

EXPLORING THE USE OF SIGNS DURING PROTEST ACTIVITIES THROUGH
SOCIAL MEDIA DATA INTEGRATION: THE CASE OF OCCUPYWALLSTREET

by

Kathryn M. Kash
A Thesis
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Master of Science
Geoinformatics and Geospatial Intelligence

Committee:

_____	Dr. Arie Croitoru, Thesis Director
_____	Dr. Anthony Stefanidis, Committee Member
_____	Dr. Andrew Crooks, Committee Member
_____	Dr. Anthony Stefanidis, Department Chair
_____	Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science
_____	Dr. Peggy Agouris, Dean, College of Science
Date: _____	Spring Semester 2016 George Mason University Fairfax, VA

EXPLORING THE USE OF SIGNS DURING PROTEST ACTIVITIES THROUGH
SOCIAL MEDIA DATA INTEGRATION: THE CASE OF OCCUPYWALLSTREET

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of
Science at George Mason University

by

Kathryn M. Kash
Bachelor of Science
James Madison University, 2009

Director: Arie Croitoru, Associate Professor
Department of Geography and GeoInformation Science

Spring Semester 2016
George Mason University
Fairfax, VA



This work is licensed under a [creative commons attribution-noncommercial 3.0 unported license](https://creativecommons.org/licenses/by-nc/3.0/).

DEDICATION

This is dedicated to my parents, Peyton and Denise, and my soon-to-be husband the day after my thesis is due, Mike. Thank you for always having confidence in me and supporting this journey while I finished my thesis and planned our wedding at the same time. And finally, this is dedicated to our dog Comet. We've only had you a short time and our journey with you is going to be cut even shorter, but thank you for being at my feet the entire time, and encouraging me to take breaks.

ACKNOWLEDGEMENTS

I would like to thank my committee, Dr. Arie Croituru, Dr. Anthony Stefanidis, and Dr. Andrew Crooks, for taking the time to work with me and provide guidance. Special thanks goes out to my place of employment, the Army Geospatial Center, for fostering the work and setting up the certificate and master's classes on location. I would also like to thank my dear friends who have supported me along the way: Ms. Robin Rodgers; Mrs. Jackie Hunke; and Ms. Heather Speight. I would be lost without your guidance, support, and encouragement. And thank you to my amazing wedding vendors and bridesmaids (Mrs. Ainsley Pittman, Mrs. Jennifer Shankel and Ms. Reihle Kash) who took on a lot the last 6 months! I can't wait to celebrate all the things.

TABLE OF CONTENTS

	Page
List of Tables	vii
List of Figures	viii
List of Abbreviations and Symbols.....	x
Abstract	xi
Chapter One	1
1.1 Introduction	1
1.2 Thesis Structure.....	4
Chapter Two.....	5
2.1 Protests, the Use of Signs, and Location.....	5
2.2 The Occupy Wall Street Movement.....	12
2.3 The Use of Social Media in Protest Activities	14
2.3.1 Twitter and OccupyWallStreet	20
2.3.2 Flickr.....	24
2.4 Text Detection	27
2.4.1 Texture Based Methods	29
2.4.2 Connected Components.....	32
2.4.3 Maximally Stable Extremal Regions (MSERs).....	33
2.4.4 Stroke Width Transform.....	35
2.4.5 Hybrid Methods	37
Chapter Three.....	39
3.1 Objective of Research	39
Chapter Four	42
4.1 Analysis Overview	42

4.2 Data	44
4.3 Text Detection	46
4.3.1 Detecting Text using MSERs	47
4.3.2 Geometric Filtering	49
4.3.3 Stroke Width Transform	51
4.3.4 Merge Text Regions	53
4.3.5 Optical Character Recognition	56
4.4 Spatio-temporal analysis	57
4.5 Spatial Data Visualization and Analysis	58
Chapter Five	60
5.1 Text Detection Accuracy Assessment	60
5.2 Spatial Data Visualization and Analysis Results	68
5.3 Discussion	90
Chapter Six	96
6.1 Conclusions	96
6.2 Future Work	97
References	101

LIST OF TABLES

Table	Page
Table 1: Confusion Matrix on subset of images for text detection and OCR process.....	61
Table 2: Cluster ID, Location, and Number of Points Per Cluster	69
Table 3: Number of Records Per City and Percent Detected vs Not Detected.....	70

LIST OF FIGURES

Figure	Page
Figure 1: Pro Fox Hunting Poster (Burrige, 2008)	9
Figure 2: Overall Text Information Extraction Process (derived from Jung et al., 2004)	28
Figure 3: Connected Component Analysis Overview (derived from Pan et al., 2011).	33
Figure 4: Overall Process Workflow	43
Figure 5: Text Detection and Recognition Workflow	47
Figure 6: Original Input Image	48
Figure 7: Detect MSER's	49
Figure 8: After Removing Non-Text Regions Based on Geometric Properties	50
Figure 9: Region Image and Stroke Width Image	52
Figure 10: After Removing Non-Text Regions Based on Stroke Width Variation	53
Figure 11: Expanded Bounding Boxes	55
Figure 12: Detected Text	56
Figure 13: OCR Results for Example Image	57
Figure 14: True Positive Examples	62
Figure 15: True Negative Examples	63
Figure 16: False Positive Examples	64
Figure 17: False Negative Examples	65
Figure 18: Examples of Incidental Text	67
Figure 19: Overview Map of all Images Harvested from Flickr relating to Occupy Wall Street in Fall 2011	68
Figure 20: Washington, D.C. Inset Map	71
Figure 21: Los Angeles, California Inset Map	72
Figure 22: San Francisco and Oakland, California Inset Map	73
Figure 23: Chicago, Illinois Inset Map	74
Figure 24: New York City, New York Inset Map	75
Figure 25: Urban Centers with OWS Activity Clustered by DenStream	76
Figure 26: Flickr Images with Text Detected and Twitter Point Density	78
Figure 27: Flickr Images with No Text Detected and Twitter Point Density	79
Figure 28: Manhattan Subset of Twitter Density and Flickr Images with no Text Detected	81
Figure 29: Manhattan Subset of Twitter Density and Flickr Images with Text Detected	82

Figure 30: Washington Square Park Inset Map with 67 Text Images Detected and Examples.....	84
Figure 31: Brooklyn Bridge Inset Map with 9 Text Images Detected and Examples	85
Figure 32: Union Square Park Inset Map with 48 Text Images Detected and Examples.	86
Figure 33: Times Square Inset Map with 128 Text Images Detected and Examples	88
Figure 34: Zuccotti Park Inset Map with 2587 Text Images Detected and Examples	89

LIST OF ABBREVIATIONS AND SYMBOLS

&	and
%	percent
Application Programming Interface	API
Bag of Visual Words	BOVW
Connected Components	CC
Convolutional Neural Network	CNN
Exchangeable image file format	EXIF
Extremal Regions	ER
Global Database of Events, Location and Tone	GDELT
Global Positioning Systems	GPS
Histogram of Oriented Gradients	HOG
International Conference on Document Analysis and Recognition	ICDAR
Maximally Stable Extremal Regions	MSER
Occupy Wall Street	OWS
Optical Character Recognition	OCR
Stroke Width Transform	SWT
Support Vector Machine	SVM

ABSTRACT

EXPLORING THE USE OF SIGNS DURING PROTEST ACTIVITIES THROUGH SOCIAL MEDIA DATA INTEGRATION: THE CASE OF OCCUPYWALLSTREET

Kathryn M. Kash, M.S.

George Mason University, 2016

Thesis Director: Dr. Arie Croitoru

Signs have long been used extensively in protest activities, such as political rallies, social unrest gatherings, picket lines, company boycotts, and marches. Consequently, recent studies have explored the use of posters in protests in terms of their role in the overall protest rhetoric, both visual and textual. While such studies articulate the central role signs have in protest activities, the locations of protest signs and their relationship to the spatial characteristics of protest activity often remain unexplored. Social media offer a new lens through which the location of signs in protest activities could be explored. In particular, the use of geotagged social media (such as Twitter and Flickr) contributions during protest activities can provide rich information about where signs are located, what narratives emerge from them, and how they are integrated into the

overall activity. Toward this goal, this research proposes an approach for integrating geotagged Flickr images and Twitter messages for examining the relation between sign locations and protest activity locations. In this approach, the Stroke Width Transform and Optical Character Recognition are used to detect text in protest-related images. The location of these images is then compared to geotagged Twitter messages relating to the same protest activity, and patterns of interest are detected. The utility of this approach was examined through the analysis of the 2011 OccupyWallStreet protests across the United States. The results suggest that, overall, signs are immersed in the protest activity, but they tend to concentrate in specific locations that are likely to have a more central role in context of the protest.

