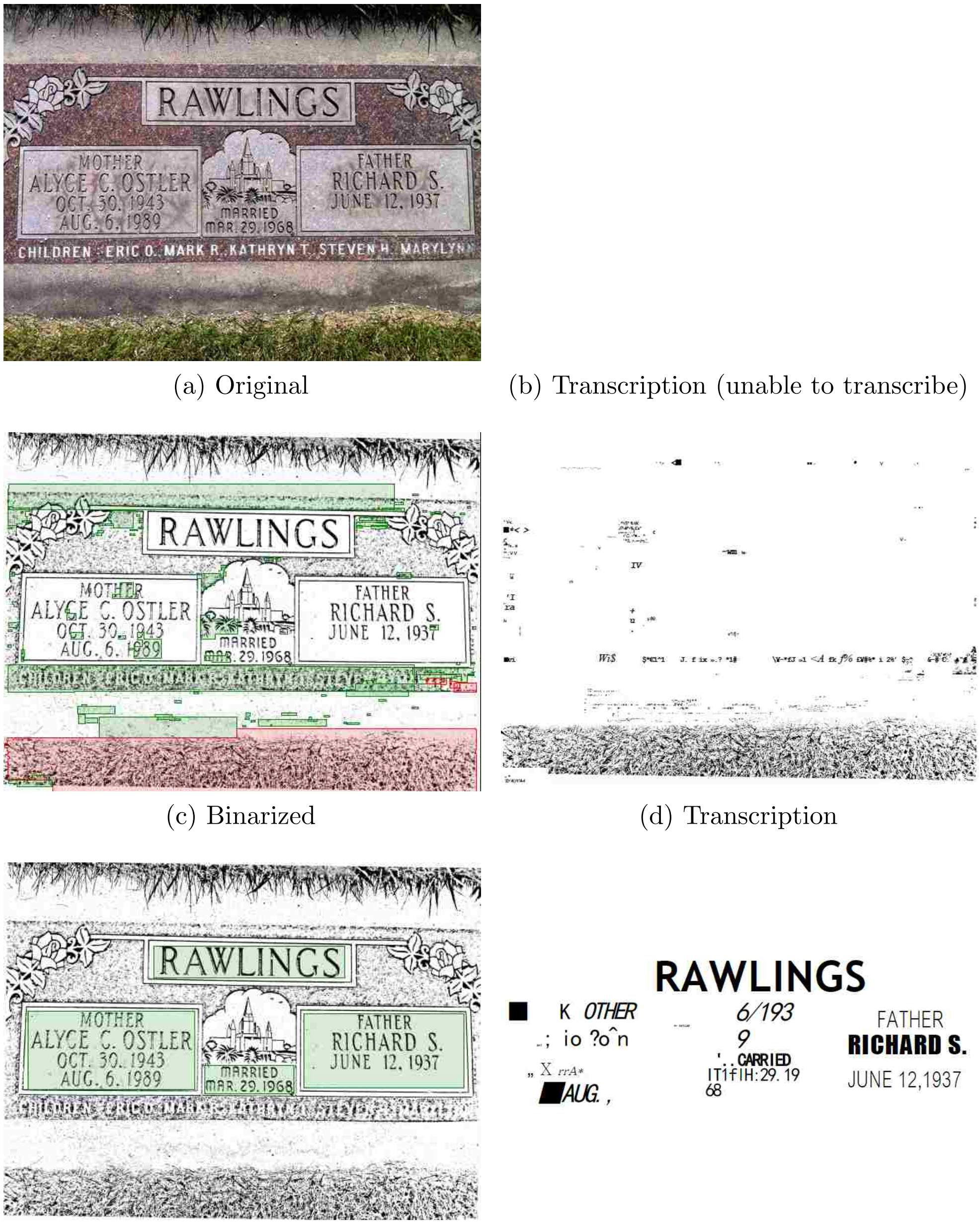
Figures 1.2c and 1.2e both show the text zoning done prior to recognition (the green overlays), the difference being in Figure 1.2c the text zoning was performed with no human intervention whereas in Figure 1.2e the text zoning was selected manually. Only through manual intervention is it possible to produce a semi-accurate transcription using ABBYY FineReader, furthering the motivation for this thesis.

Current systems make one or more of the following assumptions: a clean and consistent background texture, high-contrast text, consistent-contrast text, or consistent-color text. Such assumptions fail when applied to the domain of headstones.

1.2 Approach

To recognize and transcribe headstone data, a system that can accurately zone and segment the noisy three-dimensional text from a cluttered and highly textured background is needed.

The proposed system outlined in Figure 1.3 consists of three main processes preceded by a preprocessing step and followed with a transcription process used for quantitative analysis and validation. Pre-processing consists of the creation of an image pyramid to improve speed and accuracy and the computation of the image gradients that will be used in the zoning and segmentation processes. An additional pre-processing step uses graph cut to segment the headstone from the surrounding background. Zoning is accomplished by using trained neural networks on 10x10 regions of the image to discriminate between background stone texture and engravings which constitute the text and surrounding artwork. Segmentation is accomplished through Otsu binarization. Artwork removal performs connected component



(a) Original

(b) Transcription (unable to transcribe)

(c) Binarized

(d) Transcription

(e) Binarized and Manually Zoned

(f) Transcription

Figure 1.2: Three examples of transcriptions provided by the ABBYY FineReader OCR engine. The original headstone image (1.2a) was not recognized as containing text, thus producing no transcription (1.2b). Passing in the same image binarized by the Sauvola algorithm (Figure 1.2c) resulted in poor text zoning (text regions shown in green, image regions are shown in red) and thus a poor transcription (Figure 1.2d) of many small, seemingly random characters. Passing in the same image as Figure 1.2c, but with manual text zoning (note the green regions in Figure 1.2e) improved the transcription (Figure 1.2f), but still is unable to accurately transcribe the majority of the headstone.

analysis using trained neural networks and graph cut to separate noise and artwork from the text. The resulting text image is passed to an OCR engine to produce a transcription, and is corrected based on context (e.g. names, dates, etc.).

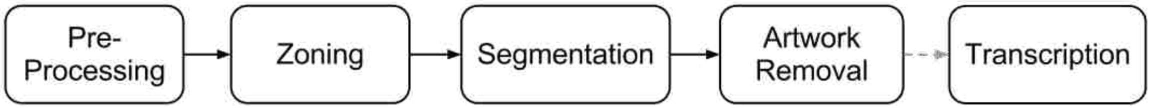


Figure 1.3: Proposed system for zoning, segmenting, and recognition of headstone text.

The problem is constrained geometrically through zoning, segmentation, and artwork removal by removing non-textual image data that would potentially interfere with the transcription process due to the large amount of noise. Additionally, in the transcription process, the problem is constrained contextually by reducing the possible transcriptions to the vocabulary found on headstones. This thesis demonstrate the usefulness of constraining the domain geometrically (Figure 3.9) and contextually (Table 4.5).

This work is published in Christiansen and Barrett [2013] and is described in more complete form in this thesis.

Chapter 2

Previous Work

Research in the field of textual information extraction from a surrounding environment has been active in the last decade due to the increasing availability of cameras on mobile computing devices. This has provided an increased demand for technology that can capture, transcribe and facilitate access to the textual information.

However, no previous research has been found that focuses on recognition of text on cemetery headstones and little (Garain et al. [2008], Thillou and Gosselin [2006]) is found with respect to engraved characters. A variety of systems have been developed to recognize scene text (billboards, signs, etc.), graphic text (text overlaid on an image), and scanned document text, many of which are discussed in previous surveys (Jung et al. [2004], Liang et al. [2005]). Additional work in text recognition is discussed below through focusing on its three common areas separately (scene text, graphic text, and document text). Related work is then addressed in terms of gradient orientation histograms, graph cut, neural networks, projection profiling, optical character recognition, and data extraction.

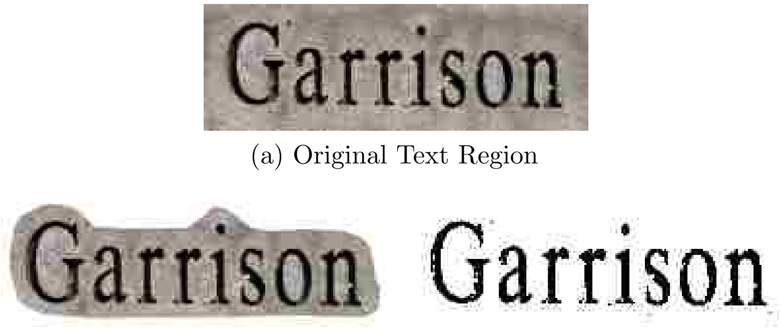
2.1 Text Recognition

Previous work in text recognition has been applied to scene text (real-world imagery), graphic text (text synthetically overlaid on an image), and scanned documents. In each application the zoning, segmentation and recognition of the text are addressed. The following definitions apply:

1. **Zoning** is the process of coarsely locating regions of interest within an image to reduce the problem to specific areas. Text zoning is therefore the ability to localize a region of the image so that it contains only text and the immediate background (i.e. no graphics or other features, Figure 2.1b). This process is also termed as text localization. Previous systems perform text zoning on an image by either zoning the text directly as a single process, or by first zoning the foreground, followed by the removal on non-textual components. The proposed system uses the latter where non-textual components are removed after segmentation in the artwork removal process.

2. **Segmentation** is a pixel labelling process in which a pixel that is part of a text character is binarized and labelled as text. Otherwise the pixel is thresholded away and labelled as background (Figure 2.1c).

3. **Recognition** is the process of classifying each segmented group or connected component of text pixels as a given character. The correct result of the recognition process for Figure 2.1 is the text string “Garrison”.



(a) Original Text Region

(b) Text Zoning

(c) Text Segmentation

Figure 2.1: A region of a headstone image (Figure 2.1a) that encompasses a collection of characters with the resulting zoning (Figure 2.1b) and segmentation (Figure 2.1c).

2.1.1 Scene Text Detection/Recognition

Of the three mentioned areas, scene text has the greatest potential for noise-related challenges. Noise can come from inconsistent lighting, perspective skew, and inconsistencies that exist

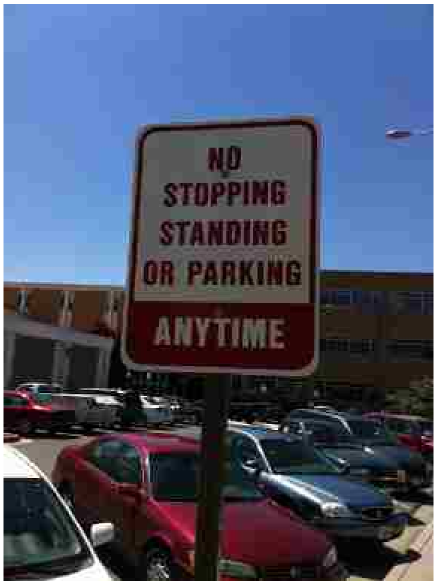


Figure 2.2: An example image that contains scene text where perspective skew is visible (note the non-parallel edges on the top and bottom of the sign) and noise is present in the bolt between the "NO" and "STOPPING" whose color is similar to the text.

simply because the the text "lives" in the real world (i.e. weathering). Therefore a wide spectrum of approaches have been applied to recognize scene text. An example image of scene text is shown in Figure 2.2.

Perspective skew is a difficult challenge in scene text, and much of the research places significant emphasis on this. A common approach (Clark and Mirmehdi [2002, 2003]) estimates vanishing points through projection profiles from varying skew angles. The angle for which entropy is at a minimum is then selected. The angles for both vertical and horizontal vanishing points are used to de-skew the text.

An alternative approach (Myers et al. [2005]) expects one or more quadrilaterals containing text (such as a note card or sign) to be found within the image. The edges of the quadrilateral are used to determine and correct the skew of the region and subsequently the text.

In this thesis, liberty is taken to require that the image be captured with the image plane as parallel to the plane of the headstone as possible. This is a reasonable requirement given that headstones are stationary and at ground level, giving flexibility to the angle at which the image is captured.

A connected component method described by Gatos et al. [2005] uses Sauvola binarization (Sauvola and Pietikäinen [2000]) to create connected components, zoning the foreground objects. These connected components (foreground objects) are removed from the original image through interpolation to create what the authors refer to as the background image. This estimated background is used to threshold the original image for segmentation of the potential text regions. Each resulting connected component is then passed through a number of tests to determine whether it is a character or not.

The quality of each connected components is crucial in this system as disjointed characters would fail in the connected component analysis. The Sauvola binarization of a headstone image region is shown in Figure 2.3. The varying contrast within the engraved characters with respect to the background have caused characters to be disjointed and incomplete. Additionally, headstones often contain artwork and require a more robust connected component analysis.

Through zoning regions of text, a more accurate binarization is possible. In addition, through use of graph cut, a proximity-aware connected component analysis is performed to minimize disjointedness or the loss of other connected components.

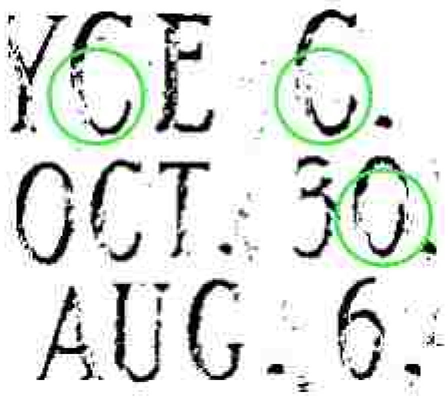


Figure 2.3: The text region of a noisy headstone after Sauvola binarization. The circles denote breaks in characters due to inconsistent contrast and coloration within

Kasar et al. [2007] describe a method that uses the Canny edge detector (Canny [1986]) to zone potential text regions and produce bounding boxes. Neighboring pixels of the bounding boxes are then used to estimate and segment the text from the background within the bounding box. This approach works well with a consistently-colored background. However, Kasar's approach is not designed for a non-homogeneous background.

A solely texture-based approach to locating text is described by Li et al. [2000]. Neural networks are used with wavelet features and focus on the frequencies found in individual regions to zone the text. Noisy stone and occasional artwork that exist on headstones images increase the difficulty of distinguishing between text and background solely from texture.

Documents that suffer from challenges similar to those found on headstones are discussed with the system proposed in Takakura et al. [2010], however the system is not automated in that user interaction is expected to produce an acceptable result.

In this thesis, the edges and features within a text region are combined with the edge-orientation to address the described challenges.

Approaches using the ultimate opening (Retornaz and Marcotegui [2007]) and independent component analysis (ICA) (Garain et al. [2008]) have been explored, but appear to be insufficiently robust or adaptable for our application.

2.1.2 Graphic Text Detection/Recognition

Graphic text recognition consists of text overlaid on a natural scene or complex image with a possible noisy background (see Figure 2.4). Much of graphic text is placed over video and therefore has a set of challenges and solutions that are specific to video such as text tracking and enhancement through multiple frames. In the proposed system only individual images are considered and therefore only techniques used for zoning and segmentation are considered. As headstone text is often placed on a noisy and complex background, it is of value to consider research on graphic text detection.

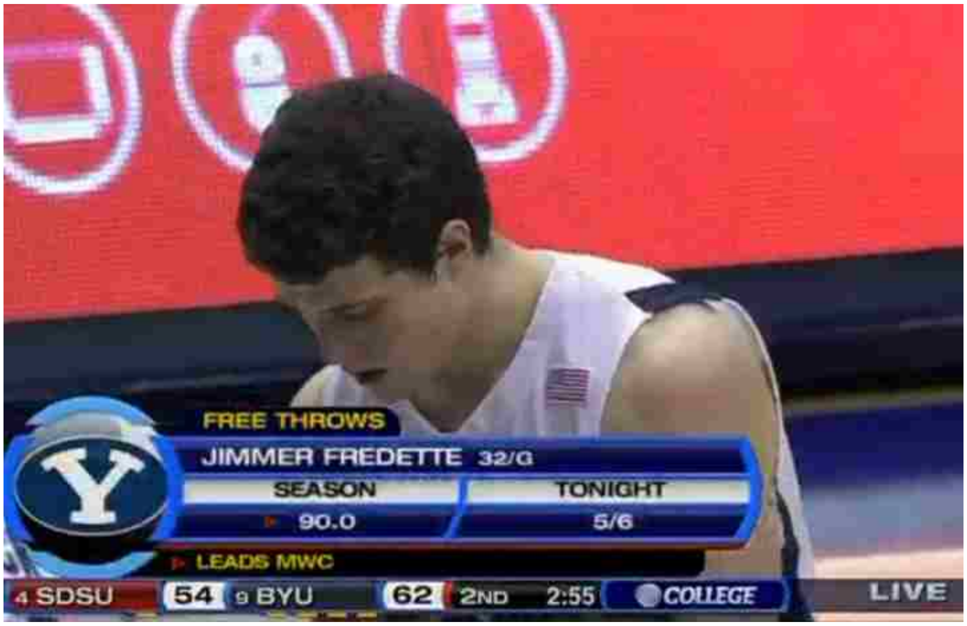


Figure 2.4: An example image that contains graphic text where text is placed on both a solid background and transparent background (behind the text "live" on the bottom-right) which is then overlaid on a complex image.

A helpful feature in graphic text is that the text of interest was intentionally placed in view to be easily recognizable. The algorithms for recognizing graphic text can therefore rely on a strong contrast or difference in color with the background.

Multiple authors have proposed methods that quantize the existing colors found within the image to increase contrast and reduce complexity (Chen and Chen [1998], Jain and Yu [1998], Wang et al. [2005]). Due to potentially low contrast and color differences however, color quantization may result in both background and text being partially (or completely) quantized together and is therefore insufficient for zoning and segmentation of text found on headstones, where the signal-to-noise ratio is often low.

Alternate methods use the image's histogram to segment colors from each other (Hase et al. [2001], Puzicha et al. [1999]). However, intensity characteristics of engraved characters often overlap with much of the background and suffer from shadows and blemishes that result in multi-colored characters, thus histogram-based color segmentation is incapable of differentiating engraved text from the background.

The method described by Thillou and Gosselin [2006] creates a color map from which all the mappings are hierarchically merged into three bins. These three bins are then used as the initial centroids for k-means clustering where $k = 3$. Each resulting cluster is to represent one of the three: foreground, background, and noise, where the foreground cluster represents the zoned and segmented text.

The characters on a headstone often contain small areas which correspond in color to the background. Also, the noisy background often contains flecks of stone that match the color of the engraved characters. Therefore, an approach to segment text solely on color would result in unwanted noise both within the character and without. Additionally, the majority of headstones contain multiple colors for background while many contain text of multiple colors (see Figure 2.5).

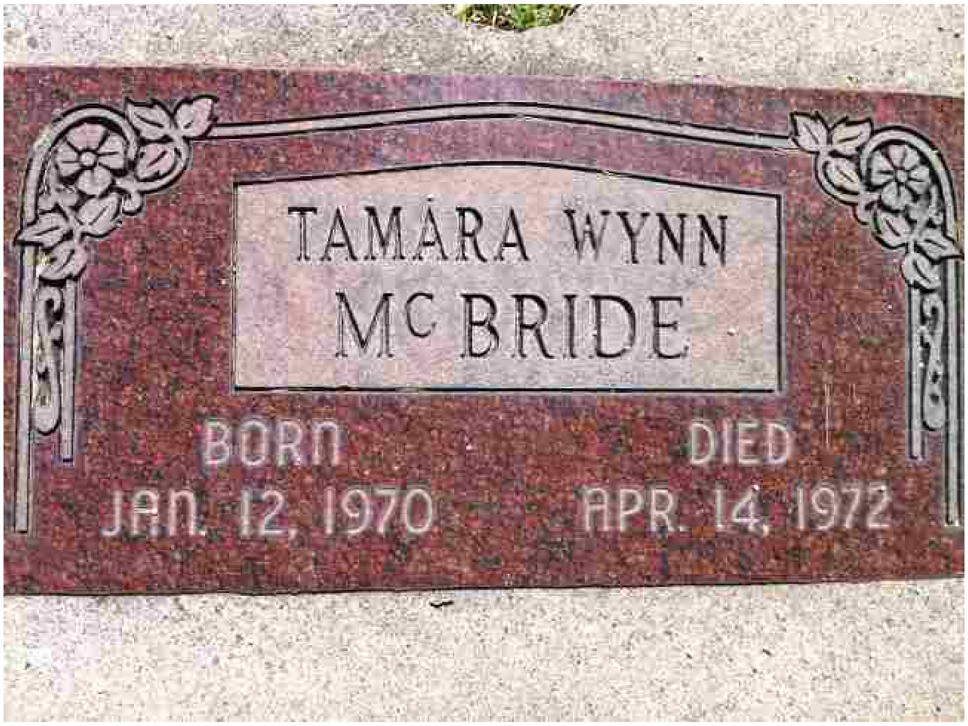


Figure 2.5: An example headstone image that contains a background of multiple colors and text of multiple colors.

The aspect of multi-colored text has, however, been addressed by Hase et al. [2003], but the method proposed is limited to a small range of colors which are visually equal.

Headstones potentially face noise and shadows causing dramatic inner character variation of color, for which this method is not applicable.

The assumptions made in the literature for recognizing graphic text such as consistently colored background and text, as well as text with strong contrast, provide for good results for zoning and segmentation of graphic text. However these assumptions are not applicable to cemetery headstones.

2.1.3 Scanned Document Recognition

Zoning, segmentation and recognition of clean-scanned, first-generation documents can be considered a mostly-solved problem. However, research continues in the area of noisy OCR, with an emphasis on binarization. Specialized binarization techniques continue to be developed in difficult domains in which the traditional algorithms of Otsu [1975] (global threshold), Niblack [1986] (adaptive threshold), and Sauvola and Pietikäinen [2000] (adaptive threshold) are insufficient. Additional approaches have been proposed by Wolf et al. [2002] and Nina et al. [2011]. The system described by Wolf et al. adapts the Sauvola method for the purpose of removing additional noise that is missed with the Sauvola algorithm when binarizing video frames. Nina et al. use the Otsu algorithm recursively on pixels classified as the background to capture more completely strokes of varying strength and intensity.

The binarization methods mentioned above are contextually unaware in that no contextual information is used. By specifying text as the desired target, binarization can become more effective. This has led to approaches that use variance in pixel intensity as a crucial identifier of text (Seeger and Dance [2001], Gatos et al. [2006]) allowing for a more accurate zoning.

The background found in images of headstones will be noise-heavy, causing difficulty in identifying and separating the text using only variance. Additionally, the engraved artwork that is placed alongside the engraved text (such as the flowers in Figure 2.5) will also be difficult to distinguish from text based on variance alone.

A specialized approach was proposed by Milewski and Govindaraju [2006] to recognize text from Pre-Hospital Care Reports which suffer from extreme carbon mesh noise and various inconsistencies in signal intensity, somewhat similar to some headstones. Milewski's approach uses masks to survey the surrounding area and based on mean values found through a trajectory path and zones the text by determining if the 5x5 region of interest is text.

Although this method is able to identify and zone text regions in a noisy environment, variations in lighting and noise within the text itself is not addressed.

2.1.4 Summary

Binarization algorithms can be used as crucial building blocks in zoning and segmenting scene text, but by their nature are not sufficient in their entirety to solve the problem at hand.

The zoning, segmentation, and recognition of text are areas in which general and robust solutions are heavily sought after and has a large base of published literature to its name. A general solution that reaches across the three areas (scene text, graphic text, and document text) has yet to be presented. Given the broad range of challenges faced among them it is unlikely that one will be developed. Such a solution may represent the "grand challenge" of text recognition in document image analysis.

Authors (such as those mentioned previously) have specialized their approaches to one of the three described areas, suggesting that more domain specific approaches are needed. This thesis claims that not only an approach specific to an area of text recognition is needed, but that an approach specific to an even smaller domain (i.e. that of engraved text on headstones) will be of greatest benefit. Through constraining the domain (both spatially and contextually in the proposed system) performance can be significantly improved (see Chapter 4).

2.2 Gradient Orientation Histogram

Calculating the gradients within an image is frequently used to find edges and structure within an image. In the domain of image matching, the SIFT features proposed by Lowe [2004] use a histogram of gradient orientations as a distinguishing feature. While the gradients measure the rate of change in pixel value (i.e. the derivative), the orientation is the angle at which the pixel change occurs and is bidirectional (for example: 0 rad = π rad). The histogram is created through adding the weighted gradient magnitude of each pixel in a 16x16 window into one of eight orientation histogram bins. An example gradient orientation histogram can be seen in Figure 3.4.

Lowe's work is adapted for use in our system and acts as a key feature discriminant in zoning headstone engravings and identifying text in the presence of noise and artwork.

2.3 Graph Cut

Graph cut is commonly used in computer vision for segmentation. A widely used implementation by Boykov et al. [2001] efficiently computes the minimum cut on the graph and is used in the proposed system. Graph cut provides for a globally optimal segmentation in which the proximity of the pixels is considered as well as the traditionally used pixel similarity. By using both factors the segmentation is less likely to be affected by local minima caused by noise.

To perform graph cut, an initialization is required for the segmentation of the foreground and background. What constitutes the foreground and background is determined by the application.

Interactive graph cut is commonly used for the segmentation of objects within an image. For initialization, however, the user is required to manually place foreground and background seeds. For segmenting text, the user would be required to place foreground seeds on each character individually, an inefficient and impractical process when hundreds of millions of headstones are involved. Because of this, previous work has been done to determine how to automatically initialize the graph.

A recent approach proposed by Howe [2011] uses the Canny edge detector (Canny [1986]) to cut between nodes where edges exist while assigning weights to the source and sink based on Laplacian values. There is a large potential for noise when using the Laplacian. Additionally, the Canny edge detector contains a number of adjustable parameters, for which manual adjustment is not feasible and could cause variability in its effectiveness across various headstones.

A Bayesian approach is taken by Kuk et al. [2008], which makes a soft decision opposed to the typical binary decision between foreground and background. However, the run-time efficiency of such an approach is inappropriate for our system.

In the proposed system, graph cut is used in two separate processes. The first uses interactive graph cut and based on assumptions on the location of the headstone within the image, places pre-defined seeds on the headstone (foreground) and the surrounding area of grass, tress, etc. (background) to automatically segment the headstone from the surrounding area (Section 3.1.3). The second process that uses graph cut separates the text (foreground) from the artwork and noise (background). Initialization is performed through a connected component analysis. Once the initialization has been performed, the graph is cut to remove the unwanted artwork and noise, resulting in a clean, binarized image (Section 3.4.3).

2.4 Neural Networks

Artificial neural networks are used in many domains for learning a problem given a collection of labelled data. The concept of a neural network is derived from modeling the biological pattern of neurons in which signals are passed through the system and produce a result. This idea is abstracted and simplified to focus on a specific problem and is used for classification purposes. The tutorial given by Jain et al. [1996] explains artificial neural networks in further detail.