CHAPTER ONE

1.1 Introduction

Merriam-Webster (2016) defines social media as “forms of electronic communication (as websites for social networking and microblogging) through which users create online communities to share information, ideas, personal messages, and other content (as videos).” In the last decade, social media platforms and usage have increased exponentially. All over the world, people are posting to their various sites (e. g. Facebook, Twitter, Instagram, Flickr), from the mundane to the thrilling. Natural disasters, terrorism events, national uprisings, and protests, to name a few, are all being tracked via social media to be seen by the entire world. With the advancement of smart phones (with an embedded camera and global positioning system [GPS]), consumer digital cameras, readily available and low cost internet, and the ability to share an image in a few seconds, people are posting on social media at an alarming rate, often including images or video. Information and images are spreading at the push of a finger tap, which can provide a wealth of information about events around the world. Being able to automatically ingest such images and extract information from them is therefore becoming an important task in many application areas (traffic panel detection, license plate detection, or blind and visually impaired assistance), while also providing greater intelligence and knowledge of an event or activity. The amount of data that is being

generated is large and only getting larger (known as big data), and it offers a new way to study complex human systems (Croitoru et al., 2014).

One such event or activity that has been using social media as a platform, organization and logistics coordination tool, and for news sharing is protest activities or contentious politics. Contentious politics can be defined as “concerted, counter-hegemonic social and political action, in which differently positioned participants come together to challenge dominant systems of authority, in order to promote and enact alternative imaginaries” (Leitner et al., 2008, p. 1). It is used interchangeably to describe social movements, protests, revolutions, and demonstrations. Posters and their use in social movement literature is scarce, yet the need to study them to gain a full understanding of protests has been determined (Wildermuth et al., 2014; Burrridge, 2008). In order to identify which protest signs have text contained in them (relating to the protest), a text detection and recognition process has to be used, due to the large quantity of information.

Traditionally, computer vision used optical character recognition (OCR) to ingest text on images to be machine readable. OCR has made considerable advances in the last few decades. By scanning pages or documents of well-formatted, printed text, OCR algorithms are very effective at text recognition. The text is segmented to differentiate between background pixels. Applications include assistance for the blind and visually-impaired, automated business data entry, license plate identification, conversion of printed documents to soft copies (which can then be edited), and converting handwriting to a digital format. But when those same algorithms are applied on natural scenes (such

as a photograph), detecting text is still a challenge (Ikica & Peer, 2014; Pan et al., 2011). Factors in natural scenes, such as extreme size and font variation, multiple languages, blur, color, and strong background clutter and noise, prove almost impossible for traditional OCR to perform effectively. Variations in font, style, scale, orientation, geometric and photometric distortions, thickness, color, size, texture, lighting, partial occlusions, image resolutions, and complex background prove to be the greatest challenge in automatic text detection and recognition algorithms (Wang et al., 2015; Yin et al., 2014; Gonzalez et al., 2012; Li & Yu, 2012; Chen et al., 2011; Jung et al., 2004). Due to all these differences in natural scene images, designing a ‘one size fits all’ system that can accurately detect text proves genuinely challenging. A variety of methods have been developed and improved upon to assist with these issues in natural scene text detection.

Many improvements have been made to text recognition, specifically in natural scenes. By using various methods of edge detection and image processing, many algorithms that improve this process have been developed. This research focuses on one suite of text detection algorithms to extract text from natural scenes, using the Occupy Wall Street (OWS) movement as a case study. Flickr images relating to OWS are harvested, then the text recognition process is used to determine if text is contained. Once text is detected, the spatial relationship of images with detected text is examined and compared with Twitter data from the same spatial location and place in time. The Twitter data serves as the baseline for the overall protest activity. By comparing text detection methods on protest signs and then further visualizing that data to see the overall

relationship between social media and protest activity, understanding of these events can be enhanced and intelligence can be gathered over that space and the event.

1.2 Thesis Structure

Chapter 2 provides a literature review analysis of some previous research related to protests, the use of signs, and the importance of spatial considerations when studying contentious politics, the Occupy movement, the use of social media in protests, and text detection algorithms. Chapter 3 gives an overview of the research objectives. Chapter 4 elaborates on the data and methods used for analysis. Chapter 5 discusses the results of the analysis, and Chapter 6 concludes the research and provides recommendations for future work.

CHAPTER TWO

2.1 Protests, the Use of Signs, and Location

Since the advent of modern printing and the technological improvements in printing for mass production in the late 19th century (Eskilson, 2012), posters have been a mainstay in human culture. Initially used primarily for advertising, poster usage is ubiquitous throughout almost all types of groups and events. Dominated by advertising, political campaigns, and other propaganda in its inception, artists created visual pictures to inform the masses and stir up attention on various issues (New York History Museum, 2015). Protestors utilize signs as a visual and textual way to convey their grievances. The low cost, ease of production, visibility, and ability to communicate a message quickly make posters the “medium of choice” for most activists and protestors (Wildermuth et al., 2014; Irwin, 2007). “From picket lines, to company boycotts, to marches on Washington; the poster plays an essential role in articulating the voice of the people” (Wildermuth et al., 2014, p. 16). The combination of words and images on a poster can show a powerful message (Tan & Tollenaar, 2004). However, poster usage in protests goes undocumented (Bailey & McAte, 2003). Some social scientists who study protests and other social

movements often do not analyze or recognize the visual methods and materials used, save for documenting images in books or articles (Philipps, 2012).

It is important to note that throughout this research, the terms ‘signs’ and ‘posters’ are used interchangeably. However, an argument can be made that ‘signs’ are homemade/handwritten do-it-yourself (DIY), while ‘posters’ are made by graphic designers and professionally printed and produced. Whichever way these signs or posters are produced (amateur vs. professional), it doesn’t change the fact that they are a powerful, immediate, and succinct form of public speech (Poynor, 2012).

Protest studies have seen an increase in the literature in examining protest rhetoric (what the protest is trying to achieve). Examples include the Occupy movement (Deluca et al., 2012); the Arab Spring protests (Al-Ali, 2012) and the online campaign against Susan G. Komen (Watt, 2012). They examine the framing of the protest (Deluca, 2012), the gendered implications of a political uprising and how women play a key role (Al-Ali 2012), and the post-feminist rhetoric the Susan G. Komen Foundation used when it distanced itself from Planned Parenthood in 2012 (Watt, 2012). However, none of these examples focuses specifically on the protest posters and the prominent display of the visuals during a protest (Wildermuth et al., 2014).

One of the only examples in the literature of examining the use of signs in protests is by Wildermuth et al. (2014), which focuses on the role the poster played in the social protests against the Wisconsin Act 10 from 2011. Their goal was to understand the strategies used in the posters to attempt to persuade the protestor’s audiences, rather than to evaluate the success or failure of the protest poster. Their argument states that the

“tones, themes and messages” are an important element to the overall rhetoric of the protest (Wildermuth, et al., 2014). They employed Irwin’s (2007) strategy of examining protest image rhetoric as a framework for evaluation of the images. Irwin grouped visual protest material into three categories: a primary argument (or enthymeme); use of emotional appeals; and finally, the use of iconic images. Wildermuth also focused on the quality of the protest signs (for example, whether they were handmade or professionally printed) and found that the vast majority of the signs used in Wisconsin Act 10 protests were handmade. The researchers concluded that visual protest materials not only illustrate the campaigns goals, but also that the analysis of the materials was the key in “unlocking the deeper persuasive strategies and cultural significance of the Wisconsin protest” (Wildermuth et al., 2014).

Some work in social movement research aims to analyze how visual protest materials can aid in improving the protest research and the outcomes. Philipps (2012) argues that by using visual protest material and doing visual interpretations, such as summarizing content analysis, iconological interpretation, and cluster analysis (the type of cluster analysis he does is on the layout, message content, design, etc.), it is possible to accurately describe the protest, understand the producer of the visual protest material, and understand the makeup of the types of people participating in the particular protest. An analysis on the 2004 German welfare system in Leipzig protests had two clear and distinct groups. The first group included the experienced demonstrators who were organized and had materials that were clear and appropriate for the demonstrations. The author called these types of protestors “professionals.” They could participate not because

they were affected personally, but because they believed in the overall cause of the protest. The second group was the “spontaneous or unorganized” type of protestor. Those people were joining the protest because of how it affected their own lives, and materials were often hastily made in a short amount of time with whatever was available to them. This type of analysis and the conclusions drawn show the benefits of analyzing visual materials and how they play a role in the overall protest movement (Phillips, 2012).

Another example of analyzing protest posters to examine protest rhetoric was performed by Burrridge (2008). Figure 1 is an example of the signs Burrridge was analyzing. This study examined seven posters used by the British Countryside Alliance during their “Liberty and Livelihood” march in September of 2002. The group and the posters were challenging the ban of hunting with dogs (which ultimately was unsuccessful). However, the authors weren’t interested in the outcome of the protest. The author was interested in how the posters co-articulated visual and textual elements for a rhetorical purpose. They state that visual materials matter, they contain arguments, and that visual and textual materials should be taken seriously, given the way in which they are used creatively to achieve a particular rhetorical effect. Future work for social scientists should engage in “analysis of materials in which visual *and* textual elements are co-articulated for particular purposes, and attempt to map the range of relationships that can exist between those elements” (Burrridge, 2008).



Figure 1: Pro Fox Hunting Poster (Burrige, 2008)

Another facet of protest posters is their abundance. The more protest posters in a location, the more the presence of the protest is felt (Dumitrescu, 2010; Sewell, 2001). Many posters with similar messages in the same location can amplify the message because of the multiplication effect. Higher numbers of posters and protestors in the same location show that a campaign has credibility (Wildermuth et al., 2014). The massive amount of people in a protest or campaign can have a positive effect on the campaign, by

giving publicity to the group directly by physically occupying public spaces and indirectly by accounts from the media, newspapers, or social media. Additionally, a large group of people mass demonstrating for a movement increases that group's solidarity (Sewell, 2001).

The location of protest posters and their relationship to the overall protest activity is one that is scarce in the literature, though not entirely ignored. Rather, space is only discussed as a case study or background, relative to the specific activity, and not as a prime focus (Sewell, 2001) or downplayed spatial context and concepts in social movement research (Martin & Miller, 2003). In recent years, some research discussing location and protest activity referring to the actual physical location of where the protest is taking place has emerged (Endres & Senda-Cook, 2011). Geographers have been studying the implications of place in social movements for the last 15 years (Martin & Miller, 2003; Sewell, 2001). McCarthy & McPhail (2006) look at the types of places where protests occur, such as public, private, and semi-private. However, rhetoricians are just starting to understand the implications that place has and the implications of place on a protest. Endres & Senda-Cook (2011) argue that "place can serve as a unique heuristic for rhetorical studies of social movements" (p. 258).

The concept of place-based arguments invokes images of meaning or memories to support that argument. Spaces and places have meaning and symbolic values that are important to consider in protest movements (Sewell, 2001). An example would be an environmental campaign to "Save the Grand Canyon" or "Save the Glaciers." The activists call upon the images of those places, even though the actual protest is not

happening at that location. Place-based rhetoric is the very place in which a protest is happening is part of the overall rhetorical message (Endres & Senda-Cook, 2011). The National Mall in Washington, D.C. is a good example of an area that has a pre-existing meaning in place-based rhetoric. The proximity to the federal government and the fact that so many other protests have occurred on the Mall have embedded in the meaning of “the Mall” that this is a place where protestors can come and have their voices heard. The Wisconsin Act 10 protests occurred at the Wisconsin State Capital, because the state legislators were the ones who could enact the change. The protest would not have the same effect if it were held in a less meaningful place. In a study of all protest activity mentioned by the New York Times from 1968 to 1973, in 83% of protests that were directed or targeted toward a school, the protest physically occurred at a school (McCarthy & McPhail, 2006). In the Civil Rights movement in the 1960s, there were many sit-ins, demonstrations, freedom rides and other actions occurring in public locations in order to protest segregation. These challenges against authority had more context and stakes because of the location in which they occurred (Sewell, 2001).

Endres & Senda-Cook (2011) and Sewell (2001) discuss three types of place-based rhetoric as a framework to use when studying the relationship of place and social movements. Protestors: 1) build on an existing building of a place; 2) temporarily reconstruct the meaning of a place; and 3) hold repeated demonstrations over time in a place and can change the meaning of that location. Urban areas or university locations with parks, plazas, and squares are ready-made for political protests because of the audience and access to public spaces and locations, as opposed to suburbs, where the

majority of gathering locations are private properties, such as malls and shopping centers, without the type of audience and accessibility to free speech and assembly (Sewell, 2001). Sewell's (2001) concepts of spatial agency, co-presence, time-distance, and the spatiality of power all play a role in studying and understanding the spatial dimension and how it relates to political protests as an actor in the protest itself and not a byproduct. Attention to spatial relations is needed when studying contentious politics and the forces driving those contentions (Martin & Miller, 2003). Indeed, Leitner et al. (2008) argue that the different aspects of space (place, scale, networks, positionality, and mobility) should be considered equally, instead of one over the other, when examining contentious politics, as they are intertwined and complex.

2.2 The Occupy Wall Street Movement

In July of 2011, inspired by the protest and Arab Spring in Egypt and the *Indignados* in Spain, the online magazine Adbusters posted on their blog a call for 20,000 people to flood lower Manhattan and not leave until their demands were met, in essence, 'occupying' the space. Grievances were published on a website created by the organizers, the New York City General Assembly, which outlined they wanted to "gather together in solidarity to express a feeling of mass injustice" (NYC General Assembly, 2011). The way in which "corporations, which place profit over people, self-interest over justice, and oppression over equality, run our governments" was unacceptable (New York City General Assembly, 2011). Adbusters emphasized the physical occupation of Wall Street but they also encouraged the discussion online using the hashtag #occupywallstreet

(Adbusters, 2011). The initial call to action was for September 17, 2011 in Liberty Square (also known as Zuccotti Park), a half-acre site in lower Manhattan's financial district (Gillham et al., 2013). The movement on the ground lasted for more than two months, with events staged around New York City, NY. The movement and ideals spread, and by October 15, 2011 (a month before the NYC protestors were forced out of Zuccotti Park), similar demonstrations had occurred in 951 cities in 82 countries (Bastos et al., 2015).

Castañeda (2012) argues that the *Indignados* (loosely translated to “the Outraged”) movement in Spain in the spring and summer of 2011 was a direct precedent to the Occupy Wall Street movement in NYC in the fall of 2011. Spanish lawmakers had put in place cuts to education, welfare, and social programs, and many felt it was unjust. Activists started camping in the *Toma la Plaza* (a popular area in Barcelona) to “occupy and liberate the square” (Castañeda, 2012). The activities wanted the average citizen's voice to be heard and not dismissed in favor of financial interests. Like the Spanish protest, the OWS movement wanted to call out income inequality and the 1% of Americans who have the majority of the wealth in America. Protestors used the slogan “We are the 99%” as a popular message that refers to the income inequality in America, conveying that the top 1% of Americans have the majority of the wealth and political influence. Additional items of protest action included the corporate influence in American politics and the federal response to the 2008 financial crisis (Gillham et al., 2012).

OWS is a unique protest because it was one of the first times social media was used in such a large way in an American protest. The uniqueness and effectiveness of using social media to spread awareness, logistics, and coverage of the event (outside traditional media) makes for an interesting case study of social media and protest activity (Croitoru et al., 2015).

2.3 The Use of Social Media in Protest Activities

The popularity and available access of social media and smartphones provides a quick, powerful, and easy tool to be used in protests. The barriers to entry are very low to joining social media, allowing virtually anyone to create an account and start sharing, viewing, finding information, and connecting with others that have like-minded views (Steinert-Threlkeld et al., 2015). Social networks facilitate communication with people who otherwise would not have connected (Granovetter, 1973). There is no longer a question on whether or not social media plays a role in protest activities and contentious politics. A myriad of studies show that social media use is related to protest activity, and that users who engage in those events are users of social media – in developed and developing countries (Bekkers et al., 2011; Earl & Kimport, 2011; Pearce & Kendzior, 2012; Valenzuela et al., 2012; Yun & Chang, 2011). Bastos et al. (2015) used the Granger causality test to show Twitter and Facebook activity (online) predicts offline protest activity for the OWS movement, as well as a feedback loop from onsite and online activity. The question then shifts from whether or not social media and protest activities are linked at all to *how* social media are used in protest activities (Kharroub & Bas, 2015; Valenzuela, 2013).

Social media played a vital role in “triggering, organizing, facilitating, accelerating, documenting and broadcasting the protests in Egypt in early 2011” (Kharroub & Bas, 2015). Before the revolution started, users discussed and shared conversations about the political conditions (Eltantawy & Wiest, 2011; Khamis & Vaughn, 2011), bringing users together who were otherwise distanced (Lim, 2012). By documenting everything on social media, such as police brutality, momentum for the movement and offline protest grew (Cottle, 2011; Khondker, 2011). Once the revolution began, social media were used to coordinate and announce events, and support was gained throughout Egypt and internationally (Eltantawy & Wiest, 2011; Khamis & Vaughn, 2011; Lim, 2012; Tufekci and Wilson, 2012; Zhuo et al., 2011).