The use of artificial neural networks (hereafter termed as neural networks) for image analysis and segmentation is not new and is discussed in a number of papers (Manjunath et al. [1990], Ho and Osborne [1991], Pal and Pal [1993]).

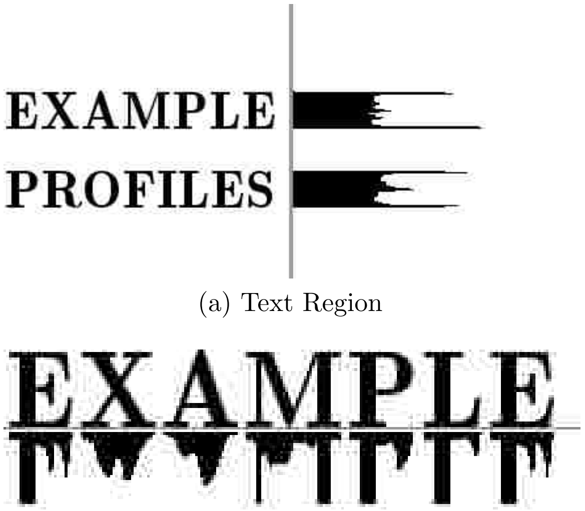
Neural networks are used in the proposed system as part of two separate processes. The first discriminates between the stone texture and engravings (zoning) while the second discriminates between the text and everything else (artwork and noise removal). Features based on the gradient orientation histogram are used in both neural networks with the addition of connected component based features for the artwork removal process.

2.5 Projection Profiling

Projection profiling is a technique used for detecting the location and orientation of a line of text. Early use of projection profiles consisted of rotationally de-skewing scanned documents (Postl [1986], Nakano et al. [1990], Baird [1995]). A projection profile makes use of directional line integrals that provide a description of the text layout in any given direction, particularly in the x and y directions (Figure 2.6). This type of analysis is attractive because of identifiable textual features, such as a common vertical position for the character base and common height. If multiple lines exist, there will also be a clear vertical space separating the two lines. These features are discernible through a projection profile in a horizontal direction (Figure 2.6a).

The leading and trailing edge of the profile delimits and localizes the text. Vertical profiles can then be used to localize each of the characters (Figure 2.6b). A key advantage is the method's simplicity and efficiency.

By analyzing the projection profiles of a text region at multiple angles in the horizontal direction, the image can be rotationally de-skewed. This is done by rotating the image to the angle at which the entropy of the profile is at a minimum. This method is described in Clark and Mirmehdi [2002, 2003] where projection profiles are also used to correct additional skew (azimuth and shear).



(a) Text Region

(b) Text Line

Figure 2.6: Projection profile of an example text region in the horizontal direction (Figure 2.6a) and of an example text line in the vertical direction (Figure 2.6b).

As the layout of cemetery headstones vary greatly, text regions that are contained within the same region (i.e. column) must be located in order to perform skew correction based on projection profiles. To do this, projection profiles are used in a recursive division of the image by alternating the direction (horizontal and vertical) of the profile. Once the text lines are located, rotational skew correction is performed.

2.6 Optical Character Recognition

Although the focus of this thesis is to zone and segment the textual data on a headstone image, the end goal of such processes is to obtain a digital version of the textual content by passing it through an OCR engine. OCR provides a qualitative evaluation of the proposed methods. The vast majority of previous research focuses on zoning and segmenting the text. Scene and graphical text rarely contain non-machine printed characters or uncommon fonts, and therefore, if the text zoning and text segmentation is successful, then the OCR process is greatly simplified, as mentioned previously in section 2.1.3. The success of the text segmentation, however, varies greatly with the inherent difficulty of the image. Little of the existing literature customizes the OCR engine within their system and for their application, but simply uses an out-of-the-box solution with no additional processing.

Understanding the domain of an image and leveraging that knowledge during processing can effectively improve performance. Therefore, the proposed system uses TesseractOCR (Smith [2007]) with a reduced set of classifiable characters. After the transcription is produced, error correction is performed using the reduced vocabulary of headstones. The error correction process is further described with the previous work in data extraction systems in section 2.7.

2.7 Data Extraction

In a contextually aware system, additional pertinent research focuses on data extraction methods. An application to data extraction from Web pages is described by Embley et al. [1999] in which an ontology is created and gives structure between various pieces of information, showing relationships between fields. Each field is then able to claim the information that belongs to it through regular expressions. The same idea can be used to structure information found on headstones while validating and correcting the recognized text based on the defined ontology, thus using the context of headstones for improved results.

A system is described by Packer [2011] in which the use of similar contextual information allows for improved OCR error detection. In the proposed system a post-processing step is performed for contextual validation, using the reduced vocabulary of headstones to correct errors in the transcription.

2.8 Summary

Three areas of literature for recognizing text have been discussed which propose solutions for scene text, graphic text, and document text. Previous work for scene text generally focuses on having a clean background, while previous work in graphic text generally focuses on having a strong contrast between text and background. Neither a clean background nor a strong contrast can be expected when analyzing headstone images and therefore imply the need for a specialized system in which low contrast between text and a noisy background can be accurately zoned, segmented, and recognized.

Such a system is proposed in which the problem is reduced over multiple steps to produce a more workable solution. Gradient orientation histograms are used to zone the headstone engravings and remove the noisy background allowing for a more accurate segmentation using Otsu binarization. Gradient orientation histograms are also used in conjunction with graph cut to remove additional noise and artwork after segmentation. The resulting binarized image is then passed to TesseractOCR followed by contextual validation to correct errors using the reduced vocabulary of headstones. Thus, the proposed system leverages the previous work of gradient orientation histograms, graph cut, and neural networks to address the difficult challenges of zoning and segmenting headstone data.

Chapter 3

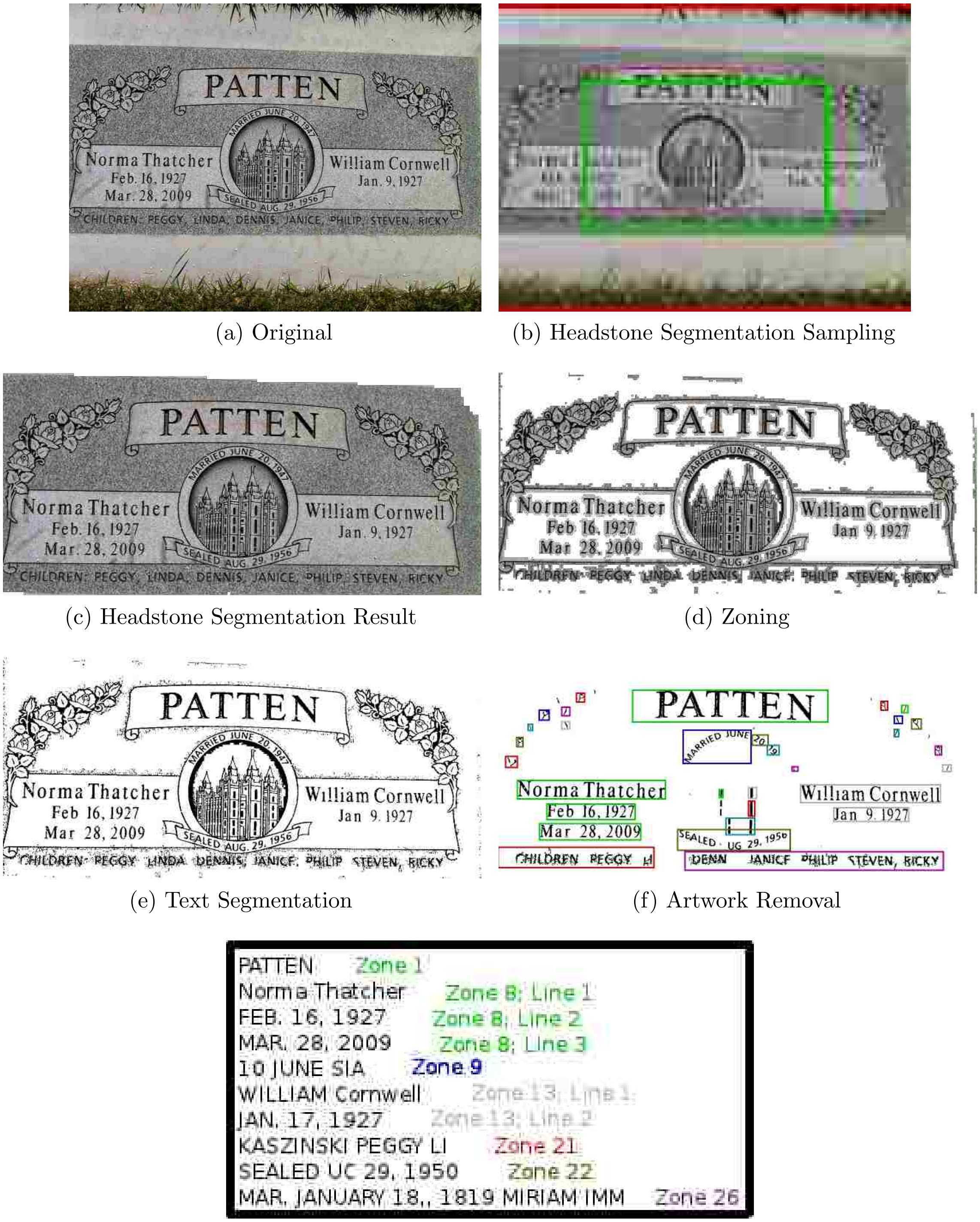
Methods

Headstone images are inherently noisy as shown in Figure 1.1. To index the valuable information found on headstones, a process is needed in which the various aspects of noise, including artwork, must be removed to provide a transcribable image to the OCR engine. This chapter discusses a novel approach that uses gradient orientation histograms, neural networks and graph cut to zone and segment textual information on headstone images.

Additionally, a simple process is performed to error-correct OCR results to improve recognition accuracy and to demonstrate the added benefit of constraining the vocabulary to the domain of headstones.

3.1 Pre-Processing

To provide for an efficient process of zoning and segmentation of the text found on a headstone image, two preprocessing steps are applied. The system first creates an image pyramid allowing for processing on a lower resolution image with easy mapping to the original high resolution image. Secondly, the headstone is separated from the surrounding background using a minimum graph cut algorithm with automated seeding of the foreground and background, thus removing unnecessary data while isolating the headstone from the surrounding environment (Figure 3.1).



(a) Original

(b) Headstone Segmentation Sampling

(c) Headstone Segmentation Result

(d) Zoning

(e) Text Segmentation

(f) Artwork Removal

```
PATTEN      Zone 1
Norma Thatcher      Zone 8; Line 1
FEB. 16, 1927      Zone 8; Line 2
MAR. 28, 2009      Zone 8; Line 3
10 JUNE SIA      Zone 9
WILLIAM Cornwell      Zone 13; Line 1
JAN. 17, 1927      Zone 13; Line 2
KASZINSKI PEGGY LI      Zone 21
SEALED UC 29, 1950      Zone 22
MAR. JANUARY 18,, 1819 MIRIAM IMM      Zone 26
```

(g) OCR with Contextual Validation

Figure 3.1: A visual summary of the proposed system: the original headstone image (Figure 3.1a), the foreground (green) and background (red) samples taken for headstone segmentation (Figure 3.1b), the headstone segmented from the surrounding background (Figure 3.1c), the zoned engravings of the headstone (Figure 3.1d), the headstone engravings segmented from the stone texture (Figure 3.1e), the artwork engravings removed (Figure 3.1f), and the transcription after a contextual validation (Figure 3.1g) where text lines are mapped to the image regions in Figure 3.1f.

3.1.1 Image Pyramid

Image pyramids are commonly used in computer vision algorithms to increase efficiency. The image pyramid is created by reducing the original image by half its size repeatedly until either the width or the height measures less that 100 pixels (see Figure 3.2). Any smaller headstone image will not contain sufficient information for processing.

The image at the top of the pyramid (lowest resolution) is used for headstone segmentation (section 3.1.3) and multiple levels are used for zoning (chapter 3.2).

Using the lowest resolution image for headstone segmentation dramatically improves performance reducing the process in general from eight seconds on the original image to less than a second. Additionally, the use of multiple levels of the pyramid facilitates clean zoning and is described in chapter 3.2.

In both cases, the loss in resolution at the top of the pyramid causes the stone texture to reduce to a more consistent texture and to become simpler to segment and classify, resulting in a more accurate segmentation and zoning.

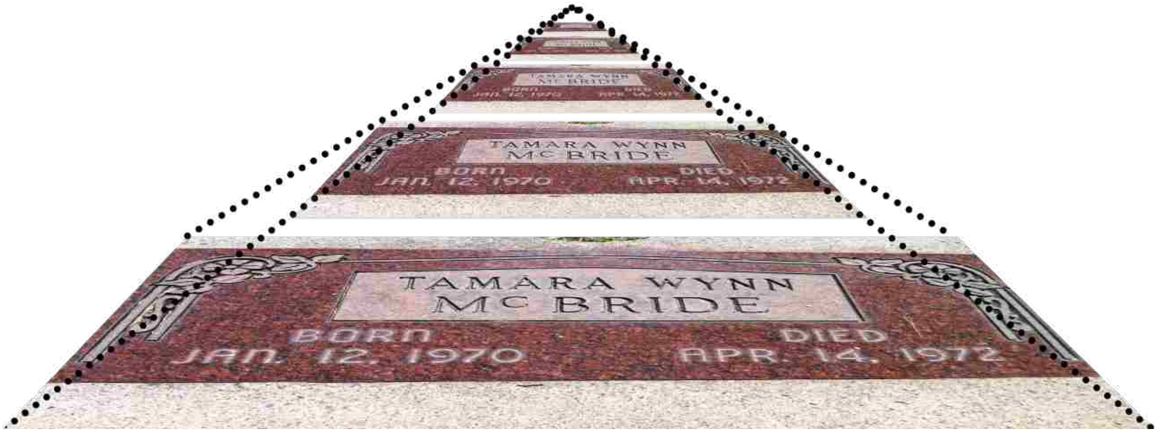


Figure 3.2: A representation of an image pyramid in which each level of the pyramid was created by scaling the previous level by 50 percent. The shown pyramid was created from the image used in Figure 2.5.

3.1.2 Computation of Gradients

The gradient orientation histogram is a key feature used in the neural networks for both zoning and artwork removal. Although the image regions that are used to create the histogram in each of these processes differ in size and shape, the original gradient magnitude and phase are needed to compute the histogram in each case. Therefore, as part of preprocessing the gradient magnitude and phase are computed for the full image and stored, rather than be computed multiple times at the zoning and artwork removal steps.

To calculate the gradient orientation histogram of a given area, 3x3 Sobel kernels are convolved with the image to find gradients in the x and y directions. If the input image is a gray-scale, single-channel image, no additional logic other than typical convolution is needed. However, input images are likely to be color images with three channels (red, green, and blue). The three channels provide additional information that give greater ability to detect gradients, therefore, the 3x3 Sobel kernel is applied to each channel individually. At each pixel the largest of the three computed gradients is preserved as that pixel's gradient.

After the gradients are individually computed in both the x and y dimensions, the magnitude and orientation are computed with the following equations, respectively:

$$\|\nabla I\| = \sqrt{\nabla I_x^2 + \nabla I_y^2} \tag{3.1}$$

$$\phi(\nabla I) = \tan^{-1} \frac{\nabla I_y}{\nabla I_x} \tag{3.2}$$

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

3.1.3 Headstone Segmentation

Images of headstones often contain surrounding objects that are not part of the headstone. For example, the surrounding landscape or surrounding upright headstones are visible (Figures 3.3a, 3.3d) in many of the captured images found on the existing systems (i.e. BillionGraves

[2011], Find A Grave [2011]). The surrounding background is unnecessary for the segmentation of text and is therefore discarded. To accomplish this, automated graph cut segmentation is performed at the top of the image pyramid on a low resolution image of the headstone.

As discussed previously in section 2.1.1, requirements are made that the headstone be the object of focus when the image is captured. Therefore, the assumption can be made that the headstone is centered and that the desired data is completely bounded within the image. With these assumptions foreground and background seeds are placed automatically.

Foreground seeds are placed using four connected lines that form a quadrilateral that is centered with respect to the image (Figures 3.3b and 3.3e). To ensure that no essential information (such as text) is lost when performing this automated graph cut, a generous width and height for the quadrilateral is chosen. The length of the lines forming the quadrilateral are calculated according to the orientation of the image. If the image width is larger than the image height, the following calculations are used:

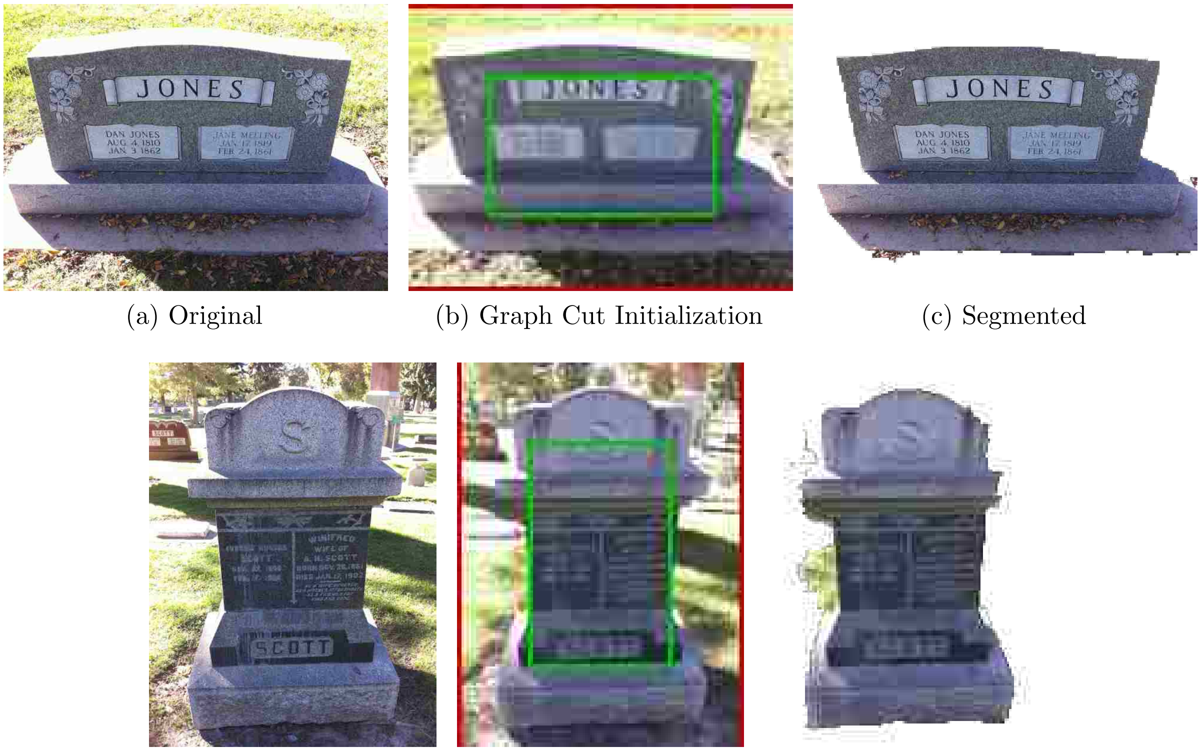
$$l_x = P_{long} \times I_{width} \tag{3.3}$$

$$l_y = P_{short} \times I_{height} \tag{3.4}$$

where l_x and l_y are the lengths of the lines in their respective dimension, I_{width} and I_{height} are the image's width and height respectively, and P_{long} and P_{short} are percentages specified by the application and are reversed when the image height is larger than the width. In the current implementation $p_{long} = 0.60$ and $p_{short} = 0.50$.

Background seeds are placed on the two long edges of the image. No background seeds are placed on the short edges as the headstone will likely fill the image from side to side if captured to fill the image plane.

Once the seeds are in place on the lowest resolution image, graph cut is performed to segment the headstone. As this operation is performed only at the top of the pyramid, the results are propagated through the remainder of the pyramid. This is done by using the segmentation of the previous level as a mask for the following level. The segmentation mask from the previous is enlarged to align with the image and all pixels labelled as background are set to be transparent. Results of this process are shown in Figure 3.3.



(a) Original (b) Graph Cut Initialization (c) Segmented

(d) Original (e) Graph Cut Initialization (f) Segmented

Figure 3.3: The removal of a headstone's surrounding background based on automated seeding of the interactive graph cut algorithm. In figures (b) and (e) the green lines represent the foreground seeds and the red lines represent the background seeds.

Through propagating the segmentation from the top of the pyramid down to the bottom, the process is reduced from 10.749 seconds (performing graph cut on the bottom of the pyramid) to 0.093 seconds on average.

3.2 Zoning

Zoning is performed to simplify the problem and gain knowledge of where the engravings are located within the headstone image. Much of the stone texture can be removed through a texture based approach that differentiates between engravings and the stone upon which they were engraved.

Texture based approaches to text zoning rely on the unique characteristics of text versus the surrounding objects. Gradient orientation histograms capture the characteristics that can distinguish the engravings from the stone texture. Such features are used as inputs to a neural network to calculate a score as to how similar that region is to either a text region or a background region. The feature extraction process, use of the neural network, and final classification through cascaded zoning are discussed in their respective order.

3.2.1 Feature Extraction

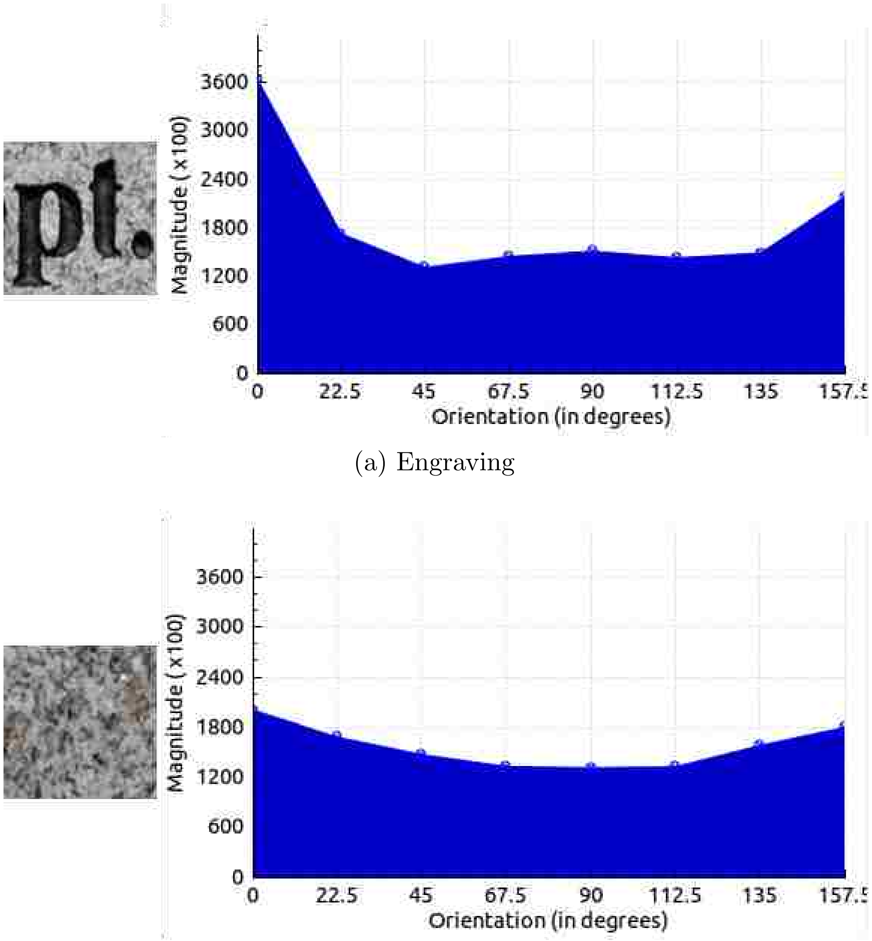
The extraction of features is a useful means to reduce the often redundant data represented in an image. Features also allow for flexibility in choosing which image areas are important to the given system. The proposed system uses features derived from the gradient orientation histogram of an NxN image region.

Gradient Orientation Histogram

A key feature that separates the headstone engravings and background is found to be the gradient orientation histogram. Inspired by Lowe's use of a gradient orientation histogram (Lowe [2004]), the proposed system uses a histogram in which the orientations are weighted by the gradient magnitude. It is found that in quantizing the weighted orientations, large peaks appear and allow for discrimination between foreground (engravings) and background (stone).

The histograms for background areas consist of a wide range of gradients, but of random orientation, causing a nearly flat distribution across the histogram (Figure 3.4b).

Engravings are characterized by consistent gradients at particular orientations, causing peaks to form within the histogram (Figure 3.4a).



(a) Engraving

(b) Background

Figure 3.4: A comparison between the two classifications in the text zoning process where Figure 3.4a shows a region of a headstone image that contains engravings and its resulting gradient orientation histogram. Figure 3.4b shows a region of the same headstone image that contains only the stone texture.

The orientation ($\phi(\nabla I)$) will be a value in the range of $(-\pi, \pi]$. To compute a gradient orientation histogram the orientation must be discretized to a reasonable scale. The work presented by Lowe [2004] uses a histogram of eight bins: one bin for every $\frac{\pi}{4}$ radians. In the proposed system, the sign of the gradient orientation is not necessary and therefore any negative orientations are changed to their corresponding positive orientation:

$$\phi(\nabla I) = \phi_{neg}(\nabla I) + \pi \tag{3.5}$$

where $\phi_{neg}(\nabla I)$ represents a negative orientation value. It is necessary to add π radians rather than a simple absolute value to preserve the correct orientation.

The gradient orientations are quantized into n bins for creation of the histogram. To determine the quantized orientation ($\phi_q(\nabla I)$) the following equation is used:

$$\phi_q(\nabla I) = \lfloor n\frac{\phi(\nabla I)}{\pi} + \frac{\pi}{n}\rfloor \tag{3.6}$$

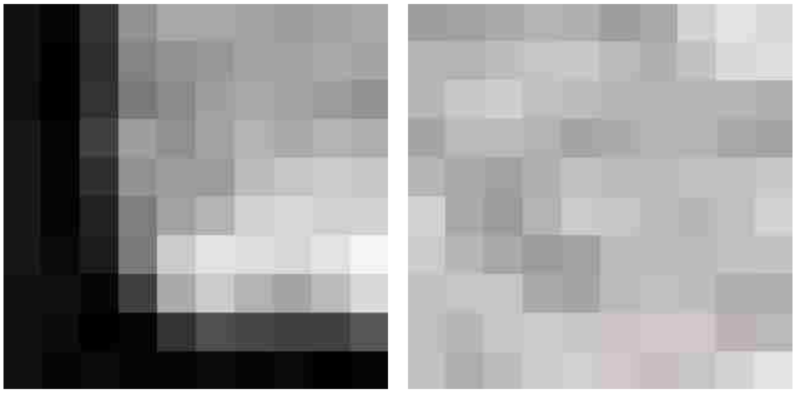
The first term in equation 3.6 converts the orientation values to the range [0, n] and the second term accounts for a shift that centers each bin at a factor of ($\frac{\pi}{n}$), thus a common value such as $\frac{\pi}{2}$ will be the center of a bin rather than an edge case. Any overflow caused by shifting the values is appropriately wrapped to bin 0. The number of bins n for the proposed system is set at eight, however this may be adjusted if more granular histograms are deemed beneficial.

Additional Feature Selection

Experiments were performed to determine additional features that can be included with the gradient orientation histogram for improved performance. The experiments consisted of training separate neural networks for each combination of additional features according to the method described in section 3.2.2. An average accuracy was computed from one hundred trained networks each created through the training process described in section 3.2.2.

The experimental features consisted of color information, a single gradient vector for the region (with separate parts of magnitude and orientation), a global gradient measure of the headstone image, the cumulative histogram magnitude, and the histogram magnitude variance. Only the cumulative magnitude ($\Sigma\|\nabla I\|$) and variance (σ^2) of the histogram magnitude showed improved performance and are the only features used in addition to the gradient orientation histogram. Although both the cumulative magnitude and variance are based on values already provided separately by the histogram bins, they supply a meaningful combination of the histogram values that individual histogram bins do not, thus improving the neural network.

Two example 10x10 regions of a headstone image are shown in Figure 3.5 with the corresponding feature vector shown in Table 3.1. Figure 3.5a has strong and consistent gradients resulting in large values at 0° and 90° (common orientations for engravings) when compared to Figure 3.5b. Additionally, both the mean and variance of the gradients in Figure 3.5a are significantly larger those in Figure 3.5b. This distinction among the various regions of a headstone image allows for classification between engraved regions and the stone texture.



(a) Engraving

(b) Stone Texture

Figure 3.5: Two regions of a headstone image. Figure 3.5a contains the edge of an engraving and Figure 3.5b contains no engravings (stone texture). The feature vector extracted from each image is shown in Table 3.1.

Figure	**0°**	**22.5°**	**45°**	**67.5°**	**90°**	**112.5°**	**135°**	**157.5°**	$\mathbf{\Sigma\|\nabla I\|}$	$\boldsymbol{\sigma^2}$
3.5a	5038	724	1002	2120	7313	2225	2116	4135	24673	4414490
3.5b	819	278	624	2621	992	1660	1065	750	8809	467981

Table 3.1: The original feature vectors generated from the headstone image region shown in Figure 3.5a (row 1) and Figure 3.5b (row 2).

3.2.2 Neural Network

A neural network is a powerful classification tool that is able to learn a decision boundary given a collection of labelled training data. This allows the system to learn how to best classify the supplied data with no additional training needed at runtime. The training and testing of the neural network are described in detail below.

Training

The neural network takes as input the ten features discussed in the previous section: the gradient orientation histogram bins, the cumulative histogram magnitude, and the histogram variance. The cumulative magnitude and variance will both be larger than any of the histogram bins (see the example feature vector shown in Table 3.1), resulting in a skewed feature vector. Therefore, the feature vector is normalized to the range [0,1] to avoid dominance by the variance values in neural network training. To normalize the histogram bins, the largest single bin value of the training set is used as the divisor to all other bin values. To normalize the cumulative magnitude and the variance, the same technique is used where the maximum value for that feature is used as the divisor. The normalized version of the feature vectors shown in Table 3.1 is shown in Table 3.2 where only the two feature vectors shown are considered for the normalization process.

Figure	**0°**	**22.5°**	**45°**	**67.5°**	**90°**	**112.5°**	**135°**	**157.5°**	$\boldsymbol{\Sigma\|\nabla I\|}$	$\boldsymbol{\sigma^2}$
3.5a	0.689	0.099	0.137	0.290	1.000	0.304	0.289	0.565	1.000	1.000
3.5b	0.112	0.038	0.085	0.358	0.136	0.227	0.146	0.103	0.357	0.106

Table 3.2: The scaled feature vectors generated from the feature vectors shown in Table 3.1.

The number of hidden nodes is determined by the common practice of doubling the number of input nodes (ignoring the bias node). Since ten input nodes are used in the current implementation, the number of hidden nodes is equal to twenty. A single output node is used for classification.

A neural network is created and trained for every level in the image pyramid and uses the same 10x10 size for the image regions. In preparation for training, a dataset of labelled headstone regions is randomly separated into three separate datasets: 10% test, 10% validation, and 80% training. The test set is used to determine the final accuracy of the trained neural network. The validation set is used to determine the accuracy of the network after each iteration of training. This iterative measure of accuracy over the validation set is used to determine when training should end, ensuring that sufficient training is performed as well as avoiding over-fitting on the training set. The training set is supplied to the neural network where the output is compared to the labelled (ground truth) classification for that instance. The error is then propagated backwards (back-propagation) through the network to update its internal weights, thus learning the correct weighting within the network to minimize the classification error.

The neural network uses every data point (image subregion) to compute an output and self-correct through back-propagation where one iteration through all the data points is called an epoch. After every epoch, the accuracy is computed on the validation set. When the validation set does not improve within n iterations, the training ends. In the current implementation, $n = 20$.

Classification

Given a headstone image, the task of text zoning is to coarsely determine the regions wherein text lies. In the proposed system, all headstone engravings are zoned (including artwork), but in a later step are refined to include only text. This is done to enhance the proposed system's ability to separate text in a noisy and unpredictable environment. A separate neural network is trained to focus on the task of artwork removal (section 3.3).

The headstone image is sub-divided into 10x10 regions (the same region size as that defined in training) with 50% overlap both vertically and horizontally (Figure 3.6). The overlapping regions cause, with the exception of the border cases, every pixel in the image to be part of four separate regions and are thus classified four separate times. Every region has the defined features extracted (section 3.2.1) which are then supplied to the trained neural network (3.2.2). For every pixel the four outputs from the neural network are averaged for a final classification. A threshold t is specified where $background < t \leq foreground$. The typical use of a neural network is to set $t = 0.5$ where the outputs range is $[0, 1]$.

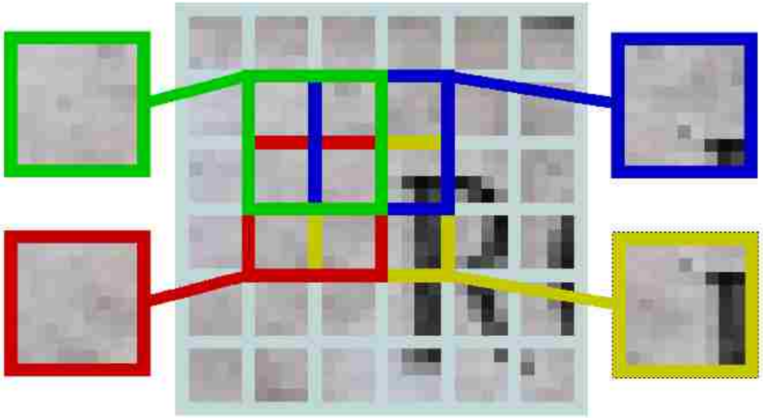
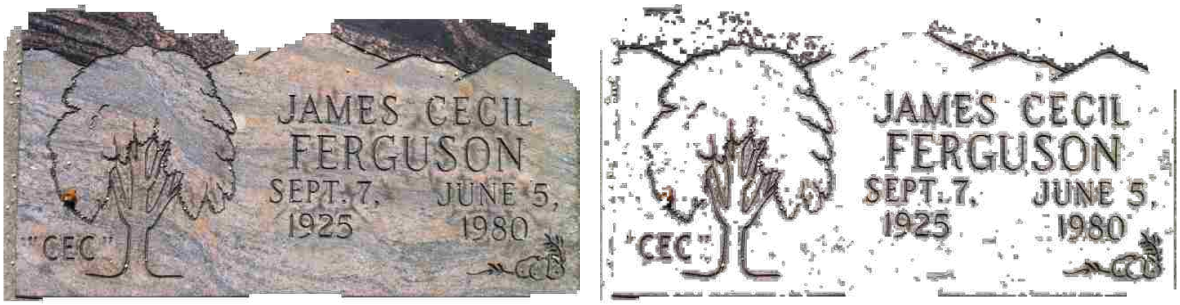


Figure 3.6: An example of the overlapping regions used for classification in the Text Zoning process.

The overlapping regions result in smaller separately classifiable regions of 5x5 ($\frac{10}{2}$x$\frac{10}{2}$), increasing the granularity of the classification. An example classification of an image is shown in Figure 3.7 where regions classified as $foreground$ retain their original pixel values and those classified as $background$ are removed (white).



(a) Original

(b) Zoned (Without using the cascade)

Figure 3.7: A comparison of the original image (Figure 3.7a) and the resulting zoned image (Figure 3.7b) without using a cascade (discussed in section 3.2.3).

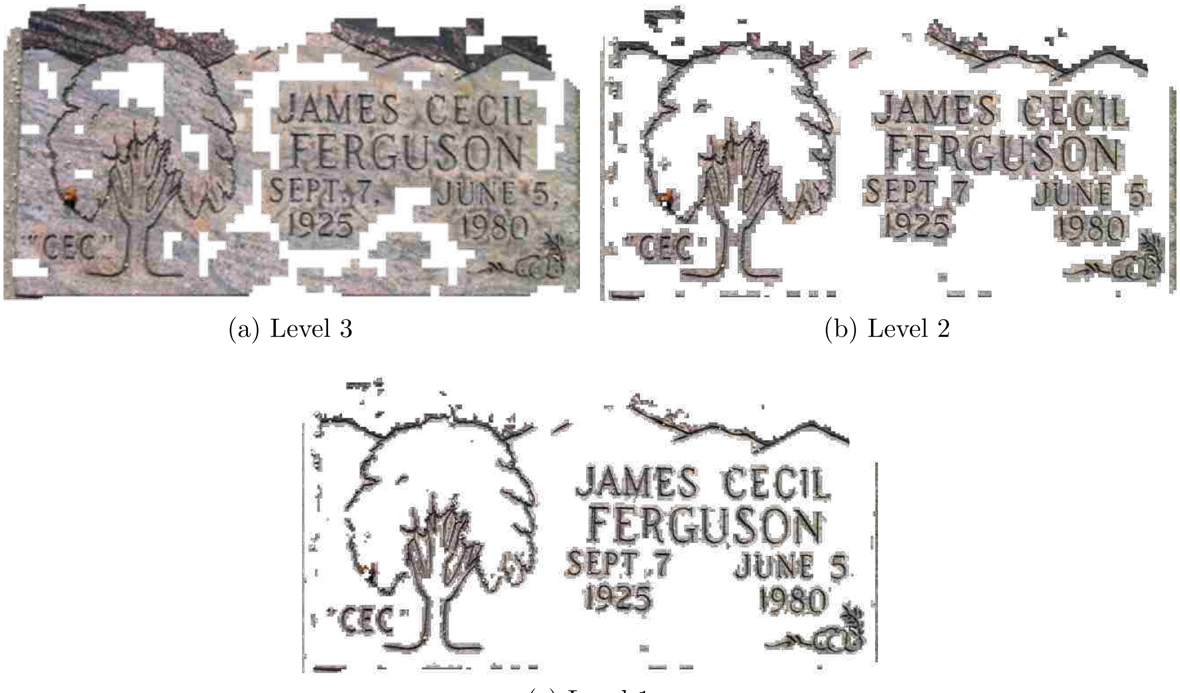
3.2.3 Cascaded Zoning

In Figure 3.7b we see that much of the background was removed, however, a large number of flecks of stone were preserved. At every level within the image pyramid, the gradients within the image are altered producing slightly varying histograms for the same relative region within a given image. The proposed system uses the Qt library (Molkentin [2007]) for scaling the images which uses an adaptive interpolation algorithm. As the original image is reduced in size, regions where the gradients are of a smaller magnitude are blurred whereas the regions with larger gradients are maintained. This has the effect of reducing the noisy stone texture as the image is reduced in size while preserving true edges. The proposed system removes the extra noise through zoning the image in a cascaded fashion from the top of the image pyramid down.

The levels of the image pyramid used for the zoning process is reduced to three levels. The top of the pyramid, which was used for headstone segmentation, is found to be too low of a resolution to recognize engraved regions, resulting in many false negatives. The bottom of the pyramid (the original full resolution image) is found to be too high of a resolution in that too much detail of the stone and characters cause many incorrect classifications. Therefore the reduction in the image pyramid from five levels to three (removing the top and bottom pyramid images) in the zoning process improves classification accuracy as well as efficiency.

Although the image pyramid was reduced, using multiple levels of images in the pyramid is necessary. If only the bottom level were to be used, added noise would exist in the result such as that in Figure 3.7b. If only the lowest or middle levels of the pyramid were used, the zoning would be too coarse, retaining a large amount of stone texture. If strong flecks of stone are found in the stone texture between characters, it is possible that the characters will be joined, reducing the OCR engine's ability to accurately recognize it. This can be seen in Figure 3.9 where the stone texture between the characters was not removed and thus connects the characters. As mentioned previously in section 3.2.2, a separate neural network is trained for each level in the image pyramid.

The smallest image in the image pyramid passes through the classification process described previously. All regions classified as $foreground$ are retained where regions classified as *background* are removed from any further processing. Being the first image in the cascade, the proposed system is careful to not remove any potential text regions. It may be the case that through the scaling of the image, low contrasting regions of text may have had their gradient strength reduced as well. Therefore, the threshold t is increased to allow for a greater amount of stone texture to be retained and reduce the risk of erroneously removed text. At each level in the image pyramid, this threshold is relaxed slightly until the bottom of the pyramid is reached, where the threshold is set to the typical value of $t = 0.5$. The images at each step in this process is shown in Figure 3.8.



(a) Level 3

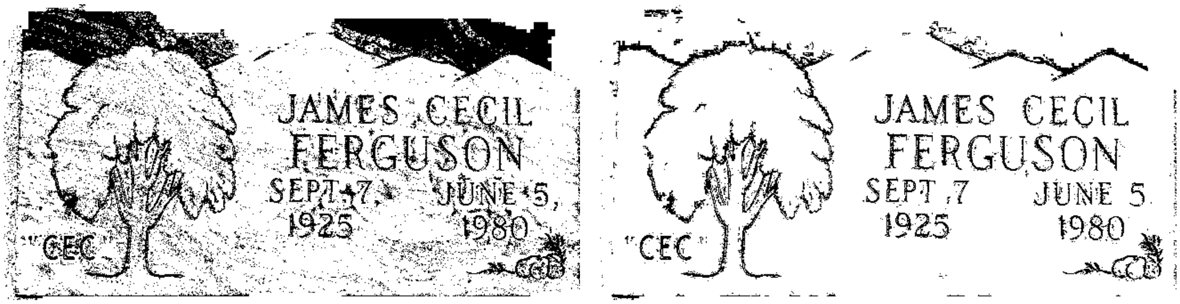
(b) Level 2

(c) Level 1

Figure 3.8: The images at each level of the image pyramid as the cascaded zoning is applied. The result is shown in 3.8c and can be compared with Figure 3.7b to show the advantage of using the cascade.

3.3 Text Segmentation

The segmentation of the text is done through binarization of the image. The traditional Otsu algorithm is used. After the background has been removed, a more accurate binarization is possible (Figure 3.9).



(a) Binarized with no zoning

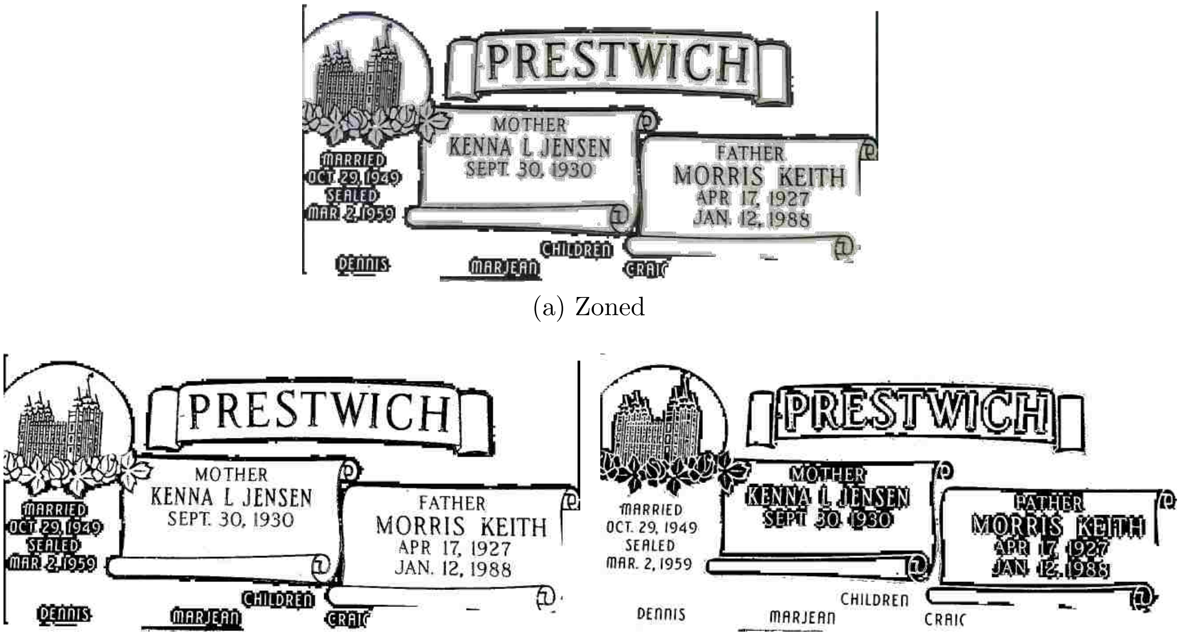
(b) Binarized with zoning

Figure 3.9: A comparison of the same image binarized with no zoning (3.9a) and binarized after zoning (3.9b) both using the standard Otsu algorithm.

Binarization is performed only on the highest resolution image in the pyramid (of the reduced pyramid discussed in the previous section). This allows for greater efficiency with no cost to accuracy. The remaining processes likewise have no need to use multiple resolutions of the image, thus only one image is retained for further processing.

For a number of headstone images, it is the case that the headstone contains text that is lighter in color than the background. A single pass of the Otsu binarization algorithm is insufficient to capture both dark text on a light background and light text on a dark background. Therefore, a second pass is used to invert the Otsu binarization. Thus, two images are produced in the text segmentation process, both of which are used in the remaining process. Figure 3.10 shows a headstone image in which both light and dark text are present with the resulting binarized images.

The reader may note that the areas seen as clear text in one image such as “PRESTWICH” in Figure 3.10b is also present and readable in Figure 3.10c. However, the inverse binarization has caused the individual characters to be connected, resulting in a large single



(a) Zoned

(b) Binarized

(c) Inverse Binarized

Figure 3.10: The images resulting from the text segmentation process. The zoned image (3.10a), the binarized image (through the traditional Otsu binarization) (3.10b), and the image binarized through an inverted Otsu binarization.

connected component which will likely be classified as artwork in the process described in section 3.4, thus retaining only the clean text in each of the images.

3.4 Artwork Removal

Through zoning and segmenting the text and artwork in the headstone image, the amount of noise has been greatly reduced with only text and artwork remaining. However, If such an image were passed to an OCR engine “as-is”, it is likely that, due to the artwork, the OCR engine will produce a number of phantom characters in the transcription. Thus, the artwork also needs to be removed to provide for an accurate transcription of the text. The artwork is removed through use of connected components and a combination of a neural network and graph cut.

3.4.1 Connected Components

The analysis performed for removing artwork is done on a connected component basis. Connected components are created and labelled through the basic flood-fill algorithm. Both the bounding box and the pixels included in the connected component are used in the artwork removal process.

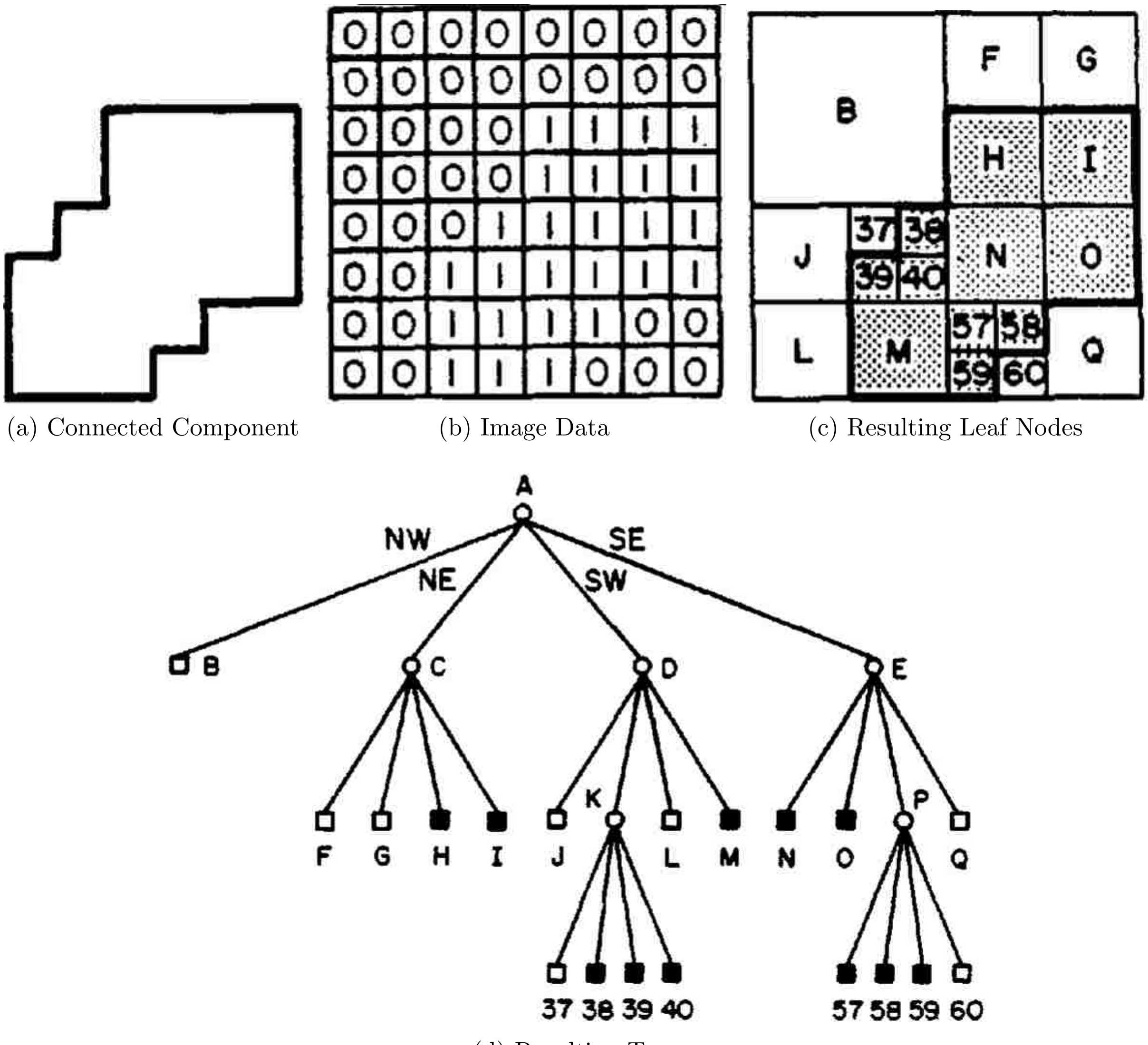
3.4.2 Quadtree

Graph cut acts as the final classifier between artwork and text. To create a graph upon which graph cut can operate, a quadtree is used. The tree is formed recursively using the full binarized image as the root node in the tree. If more than a single connected component is contained within that node, it is subdivided into four equal parts (forming the children of that node, see Figure 3.11). Thus, every leaf node in the tree contains either zero (node 'B') or one connected component (node 'H'). Additionally, it is noted that a single connected component may span across multiple leaf nodes (i.e. in Figure 3.11, all nodes colored gray contain a portion of the same connected component).

Using only the resulting leaf nodes, a graph is formed where each node is of variable dimensions, but guaranteed to be a rectangle. In forming a graph it is necessary to have an understanding of who the neighbors of each node are. In a quadtree, this is possible through using the structure of the quadtree to find the lowest common ancestor within the tree. When that ancestor is found, a mirrored traversal back down the tree can be used to find the neighboring node. This approach is described in Samet's survey on quadtrees (Samet [1984]).

3.4.3 Graph Cut

The graph cut algorithm requires that an initialization be made in the weighting of both the t-links (between the source and sink) and the n-links (between neighboring nodes). All measures used for the initialization of the weights between nodes are based on the connected



(a) Connected Component

(b) Image Data

0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	1	1	1	1
0	0	0	0	1	1	1	1
0	0	0	1	1	1	1	1
0	0	1	1	1	1	1	1
0	0	1	1	1	1	0	0
0	0	1	1	1	0	0	0

(c) Resulting Leaf Nodes

(d) Resulting Tree

Figure 3.11: An example connected component (Figure 3.11a), the image data containing the connected component (Figure 3.11b), the resulting image regions (leaf nodes) generated by quadtree creation (Figure 3.11c) and the generated quadtree (Figure 3.11d). The figures shown are taken from Samet [1984].

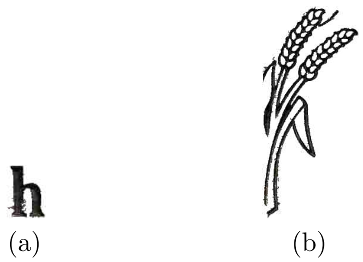
component found within the node of interest and their similarity to neighboring connected components.

In this and following sections, a visual representation of the graph cut weightings are shown (such as in Figure 3.13). In these images the t-link weightings are colored green if classified as foreground and red if classified as background. The intensity of the colors represents the strength of that weighting where black represents a zero weight for both foreground and background. The n-links are shown as the edges of these nodes where the edge between two nodes represents the n-link weighting between the them. The intensity of the edge inversely represents the weighting. When white, the nodes are maximally dissimilar and when black, the nodes are very similar or the same.

t-links

The weightings given to t-links represent how similar the node is to either the source (text) or the sink (artwork). To classify the nodes, a neural network is used to compute confidence scores. Individual connected components from the binarized image are used as the sample regions from the image, however, the points belonging to the connected component are mapped back to the original image allowing for feature extraction from the full color image. This mapping to the original image preserves the gradient information and is subsequently used for the feature extraction process.