Kharroub and Bas (2015) contribute to the literature on the contents, uses, and effects of social media in political activism by evaluating visual content in the 2011 Egyptian revolution. They found that the majority of images on Twitter relating to the revolution contained a more efficacy-eliciting content (crowds, protest activities, and national and religious symbols) than emotionally arousing images (such as violence). Conclusions were drawn that this was an intentional tool to motivate and gain more followers and encourage political participation through social media.

In the case of the OWS campaign, the use of social media aided in the quick spread and geographically dispersed network that supported the cause, throughout the United States and internationally (Penney & Dadas, 2013). The mainstream media all but ignored the Occupy movement, with the first report of the event three weeks after the initial call to action in Zuccotti Park. Deluca et al. (2012) evaluated traditional

(old/mainstream) media and social (new) media and how it framed the Occupy Wall Street coverage. In traditional (television, newspapers) media, there was a coverage blackout for the first eight days. After the first eight days, only small articles were in major publications. In the first month of the OWS protests, only 104 stories appeared in the top five newspapers in the United States (The New York Times, The Washington Post, The Los Angeles Times, the Wall Street Journal, and USA Today). Television media largely ignored the Occupy movement until after a few weeks of coverage. The media that did cover it framed it in a negative light. The evolution of how mainstream media covered OWS was “stillborn, first neglected, and then frivolously framed” (Deluca et al., 2012, p. 500).

Two early New York Times articles painted the activists as “hippies” and “flakes” and the OWS movement as frivolous and aimless. This early, negative framing by mainstream media didn’t do much to help discuss these issues that OWS was trying to protest. Issues such as income inequality, greed on Wall Street, and corruption from politicians and corporations were not being discussed and brought to light.

Social media, on the other hand, provided “a different space for different voices to create a diversity of framings” (Deluca et al., 2012, p. 491). Over 10 million results in a Google Blog search mentioned OWS in the first 30 days of the protest (17 September - 17 October, 2011). Blogs from the political left and political right were analyzed and categorized, and while they had vastly different views on OWS, it at least provided a medium that was against the mainstream traditional news media. On Twitter on the first day alone, there were over 4,000 mentions of OWS. After almost a month into the event,

over 40,000 mentions of OWS were counted on Twitter in one day (Deluca et al., 2012). The numbers do not alone reveal the entire story, but they show how activism has changed, with OWS being one of the first large scale protests in the United States with the benefit of the smartphone and the internet at everyone's fingertips. Smartphones and the internet are no longer a tool within an environment, but the environment itself. The OWS movement was one of the first large movements in the United States where the majority of users had smartphones. On the screens of social media, OWS was vibrant, and each side debated and discussed it, unlike the negative framing from traditional media. Social media (blogs, and Twitter/Facebook, etc.) created a new context for activism that does not exist in the world of mass media. Through social media, "the grounds of possibility for activism have been multiplied and transformed" (Deluca et al., 2012, p. 500).

The role played by social media in political and social protests has been increasing across the world (Earl & Kimport, 2011; McCaughey & Ayers, 2003). The Arab Spring (which started in Tunisia and then spread to other countries, with Egypt having a large protest) and the Vinegar protests in Brazil all used social media as a key component of the protest. One activist in Tunisia said "social media has created bridges, has created channels between individuals, between activists, between even ordinary men, to speak out, to know that there are other men who think like me. We can work together, we can make something together" (Pollock, 2011). Facebook in particular was very popular in the Tunisian revolution; it acted as the streets, where people could post videos,

messages, and organize meet-ups, as activists once did on the streets with pamphlets and blogs. It became a central location and meeting place to get and share information.

Steinert-Threlkeld et al. (2015) carried out a study using 13.8 million geo-located Tweets (not pulled based on specific hashtags and keywords) and protest data from a publically available machine-coded data set, the Global Database of Events, Location and Tone (GDELT) from 16 countries involved in the Arab Spring from November 2010 to December 31, 2011. Hashtags were analyzed and the Gini coefficient is used to measure extent of coordination. Coordination is defined as “converging on a few hashtags and using them intensively” (Steinert-Threlkeld et al., 2015, pp. 3-4). A Gini coefficient of zero (0) indicates complete equality with all hashtags used the same number of times. Conversely, a Gini coefficient of one (1) indicates that everyone used one single hashtag and nothing else was used. The higher the coefficient, the more coordination being done about an event that that hashtag represents. Their findings show evidence that coordinating messages on Twitter is associated with increased protest activity the following day. Egypt had an average Gini coefficient of 0.59, Syria was 0.6. Kuwait was 0.09, and Oman came in at 0.02. This work also supports the idea that decentralized players (people with weak ties) can facilitate a mobilization or protest (Granovetter, 1973), challenging the notion that the people who protest only do so if friends or family protest (Gould, 1991, as cited in Steinert-Threlkeld et al., 2015) or are people with prior protest experience (McAdam, 1986, as cited in Steinert-Threlkeld et al., 2015). While these claims are true for the countries involved in the Arab Spring, the authors do caution

that other countries with free speech and free media might not need to do as much coordination online (Steinert-Threlkeld et al., 2015).

Valenzuela (2013) examines the relationship of using social media and increased protest activity among the adult population and gives three explanations for that relationship: information; expression; and activism. Survey data were collected from the adult urban population in Chile in winter of 2011, a period of time where citizens were demanding changes in education and energy policies. They found that expression and activism did have a positive and statistically significant relationship, but the use of social media as a news source (information) was not statistically significant. This could be because of the nature of Chilean media providing most of the news and should be looked at further in studies of other countries. Valenzuela argues that social media is not creating a new form of protest, but augmenting traditional forms of protest. Offline and online are not separate and parallel worlds of activism, but intertwined and related to each other (Valenzuela, 2013).

A sign or poster in use in a protest can be thought of as an analog tweet – sharing a short message with followers and trying to persuade onlookers of the voice to see the activist's perspective. Where a sign is only viewed by the immediate audience, Tweets are shared globally to a much wider audience. Taking a picture of the protest signs and posting it on Flickr is a 'message within a message,' and protests and signs are being augmented with the spread and use of social media to share the message.

2.3.1 Twitter and OccupyWallStreet

Twitter is a social networking platform that enables users to openly share short messages no longer than 140 characters. Each short message is called a *tweet*, and each tweet can be “re-tweeted” by other users, sharing and forwarding on the message to their network of followers. With over 320 million monthly users (Twitter, 2015), Twitter is one of the largest social media platforms around the world. Virtually every aspect of modern life - news, sports, television, natural disasters and other current events - are all documented real-time on Twitter. Trending topics are identified by Twitter’s algorithms to determine the most popular topics in a user’s area, based on their location and who they follow (Twitter, 2016), though it is important to note that Twitter does not require users to register for an account to view others Tweets. Unlike Facebook, Twitter is an opt-in to privacy; all tweets are public by default. The short messages (140 characters) with links to images, videos, and other blogs and media facilitate conceptual understanding of content and the ability to view many sources very quickly to learn and gain information (Gleason, 2013).

With the initial call to action from Adbusters and by using the hashtag #occupywallstreet, Twitter became the de facto preferred platform associated with the OWS movement. Activities were coordinated, messages were dispersed, and editorial commentary was distributed. Twitter was used to bypass the mainstream media (Penney & Dadas, 2013).

According to research by Penney and Dadas (2013), there are seven overlapping ways in which Twitter was used in relation to the OWS movement. By interviewing 17 people with varying backgrounds, geographic locations, and involvements in the OWS

movement, they came up with seven themes on how Twitter is utilized in protest movements. While they admit their research is not an exhaustive look at how every user of Twitter in the OWS movement feels and behaves, these firsthand accounts provide rich information on how Twitter was used and can inform future studies in examining Twitter and protest usage.

The first theme is the most obvious: facilitating the face-to-face, offline protest in an online forum. Earl and Kimport (2011) referred to this as “e-mobilization,” where the “internet is used to facilitate the sharing of information in the service of an offline protest action” (p. 12). Giving out information in a tweet, such as the location and time of a meeting, is the most basic representation of e-mobilization. This is the same as using an advertisement, a poster, or commercial to give out information. The difference with Twitter is that the tweets can be shared quickly and reach a large number of people, thereby hopefully attracting a greater turnout of people at the offline protest location.

This face to face “e-mobilization” theme can also apply to coordinating the logistics during an event as it happens (as opposed to before the event occurs, such as the previous example). Someone can Tweet “the food tent needs supplies” while the event is happening, and followers will see that tweet and provide help. The real-time, on the ground needs of the event could be communicated in an online forum, so that other event members see that and react in an offline way.

Live reporting, or citizen journalism, is the act of allowing users on the ground to become journalists and “live-tweet” an event as it is occurring. The use of smartphones to share video, photos, and messages allows users who are not participating in person to

view the event and feel as though they are on the ground, or to gain interest from people not already engaged in the discussions. This real-time, in-situ reporting is also observed by Wayant et al. (2012) in their spatiotemporal analysis of Tweets and OWS activity. Sharing real-time information is not exclusive to Twitter; the new social network platform Periscope allows users to film video and simultaneously broadcast it out to followers (rather than filming something, saving it, and then uploading it afterward to a platform such as Flickr, YouTube, or Facebook).

Twitter allows users to retweet a message (or Tweet), which is essentially forwarding the message from one user out to another user's followers. This increases the number of people seeing the message and enables even more users to retweet that message. Many of the Twitter activists interviewed (Penney & Dadas, 2013) claimed that retweeting was one of the central and most important functions to participating in the OWS movement online. Users can also add context to someone's tweet, adding words, a headline, or a link to an article that offers more than the character limit allows.

Twitter allows users to provide editorial commentary on a topic and engage in online deliberation with others who might have differing viewpoints (though some users in the interviews claimed that Twitter's architecture doesn't make it conducive to have good debate with people of differing viewpoints). Conversely, it also allows users who are in agreement become stronger and create a sense of community. By engaging in informal communication, the bond of the group is stronger, and many said that there was a "sense of community, solidarity, and group identity" (Penney & Dadas, 2013). Building these ties is important to social movements (McAdam & Paulsen, 1993).

Finally, the seventh theme presented by Penney and Dadas (2013) is the use of Twitter to facilitate online actions. This includes emailing and other lobbying campaigns, calling government offices to voice opinions on a piece of legislation, and signing a petition online. The ability to participate in a campaign or movement, without having to be physically present, allows for activists to continue to shape and amplify and continue to grow the rhetoric and movement “across physical boundaries” (p. 89).

While there were many different protest locations throughout OWS (a movement that spanned the country, and locally in NYC, there were multiple sites of action), Twitter was the constant theme throughout. Twitter provided ways for users at an offline protest in one location to stay connected with offline protestors in another location, thereby creating an online presence. Croeser and Highfield’s (2014) work on the Occupy Oakland movement expanded on research from Juris (2012) and Gerbaudo (2012) that examined the relationship between online aspects of a social movement and the physical location of the protest. They determined that the relationship is deeply entwined, with regard to the online and offline spaces of protest, and both rely on each other. Twitter allows for users who are not physically present at the protest to feel as if they are there. Indeed, Bastos et al. (2014) confirm this idea with their work in the Vinegar protests in Brazil. By harvesting geospatial information from Tweets (ambient or exact), they claim that users tweeting about the location aren’t necessarily physically present at the place of protest, but are still involved in the movement. Activists shift between the physical and the online spaces in order to balance the constraints and affordances of each space (Croeser & Highfield, 2014).

While there are many advantages for using Twitter in a protest campaign, some users are wary of the disadvantages, such as censorship from the website itself and being watched by law enforcement and other government entities. The audience of tweets is worldwide, and law enforcement can monitor tweets just like fellow protest participants and head off any illegal activity (Penney & Dadas, 2013).

Performing spatiotemporal analysis on tweets relating to OWS activity, Wayant et al. (2012) were able to get an overview of the activities on the Day of Action (November 17, 2011) by simply using the Twitter data, showing that Twitter is a good proxy of activity without having to have sensors on the ground.

2.3.2 Flickr

Flickr is an online image and video hosting platform that allows users to upload images and videos and categorize and share those images using optional keyword tags (freely chosen words). With over 115 million users (Flickr, 2016) and an estimated 10 billion images uploaded (Flickr, 2015), it is clear that Flickr is one of the leading photo sharing platforms in the world. In addition to storing personal photos, users can make connections and networks to other users and create a community of sharing images in user groups.

In addition to the exchangeable image file format (EXIF) data, which include technical information about the photo (date and time taken, aperture, type of camera, ISO, shutter speed, etc.) from the camera that took the photograph, users can add contextual information, such as tags, to images. User-added tags are a manual process performed by the owner of the image to provide additional information, allowing for

richer semantic searches from the Flickr community and adding context to a photo by adding descriptive keywords. Flickr imposes a limit of 75 user-created tags per image (Flickr). Images can only be tagged by the owner, unless that owner changes privacy settings to allow for their network to crowdsource tags. Marlow et al. (2006) found this feature to be underutilized; most tags were generated from the owner out of the 58 million observed tags they studied. While most tags are generally added for personal classification for the user to find their own images again later, tagging also introduces new social communication methods and opportunities for data mining and acts as a primary navigational tool within Flickr to find people or images with like-minded interests (Marlow et al., 2006; Ames & Naaman, 2007).

An example of tags added to a photo of Nationals Park baseball stadium in Washington, D.C. would be “baseball,” “stadium,” “MLB,” or “Washington Nationals.” Work done by Sigurbjörnsson and Van Zwol (2008) explored the different ways to add recommended tags using co-occurrence and voting/promotion methods. Co-occurrence between two tags is the number of photos (in their collection) where both tags use the same annotation. The raw co-occurrence is then normalized against the overall frequency of the tags using asymmetric measures, which showed a better result than symmetric normalization. After a list of possible tags is generated, aggregation and promotion are computed to come up with a list of recommended tags to add to an image (Sigurbjörnsson & Van Zwol, 2008).

In May of 2015, Flickr introduced automated tagging to its site, which uses pattern recognition, such as convolutional networks, to view the content of the image and

automatically add tags to images that have a high confidence threshold, such as 95% (Flickr, 2014). Users can decide to remove automatically added tags easily, just as they would add or remove their own user-generated tags. Usually these automated tags tend to be very general, such as ‘car’ or ‘dog,’ if the contents of the photo included a car or a dog. If a user recognizes that a tag was incorrectly added (such as your grandmother being tagged as ‘cat’ instead of ‘people’), by removing a tag, the algorithm is trained and learns that that photo does not have a cat in it, but rather a person, improving the future use of the algorithm (Flickr, 2015).

If the camera allows for geotagging, the coordinates are included in the EXIF data that are submitted with the photo. This gives location information to an image and can be plotted on a map. As of 2011, Flickr reported having over 300 million images geotagged (Flickr, 2011). If the exact coordinate information is not available, but users know the relative date and/or time of when and where the photo was taken, users can manually add location information to that photo through the Flickr interface, giving that image geospatial context.

Flickr introduced machine tagging in 2007, giving even more context to tags and the image it is associated with. Using the **namespace:predicate=value** syntax, users can add more layers of information about a tag and an image. An example could be **medium:paint=oil**, or **flora:tree=coniferous**. These tags can be added just as a regular user generated tag; however, they are only able to be queried using the Flickr application programming interface (API; Flickr, 2007).

Hollestein and Purves (2010) used tags from 8 million Flickr images to determine how people describe spaces and city core areas. They were able to describe the use of the term “Downtown” and explore border boundaries of neighborhoods by analyzing Flickr data and metadata. Terms such as “downtown”, “city center,” and “near Central Station” describe vernacular geography which “encapsulates the spatial knowledge that we use to conceptualize and communicate about space on a day-to-day basis” (Hollenstein & Purvis, 2010, p. 22). The increase in digital libraries that include georeferenced data has been beneficial to researchers who wish to explore the boundaries and locations of vernacular regions. Users can upload an image to a library, and based on the coordinates, systems can provide suggestions for the place or tags based on the coordinates (Kessler et al., 2009; Grothe & Schaab, 2009; Ahern et al., 2007). Those studies focused on larger areas or neighborhoods where the names of those areas are already defined and known. The work by Hollestein and Purves (2010) expands that concept by exploring how georeferenced tags (not the same as geotags, which are coordinates associated with an image) can be used to “define and compare the usage of different city core areas.” Their findings conclude that this user generated content of adding tags to images is a viable way to explore areas that are seen as more common in everyday lexicon.

2.4 Text Detection

For the purposes of this research, it is important to differentiate between text detection and text recognition. Text detection involves determining if an image contains text and locating the text regions in an image, as opposed to recognizing the text (Ikica & Peer, 2014; Wang et al., 2013; Jung et al., 2004). Text recognition is the extraction of the

detected regions of text into readable digital format through an OCR program. Text has to be detected before it can be recognized and retrieved (Yin et al., 2014). Wang et al. (2015) described the difference: “text detection is to extract text regions from a given image, and text recognition is to translate pixel-based text into readable code” (p. 1).