Similar to the zoning process, the gradient orientation histograms prove to be a useful tool for distinguishing text from surrounding regions. Textual gradients tend to accumulate in the histogram bins representing the straight vertical gradient (90°) or the straight horizontal gradient (0°) whereas artwork gradients form peaks, but of a more random nature (artwork is not constrained to a set of alphanumeric characters). Additionally, artwork is typically larger in size than text while having the same gradient strengths resulting in a greater cumulative magnitude. Therefore, the same 10 features (each gradient orientation histogram bin, the cumulative histogram magnitude, and the histogram variance) are also used in the neural



(a) (b)

Figure 3.12: A comparison of two connected components of text (Figure 3.12a) and of artwork (Figure 3.12b). The resulting feature vector is given in Table 3.3.

network for removing artwork. However, five additional features are also used to leverage the information found from connected components.

The connected components provide information that previous systems such as Gatos et al. [2005] have used for similar analysis. Likewise, the following connected component features are used in addition to those based on the gradient orientation histogram:

- width (w)
- height (h)
- aspect ratio ($r_a = \frac{w}{h}$)
- area ($A = w \times h$)
- density ($D = \frac{n}{A}$, where n is the number of pixels in the connected component)

Therefore, fifteen features in total are used in the neural network. Similar to the process described in section 3.2.1, feature selection was based on a number of experiments and their resulting accuracy.

The neural network is trained in the same manner described in section 3.2.2. Additionally, the features are extracted from the connected components as a whole (variable dimensions and shape) as opposed to the fixed 10x10 region used for zoning. Two example connected components from the same headstone image are shown in Figure 3.12 and their corresponding feature vectors are shown in Table 3.3.

Figure	**0°**	**22.5°**	**45°**	**67.5°**	**90°**	**112.5°**	**135°**	**157.5°**	$\boldsymbol{\Sigma\|\nabla I\|}$	$\boldsymbol{\sigma^2}$
3.12a	601.9	155.0	106.7	144.4	301.2	174.8	198.4	337.8	2020.2	2.29×10^{08}
3.12b	6350.6	7151.5	4790.9	2157.6	1387.0	993.4	1210.0	2621.9	26662.9	5.18×10^{10}

Figure	$\boldsymbol{w}$	$\boldsymbol{h}$	$\boldsymbol{r_a}$	$\boldsymbol{A}$	$\boldsymbol{D}$
3.12a	44	61	0.72	2684	955.74
3.12b	161	357	0.45	57477	5211.53

Table 3.3: The feature vectors generated from the two connected components shown in Figure 3.12. Each value is scaled to $\frac{1}{100}th$ of its original value.

Features are extracted for every connected component and passed to the neural network for the same reasons described in the zoning process (section 3.2). Similar to the neural network used in the zoning process, the feature vector is scaled to the range of $[0, 1]$.

The output of the neural network will be in the range $[0, 1]$. When the output nears 0 or 1, more confidence is had by the neural network that the component is artwork or text, respectively. An important aspect in the artwork removal process is that the output is not the final classification. The confidence scores are used to determine the t-link weightings for graph cut as follows:

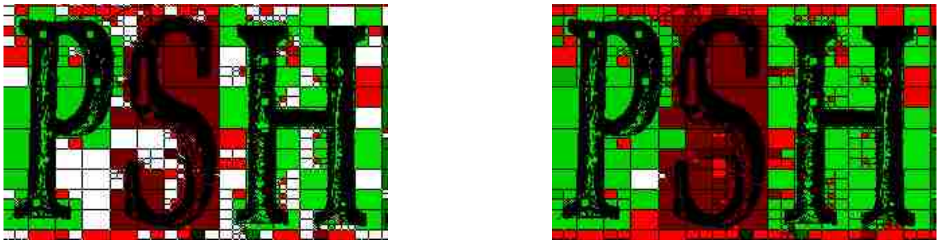
$$\omega_{t_{n_1}} = max\{2\omega_{max_t} * (O - 0.5), 0\} \tag{3.7}$$

$$\omega_{a_{n_1}} = max\{-2\omega_{max_t} * (O - 0.5), 0\} \tag{3.8}$$

where $\omega_{t_{n_1}}$ and $\omega_{a_{n_1}}$ represent the text weighting and artwork weighting respectively. In both cases, the output is scaled to the range $[0, \omega_{max_t}]$, with a distinction as to where the weight is given. If the output is greater than 0.5, the output is scaled and given to the text weighting ($\omega_{t_{n_1}}$), if the output is less than 0.5, then the reverse is true.

The creating of nodes through a quadtree, however, will leave many nodes that are empty (contain zero connected components). Since no classification can be done for empty nodes that have no features, there is the question of how to initialize such nodes in the graph. Any weighting given will favor either the background or the foreground. Therefore, they are treated innocuously by assigning the n-link weightings with the minimum weight (0).

However, innocuous nodes will form large gaps between neighboring connected components (see the white nodes in Figure 3.13a). These gaps cause a loose connection between neighboring connected components. For graph cut to be effective in retaining misclassified text, the nodes containing the characters must be strongly linked together.



(a) Weightings of nodes

(b) Weightings of merged nodes

Figure 3.13: A comparison of the original t-link weightings (Figure 3.13a) and the t-link weightings after empty nodes are merged with the neighboring node with the strongest confidence in its classification (Figure 3.13b).

To solve this problem, if a node is empty, an analysis on the t-link weighting of each of its neighbors is performed to find which node to merge with:

$$N' = \arg\max_{\eta \in \mathbf{N}} \{\omega \times \eta_n\} \tag{3.9}$$

where N' is the neighbor to merge with, $\mathbf{N}$ is the set of all neighboring nodes, ω is the t-link weighting regardless of its classification, and η_n is the number of pixels in the connected component of node η.

The merging of two nodes is a simple assignment of the connected component from node N' to the empty node. A flag also must be set on the previously empty node to indicate that the node does not have a true connected component to avoid selecting a previously empty node for N'.

It is important that assigning t-link weights and merging empty nodes be done prior to assigning n-link weights. By so doing, the n-links will be assigned by also regarding the previously empty nodes.

The result is a graph that is more tightly connected and representative of the objects desired for segmentation (Figure 3.13b). This process could be repeated to remove all nodes originally without a connected component if needed. However, the characters found on cememtery headstone images are near enough in proximity that this is not needed.

n-links

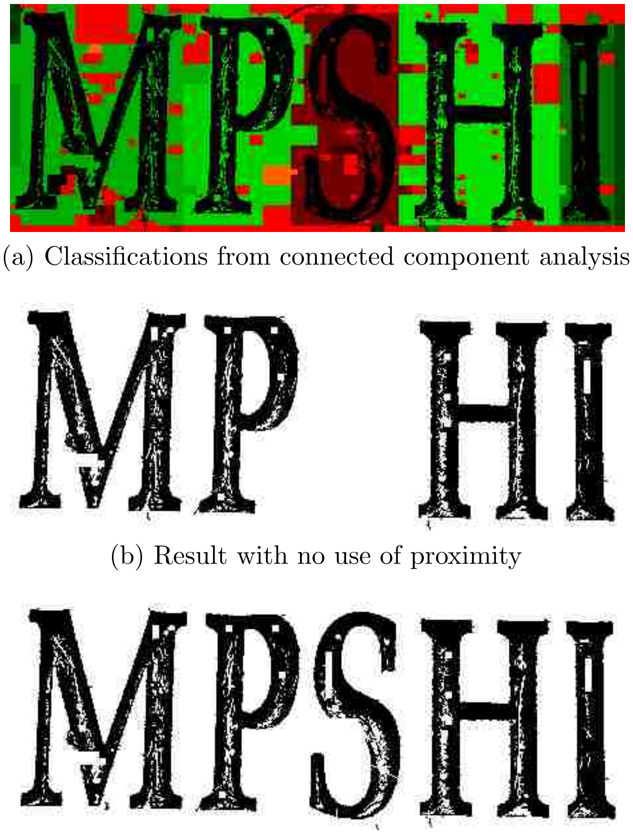
Traditionally the distinction between text and artwork/noise is done through a connected component analysis such as that described in Gatos et al. [2005]. The proposed system is unique in that the globally optimal graph cut algorithm is used in conjunction with a connected component analysis. Such an approach considers not only the component's similarity to text, but also the textual similarity of neighboring components. Thus, a misclassification in a traditional connected component analysis (Figures 3.14a and 3.14b) is more likely to be retained in the presence of surrounding correct classifications (Figure 3.14c).

The initialization of the n-links is done through calculating a similarity score between a node and its neighbor. The similarity score is based on the similarity between the contained components in four areas: pixel count (S_n), baseline (largest y-value of within the component) (S_y), width (S_w), and height (S_h). All four of these similarity scores are computed relative to the connected components being compared. They are computed as follows:

$$S_n = 1 - \frac{\Delta n_{c1,c2}}{\min\{n_{c1}, n_{c2}\}} \tag{3.10}$$

$$S_y = 1 - \frac{\Delta y_{c1,c2}}{\min\{y_{c1}, y_{c2}\}} \tag{3.11}$$

$$S_w = 1 - \frac{\Delta w_{c1,c2}}{\min\{w_{c1}, w_{c2}\}} \tag{3.12}$$



(a) Classifications from connected component analysis

(b) Result with no use of proximity

(c) Result with use of proximity

Figure 3.14: A visual representation of a connected component analysis is shown in Figure 3.14a where the character 'S' is misclassified as non-text (shown with a red background) in the presence of surrounding characters that were classified as text (shown with a green background). The resulting image using only the connected component classification is shown in Figure 3.14b and the resulting image using graph cut is shown in Figure 3.14c.

$$S_h = 1 - \frac{\Delta h_{c1,c2}}{\min\{h_{c1}, h_{c2}\}} \tag{3.13}$$

where $\Delta n_{c1,c2}$ represents the difference in n between connected components 1 and 2. The term $\min\{n_{c1}, n_{c2}\}$ results in the minimum n between the connected components 1 and 2. The same is true in all equations only that n is replaced by y, w, or h according to the similarity score being calculated.

The combined similarity score is a simple summation:

$$S = S_n + S_y + S_w + S_h \tag{3.14}$$

Each of the similarity scores S_n, S_y, S_w, and S_h have the potential to result in a negative value (whenever the difference is greater than minimum value). Therefore, any similarity scores found to be less than zero are clamped at 0. This prevents the combined similarity score from becoming overweighted by an extreme negative score. Such an example may be a broken character 'n' in which the left vertical stroke is isolated from the rest of the character. The score S_w (based on the width) would be drastically negative (i.e. -10) due to the difference in width and therefore easily overpower the S_h (height) and S_y (vertical location) scores which at most would each be 1.

After the similarity score is computed between the neighboring nodes, the score is scaled to fit an appropriate weight for the graph. The final weighting is computed as follows:

$$\omega_{n_1,n_2} = \min\{S * \omega_s, \omega_{max_n}\} \tag{3.15}$$

where ω_{n_1,n_2} represents the weight between node 1 and node 2, ω_s represents a weighting on how strong the similarity score is in relation to the graph cut weightings, and ω_{max_n} represents the maximum weighting for the n-links.

The weight for ω_s is used to strengthen the similarity score in relation to the graph. Since each of the four similarity measures have the potential to equal 1, the combined score can at most be equal to 4. Thus, for ω_s to be innocuous a value of $\omega_{max_n}/4$ must be used. However, a stronger weighting is desired and is set at $\omega_{max_n}/2$.

In many cases the neighboring nodes will contain the same connected component, or both will contain no connected components. In these cases, $\omega_{n_1,n_2} = \omega_{max_n}$. Alternatively, if one of the nodes contain a connected component and the other does not, then $\omega_{n_1,n_2} = 0$.

To give sufficient strength to the n-links for correct segmentation, the following ratio is used by the system:

$$\omega_{max_n} = 4 \times \omega_{max_t} \tag{3.16}$$

where ω_{max_t} is the maximum weight given to t-links.

Prior to finalizing the n-links, additional use of contextual knowledge is applied. Graph cut is a general segmentation algorithm, therefore the system applies the knowledge that segmenting text is the goal. Characters of a word in a typical layout span from left to right. Therefore, the weights of the n-links between nodes that neighbor horizontally are doubled in strength than those that neighbor vertically.

The initialization of graph cut is complete with the assignment of weights for the n-links and the t-links. Figures 3.15 and 3.16 give a visual representation of the weightings used for graph cut.

Graph cut is performed using the assigned weights to make the final classification of what is text and what is artwork/noise and retains only the components classified as text. The resulting images from the weightings shown in Figures 3.15 and 3.16 are shown in Figure 3.17.

Note that because proximity is considered (as well as a textual similarity measure), the letter ‘S’ in Figure 3.16 was retained whereas without considering proximity, it would have been removed. Such a case demonstrates the power of using graph cut for artwork removal.

Through the use of a neural network the system is able to learn how to best discriminate between text and non-text (artwork and noise). The confidence of the neural network in its

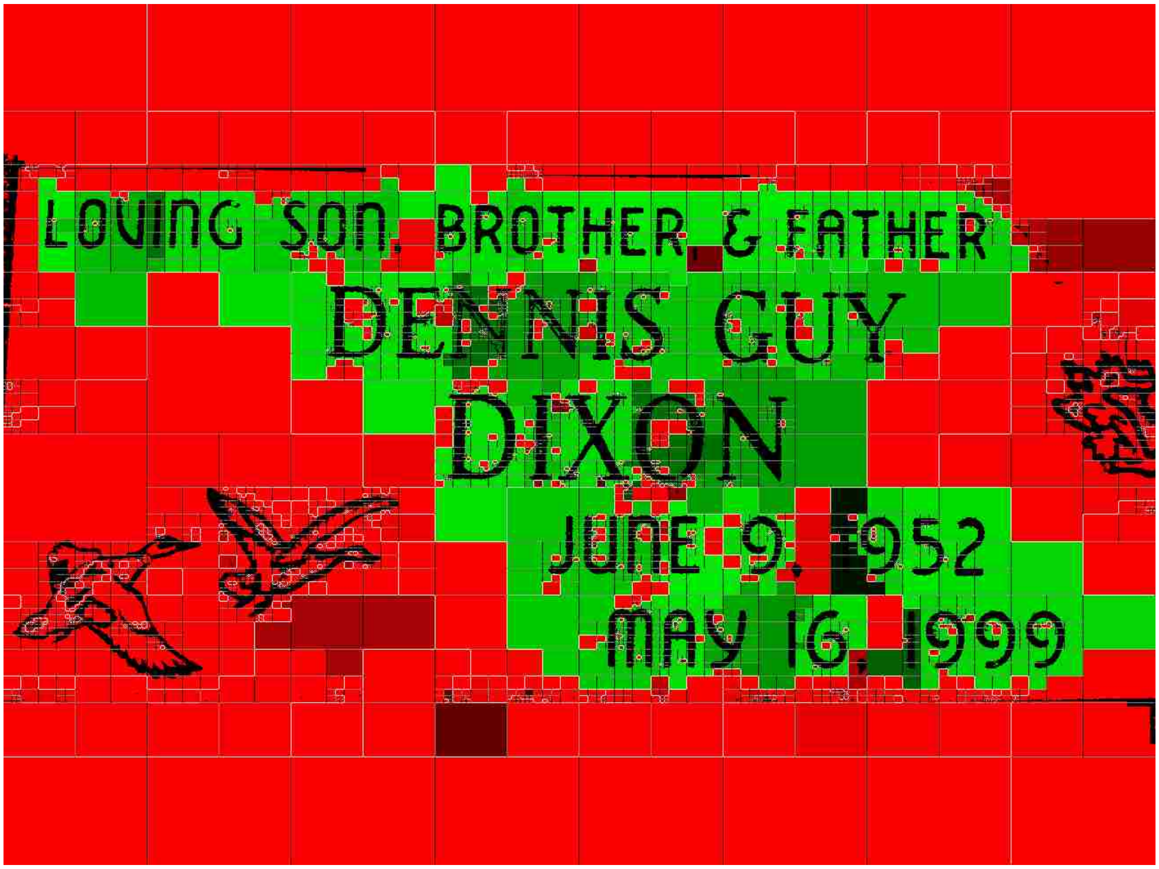


Figure 3.15: A representation of the weightings given on the graph formed by the binarized input image to the artwork removal process (Figure 3.17a). An explanation of the colors shown is given at the beginning of section 3.4.3.

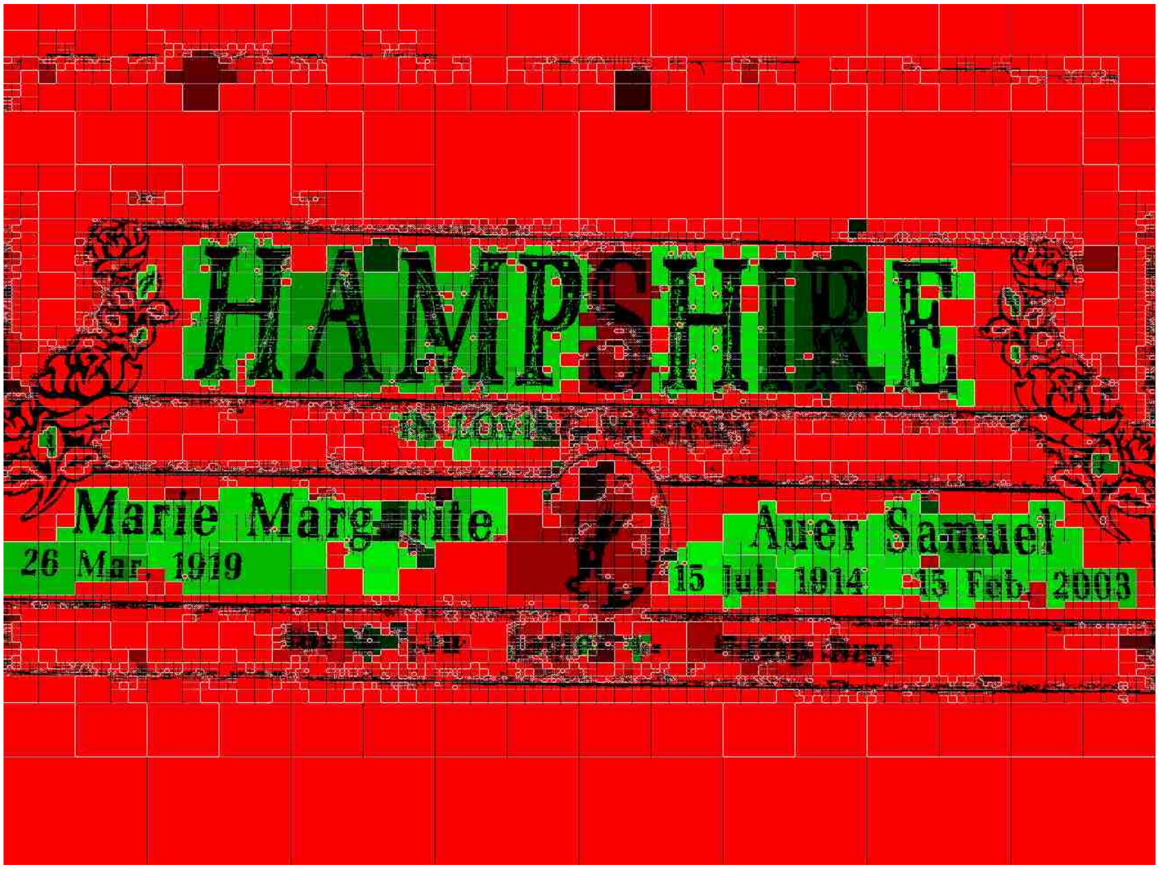
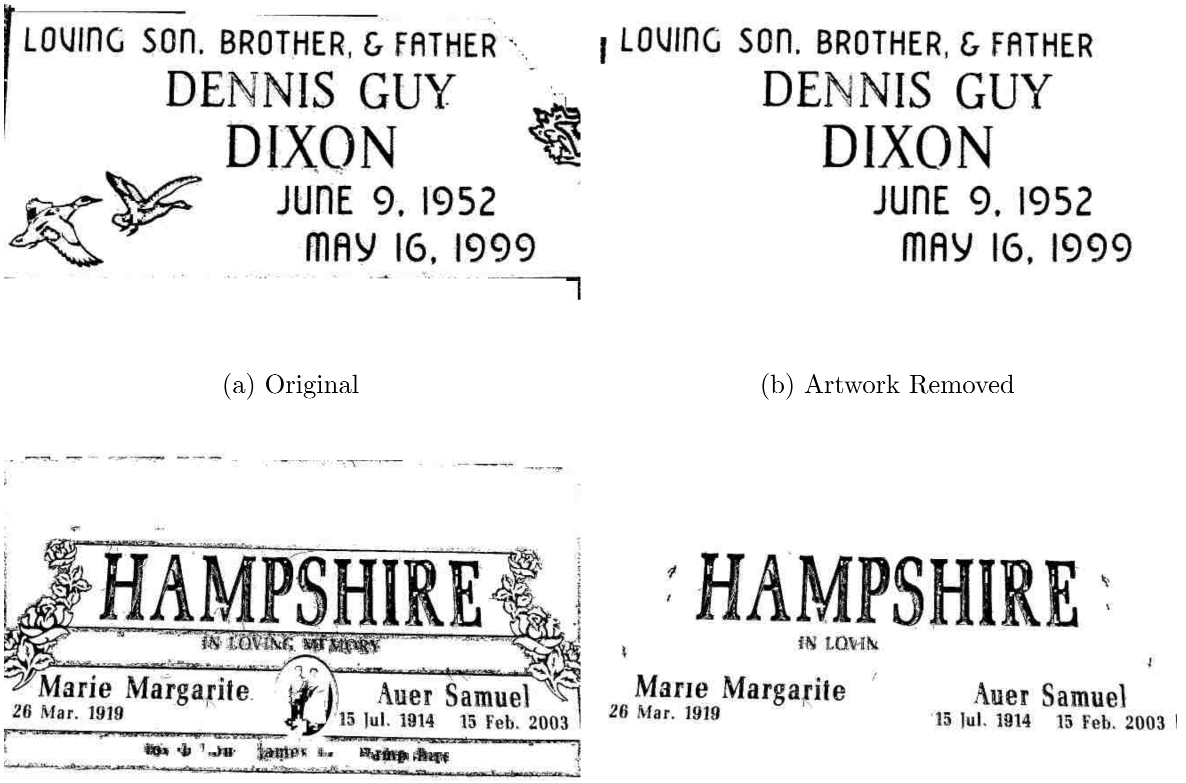


Figure 3.16: A representation of the weightings given on the graph formed by the binarized input image to the artwork removal process (Figure 3.17c). An explanation of the colors shown is given at the beginning of section 3.4.3.



(a) Original

(b) Artwork Removed

(c) Original

(d) Artwork Removed

Figure 3.17: The images resulting from the artwork removal process. The originals (3.17a, 3.17c) are shown in comparison with the final images (3.17b, 3.17d).

classifications represent each connected component's textual resemblance and are used as the weightings in the graph cut algorithm. Through graph cut both the textual resemblance and proximity are considered. Vertical proximity is given two times as much weight as the textual resemblance, while horizontal proximity is given four times as much weight. This additional weighting for proximity can correct misclassification by the neural network. As text will always have neighboring characters when forming a word or name, use of proximity will improve results in the presence of misclassified connected components.

3.5 Text Recognition

In the previous sections, steps have been described in which the problem of transcribing headstones has been reduced. Through the zoning and segmenting of the text, a clean binarized image is produced (Figures 3.17b and 3.17d), representing the core work of this thesis. However, the quality of text segmentation is subjective and difficult to quantify. Additionally, to have value in a real world system, a final transcription is necessary output. Therefore, a transcription process is described. Consistent with a major theme in this thesis, the domain is constrained to improve accuracy of the final transcription.

3.5.1 Text Region Creation

Certain layouts, such as where multiple columns are used, are difficult to accurately transcribe for many OCR systems. Decomposing scanned text documents into zones has received significant attention due to its value and difficulty (Antonacopoulos et al. [2007]). The layout of the text on a cemetery headstone varies greatly in that text is often separated into multiple regions and can cause errors for the OCR engine (Figure 3.18). Therefore, the proposed system further decomposes the image into text regions.

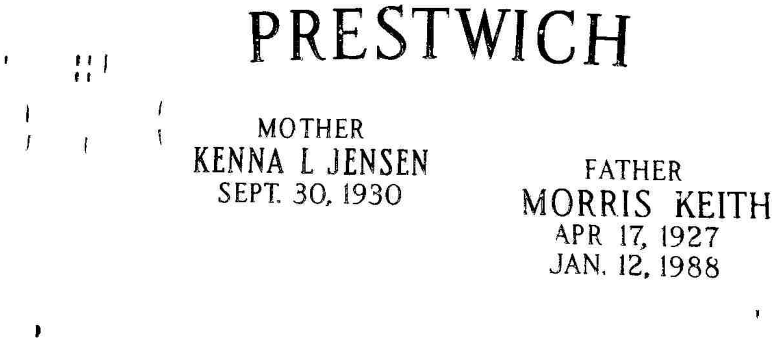


Figure 3.18: Headstone image in which the layout of the text forms multiple columns and regions.

The graph cut algorithm used for removing artwork produces a segmentation between nodes containing textual components, and those that do not. To form text regions, the graph cut segmentation is used to form connected components based on neighboring nodes that were classified as text. The resulting connected components act as the text regions.

The regions formed are used to group neighboring information to allow for a more complex contextual validation. By forming regions, the system gains a better understanding of which data is grouped together, thus constraining further the domain and reducing the possible solutions for that region.

Further processing is needed, however, as data that should be grouped together may have been placed into two separate regions (due to noise, misclassification, etc.). Therefore a simple process is used to merge two separate components (c_1, c_2) when one or both of the following conditions are met:

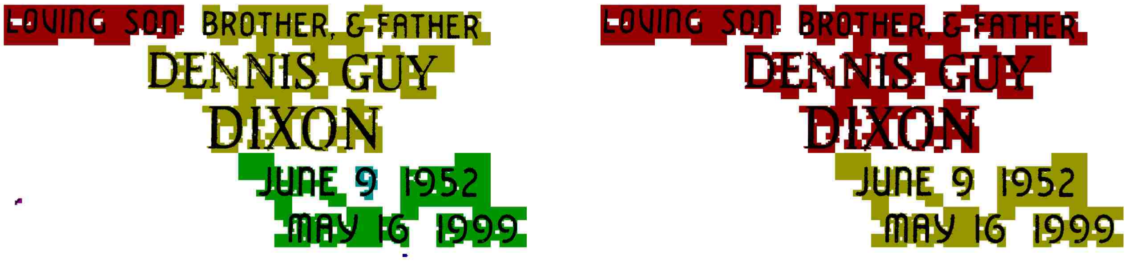
1. The bounding boxes of c_1 and c_2 overlap on the y axis AND on the x axis.
2. c_1 and c_2 overlap by at least h_{vert} pixels on the y axis AND are within w_{space} pixels of each other on the x axis.

where h_{vert} is the amount of vertical alignment expected between c_1 and c_2 and w_{pixels} is the maximum size expected for a space between two words.