Jung et al. (2004) synthesized the overall approach to text detection and recognition. It consists of four stages: text detection; text localization; text extraction and enhancement (this is sometimes referred to as binarization); and recognition, or OCR, as shown in Figure 2. Almost all text detection and recognition algorithms can be binned into this overall approach, though there are varying methods of reaching the same conclusion. This type of method of detection is sometimes referred to as “End-to End,” with the objective being to “simultaneously localize and recognize all of the words in the image or video sequence, modelling complete systems for text understanding” (Karatzas et al., 2015, p. 1).

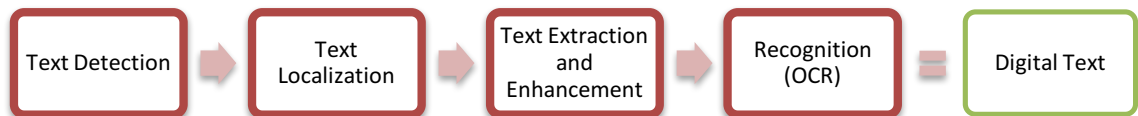


Figure 2: Overall Text Information Extraction Process (derived from Jung et al., 2004)

Text detection methods are generally divided into three major categories: texture-based; connected-components (CC); and a hybrid of both texture and CC-based. Due to the computationally complex nature of texture-based methods, a CC approach, which is more efficient for real time analysis of natural scene images, is used in this research. Real

time is defined as an algorithm processing time that is comparable to a human reading the text (Neumann & Matas, 2012). A more comprehensive look at the different methods and their advantages and disadvantages is discussed below.

2.4.1 Texture Based Methods

Texture methods can also be referred to as sliding window or learning-based methods. Some authors also use the term ‘region-based’ to describe sliding window and learning methods; however, there is a subset of authors that use ‘region-based’ to refer to a connected component analysis (Section 2.4.2). In this research, any reference to region-based methods are referring to a texture method, using sliding windows to detect text. This is different than the regions that are detected in an analysis that are potential text candidates, such as Maximally Stable Extremal Regions (MSER), which is discussed in Section 2.4.3.

A sliding window is used to search for possible texts in the image, and then machine learning is used to identify the texts (Yin et al., 2014). Images are scanned at different scales, looking for text-like features, and are then classified as text or non-text. A pre-trained classifier is used to determine if that region has text or not (Gao et al., 2013). The basic premise behind texture or learning methods is that the processes assume that text has a “special” texture and is uniform compared to its background, from which it can then be distinguished (Shi et al., 2013; Chen et al., 2011). Fast Fourier Transform and wavelet decomposition methods extract the textural regions, and then those regions are fed into a classifier such as support vector machine (SVM) or AdaBoost to further specify regions as text or ‘non-text’ (Shi et al., 2013; Li & Hu, 2012; Chen et al., 2011).

It is generally accepted that the learning and texture-based methods are less efficient and have a high computational complexity, as opposed to connected component based methods for text detection. (Wang et al., 2015; Ikica & Peer, 2014; Chen et al., 2011). The sliding window needs to be run at multiple scales to search for all the texts in the image, despite the fact that this procedure slows down processing, though it does make it robust to noise in images (Wang et al., 2015; Yin et al. 2014). Text regions have distinct textural properties from non-text ones, which allows for more accurate text detection in noisy images (Pan et al., 2011). Training data need to be fed to the learning classifiers of what exactly are text and non-text regions before the algorithms can be processed. If the training data supplied are not sufficient in detecting all types of text, the output is not as strong. It is difficult to get a strong representation of all text and non-text regions (Shi et al., 2013).

Gao et al. (2013) proposed a region-based method that employs an adapted, pre-trained AdaBoost classifier to determine if sliding window regions had text in them. They use transfer learning, which assumes that different scenes have different features. The weak learners for the AdaBoost classifier are reweighted based on each scene and the confidence level of each scene. This helps eliminate false positives, but keep true positives. Results are competitive; however, it still experiences issues when complex images are being read and processing time is not ideal (Gao et al., 2013).

González et al. (2014) present a method for traffic sign and panel detection and recognition using Google street view imagery as an application to intelligent transportation systems. The goal of their research was to automatically create an

inventory of traffic sign panels to support maintenance and assist drivers. While the images used had traffic signs that are relatively uniform in shape and color, the authors point out the images were obtained at various weather conditions, times of day, and landscapes. However, their approach begins with a blue or white color segmentation of all the images, thereby only using images that have a traffic panel in them. Then, the “bag of visual words” (BOVW) methodology is applied to detect the text. Finally, the images are classified using either Naive Bayes or support vector machine (SVM), and the paper compares the differences. By using the BOVW approach, they don’t use edge or geometrical characteristics, which most other text recognition algorithms use. Their model depends on a fixed dictionary that contains common words and a dynamic dictionary that is region specific for where the panel is located. While this experiment was successful, there were numerous limitations that prevented it from being applied to another application without tweaking the algorithm and underlying dictionary (González et al., 2014).

Wang et al. (2012) used an unsupervised learning algorithm to extract features (or regions) from images, and then those learned features were input into a convolutional neural network (CNN). Their end-to-end system combines a lexicon (a list of words that could be detected from an image) to determine the words in an image and achieved competitive results by using the CNN. Although this system is described as ‘simple,’ it still requires processing time and a lexicon of possible words that the algorithm can choose from to determine what the text in the image says (Wang, et al., 2012). There are certain advantages to using a lexicon, such as a list of names from a sports match when watching

video (Ballan, et al., 2010) or having a list of products in a supermarket, which could be used in aiding the visually impaired (Phan et al., 2013, Wang, et al., 2011). For images where no prior knowledge of the subject area is known, having an algorithm that requires a predetermined list of words could be a disadvantage.

2.4.2 Connected Components

The other, more widely used types of text detection algorithms utilize a connected component (CC) approach. CC methods have the same premise: text is found by grouping characters into pixel regions, because it is assumed that pixels belonging to the same character have similar properties (Neumann & Matas, 2012; Li & Lu, 2012). The properties on which pixels are grouped could range from edge detection, stroke width, color, image intensity, and geometry (Wang et al., 2015; Shi et al., 2013; Neumann & Matas, 2012). This ‘bottom-up’ approach (Ikica & Peer, 2014; Gomez & Karatzas, 2013) selects pixels via image segmentation into regions, which are then grouped into connected components. The regions are then determined to be text or ‘non-text,’ based on geometric properties, classifiers, and other heuristics. The ‘non-text’ regions are then discarded from further analysis (Li & Lu, 2012; Chen, et al., 2011; Pan et al., 2011). After removing these false positives, regions are grouped into words or lines (Ikica and Peer, 2014; Shi et al., 2013). Figure 3 shows a diagram of the overall CC analysis, which is derived from Pan et al. (2011). These derived words/lines can then be fed into a commercial OCR program to determine the text from the image (unlike texture based methods, which needs further processing).



Figure 3: Connected Component Analysis Overview (derived from Pan et al., 2011).

There is a smaller number of CC regions detected than regions textured for a texture-based approach, which is why the CC-based method is favorable; it is computationally less complex, which means a shorter processing time (Wang et al., 2015; Pan et al., 2011). CC methods also don't depend on properties of text, such as text orientation and size, font, and scale (Neumann & Matas, 2012; Su & Xu, 2015), though they are sensitive to blur, skew, low resolution, and illumination problems (Wang et al., 2015) and complex backgrounds or clutter (Neumann & Matas, 2012; Li & Lu, 2012). Pan (2011) suggests that CC methods can't accurately segment components without prior knowledge of the text location and scale (Pan et al., 2011). Even though there are many advantages to using a connected component analysis, it is still challenging to develop a perfect system for connected component analysis that eliminates false positives without losing actual text character candidates as well (Shi et al., 2013).

2.4.3 Maximally Stable Extremal Regions (MSERs)

MSERs are used in many text detection processes and methods in the literature (Neumann & Matas, 2010; Chen et al., 2011; González et al., 2012; Li & Lu, 2012; Shi et al., 2013; Yin et al., 2014). First proposed in 2004 by Matas, Chum, Urban and Pajdla, it was used for stereo image and reconstruction of 3D scenes. In recent years virtually every

type of text detection using a connected component analysis utilizes the MSER method to detect possible text candidates. In fact, in the 2015 ICDAR Competition on Robust Reading, almost every method proposed had utilized the MSER segmentation algorithm (Karatzas, et al., 2015). MSERs are a CC-based method that use the detected MSERs as the character candidates.

The MSER is a type of extremal region (ER) whose size remains the same over a range of thresholds on intensity values (Phan et al., 2013; Neumann & Matas, 2012). MSERs are robust to varying degrees of geometric, view point, illumination, and lighting conditions (Neumann & Matas, 2010; Shi et al., 2013; Mikolajczyk et al., 2005), and have a high character detection recall (Wang, et al., 2015), making them a perfect choice for text detection. Text is usually the same color, intensity and in contrast with its background, which are ideal conditions for MSER segmentation (Shi et al., 2013). MSER detection is also very efficient, with a near linear complexity (Nistér & Stewénus, 2008; Matas et al., 2004).

Shi et al. (2013) use MSERs to detected character candidates and a graph cut model to detect text. Region based methods and context information are used in a cost function that remove non-text MSERs, improving the text grouping process. This research merges texture methods and CC-based methods, also known as a hybrid process.

A distinct disadvantage of the MSER segmentation is the sensitivity to blur on images with low contrast (Neumann & Matas, 2012; Li & Lu, 2012; Chen et al., 2011). Images that have smaller text that aren't focused, or if there was too much motion in the image, won't be detected using the out of the box MSER process (Chen et al., 2011). Li

and Lu (2012) overcome this by incorporating a contrast enhanced intensity information on the boundary between text and background. Chen et al. (2011) use an edge enhanced MSER process, by running the image through a Canny edge detection method before MSER segmentation.

The MSER segmentation does an adequate job to detecting text candidates, but it also detects more than just the text in an image. Further processing is needed on the MSERs to narrow down and remove the non-text candidates, which can be difficult (Wang et al., 2015). Properties such as stroke width transform (Epshtein et al., 2010), geometric filtering (Chen et al., 2011; Li & Lu, 2012), adjacency relationships and distance metrics that compute distance between MSER regions (Yin et al., 2014) and AdaBoost classifiers to learn relationship between MSER regions by clustering (Koo & Kim, 2013) are used to refine and remove false positive MSER regions.

2.4.4 Stroke Width Transform

One of the most common properties in a connected component analysis to refine potential text is the Stroke Width Transform (SWT), first proposed by Epshtein et al. (2010). The SWT transforms the image data from containing color values per pixel to containing the most likely stroke width. The SWT leverages the constant stroke width that separates text from other elements of a natural scene. The simplicity of the algorithm allows for fast and robust text recognition and the versatility to detect text in many languages and fonts. It is localized and data dependent, meaning it reduces the need for multi-scale computation. Combining the SWT with geometric filters proves to have promising results in text detection (Li & Lu, 2012; Chen, 2011).

The SWT algorithm doesn't require searching for separating features per pixel (such as color or gradient). In addition, there is a lack of a scanning window over a multi-scale pyramid, separating the SWT from other text recognition algorithms. The experiment they conducted to test the operator was on a standard set of images from the IDCAR dataset which is a common benchmark in testing. The authors also used another 'harder' set of images, which contain much vegetation and repeating patterns (such as windows), to test the algorithm. The authors conclude it is more accurate and performs 15 times faster than other text recognition algorithms.

Ben-Ami et al. (2012) introduced an automatic system of recognizing racing bib numbers to determine which competitors are in a specific photo. Due to the large number of photographs taken at running races, this is a significant challenge. The authors use facial recognition software to locate a person's face and narrow down the region of text to be detected to the torso, as all racing bib numbers are on the front of each competitor. This removes other text from being processed, such as billboard signs, text on clothing, and other background information that is not relevant. Then, they utilize the SWT operator to detect the racing bib numbers, with a few enhancements. They implemented a maximum stroke width size, which limits the character candidates scale, and for images where the bibs were low resolution (1-2 pixel width for the text), resolution is scaled up to allow for edge extraction. By limiting the search area so significantly, the SWT operator performs very well, with very few false positives and a high precision rate (Ben-Ami et al., 2012).

2.4.5 Hybrid Methods

To overcome some of the pitfalls of both texture and CC-based methods, many researchers are now using a hybrid method to achieve state-of-the-art results. By using a combination of texture/region based and CC-based methods, the advantages from both can be exploited. A hybrid method can combine texture detection methods to determine the text segments to be analyzed, and then a component candidate analysis can be performed using a mixture of geometric and stroke width classifiers, along with machine learning techniques to filter out false positives. Conversely, a hybrid approach could also mean doing a CC analysis to segment an image and then using texture or region-based methods to filter out false positives.

Pan et al. (2011) use a texture to CC approach. A texture-based method to detect the text candidates is used to segment regions of an image, and then CCs are extracted as text candidates using local binarization. Non-text regions are eliminated using a Conditional Random Fields (CRF) model with supervised learning, and then characters can be grouped into words or lines.

Fabrizio et al. (2013) use a hybrid method of merging the traditional CC strategy for the first step of the decision stage and a texture/region based validation stage to filter out false positives. Their method collects potential text boxes using a CC segmentation algorithm, followed by classification and grouping of the text boxes. False detections are then filtered out through a validation step, based on a global SVM of the text box content adapted from the Histogram of Oriented Gradients (HOG) approach. The authors use an ITOWNS dataset (similar to Google street view), as well as the International Conference on Document Analysis and Recognition (ICDAR) dataset, which is used as a benchmark

in text recognition. The ICDAR competition is held every two years to challenge and improve researchers in the field of computer vision. The results on the ICDAR database show their method is competitive, and their algorithm has been adopted for use for the ITOWNS project.

CHAPTER THREE

3.1 Objective of Research

Prior literature affirms that social media posts are harnessed and utilized in various ways in modern life, particularly in protest activities. In addition to posts online, signs and posters have been used extensively at protest events, including political rallies, social unrest gatherings, picket lines, company boycotts, and marches. By studying the location of a protest and where signs are placed in relation to the protest, intelligence over that space can be acquired. Researchers of space and contentious politics claim that space has meaning, and protests occur at spaces that are symbolic (Endres & Senda-Cook, 2011; Sewell, 2001). Alternatively, the place may not be significant, but the meaning of that space can be transformed because of the movement held there.

In order to determine the location of signs and posters in relation to the actual protest, the initial challenge lies in detecting the text on protest signs and posters. A further challenge lies in actually recognizing the message conveyed by the text. A plethora of research has been conducted in the computer vision field of text detection and optical character recognition over the last decade and beyond. Improvements in technology and the ICDAR Robust Reading competitions have challenged the status quo

on text detection and recognition. However, most of the algorithms only published their work on the baseline image sets from the reading competitions, usually in addition to one other dataset, such as a street view set, traffic signs, license plates, or other natural scenes. Another popular text detection use case is for the blind and visually impaired, from navigation in the street to navigation inside a grocery store.

At the time of this research, text detection that provides an interesting subject area in content-based image retrieval (such as protest signs and their locations) has not been performed. As referenced by Burrige (2008), Wildermuth et al. (2014), and Bailly & McAtee (2003), it is evident that there is a gap in the research of signs in protest activities, with the majority of the research focused on protest rhetoric. A subset of the literature confirms that studying visual protest material is important in order to glean intelligence of the motives of the protest, who is participating in the protest, and how it relates to opponents' claims regarding the protest. While this research will not analyze the rhetoric of protest posters, it does highlight a way to analyze a large set of posters from events and highlights reasons why studying rhetoric of protest images is important.

Based on the literature reviewed, it is apparent that text recognition continues to be an active research area, most likely due to the complexity and difficulties in deriving an adequate algorithmic solution. This is particularly true for natural scenes containing freehand text. Many of the recent works published seem to suggest new methods or tweaks to methods to improve the state of the art. A trend in the literature is using an ICDAR benchmark image set to test accuracy, and many others also use some sort of urban scene or street view imagery, since those are traditionally the most complex. These

methods, while using real world images, are not using a dataset that is taken from a protest that showcases text, usually handwritten and in large and varied fonts in natural scenes.

This research aims to explore the location of protest signs in relation to the overall protest activity. Text detection algorithms, including MSER detection, geometric filtering, SWT, and OCR, will be applied to a large and varied dataset of protest images captured from the Occupy Wall Street (OWS) movement in the fall (September through December) of 2011. The resulting set of images that had detected text is compared with the location of overall protest activity, which is gleaned from Twitter data from the same time period covering the same issues. The research objective is not to study the accuracy of the text detection algorithm, but rather to use the text detection process to evaluate protest signs. The assumption is that, because these images were harvested using keywords and tags or hashtags that relate to the OWS movement, if text is detected, then that text also relates to the OWS protest activity. The text detection results lead to a consideration of the following questions: Do the Flickr images (with text detected) overlay spatially with the majority of the tweets (a proxy for the overall protest activity)? Will these findings confirm or reject the theory that space in contentious politics is important to study when examining the overall protest? In essence, this research examines the relationships between physical locations of social protests and the online presence of social protests, as determined by the detection of text in Flickr images and the locations of Twitter data.