Additionally, the assumption is made that if a labelled connected component is less than h_{min} in height (where h_{min} is the minimum height required for text to be recognizable), it is not part of the textual content and can be removed. Such an assumption is valid given that any disjointed character which may be less than h_{min} in height, will neighbor the other fragmented characters or neighboring text and be labelled as the same connected component, resulting in a height greater than h_{min}. Figure 3.19 shows an example where multiple regions were merged together.

3.5.2 Text Line Creation

To more fully leverage contextual information, the system creates individual lines of text within the text regions created in the previous step. Names and Dates are unlikely to be



(a) Labelled connected components

(b) Labelled connected components combined

Figure 3.19: Labelled connected components are shown in Figure 3.19a where text on the same text line was placed into two separate connected components. Figure 3.19b shows the updated labelling of the connected components in which text from the same text line is merged into the same component.

located on the same line as names are made to be prominent and identifiable. Likewise, dates are often grouped together with no other information on the same text line. Therefore, after a transcription is created through OCR, information extraction can be performed to classify the text and improve accuracy given the limited context. This is discussed in further detail in section 3.6.

Additionally, the reduction of data in a single, one text-line image enables the OCR engine to perform more accurately as its domain is constrained to a more specific image area.

Projection profiles are used to separate the text regions into individual text lines. The projection profiles show the structure of the text and allow for separation of individual text lines. The horizontal scan lines that contain no black pixels are removed and act as dividers between individual lines of text (Figure 3.20).

It may be the case that text regions found by the process described in section 3.5.1 contain multiple columns (Figure 3.21). To handle such cases, the text line creation process uses a recursive approach to continually subdivide the text region (alternating between a horizontal division and a vertical division) until no new divisions are created. Constraints are placed on the division process to avoid excessive subdivision of the image. For horizontal

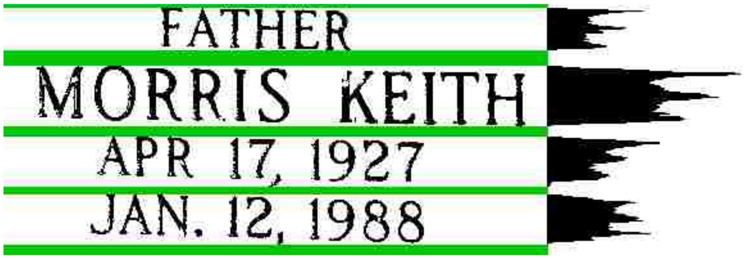


Figure 3.20: A text region where the projection profile is used to create individual text lines. The green regions represent horizontal scan lines where the profile was zero (i.e. no black pixels) and regions that will be discarded.

divisions, the value h_{min} is used to ensure that the text line is more than the minimum height required to be recognizable. For vertical division two requirements must be met:

1. The text region must be of greater width than w_{min} where w_{min} represents the minimum width required for a character to be recognizable.
2. The width between two text regions must be of greater width than $s_{w_{max}}$ where $s_{w_{max}}$ represents the maximum width for a space between two words.

The recursive division of the text region shown in Figure 3.21 is shown in Figure 3.22. Each region created retains the location information relative to the original image, thus enabling future contextual analysis to use proximity in classifying transcriptions.

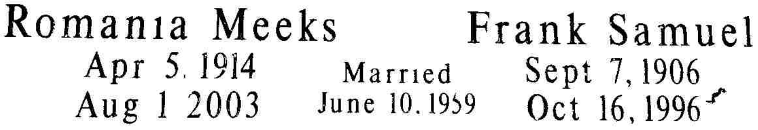
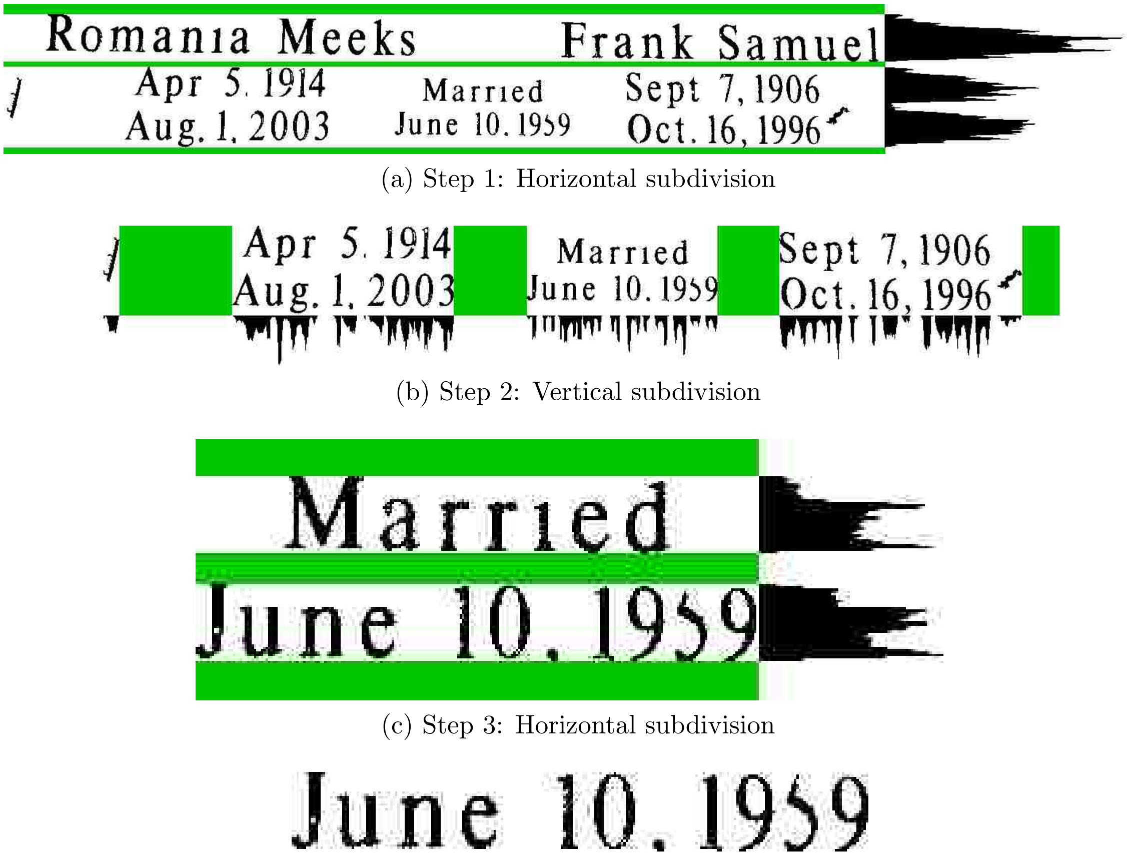


Figure 3.21: Headstone text region containing a complex layout of text.

3.5.3 OCR

Every text line created from the previously described process is passed to an OCR engine for transcription separately and is individually validated by the process described in chapter 3.6. The proposed system uses the open-source TesseractOCR engine (Smith [2007]).



(a) Step 1: Horizontal subdivision

(b) Step 2: Vertical subdivision

(c) Step 3: Horizontal subdivision

(d) Base case: no further divisions can be created

Figure 3.22: Visual representation of the recursive process used to create individual text lines. In each of the first three steps (Figures 3.22a, 3.22b, and 3.22c) multiple regions are created, however only one of the new regions are displayed in the following step for the purpose of simplifying the figure. Figure 3.22d shows one of the final text lines created through this process.

3.6 Contextual Validation

A constrained solution space provides for improved error correction, thus an improved final accuracy as well. The proposed system uses the constrained domain of cemetery headstones to limit the possible transcriptions. Although the error correction and information extraction performed in the proposed system is not the focus of this thesis, a simple process has been created to demonstrate the power of limiting the domain. A more powerful contextual validation system may be used in practice for improved OCR results.

To perform validation, the transcription of a text line is split into separate text-segments based on white-space. Multiple dictionaries and a confusion matrix are used to compute a score as follows:

$$P_d = \max_{w=W}\{p(w|d) \times \prod_{i=0}^{I} p(l_i|\boldsymbol{l_i})\} \tag{3.17}$$

where W is the set of all possible alternate words in the current dictionary, l_i and $\boldsymbol{l_i}$ are the i-th characters in the alternate word and transcription respectively, $p(l_i|\boldsymbol{l_i})$ is the probability that l_i was mistakenly transcribed as $\boldsymbol{l_i}$ (per a confusion matrix), and $p(w|d)$ is the probability of w given the current dictionary. Additionally:

$$I = \max\{|w|, |\boldsymbol{w}|\}$$

where $|w|$ and $|\boldsymbol{w}|$ are the lengths of the alternate word and the transcribed word respectively.

The result, P_d, represents the best score from the dictionary d. In the proposed system, the word that gave the best score is used as the final transcription. A minimum probability is given to characters not found in the confusion matrix (for a given alternate character) as well as the original transcription, if not found in the current dictionary.

The seven separate dictionaries that are used for validation are as follows:

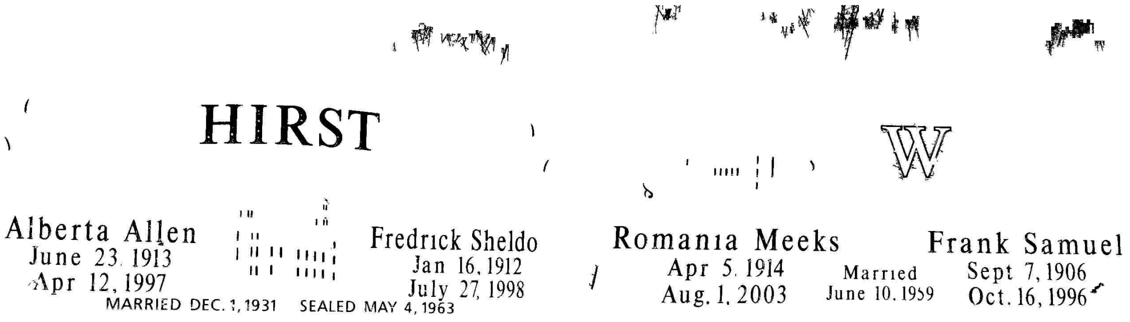
1. **Month**: all forms of a month including the full month name and all common abbreviations.
2. **Day**: all possible dates, forming a list of numbers ranging from 1 to 31 (created in memory at run time).
3. **Year**: all reasonable years, forming a list of numbers ranging from 1750 to the current year (created in memory at run time).
4. **Given Name - Female**: all female given names reported from census records (United States Census Bureau [1995]).

5. **Given Name - Male**: all male given names reported from census records (United States Census Bureau [1995]).

6. **Surname**: all surnames reported from census records (United States Census Bureau [1995]).

7. **Common Words**: common words found on headstones such as ”Mother”, ”Father”, ”Loving”, etc.

Examples of the corrections made by the described process for the input images shown in Figure 3.23 are given in Table 3.4.



(a) Example 1

(b) Example 2

Figure 3.23: Two binarized headstone images where the OCR result is shown in Table 3.4.

Original	Corrected
HIRST	HIRST
A1berta A11Cn	ALBERTA AMICI
June 23.1913	June 23, 1913
Apr 12.1997	APR. 12, 1997
Fredrlck Sheldo	FREDRICK SHELDON
Jan 16.1912	JAN. 16, 1912
July 221998	July 22 1998
MARRIED DEC.1.1931	MARRIED DEC. 1, 1931
SEALED MAY 411963	SEALED MAY 4, 1963

(a) Example 1

Original	Corrected
W	
Romama Meeks	ROMANA Meeks
Frank Samuel	Frank Samuel
Apr 5.1914	APR. 5, 1914
Aug. 1. 2003	AUG. 1, 2003
Marr1ed	MARRIED
June 10.1959	June 10, 1959
Sept 7.1906	SEPT. 7, 1906
0m.16.1996”	AM. 16, 1996

(b) Example 2

Table 3.4: Transcriptions that were corrected using the described contextual validation. The original images can be seen in Figure 3.23.

The contextual validation process can make important corrections such as correcting “SEALED MAY 411963” to “SEALED MAY 4, 1963” (Example 1) and “Marr1ed” to “MARRIED” (Example 2). However, since the contextual validation relies heavily on the original transcription, OCR errors greatly affect the final transcriptions accuracy, especially with regard to insertions or deletions. Such a case is in Example 2 where the clean image text “Oct. 16, 1996” was transcribed as “0m.16.1996” and then changed to “AM. 16, 1996.” Without a decent initial transcription, little can be done to accurately correct it within the contextual validation process.

As mentioned previously the contextual validation process is discussed to demonstrate the possible accuracy gains that can be achieved through using context, and not to be a commercial strength process. However, the reader may note that through a more complex analysis, the system could classify the text line ”0m.16.1996” as a date and appropriately recognize that “0m” would need to be corrected as a month.

We have shown that despite creating the contextual validation as a simple process, improvements in final accuracy are achieved through using context. Additionally, the potential accuracy gains through a more complex system have been discussed.

The methods described in this chapter make possible the automatic transcription of headstones through removing the inherent noise on headstones caused by stone texture, noise, artwork, etc. The benefits of using a constrained domain have also been presented with an example contextual validation process.

Chapter 4

Results

Direct analysis of the accuracy of a zoning and segmentation system is difficult. The majority of previous work has included an OCR engine as a final processing step to give a quantitative measure of accuracy. An analysis of which OCR engine produces the best results for the domain of headstones has yet to be conducted and is mentioned as future work in section 5.2.1.

To present the proposed system's results quantitatively, the accuracy is given of the system's neural networks: the three in the cascaded zoning process, and that which determines textual resemblance in the artwork removal process. To quantitatively measure the system as a whole, transcriptions produced through the OCR engine and contextually-based error corrected are used.

Much of previous work in the image zoning and segmentation domain have also relied on a qualitative analysis to demonstrate their system's usefulness. Likewise, a qualitative analysis is given for each process in order, including a number of resulting images for each.

The new headstone image data sets used for analysis is described first, followed by a quantitative and qualitative analysis.

4.1 Data Sets

Two known labelled datasets of headstone images exist due to volunteer efforts at Find A Grave [2011] and BillionGraves [2011]. The images contained within these datasets, however, are stored solely for the purpose of viewing on the internet and not for processing, thus the original images are scaled to a lower resolution than that required by the proposed system. Two new datasets have been created, named by the cemetery at which the images were captured: Provo City Cemetery and Orem City Cemetery.

The ground truth data for each of the data sets is subdivided into two categories: key and non-key data. The criteria and purpose of this subdivision is discussed after both data sets are individually discussed.

4.1.1 Provo City Cemetery

The Provo City Cemetery dataset contains many headstone images that suffer from one or more of the following:

1. **Age:** Some of the headstones date back to the late 1800s. Generally speaking, an older headstone will suffer from greater weathering than a more recently placed headstone.
2. **Orientation:** Some of the headstones are upright. Upright headstone images will often contain additional headstones in the background from the headstone of interest, causing additional noise in the transcription. Upright headstones also have greater potential for perspective skew (affecting the readability of the headstone by an OCR engine). Although it may be possible to correct the perspective skew through existing systems (Clark and Mirmehdi [2002, 2003]), such processing is out of scope for this thesis.
3. **Physical Environment:** Some of the headstones lie underneath trees, causing unpredictable shadows within the image.

Such challenges cause greater difficulty in accurately zoning and segmenting the headstone text. The Provo City Cemetery dataset contains 160 images.

4.1.2 Orem City Cemetery

The Orem City Cemetery dataset contains headstone images that date back to the 1950s and are all in-ground. Additionally, fewer trees are in the Orem City Cemetery reducing the number of headstone images containing shadows. In general, the Orem City Cemetery images suffer less from weathering, perspective skew and shadows than the images in the Provo City Cemetery. The Orem City dataset contains 208 images.

4.1.3 Headstone Data Type

Both datasets contain a subdivision of data within each transcription. The data set transcriptions were created with a distinction made between the key headstone data and the other possible data found on a headstone. The key data is limited to only the individual's names and the dates of birth and death. All other information including marriage date, children names, titles (such as mother, father), etc. are tagged as ”non-key” within the transcription.

The key data is vital to the headstone's purpose. Without the key data, the value of the headstone information is greatly reduced. The key data's importance can also be recognized through the prominence given within the headstone's layout. Often the key data area will not only include larger characters, but also consist of a higher contrast and a less cluttered background whereas other information such as children names are, on some headstones, of such low contrast that it is unreadable to the human viewer (For example, see Figure 4.4b).

The distinction between key data and other data is also important in regard to reported accuracy of the transcription. As mentioned above, key data is almost always given the most prominent placement on a headstone while the other data is often engraved with much lower readability. Thus dividing the data contained on the headstone provides for a cleaner calculation of accuracy for the key data. Otherwise a headstone in which key data was accurately recognized would be penalized due to non-key information found in less readable regions.

4.2 Quantitative Analysis

The results of the proposed system are discussed quantitatively by first looking at individual sub-processes within the system and is followed by a quantitative analysis of the system as a whole.

4.2.1 Sub-Processes

A quantitative analysis is ideal for comparing results of two distinct systems, regardless of which metric is used. Systems in which individual results can be termed correct or incorrect are capable of doing so. However, the quality of an image is subjective and the labelling of images as correct and incorrect is difficult. Therefore, only two of the processes (zoning and artwork removal) are discussed quantitatively in addition to an overall quantitative analysis based on OCR transcriptions.

Zoning

The zoning process uses neural networks trained on labelled data and therefore the accuracy can be easily computed. Since the zoning dataset was created by sampling a large number of headstone images from both the Provo and Orem City Cemeteries, the accuracy reported by a trained neural network is representative of the individual network's accuracy when presented novel data.

Three neural networks are created for each of the three levels of the image pyramid that are used in the cascaded zoning process (section 3.2.3). The image size of the sub-regions used as the data points in the neural networks are 10x10 where the image data within the sub-region becomes more detailed as the bottom of the image pyramid is approached.

To calculate the accuracies of the neural networks at each pyramid level, the neural networks are individually trained 100 times using a randomized 80% percent of their respective dataset for training, 10% as a validation set, and 10% for measuring accuracy. The average accuracy is shown in Table 4.1. Although all three datasets (training, validation, and test) are shown for reference, the test dataset represents the system's performance when given novel data.

	Level 1	**Level 2**	**Level 3**
Training	91.3375%	96.8250%	96.5207%
Validation	90.7300%	95.3000%	95.5333%
Test	90.0700%	95.5500%	95.4000%

Table 4.1: Neural network accuracy in classifying headstone regions as engravings and background (zoning). The three levels represent the levels within the image pyramid, where level 1 is the highest resolution image and level 3 is the lowest resolution image.

Comparing the three neural networks shown in Table 4.1, it is shown that although both level 2 and level 3 are comparable in accuracy, a 5% decrease in accuracy is given for level 1. Such a decrease suggests that in Level 1 the stone texture and its natural gradients have a stronger presence and are more difficult to classify while in Level 2 and Level 3 the effects of the stone texture gradients are minimized due to the reduced resolution.

The cascade used in the zoning process addresses the more difficult classification task at level 1 by using the networks at levels 2 and 3 to remove the majority of the stone texture.

Artwork Removal

The artwork removal process uses a combination of algorithms (quadtree creation and traversal, neural networks, and graph cut) to identify connected components of text and remove all others (artwork and noise). The value of this process can be seen in the percentage of connected components removed (Table 4.2).

Provo City Cemetery			**Orem City Cemetery**		
Normal Otsu	Inverse Otsu	Total	Normal Otsu	Inverse Otsu	Total
92.80%	95.55%	94.87%	85.61%	95.73%	92.25%

Table 4.2: The average percentage of connected components per headstone image that are removed in the artwork removal process on the binarized images. The percentages are reported separately for the normal otsu binarization images and inverse Otsu binarization images. The average total percentage is also given.

The average percentage of connected components removed per headstone is large regardless of which type (normal or inverse) of binarized image is being processed. This is due to the large amount of artwork and noise that is often found on a headstone that results in many small connected components. The connected components that remain are typically larger and represent a smaller proportion of the total connected component count.

The percentage is higher for the inverse binarized images and is a result of dark text on a light background which is found on most headstones. When dark text is on a light background, the inverse binarized image contains additional components representing the inverse of the text as seen in Figure 3.10c and will likely be removed.

A neural network is used to generate t-link weightings based on the component of interest's resemblance to text. Although the network is only one piece of the artwork removal process, through analyzing its accuracy, insight is gained as to the performance of identifying textual components.

Similar to the results presented in section 4.2.1, the average accuracy of 100 separately trained neural networks is computed for the training, validation, and test data sets. These results are shown in Table 4.3.

	Accuracy
Training	87.1432%
Validation	84.0823%
Test	83.9873%

Table 4.3: Neural network accuracy in classifying headstone regions as text and noise

The accuracy given for the test data set shows that the task of distinguishing text from artwork (artwork removal) is a more difficult task than distinguishing engravings from stone texture (zoning). However, an accuracy of 83% is reasonable given that it is only one piece of the artwork removal process. As demonstrated in Figure 3.14, the use of proximity through graph cut can correct a misclassification of the neural network.

Through the artwork removal process a large amount of artwork and noise is removed. The

4.2.2 Overall

This thesis presents a novel zoning and segmentation algorithm, producing a clean and OCR-able binarized image of the headstone with stone texture and artwork removed. As mentioned previously, however, a qualitative analysis of such an image is difficult. Therefore, to analyze the effectiveness of the proposed system a transcription is produced upon which metrics may be used to quantify the system's performance.

The approach used to produce the transcription and metric is discussed, followed by a discussion on the metrics of precision and recall, accuracy, error rate, and efficiency.

Approach

The transcription is produced through the two-step process of passing the image to an OCR engine (section 3.5) and error-correcting the transcription based on contextual knowledge (section 3.6).

However, the layouts possible on a headstone vary greatly and cause ambiguity as to the correct linear ordering of the different text regions such as in Figure 4.1. This complicates greatly the ability to match the transcription sequentially.

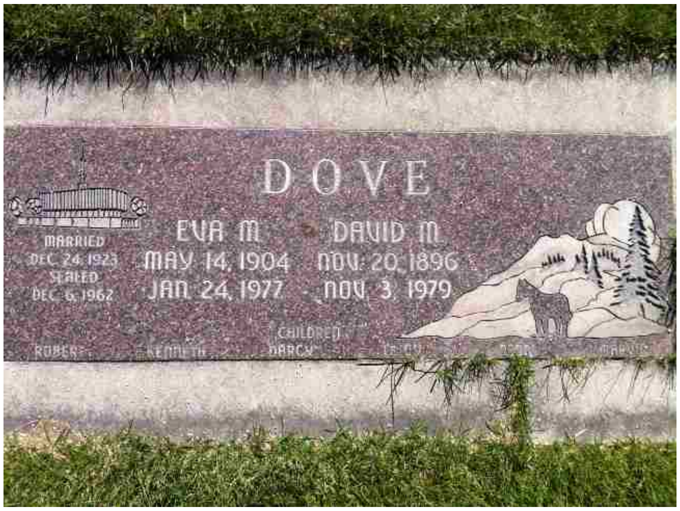


Figure 4.1: A headstone in which the linear ordering of the text regions for transcription is subjective. The marriage information and the individual's information may be ordered differently depending on scan order.

A system used to analyze the layout analysis performance of OCR systems is proposed in Kanai et al. [1995]. In this system a word matching algorithm is based in part on edit distance and is used to compare the ground truth with the OCR result. This approach is adapted and used in the proposed system to match words and perform an analysis of the produced transcription.

Word matching is performed by minimizing the edit distance (Levenshtein [1966]) to find a word-match. Due to the complicated layouts possible, a more industrial-strength layout analysis is needed to match more than individual words (i.e. text lines). However, given the scope of this thesis a simple layout analysis is used and words are matched individually. Every word (i.e. a collection of characters separated on both sides by white-space) provided by the ground truth is matched with the word from which the smallest edit distance is computed. Due to the limited amount of data and the relatively high amount of noise found on a headstone, the word matches have no ordering to avoid inaccurate constraints. Additionally, each OCR transcription word can not match more than one ground truth word. During the matching process, the system ensures that the ground truth word is matched with the best possible OCR transcription word, where if the best edit distance is shared by multiple OCR transcription words, the first word encountered is retained.

Precision & Recall

Precision and recall is commonly used in OCR tasks to capture both the sensitivity and specificity of the transcription by answering the following questions:

- **Precision:** how well can the system avoid producing erroneous characters?
- **Recall:** how well can the system produce an accurate transcription?

thus ensuring that the system provides both an accurate and useful transcription by maximizing both measures.

Precision (P) and recall (R) are calculated using the following equations:

$$P = \frac{|\{C_{gt}\} \bigcap \{C_{ocr}\}|}{|\{C_{ocr}\}|} \tag{4.1}$$

$$R = \frac{|\{C_{gt}\} \bigcap \{C_{ocr}\}|}{|\{C_{gt}\}|} \tag{4.2}$$

where $|\{C_{gt}\} \bigcap \{C_{ocr}\}|$ is the total number of characters in a matched pair that are correctly transcribed (characters that are both relevant and retrieved), $|\{C_{ocr}\}|$ is the count of characters in the OCR transcription (characters that are retrieved), and $|\{C_{gt}\}|$ is the count of characters in the ground truth transcription (characters that are relevant).

Additionally, to produce a single metric to measure the overall performance (combining both the precision and recall), the F-measure (F) is computed as well:

$$F = \frac{1}{\alpha \frac{1}{P} + (1 - \alpha) \frac{1}{R}} \tag{4.3}$$

where α is a factor which determines the weighting of precision and recall. In the proposed system no preference is given to one over the other, therefore $\alpha = 0.5$.

The average precision, recall and f-measure per headstone image is shown in Table 4.4 where the results of transcriptions from raw, unprocessed images are compared with those of the proposed system.

	Provo City Cemetery			**Orem City Cemetery**		
	Precision	Recall	F-measure	Precision	Recall	F-measure
Raw	1.76%	39.51%	3.37%	1.16%	39.10%	2.25%
Proposed System	51.85%	54.28%	53.04%	55.15%	62.14%	58.44%

Table 4.4: The precision, recall and F-measure of raw image OCR compared with those of proposed system's OCR.

The precision measures of the raw image OCR transcriptions are low due to the cluttered and noisy nature of cemetery headstones (1.76% and 1.16% for the Provo and

Orem City Cemeteries respectively). This result demonstrates that performing OCR on raw headstone images is of little to no value. However, by constraining the domain using the methods implemented in the proposed system, the cluttered and noisy environment of headstone text is greatly minimized resulting in a significantly higher precision (51.85% and 55.15%).

The recall measure for raw headstone images is at 39% for both datasets. This result is slightly misleading on how accurate the transcription is. The OCR transcription for the raw images creates a large number of spurious characters as evidenced by the 1% precision on both datasets. Due to the word matching algorithm where no ordering is considered, a number of the spurious characters match with words in the ground truth transcription albeit the word from the OCR transcription is only one character in size. Although this is not a good match (edit distance is equal to the ground truth word size minus 1), such matches do increase the recall. By using the proposed system the recall gains 15% and 23% for the Provo City Cemetery and Orem City Cemetery respectively.

Both the precision and recall are similar in each of the datasets for the raw image transcriptions, however the increase for precision and recall is greater for the Orem City Cemetery than that of the Provo City Cemetery. Such a result confirms the previous discussion on the comparative difficulty of the datasets where due to age, orientation, and physical environment, the Provo City Cemetery is more difficult to process and transcribe.

Accuracy

Given the definition of recall in equation 4.2, the accuracy ($\frac{\#correct}{\#characters}$) is equal to the recall measure and the two are synonymous:

$$a = R = \frac{|\{C_{gt}\} \bigcap \{C_{ocr}\}|}{|\{C_{gt}\}|} \tag{4.4}$$

Therefore, in the discussion that follows the term accuracy may be substituted with recall, however in this section the results are separated into the two categories of key and non-key data.

As previously mentioned, both the Provo City Cemetery and the Orem City Cemetery datasets contain labelled transcriptions in regard to the key data and the non-key data found on the cemetery headstone. When the headstone is transcribed, the system is unable to distinguish between the key data and the non-key data and therefore the denominator in the equation for precision (equation 4.1) cannot be constrained to be the individual count of either the key or non-key data. However, the accuracy faces no such restraint and can be calculated for each label individually. The accuracy for the key data and non-key data is shown in Table 4.5.

	Provo City Cemetery		**Orem City Cemetery**	
	Key Data	Other Data	Key Data	Other Data
Raw	39.95%	24.87%	40.49%	32.70%
Proposed System	58.03%	34.62%	70.86%	46.26%

Table 4.5: Accuracy of the proposed system on the Provo City Cemetery and Orem City Cemetery datasets separated into key and non-key data.

The difference in accuracy of the proposed system between the Provo City Cemetery and the Orem City Cemetery is approximately 12% and gives insight as to the varied nature of headstone readability.

Additionally, the difference in accuracy between the key data and the other data (approximately 24% for both the Provo City Cemetery and the Orem City Cemetery) gives insight as to the prominence given to the key information on a headstone.

4.2.3 Error Rate

A common metric in measuring the performance of OCR is that of word error rate. Word error rate (WER) is defined as follows:

$$WER = \frac{\#insertions + \#deletions + \#substitutions}{\#words} \tag{4.5}$$

The word error rate allows discussion towards how often the system is incorrect, the alternative to accuracy which reveals how often the system is correct.

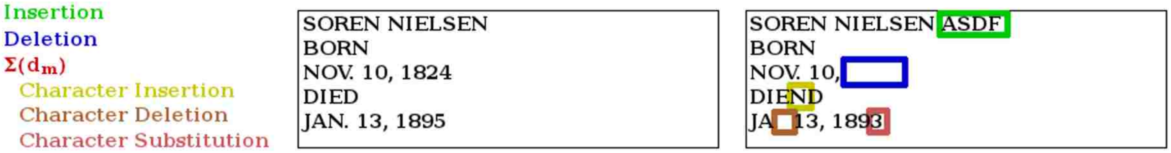
As mentioned previously, the transcriptions of headstone images contain small regions of text and due to complex layouts the system is unable to sequentially match words. Additionally, in the noisy domain of cemetery headstones a word error rate is too coarse for measuring the multiple errors that may occur within a single word. For these reasons an error rate at the character level is more reasonable.

The error rate (ER) for the proposed system is measured accordingly:

$$ER = \frac{\#insertions + \#deletions + \Sigma(d_m)}{\#characters} \tag{4.6}$$

where $\#insertions$ is the count of spurious characters not matched, $\#deletions$ is the count of non-matched ground truth characters, $\Sigma(d_m)$ is the summed edit distance of all the matched words, and $\#characters$ is the number of characters in the ground truth transcription.

The edit distance of each matched word includes all three possible errors: insertions, deletions, and substitutions of individual characters. Thus, the edit distance of matched words, the count of characters in all extraneous words (insertions) and the count of all ground truth characters not part of a word match (deletions) account for all possible errors in a given transcription. This is normalized by the number of characters in the ground truth transcription to give a rate as to how often, per ground truth character, an error is made. A visual description of each possible type of error is shown in Figure 4.2.



(a) Color Key (b) Ground Truth (c) Transcription with Errors

Figure 4.2: A visual description of the possible error types (Figure 4.2a) in the transcription (Figure 4.2c) with the ground truth (Figure 4.2b)

The total error rate, as well the error rate for each error type is given for the Provo City Cemetery dataset in Table 4.6 and for the Orem City Cemetery in Table 4.7.

	Provo City Cemetery			
	Insertions	Deletions	$\Sigma(d_m)$	Total
Raw	14953.69%	1.81%	58.68%	15014.18%
Proposed System	43.35%	17.52%	28.20%	89.07%

Table 4.6: Error rate of the transcriptions produced by the raw headstone images and the images processed by the proposed system on the Provo City Cemetery dataset.

	Orem City Cemetery			
	Insertions	Deletions	$\Sigma(d_m)$	Total
Raw	8241.40%	1.15%	59.76%	8303.30%
Proposed System	39.94%	8.25%	29.60%	77.80%

Table 4.7: Error rate of the transcriptions produced by the raw headstone images and the images processed by the proposed system on the Orem City Cemetery dataset.

As shown in the results, the insertions have no upper bound in the number of errors they may cause. Due to the noisy nature of cemetery headstones a large number of insertions are reported for the raw images in both datasets. Such high rates of insertion heavily dominate the transcription and make it of little use. The proposed system greatly reduces the insertion error rate and demonstrates the value of zoning the engravings to remove stone texture and removing artwork prior to performing OCR. Both the stone texture and artwork in most cases will cause additional spurious characters to be inserted into the transcription.

The deletions have a low error rate due to the large number of possible matches (provided by the insertions) to ground truth words. In processing the images through the proposed system, a jump in the error rate for deletions is seen. Due to the lack of image information such as gradients, the proposed system will occasionally remove regions of the headstone image that contain text during the various steps of processing. The error rate for deletions is double in the Provo City Cemetery dataset than that in the Orem City Cemetery as a number of headstones in the Provo Cemetery have suffered from excessive weathering, causing weaker gradients within the image.

The error rate of the matched words are reduced in half through the proposed system. This is due in part to the contextual validation performed that corrects words based on the domain of cemetery headstones. These corrections allow for better, if not exact, matches to the ground truth. The cause for a higher error rate in the raw images is the case when a single spurious character matches with a ground truth word. Such a match is not correct, however the edit distance is reduced due to the single character matching a character in the ground truth word and is therefore retained as a match. This results in a number of poorly matched words that increase the error rate in this category.

In both raw and processed images for both datasets, insertions cause the greatest number of errors, although they are greatly reduced by the proposed system. The error rate is reduced by more than 100 times in both datasets and demonstrates the value of the proposed system.

Efficiency

Added significance of efficiency comes as a result of the thousands of volunteers that have manually transcribed headstones at Find A Grave [2011] and BillionGraves [2011] as part of a tedious and time consuming process. Therefore effort was made to optimize the speed of the proposed system to enable an efficient transcription process.

The run times for each sub-process has been averaged over all headstone images from both the Provo City Cemetery and Orem City Cemetery datasets and are shown in Table 4.8.

Sub-Process	**Run Time (in seconds)**
Pre-processing	0.6340
Headstone Segmentation	0.1075
Zoning	0.6021
Segmentation	0.2476
Artwork Removal	1.9104
Process Total	3.5031

Table 4.8: The run times of the proposed system broken down at the sub process level.

The row labelled pre-processing consists of creating the complete image pyramid and the calculation of gradients for the three levels in the pyramid discussed in section 3.2.3. Headstone segmentation is very fast as the operation is performed only at the top of the image pyramid.

The run times of the first two operations (pre-processing, and headstone segmentation) are consistent across all images of the same size while the others vary based on the number of remaining pixels to be processed. The run times of both zoning and segmentation are directly proportional to the remaining pixels, however the artwork removal process is not. The run time of the quadtree generation in the artwork removal process is affected by the number of connected components, their size, and their proximity to each other. Therefore, the artwork removal run time varies from image to image based on the amount of noise remaining within the image.

Without regard to the post-processing of OCR and validation, the system performs at an average of 3.5031 seconds. The run-time for the OCR and validation process and the resulting transcription total run time is shown in Table 4.9.

Process	**Run Time (in seconds)**
Process Total	3.5031
OCR & Validation	4.6753
Transcription Total	8.1768

Table 4.9: The run times of the OCR and validation of the proposed system.

The OCR and validation process requires more time to run than the full process shown in Table 4.8. However, little effort was given to optimize the validation process as it is not a focus of this thesis.

In the proposed system where both the core process and post-processing is performed, an averaged total of 8.1768 seconds is required on a single core of a 3.33GHz Intel Core i7 processor.

4.3 Qualitative Analysis

A qualitative analysis is ideal for visually understanding the quality of a result. Although a visual inspection of an image to measure performance is subjective and difficult to quantify, the reader can gain a deeper understanding of the proposed system's performance through a qualitative analysis.

The proposed system contains a number of sub-processes that reduce the challenge of extracting text from a noisy and cluttered headstone into more manageable tasks by removing non-textual regions. The visual results of each sub-process is presented and discussed in order, namely: headstone segmentation, zoning, segmentation, and artwork removal followed by a discussion of overall results.

4.3.1 Sub-Processes

Each sub-process seeks to remove non-textual regions of the headstone image and is thus vital to the success of the system. Therefore it is important to discuss the results of each sub-process and demonstrate its strengths and weaknesses individually. The results of the sub-processes are presented in their respective order.

Headstone Segmentation

Headstone segmentation is a direct implementation of interactive graph cut with foreground and background seeds sampled automatically. The results of this process are shown in Figure

4.3. The processing is performed on the lowest resolution image in the image pyramid, therefore Figures 4.3c and 4.3d show the lowest resolution image with an overlay of the sampled seeds for foreground (green) and background (red).

The images on the left results in a clean segmentation of the headstone from the surrounding area. This is largely due to the shadows that surround the headstone causing a strong contrast between the headstone face and the surrounding area allowing for graph cut to give an accurate segmentation.

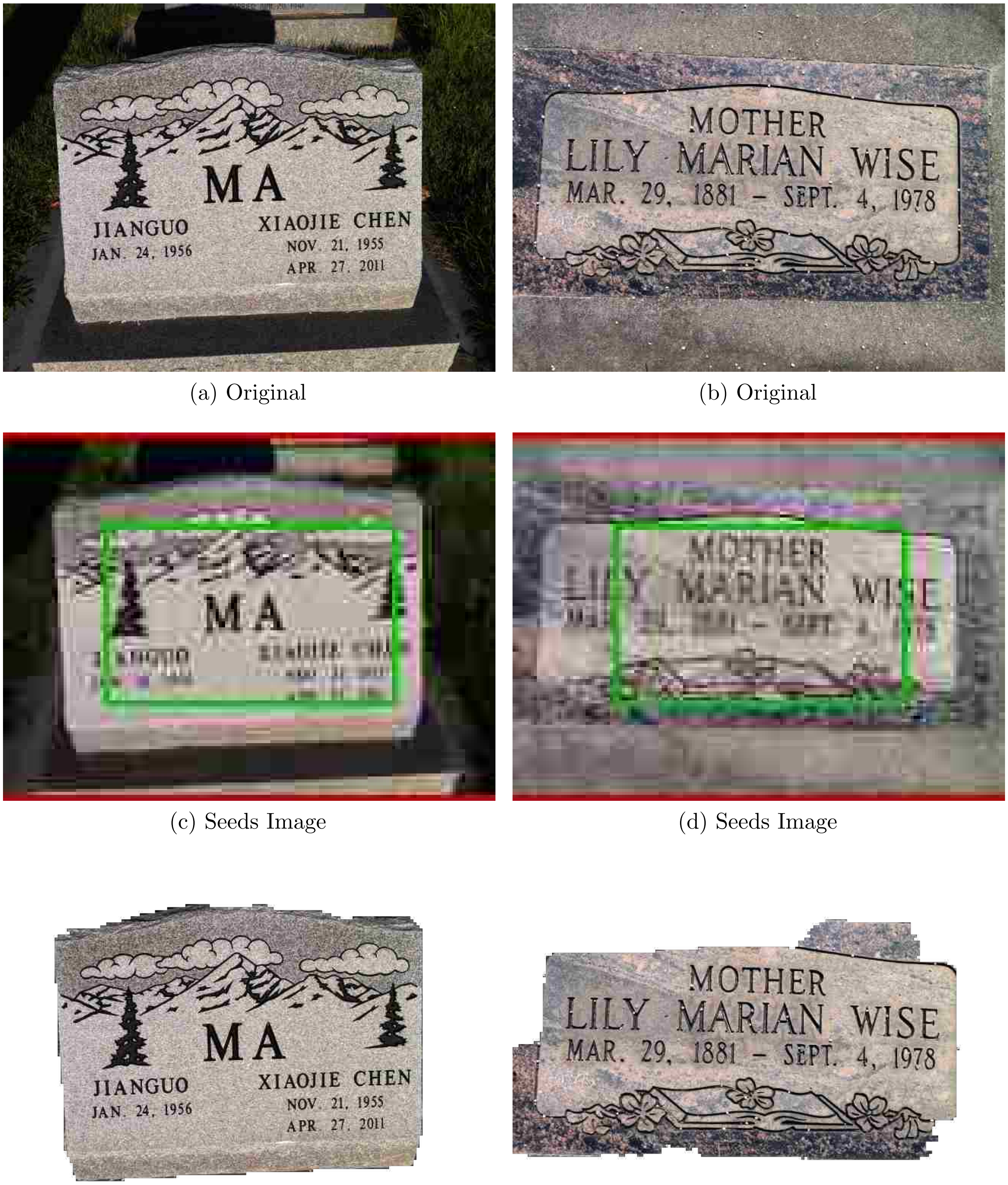
A more difficult headstone image is shown on the right where the complete headstone is not captured. In this case, the stone texture is similar in color to the surrounding area (concrete) and is difficult to segment. However, the region that contains the text was retained and therefore the segmentation is considered successful.

Zoning

The zoning process retains engravings in the midst of a noisy and cluttered environment. Three images from the image pyramid are provided as the input, all of which have been updated by the headstone segmentation process. The results are given in Figure 4.4 and show the segmented headstone image and the resulting zoning of that image.

The left image is the same headstone image used in Figure 4.3. Due to the strong contrast between the engravings and the stone texture, the zoning is successful and isolates the artwork and the text.

The right column however, shows a headstone in which contrast is low with the stone texture. Some textual information is lost on the lighter regions such as the last number in the marriage date, however in these regions the majority is retained. However, the textual information engraved without a lighter background (i.e. the children names on the bottom) is difficult to zone given the low contrast with the surrounding stone texture, resulting in the discard of desired (albeit non-key) information.



(a) Original

(b) Original

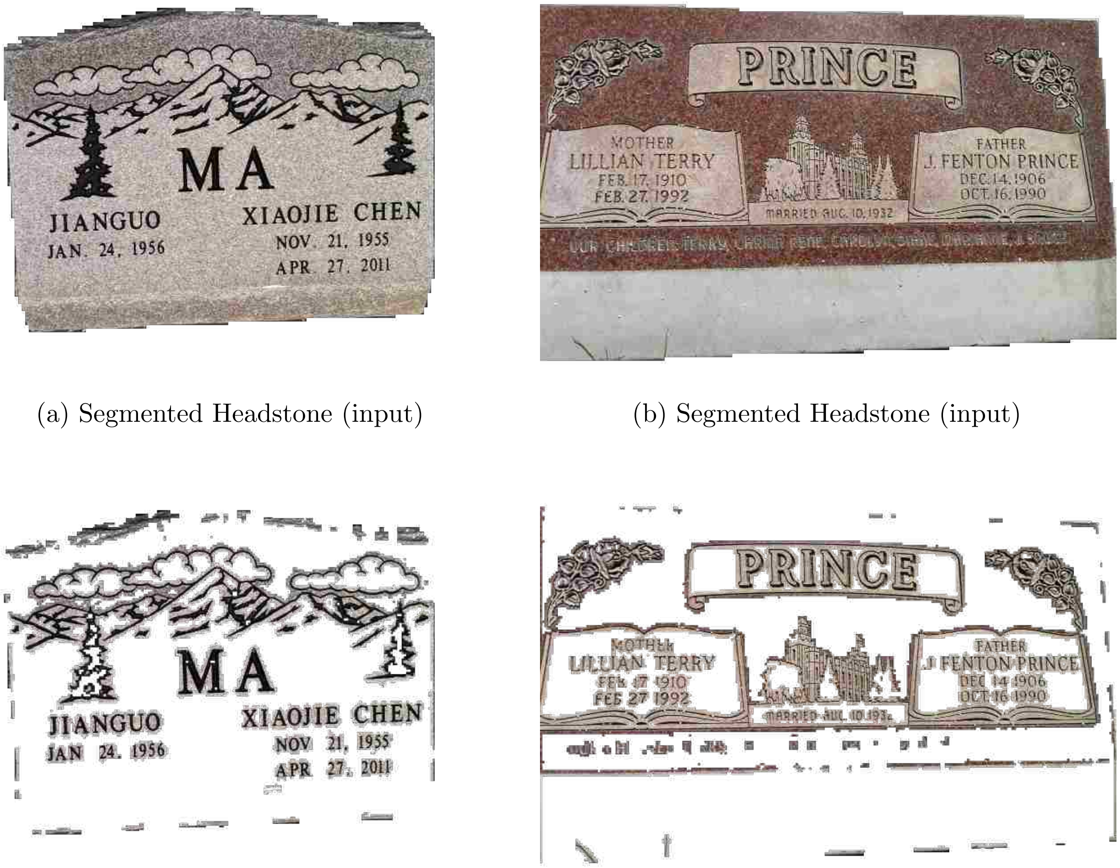
(c) Seeds Image

(d) Seeds Image

(e) Segmented Headstone

(f) Segmented Headstone

Figure 4.3: Example results from headstone segmentation. In the images (c) and (d), an overlay of the automatic placement of foreground (green) and background (red) seeds is shown.



(a) Segmented Headstone (input)

(b) Segmented Headstone (input)

(c) Zoned

(d) Zoned

Figure 4.4: Example results from the zoning process.

Segmentation

The segmentation process uses both traditional Otsu binarization and an inverse Otsu binarization. The engravings are segmented from the surrounding padding retained by the zoning process. The results of this process are show in Figure 4.5 where the input image, the traditional binarized image, and the inverse binarized image are given.

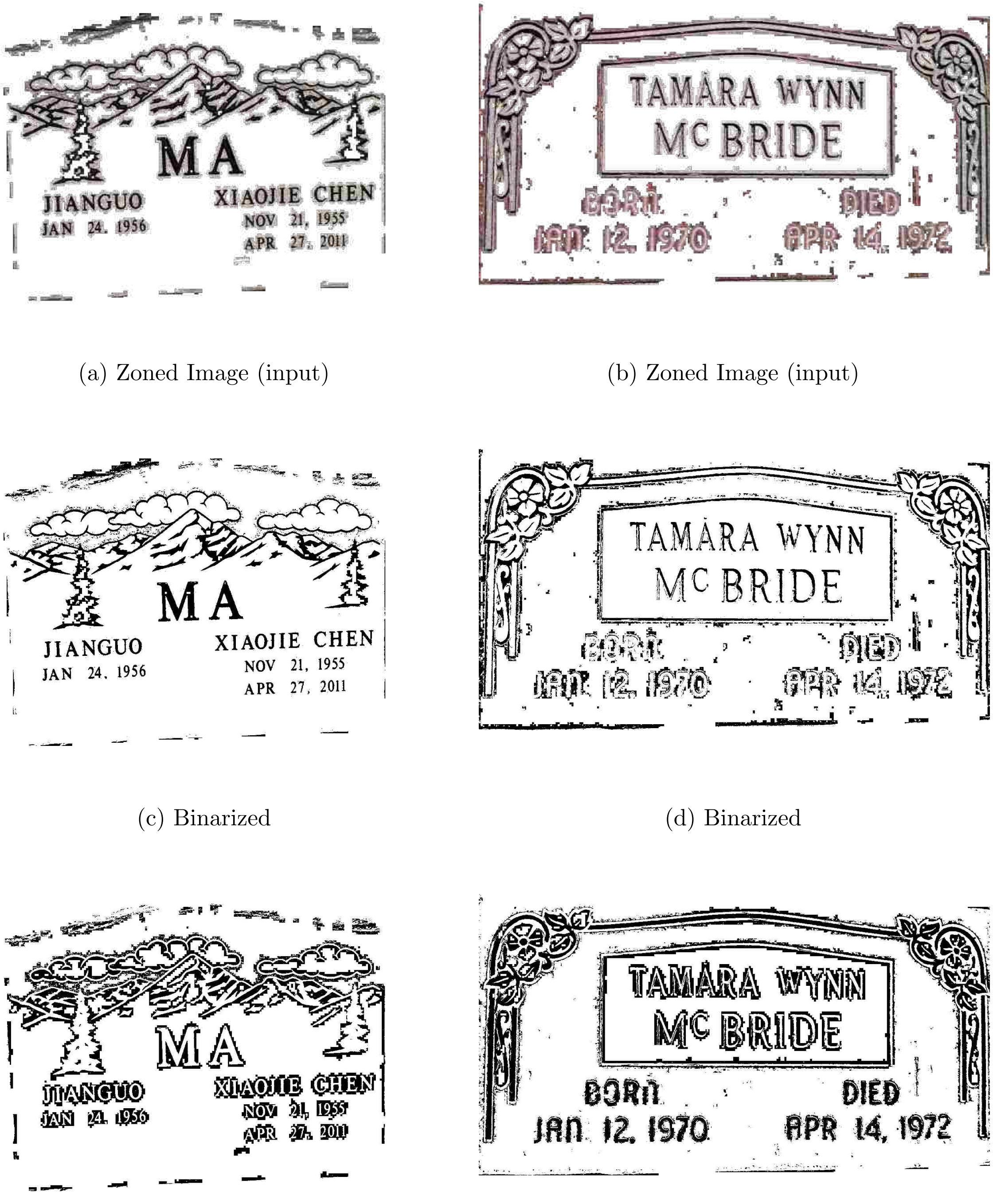
The binarized result (Figure 4.5c) shows the textual information cleanly segmented from the surrounding stone texture and shows the success of the such a binarization. As this headstone does not have any inverted text (light text on a dark background), the inverse binarized result provides no further information. However, the padding that surrounds the characters results in connected components that do not resemble text and will be removed in the artwork removal process (Figure 4.6f).

The headstone image on the right demonstrates the need for performing both a traditional binarization and an inverse binarization. Text found on this headstone image is both dark on a light background (names) and light on a dark background (dates). The traditional binarization (Figure 4.5d) cleanly segments the names from the background while the inverse binarization (Figure 4.5f) cleanly segments the dates. Both are needed to capture the full amount of information.

Although performing binarization twice (traditional and inverted) creates a number non-textual connected components, the majority of the non-textual components are removed in the artwork removal process. In many cases the padding that existed prior to binarization was connected and will thus connect the resulting components, reducing its textual resemblance.

Artwork Removal

Through use of a neural network and graph cut segmentation, the artwork removal process removes non-textual components by considering both text-like qualities and proximity.



(a) Zoned Image (input)

(b) Zoned Image (input)

(c) Binarized

(d) Binarized

(e) Inverse Binarized

(f) Inverse Binarized

Figure 4.5: Example results from the segmentation process.

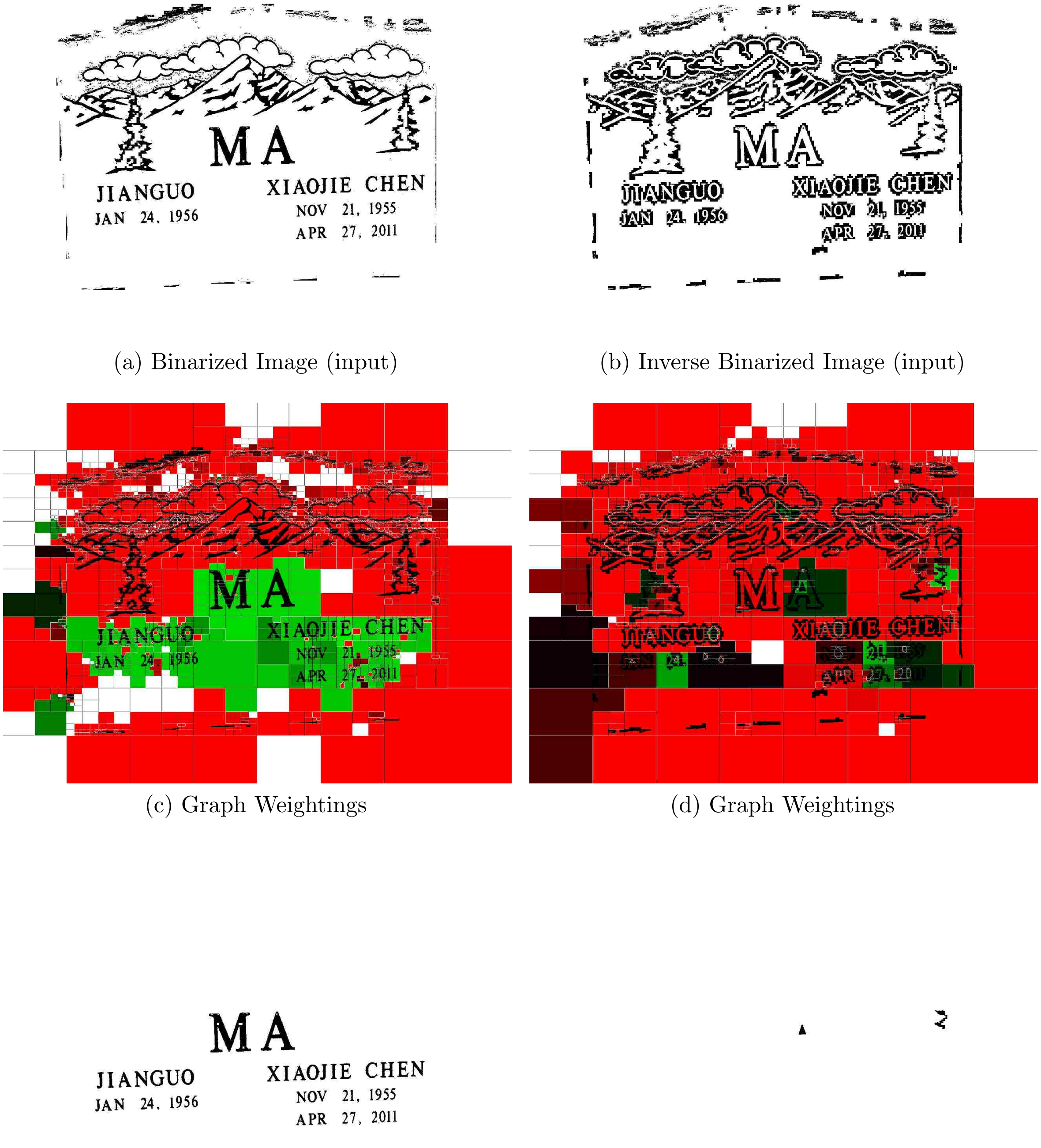
The neural network outputs the confidence score of its labelling of the component to either text or noise while a similarity score is calculated to determine how similar two neighboring regions are.

The consideration of proximity allows for correction on misclassified characters where the text-like qualities were not found. As the strength in similarity increases, the more likely the two regions will be classified to the same class in the graph cut segmentation. The results are shown for both the traditional binarized image and the inverse binarized image. The weightings are also shown where green (text) and red (non-text) represent the textual resemblance the connected component has. Additionally, similarities between two neighboring nodes are shown by the coloring of the edge between them. The darker the edge, the more similar the two regions are whereas the brighter the line, the more distinct and separate the two region are (for a more detailed view of graph weightings in the artwork removal process, see Figures 3.15 and 3.16. These weightings and the final results are shown in Figures 4.6 and 4.7.

We see in Figure 4.6f that only text was retained even though a few non-textual connected components resembled text (colored green in Figure 4.6c). Despite the large number of black pixels in Figure 4.6b, little was retained due to both textual resemblance and proximity of the resulting connected components. In addition to removing misclassified non-textual components, textual components that were classified as non-text were retained as a result of proximity, such as the commas in the JAN and NOV dates.

The headstone image seen in Figure 4.7 presents a more difficult problem than that in Figure 4.6. A large amount of noise is present in the input image and causes some of the characters to be connected. The bottom line of text on the left images suffers from such connected characters. Because of this, the neural network classifies a portion of it as non-text. Due to the size of the misclassification, graph cut is unable to correct this error.

Despite lost information however, the strength of the proposed system is shown in the final result. In the presence of much noise, the great majority of text was retained and



(a) Binarized Image (input)

(b) Inverse Binarized Image (input)

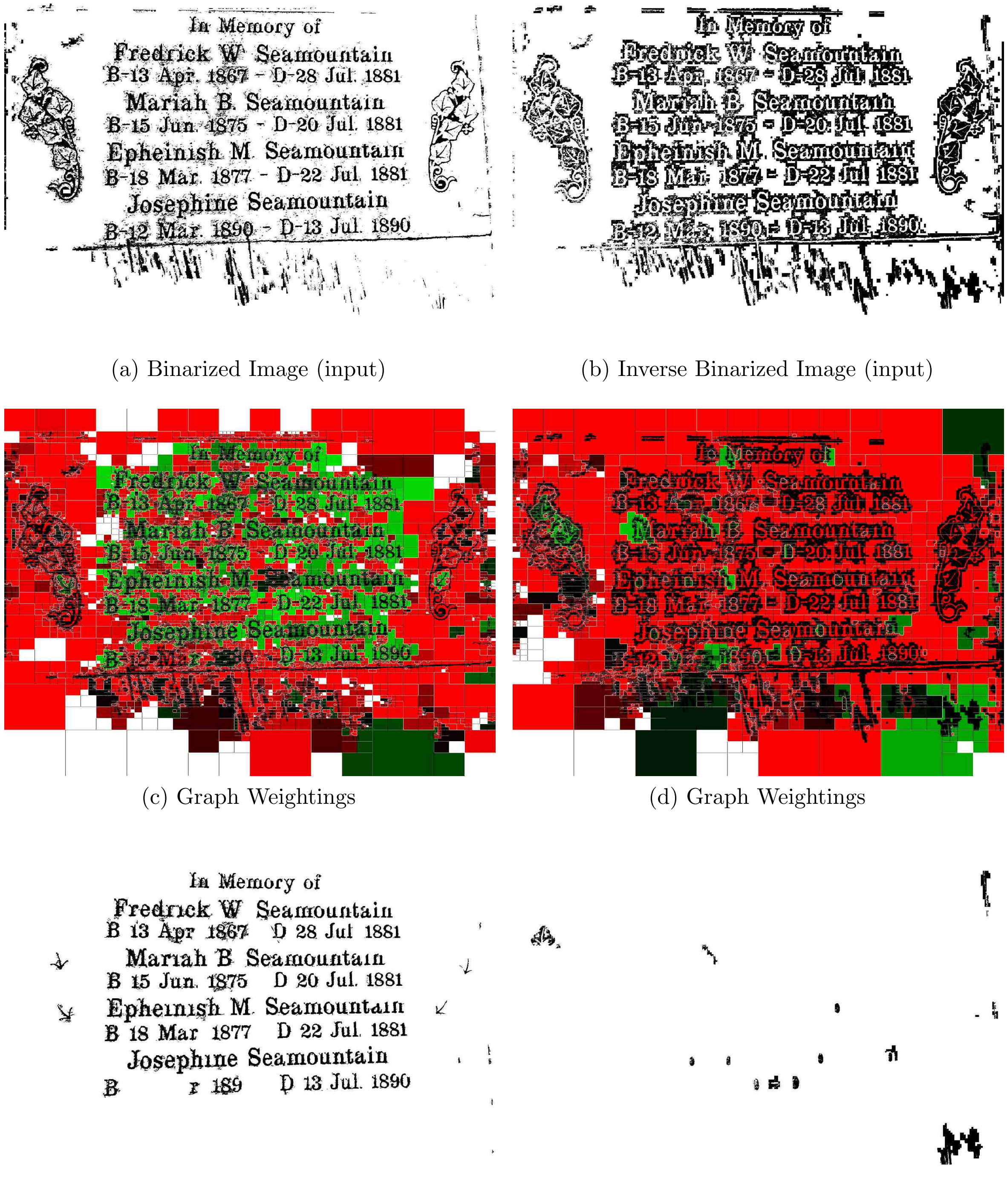
(c) Graph Weightings

(d) Graph Weightings

(e) Artwork Removed

(f) Artwork Removed

Figure 4.6: Example results from the artwork removal process.


In Memory of
Fredrick W Seamountain
B-13 Apr. 1867 - D-28 Jul. 1881
Mariah B. Seamountain
B-15 Jun. 1875 - D-20 Jul. 1881
Epheinish M. Seamountain
B-18 Mar. 1877 - D-22 Jul. 1881
Josephine Seamountain
B-12 Mar. 1890 - D-13 Jul. 1890

(a) Binarized Image (input)

(b) Inverse Binarized Image (input)

(c) Graph Weightings

(d) Graph Weightings

(e) Artwork Removed

(f) Artwork Removed

Figure 4.7: Example results from the artwork removal process.

separated from the noisy stone texture. Additionally, in Figure 4.7f, a large amount of the noise was removed.

Each of the sub-processes discussed qualitatively have been shown to remove regions of non-text resulting in a clean extraction of the textual information from the noisy and cluttered environment found on a cemetery headstone.

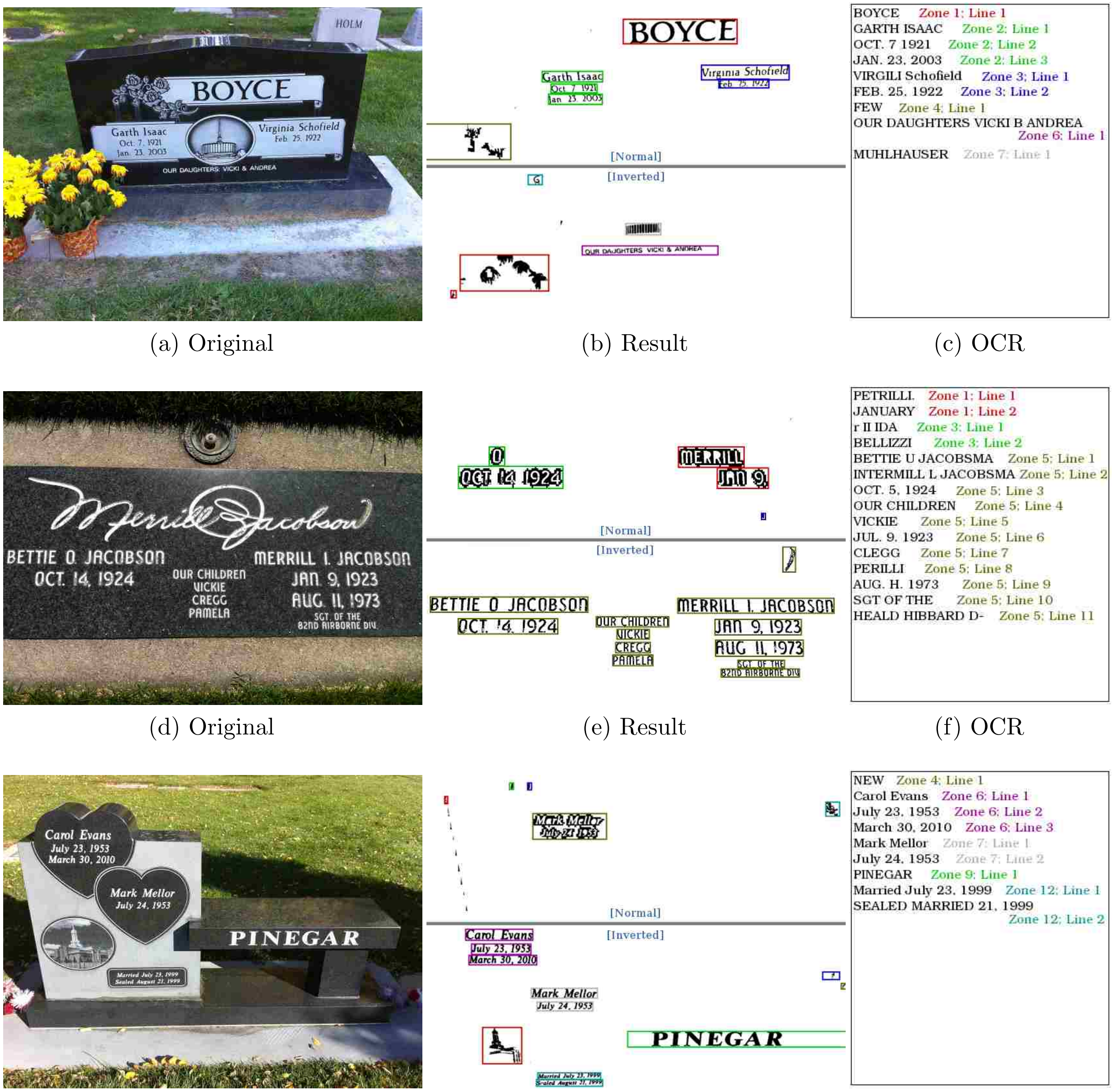
4.3.2 Overall

For a qualitative analysis of the system as a whole, the original image is given with the resulting binarized images (from both traditional and inverted Otsu binarization) with an overlay of the text lines discovered by the system (Figure 4.8, left and middle columns). Additionally, the transcription of the image is given on the right with colored labels corresponding to the text lines in the binarized images.

We see in these results that the system is able to accurately locate the text within the image, removing the majority of noise. Any noise that does remain has characteristics similar to text. Due to the headstone segmentation process, the text on the neighboring headstone found in Figure 4.8a is not retained, avoiding an incorrect transcription.

The advantage of performing both a traditional and inverted binarization is seen in Figure 4.8b and the resulting transcription in Figure 4.8c where both the key headstone information is recognized as text and transcribed. However, some additional noise is created through the dual binarization process as seen in Figure 4.8e causing incorrect insertions in the final transcription (Figure 4.8f). Additional qualitative results are found in Appendix A.

Qualitatively analyzing results on individual images reveals strengths and weaknesses that otherwise are imperceivable through a quantitative measure. We have demonstrated visually the results of each sub-process in the system as well as an analysis on the system as a whole.



(a) Original (b) Result (c) OCR

```
BOYCE    Zone 1; Line 1
GARTH ISAAC    Zone 2; Line 1
OCT. 7 1921    Zone 2; Line 2
JAN. 23, 2003    Zone 2; Line 3
VIRGILI Schofield    Zone 3; Line 1
FEB. 25, 1922    Zone 3; Line 2
FEW    Zone 4; Line 1
OUR DAUGHTERS VICKI B ANDREA
                    Zone 6; Line 1
MUHLHAUSER    Zone 7; Line 1
```

(d) Original (e) Result (f) OCR

```
PETRILLI.    Zone 1; Line 1
JANUARY    Zone 1; Line 2
r II IDA    Zone 3; Line 1
BELLIZZI    Zone 3; Line 2
BETTIE U JACOBSMA    Zone 5; Line 1
INTERMILL L JACOBSMA Zone 5; Line 2
OCT. 5, 1924    Zone 5; Line 3
OUR CHILDREN    Zone 5; Line 4
VICKIE    Zone 5; Line 5
JUL. 9, 1923    Zone 5; Line 6
CLEGG    Zone 5; Line 7
PERILLI    Zone 5; Line 8
AUG. H. 1973    Zone 5; Line 9
SGT OF THE    Zone 5; Line 10
HEALD HIBBARD D-    Zone 5; Line 11
```

(g) Original (h) Result (i) OCR

```
NEW    Zone 4; Line 1
Carol Evans    Zone 6; Line 1
July 23, 1953    Zone 6; Line 2
March 30, 2010    Zone 6; Line 3
Mark Mellor    Zone 7; Line 1
July 24, 1953    Zone 7; Line 2
PINEGAR    Zone 9; Line 1
Married July 23, 1999    Zone 12; Line 1
SEALED MARRIED 21, 1999
                    Zone 12; Line 2
```

Figure 4.8: Results of the proposed system where on the left the original images are displayed, the resulting images in the middle (using traditional Otsu [Normal] and inverted Otsu [Inverted]), and the OCR transcription on the right.

4.4 Summary

The proposed system has been shown to efficiently and accurately zone and segment textual data from a highly textured and cluttered background. This is shown quantitatively first by evaluating the OCR of the resulting image. Over the transcription of the raw headstone image, the F-measure is increased by approximately 50%, the accuracy by as much as 30%, and the error rate by more than 100 times.

The proposed system is then analyzed qualitatively by showing the resulting images of each process. Through these images the reader can visually understand the processing of the proposed system.

Chapter 5

Conclusion

In conclusion of the work presented, the contributions are detailed, the possible future work to improve the system is discussed and concluding remarks are given below.

5.1 Contribution

The contribution of this thesis is fourfold:

1. **The use of gradient orientation histograms to remove noise.** Gradient orientation histograms are commonly used as image descriptors for pattern recognition. This thesis demonstrates the power of gradient orientation histograms in identifying true edges within a noisy environment.
2. **The use of graph cut for a connected component analysis.** Connected component analysis is commonly performed for extracting text from scene and graphic text images to remove noise. Graph cut finds a globally optimal segmentation in which the proximity of neighboring and similar connected components are used. Graph cut improves the accuracy of a textual-based connected component analysis (rule based or machine learning), adding the dimension of proximity.
3. **Demonstration of improved performance through constraining the problem to a specific domain.** This work demonstrates the general principle that when a problem is constrained to a specific domain performance is increased significantly. In the

proposed system the domain is constrained both geometrically (zoning) and contextually (contextual validation of transcription).

4. **Cemetery Headstone Datasets.** Two datasets have been created to which future systems can compare accuracy and efficiency against the proposed system. Although datasets currently exist, the images are of a lower resolution than that available upon capture. Creating a new dataset in which images are of full resolution (i.e. 2592 x 1936) allows for more accurate processing and a better representation of the novel images captured in practice.

5.2 Future Work

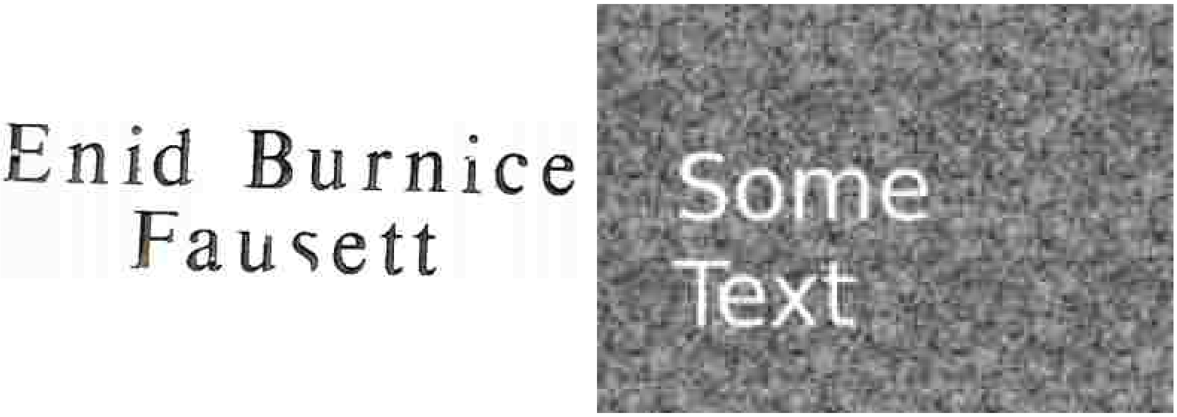
The system proposed in this thesis is foundational to future work in extracting the textual data found on headstone images. Possible future work has been grouped into four categories: OCR, contextual validation, neural networks, and mobile application development. These are each discussed individually in the following sections.

5.2.1 Improved OCR

The OCR transcriptions for the system were created using the open-source TesseractOCR engine (Smith [2007]). Experiments were performed with multiple OCR engines transcribing the same headstone images (Figure 5.1) to reveal the ability of each in noisy, headstone-like environments, the results of which can be seen in Table 5.1.

	ABBYY FineReader	**Adobe Acrobat**	**TesseractOCR**
Figure 5.1a	nid Burnice Fausett	Enid Burnice Fau ett	Enid Burnice Famett
Figure 5.1b	Some Text pr.. . -	{Not Recognizable}	xgv _Vw N ‘__w_\M¿_‘_4 v__ JV Mn _V_}__%V_T_mH_V_m _W

Table 5.1: The OCR results of Figures 5.1a and 5.1b given from ABBYY Fine Reader, Adobe Acrobat, and TesseractOCR.



(a) (b)

Figure 5.1: Images used to test OCR systems

The first experiment provided cleanly segmented engraved characters from a headstone image (Figure 5.1a) to each of the OCR engines. All OCR engines performed well: ABBYY Fine Reader (ABBYY [2012]) and Adobe Acrobat each result in an edit distance of one and an edit distance of two for the TesseractOCR engine.

The second experiment provided clean text overlaid on a noisy background (Figure 5.1b). ABBYY Fine Reader successfully transcribed the text whereas the other OCR engines failed. Such results suggest that the commercial-grade OCR engine ABBYY Fine Reader may be more robust towards a noisy environment than the TesseractOCR engine. Thus future work may include the installment of various OCR systems to explore the change in accuracy.

An additional possibility to improving the OCR results of headstone images is to train the OCR engine specifically on binarized headstone characters (i.e. a binarized version of Figure 5.1a). Such efforts may prove useful as the OCR engine will learn to overlook the stochastic noise (i.e. the white ball lodged the letter 'E' in Figure 1.1) in the sample images as well as the inner-character shadows.

Future work may also include an iterative process to refine the image zones where either a contextual correction was made or a low-confidence OCR transcription was given. The image zones may be mapped back to the original image and reprocessed to produce a higher confidence OCR result as well as confirm that the contextual validation was correct.

As the OCR transcription was not the focus of this thesis, additional work in this area may improve the quantitative results.

5.2.2 Improved Contextual Validation

A process was created for error-correcting the automated OCR of the system based on using the constrained domain of cemetery headstones. However, this process acts as a proof on concept rather than a production-grade system. Much can be gained in leveraging previous work in information extraction. The work proposed in Packer [2011] demonstrates the usefulness of such work.

Further research in information extraction and error correction in the domain of cemetery headstones will likely improve the qualitative results.

5.2.3 Neural Networks

Neural networks play an important role in both zoning and artwork removal. A key feature of neural networks is their ability to learn a solution to a given problem using training data. As more training data is provided to the neural network, a more precise understanding of the task is gained by the system. Therefore, over time as more training data is created through use of the system, additional training may be performed (upon labelling of the collected data). Therefore the neural network will likely improve its ability to classify.

A large body of literature exists for machine learning and classifying tasks. Methods to improve neural network classification may be considered in addition to completely different classifiers such as support vector machines or a combined system such as an ensemble. Many options are available, all of which must have sufficient run time efficiency to be used in a real time system.

5.2.4 Mobile Device Implementation

The proposed system is able to remove noise and artwork effectively to provide a readable image for the user or for an OCR engine. When combined with an OCR engine, the system becomes a complete headstone transcription program where a headstone image is given with a transcription returned to the user. Such a system is valuable to the thousands of volunteers at BillionGraves and at Find A Grave and would increase their throughput significantly.

The development of a complete automated-transcription system will benefit many genealogists who wish to have the data found on headstones indexed and available online.

5.3 Concluding Remarks

A novel system has been described in which noisy and cluttered text has been zoned and segmented through use of neural networks and graph cut with the specific application to cemetery headstones.

Previous and related work have been discussed in Chapter 2. While some previous literature has proposed solutions to extracting the text from low-contrasting engraved characters (Garain et al. [2008] Thillou and Gosselin [2006]), none have discussed the extraction of engraved text that suffer from headstone-specific challenges such as weathering, shadows, and stochastic noise (i.e. stone texture and foreign objects).

To accurately zone and segment textual data found on headstones, concepts from previous work have been used. Gradient orientation histograms have been used as the key features in the neural networks for both zoning and artwork removal. Additionally, graph cut segmentation is applied in the artwork removal process to find a globally optimal segmentation between text and artwork.

The task of extracting text is separated into sub-processes in which each reduces the problem to a simpler task. First, the headstone is segmented as a whole from the surrounding background. Second, the stone texture is removed to zone the engravings found on the headstone. Third, Otsu binarization is used to segment and cleanse candidate regions for text and fourth, a connected component analysis is performed in which text is segmented from all remaining artwork and noise. Additionally, OCR is performed and is followed by a validation process which error-corrects the transcription using the constrained domain to improve OCR accuracy. These processes are discussed in Chapter 3.

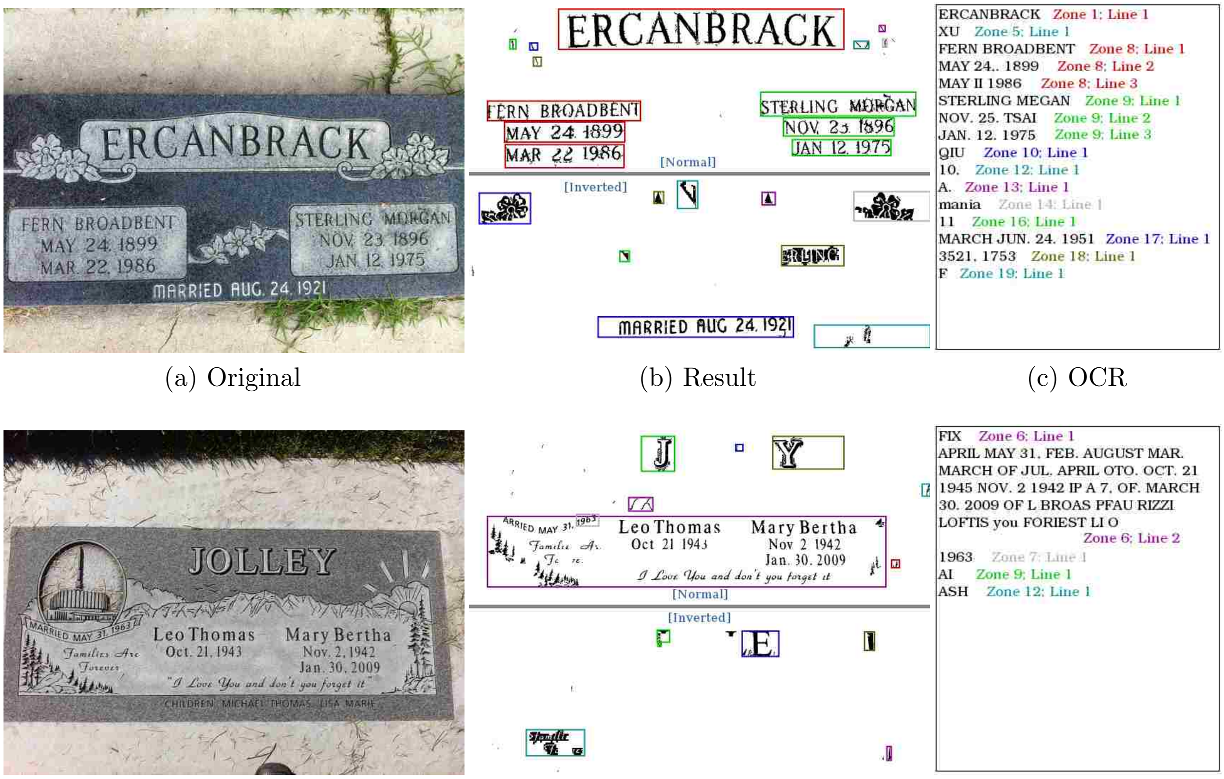
The results of each sub-process and of the complete system have been demonstrated both quantitatively and qualitatively in Chapter 4. It has been shown that an out-of-the-box OCR system can transcribe the processed images with a 70.86% accuracy on the headstone's key data. Additionally, the proposed system reduces the error rate of the transcriptions by over 100 times when compared with raw image transcriptions, as well as increase the F-measure by 50-56%.

An automated system for the transcription of cemetery headstones is of great genealogical and historical worth. In a domain where the documents (headstones) continue to deteriorate every year, the current manual process used for gathering headstone data at multiple organizations is slow, tedious, and does not scale with the large number of headstones that remain unindexed.

This thesis acts as the primary work in this domain, focusing on removing the noise on headstone images and facilitating the OCR transcription process. Due to the constrained domain of cemetery headstones, the application of a stronger contextual validation with a more accurate and specialized OCR engine may make automation possible and allow easy access for all to information that was previously accessible to only a few.

Appendix A

Additional Qualitative Results



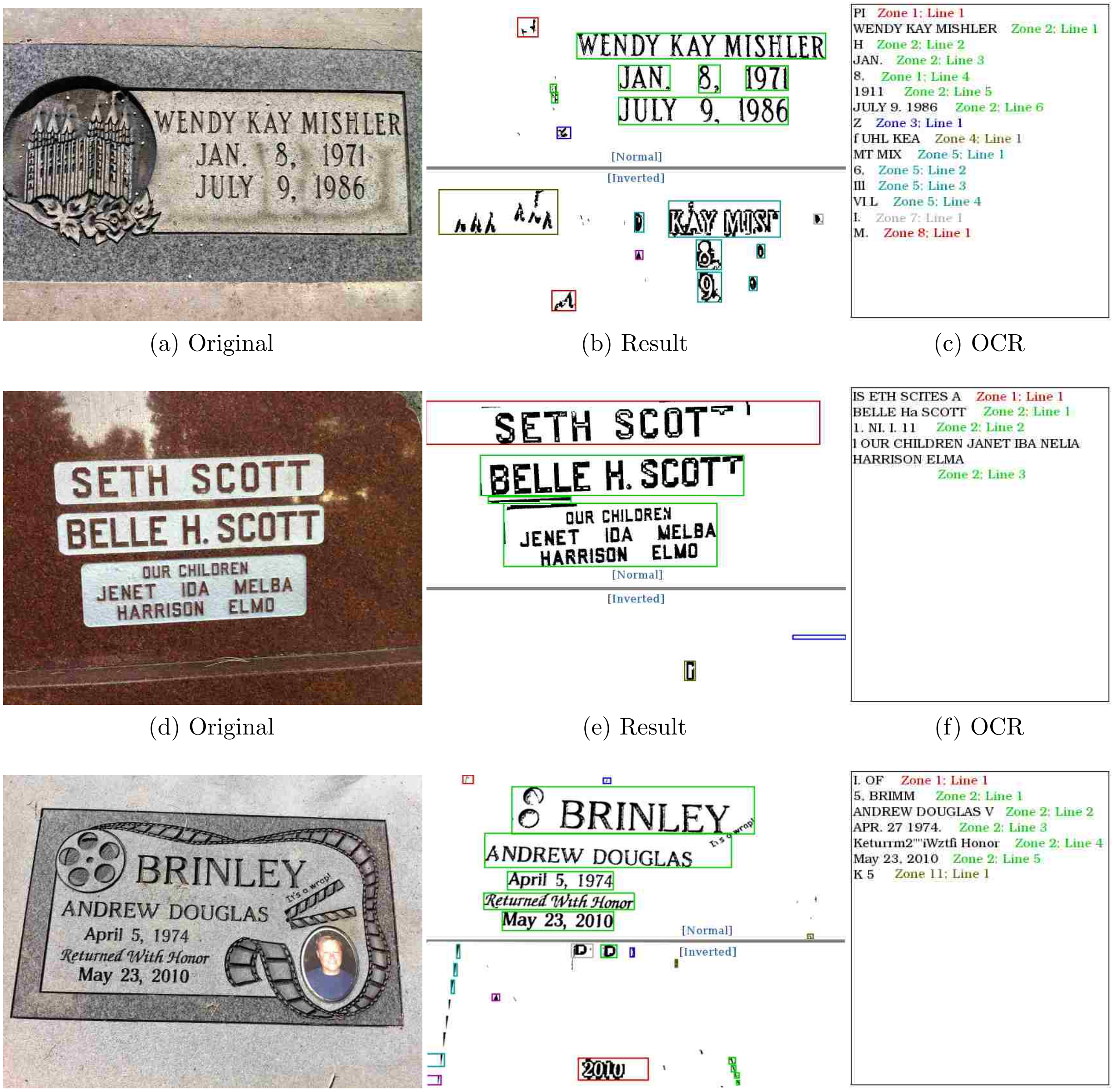
(a) Original (b) Result (c) OCR

```
ERCANBRACK   Zone 1; Line 1
XU   Zone 5; Line 1
FERN BROADBENT   Zone 8; Line 1
MAY 24.. 1899   Zone 8; Line 2
MAY II 1986   Zone 8; Line 3
STERLING MEGAN   Zone 9; Line 1
NOV. 25. TSAI   Zone 9; Line 2
JAN. 12. 1975   Zone 9; Line 3
QIU   Zone 10; Line 1
10.   Zone 12; Line 1
A.   Zone 13; Line 1
mania   Zone 14; Line 1
11   Zone 16; Line 1
MARCH JUN. 24. 1951   Zone 17; Line 1
3521. 1753   Zone 18; Line 1
F   Zone 19; Line 1
```

(d) Original (e) Result (f) OCR

```
FIX   Zone 6; Line 1
APRIL MAY 31. FEB. AUGUST MAR.
MARCH OF JUL. APRIL OTO. OCT. 21
1945 NOV. 2 1942 IP A 7. OF. MARCH
30. 2009 OF L BROAS PFAU RIZZI
LOFTIS you FORIEST LI O
                        Zone 6; Line 2
1963   Zone 7; Line 1
AI   Zone 9; Line 1
ASH   Zone 12; Line 1
```

Figure A.1



```
PI   Zone 1; Line 1
WENDY KAY MISHLER   Zone 2; Line 1
H   Zone 2; Line 2
JAN.   Zone 2; Line 3
8.   Zone 1; Line 4
1911   Zone 2; Line 5
JULY 9. 1986   Zone 2; Line 6
Z   Zone 3; Line 1
ſ UHL KEA   Zone 4; Line 1
MT MIX   Zone 5; Line 1
6.   Zone 5; Line 2
III   Zone 5; Line 3
VI L   Zone 5; Line 4
I.   Zone 7; Line 1
M.   Zone 8; Line 1
```

(a) Original

(b) Result

(c) OCR

```
IS ETH SCITES A   Zone 1; Line 1
BELLE Ha SCOTT   Zone 2; Line 1
1. NI. I. 11   Zone 2; Line 2
l OUR CHILDREN JANET IBA NELIA
HARRISON ELMA
                    Zone 2; Line 3
```

(d) Original

(e) Result

(f) OCR

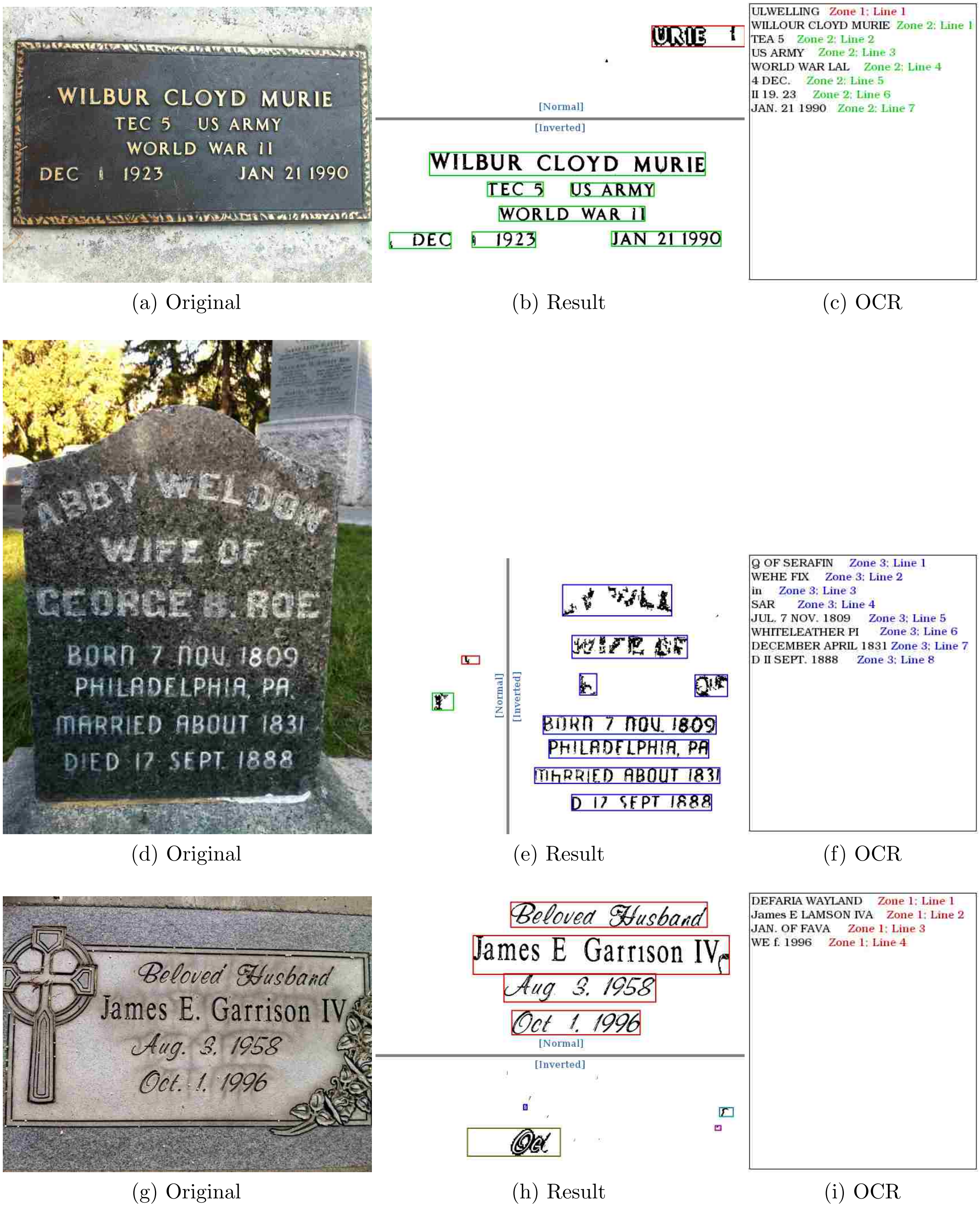
```
I. OF   Zone 1; Line 1
5. BRIMM   Zone 2; Line 1
ANDREW DOUGLAS V   Zone 2; Line 2
APR. 27 1974.   Zone 2; Line 3
Keturrm2""iWztfi Honor   Zone 2; Line 4
May 23. 2010   Zone 2; Line 5
K 5   Zone 11; Line 1
```

(g) Original

(h) Result

(i) OCR

Figure A.2



(a) Original (b) Result (c) OCR

```
ULWELLING  Zone 1: Line 1
WILLOUR CLOYD MURIE  Zone 2: Line 1
TEA 5  Zone 2: Line 2
US ARMY  Zone 2: Line 3
WORLD WAR LAL  Zone 2: Line 4
4 DEC.  Zone 2: Line 5
II 19. 23  Zone 2: Line 6
JAN. 21 1990  Zone 2: Line 7
```

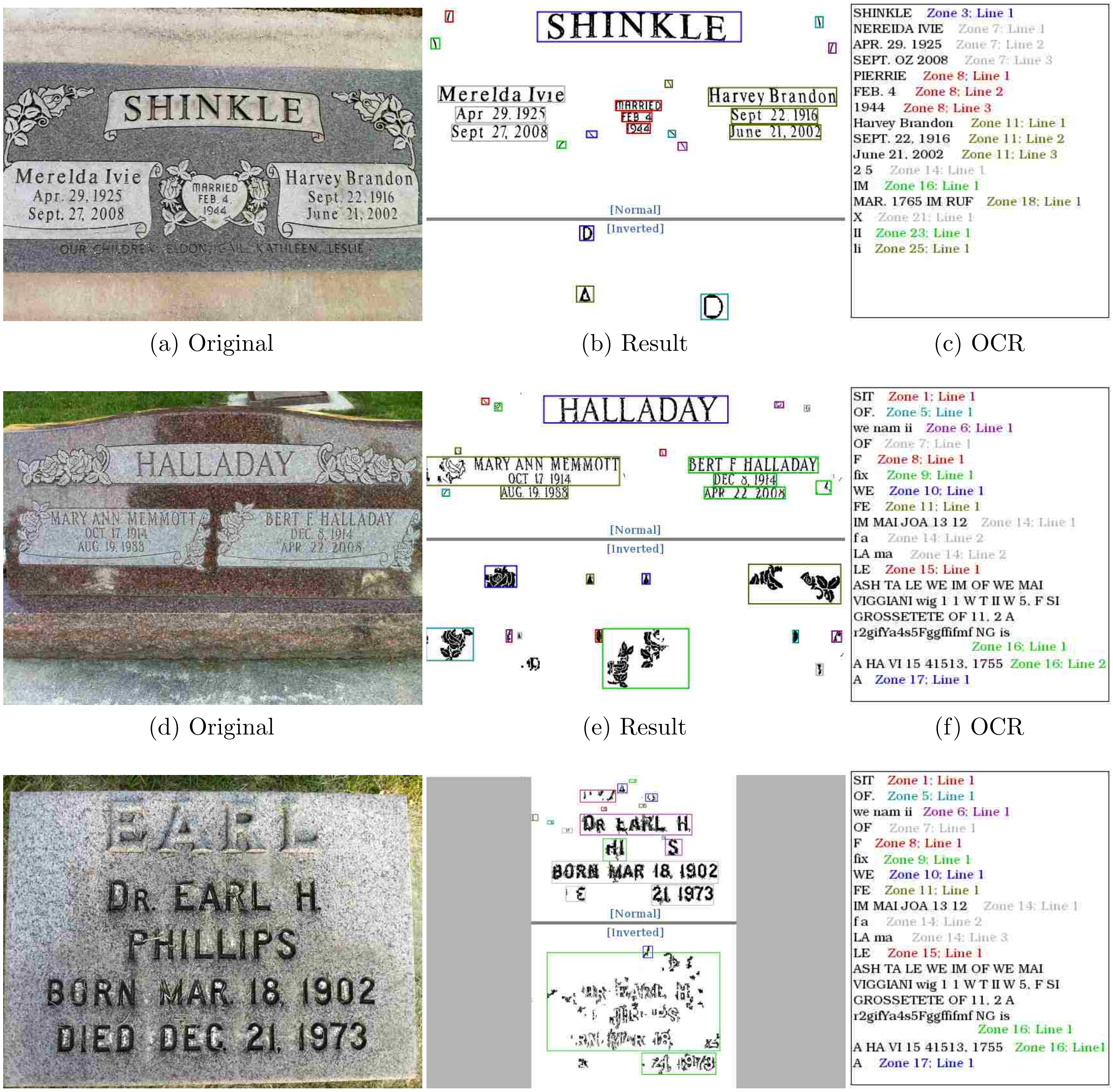
(d) Original (e) Result (f) OCR

```
Q OF SERAFIN  Zone 3: Line 1
WEHE FIX  Zone 3: Line 2
in  Zone 3: Line 3
SAR  Zone 3: Line 4
JUL. 7 NOV. 1809  Zone 3: Line 5
WHITELEATHER PI  Zone 3: Line 6
DECEMBER APRIL 1831 Zone 3: Line 7
D II SEPT. 1888  Zone 3: Line 8
```

(g) Original (h) Result (i) OCR

```
DEFARIA WAYLAND  Zone 1: Line 1
James E LAMSON IVA  Zone 1: Line 2
JAN. OF FAVA  Zone 1: Line 3
WE f. 1996  Zone 1: Line 4
```

Figure A.3



(a) Original

(b) Result

(c) OCR

```
SHINKLE    Zone 3; Line 1
NEREIDA IVIE    Zone 7; Line 1
APR. 29. 1925    Zone 7; Line 2
SEPT. OZ 2008    Zone 7; Line 3
PIERRIE    Zone 8; Line 1
FEB. 4    Zone 8; Line 2
1944    Zone 8; Line 3
Harvey Brandon    Zone 11; Line 1
SEPT. 22, 1916    Zone 11; Line 2
June 21, 2002    Zone 11; Line 3
2 5    Zone 14; Line 1
IM    Zone 16; Line 1
MAR. 1765 IM RUF    Zone 18; Line 1
X    Zone 21; Line 1
II    Zone 23; Line 1
li    Zone 25; Line 1
```

(d) Original

(e) Result

(f) OCR

```
SIT    Zone 1; Line 1
OF.    Zone 5; Line 1
we nam ii    Zone 6; Line 1
OF    Zone 7; Line 1
F    Zone 8; Line 1
fix    Zone 9; Line 1
WE    Zone 10; Line 1
FE    Zone 11; Line 1
IM MAI JOA 13 12    Zone 14; Line 1
f a    Zone 14; Line 2
LA ma    Zone 14; Line 2
LE    Zone 15; Line 1
ASH TA LE WE IM OF WE MAI
VIGGIANI wig 1 1 W T II W 5, F SI
GROSSETETE OF 11, 2 A
r2gifYa4s5Fggffifmf NG is
                    Zone 16; Line 1
A HA VI 15 41513, 1755  Zone 16; Line 2
A    Zone 17; Line 1
```

(g) Original

(h) Result

(i) OCR

```
SIT    Zone 1; Line 1
OF.    Zone 5; Line 1
we nam ii    Zone 6; Line 1
OF    Zone 7; Line 1
F    Zone 8; Line 1
fix    Zone 9; Line 1
WE    Zone 10; Line 1
FE    Zone 11; Line 1
IM MAI JOA 13 12    Zone 14; Line 1
f a    Zone 14; Line 2
LA ma    Zone 14; Line 3
LE    Zone 15; Line 1
ASH TA LE WE IM OF WE MAI
VIGGIANI wig 1 1 W T II W 5, F SI
GROSSETETE OF 11, 2 A
r2gifYa4s5Fggffifmf NG is
                    Zone 16; Line 1
A HA VI 15 41513, 1755  Zone 16; Line1
A    Zone 17; Line 1
```

Figure A.4

References

ABBYY. ABBYY FineReader, February 2012. URL http://www.abbyyinfo.com/finereaderpro/.

A. Antonacopoulos, B. Gatos, and D. Bridson. Page segmentation competition. In *Proceedings of the Ninth International Conference on Document Analysis and Recognition*, volume 2, pages 1279–1283. IEEE, 2007.

H.S. Baird. The skew angle of printed documents. In *Document Image Analysis*, pages 204–208. IEEE, 1995.

BillionGraves, April 2011. URL http://www.billiongraves.com/.

Y. Boykov and M. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In *Proceedings of the Eighth IEEE International Conference on Computer Vision*, volume 1, pages 105–112. IEEE, 2001.

Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(11):1222–1239, 2001.

J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

R. Cattoni, T. Coianiz, S. Messelodi, and C. Modena. Geometric layout analysis techniques for document image understanding: a review. Technical report, ITC-IRST, 1998.

W.Y. Chen and S.Y. Chen. Adaptive page segmentation for color technical journals' cover images. *Image and Vision Computing*, 16(12-13):855–877, 1998.

C.S. Christiansen and W.A. Barrett. Data acquisition from cemetery headstones. In *Proceedings of the SPIE Symposium on Document Recognition and Retrieval*. SPIE, 2013.

P. Clark and M. Mirmehdi. Recognising text in real scenes. *International Journal on Document Analysis and Recognition*, 4(4):243–257, 2002.

P. Clark and M. Mirmehdi. Rectifying perspective views of text in 3d scenes using vanishing points. *Pattern Recognition*, 36(11):2673–2686, 2003.

D.W. Embley, D.M. Campbell, Y.S. Jiang, S.W. Liddle, D.W. Lonsdale, Y.K. Ng, and R.D. Smith. Conceptual-model-based data extraction from multiple-record web pages. *Data & Knowledge Engineering*, 31(3):227–251, 1999.

Find A Grave, April 2011. URL `http://www.findagrave.com/`.

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 *Proceedings of the 19th International Conference on Pattern Recognition*, pages 1–4. IEEE, 2008.

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

B. Gatos, I. Pratikakis, and S. Perantonis. Adaptive degraded document image binarization. *Pattern Recognition*, 39(3):317–327, 2006.

H. Hase, T. Shinokawa, and M. Yoneda. Character string extraction from color documents. *Pattern Recognition*, 34(7):1349–1365, 2001.

H. Hase, M. Yoneda, S. Tokai, J. Kato, and C.Y. Suen. Color segmentation for text extraction. *International Journal on Document Analysis and Recognition*, 6(4):271–284, 2003.

J. He, Q.D.M. Do, A.C. Downton, and J.H. Kim. A comparison of binarization methods for historical archive documents. In *Proceedings of the Eighth International Conference on Document Analysis and Recognition*, volume 1, pages 538–542. IEEE, 2005.

W.J. Ho and C.F. Osborne. Texture segmentation using multi-layered backpropagation. In *Proceedings of the IEEE International Joint Conference on Neural Networks*, volume 2, pages 981–986. IEEE, 1991.

N.R. Howe. A Laplacian energy for document binarization. In *Proceedings of the International Conference on Document Analysis and Recognition*, pages 6–11. IEEE, 2011.

A.K. Jain and B. Yu. Automatic text location in images and video frames. *Pattern Recognition*, 31(12):2055–2076, 1998.

A.K. Jain, M. Jianchang, and K.M. Mohiuddin. Artificial neural networks: a tutorial. *Computer*, 29(3):31–44, 1996.

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.

J. Kanai, S.V. Rice, T.A. Nartker, and G. Nagy. Automated evaluation of OCR zoning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(1):86–90, 1995.

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

J.G. Kuk, N.I. Cho, and K.M. Lee. MAP-MRF approach for binarization of degraded document image. In *Proceedings of the 15th IEEE International Conference on Image Processing*, pages 2612–2615. IEEE, 2008.

V.I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet Physics Doklady*, volume 10, pages 707–710, 1966.

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

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.

D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

C. Mancas-Thillou and M. Mancas. Comparison between pen-scanner and digital camera acquisition for engraved character recognition. In *Proceedings of the 2nd International Workshop on Camera-Based Document Analysis and Recognition*, page 130. Springer, 2007.

B.S. Manjunath, T. Simchony, and R. Chellappa. Stochastic and deterministic networks for texture segmentation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38 (6):1039–1049, 1990.

Edward Mendelson. ABBYY FineReader 10 Professional Edition, November 2009. URL `http://www.pcmag.com/article2/0,2817,2355520,00.asp`.

R. Milewski and V. Govindaraju. Extraction of handwritten text from carbon copy medical form images. In Horst Bunke and A. Spitz, editors, *Document Analysis Systems VII*, volume 3872 of *Lecture Notes in Computer Science*, pages 106–116. Springer, 2006.

D. Molkentin. *The Book of Qt 4: the Art of Building Qt Applications.* No Starch Pr, 2007.

S.L. Murphy, J. Xu, and K.D. Kochanek. Deaths: Preliminary data for 2010. *National Vital Statistics Reports*, 60(4):1–68, 2012.

G.K. Myers, R.C. Bolles, Q.T. Luong, J.A. Herson, and H.B. Aradhye. Rectification and recognition of text in 3-d scenes. *International Journal on Document Analysis and Recognition*, 7(2):147–158, 2005.

Y. Nakano, Y. Shima, H. Fujisawa, J. Higashino, and M. Fujinawa. An algorithm for the skew normalization of document image. In *Proceedings of the 10th International Conference on Pattern Recognition*, volume 2, pages 8–13. IEEE, 1990.

Names In Stone, April 2011. URL `http://www.namesinstone.com/`.

W. Niblack. *An Introduction to Digital Image Processing.* Prentice Hall, 1986.

O. Nina, B. Morse, and W. Barrett. A recursive Otsu thresholding method for scanned document binarization. In *Proceedings of the IEEE Workshop on Applications of Computer Vision*, pages 307–314. IEEE, 2011.

N. Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296): 23–27, 1975.

T.L. Packer. Performing information extraction to improve OCR error detection in semi-structured historical documents. In *Proceedings of the Workshop on Historical Document Imaging and Processing*, pages 67–74. ACM, 2011.

N.R. Pal and S.K. Pal. A review on image segmentation techniques. *Pattern Recognition*, 26 (9):1277–1294, 1993.

W. Postl. Detection of linear oblique structures and skew scan in digitized documents. In *Proceedings of the International Conference on Pattern Recognition*, pages 687–689. IEEE, 1986.

J. Puzicha, T. Hofmann, and J.M. Buhmann. Histogram clustering for unsupervised segmentation and image retrieval. *Pattern Recognition Letters*, 20(9):899–909, 1999.

T. Retornaz and B. Marcotegui. Scene text localization based on the ultimate opening. In *Proceedings of the International Symposium on Mathematical Morphology*, volume 1, pages 177–188. CSIRO, 2007.

H. Samet. The quadtree and related hierarchical data structures. *ACM Computing Surveys*, 16(2):187–260, 1984.

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

M. Seeger and C. Dance. Binarising camera images for OCR. In *Proceedings of the Sixth International Conference on Document Analysis and Recognition*, pages 54–58. IEEE, 2001.

R. Smith. An overview of the Tesseract OCR Engine. In *Proceedings of the Ninth International Conference on Document Analysis and Recognition*, volume 2, pages 629–633. IEEE, 2007.

J. Takakura, A. Kitadai, M. Nakagawa, H. Baba, and A. Watanabe. Techniques to enhance images for mokkan interpretation. In *Proceedings of the International Conference on Frontiers in Handwriting Recognition*, pages 358–362. IEEE, 2010.

C. Thillou and B. Gosselin. Segmentation-based binarization for color degraded images. In K. Wojciechowski, B. Smolka, H. Palus, R.S. Kozera, W. Skarbek, and L. Noakes, editors, *Computer Vision and Graphics*, volume 32 of *Computational Imaging and Vision*, pages 808–813. Springer, 2006.

United States Census Bureau. Frequently occurring first names and surnames from the 1990 census, October 1995. URL `http://www.census.gov/genealogy/names/names_files.html`.

B. Wang, X.F. Li, F. Liu, and F.Q. Hu. Color text image binarization based on binary texture analysis. *Pattern Recognition Letters*, 26(11):1650–1657, 2005.

C. Wolf, J. Jolion, and F. Chassaing. Text localization, enhancement and binarization in multimedia documents. In *Proceedings of the International Conference on Pattern Recognition*, volume 2, pages 1037–1040. IEEE, 2002.