CHAPTER FOUR

4.1 Analysis Overview

In order to analyze the location of protest signs and its relationship to the overall protest, a text detection methodology was utilized to determine which images actually contained signs. Then, the locations of the text detected images are compared with Twitter point density data, which represent the overall protest activity.

There is no ‘one size fits all’ when it comes to text detection methods, and which type of method to use largely depends on the dataset available and the purpose of the research. For the process used in this research, a connected component analysis is used, even though hybrid results have proved to be more effective in text detection and recognition (Neumann & Matas, 2012). Because of the large quantity of images to process, the efficiency and low processing time of the connected component approach were desired. The connected component approach also requires no a priori knowledge.

The overall method of this research is shown in Figure 4. Each section is called out referencing the section number that describes that topic. Section 4.2 details how data were harvested from Flickr and Twitter. Section 4.3 discusses the text detection methodology used to determine which Flickr images contained protest signs. Section 4.4 highlights the spatiotemporal clustering methodology used to analyze the relationship

between the Flickr images with detected text and physical space, and Section 4.5 discusses some of the processes used to view and analyze the detected Flickr images and that relationship to overall protest activity using tweets.

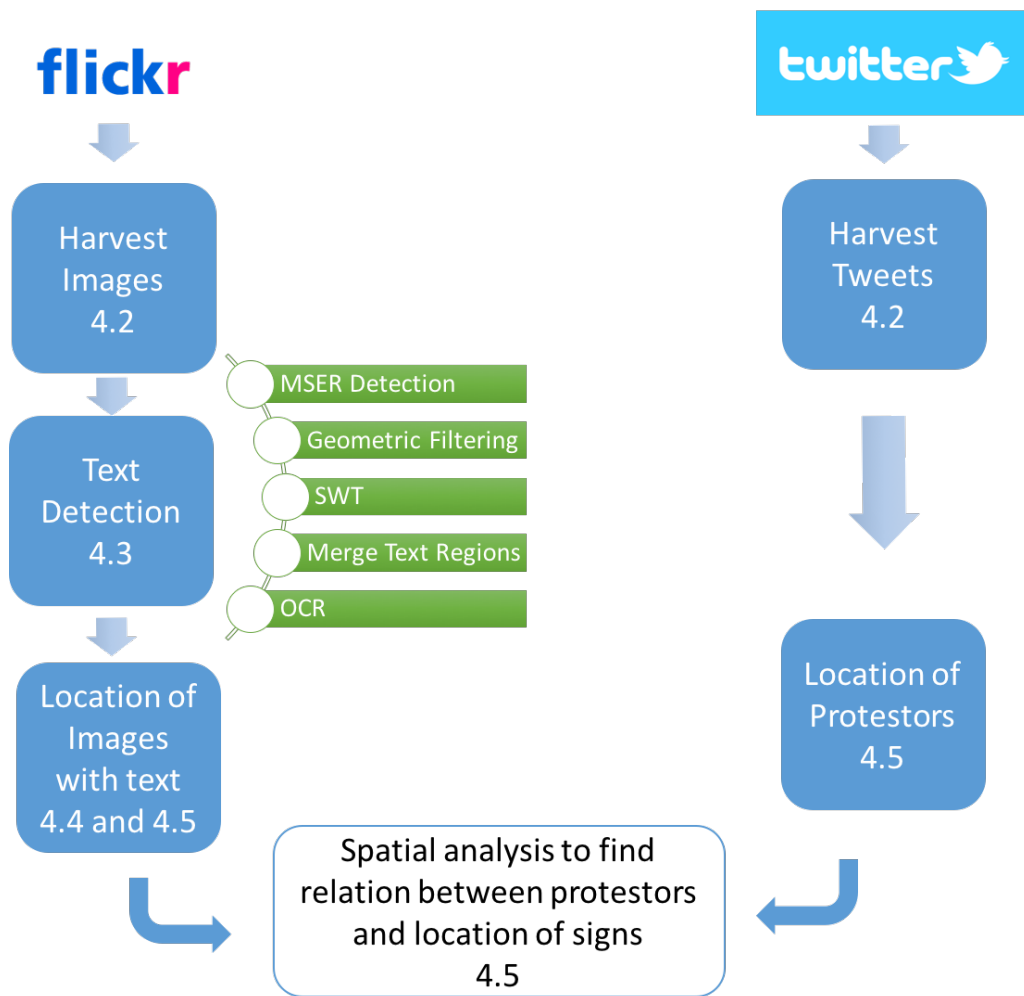


Figure 4: Overall Process Workflow

4.2 Data

Many of the test datasets in the literature are from reading competitions, such as the International Conference on Document Analysis and Recognition (ICDAR), which is held every two years, the most recent conference being in 2015. They provide test datasets for all algorithms to be tested against, so that the precision, recall, and *f-measure* are baselined. Other use cases of text detection/recognition algorithms use image datasets that are from scenes but still have a focused text, as is the case with the majority of images. Focused texts are defined by images captured with the user's intention and intervention (Yao et al., 2015), and text is the main subject. For the first time in the 2015 competition, ICDAR provided an incidental text dataset, which means that any text that appeared in a natural image is captured without the user's prior preference or intention. This introduces more complexities and difficulties in text detection, such as blur, layout, non-uniform illumination, low resolution, and cluttered background (Yao et al., 2015). The differences in focused text and incidental text are mentioned to show that images can be taken with the purpose of detecting text, thereby having a clear focus - text centered in the image - and that is the purpose of taking the photograph. Other times, natural scene images can be taken, and there happens to be text included in the image. The dataset used in this research is a mixture of focused and incidental natural scene images, which increases the complexity of text detection. The intention with which images were taken is not known.

Over 12,000 distinct images (from 644 individual users) were harvested from Flickr's publically available API, taken starting September 17th, 2011 (the start of the

Occupy movement in Zuccotti Park in New York City, NY) to December 10, 2011. Each image has metadata associated with it, such as date taken, location of capture (latitude and longitude coordinates, or geotags), and Flickr tags. No pre-processing was performed on the images before running them through the text detection process. The tags associated with each image were user-generated, as the Flickr process for automatic tag suggestion was not in place at the time these photos were uploaded. Tags such as “OWS,” “OccupyWallStreet,” “OccupyWallSt,” and “occupy” were used to get images that are relevant to the Occupy Wall Street movement and geolocated throughout the United States and Canada. For the purposes of this research, it is assumed that the Flickr image locations and the actual protest signs are very close, while acknowledging that there is some distance between the actual camera taking the photograph of the protest sign. The scale and analysis of this research assumes the location of the image is relatively the same as the location of the protest signs.

Twitter data were harvested using Twitter’s RESTful API. The search terms used included “OWS,” “OccupyWallSt,” “OccupyWallStreet,” and their hashtag (#) equivalents. The dataset was centered around the greater New York, NY region (data spanned the states of New York and New Jersey) and included over 21,000 precisely geolocated tweets. A subset of the region was taken, to include Tweets that were located in Manhattan, Brooklyn, Queens, the Bronx, Long Island, and a portion of Staten Island. The resulting subset included 14,083 tweets (from 5964 individual users). The tweets were posted between October 12, 2011 and April 19, 2012.

The scale of the analysis begins with a country-wide look at the clusters of urban centers of OWS activity. Subsets of the data from the five largest cities with OWS activity are shown to better view the locations of the Flickr data. Finally, the New York city (NYC) region is the focus of further analysis for two reason: 1) the Twitter data are located in the NYC area; and 2) the majority of the Flickr images were geolocated to the NYC area, with over 7,000 images from the entire Flickr dataset coming from the NY area. Indeed, the protests in Manhattan and Zuccotti Park were the epicenters of activity of the Occupy Wall Street protests.

4.3 Text Detection

The process used for text detection is a connected component approach with simple rule based filtering to remove non-text regions (Li & Lu, 2012; Chen et al., 2011). Then, those possible text regions are fed into an optical character recognition algorithm to read the text. The overall process workflow is outlined in Figure 5. The corresponding number is the section in which that topic is discussed more in depth. This process was applied using the Mathworks Matlab software suite 2015b Student Edition with the Computer Vision toolbox.

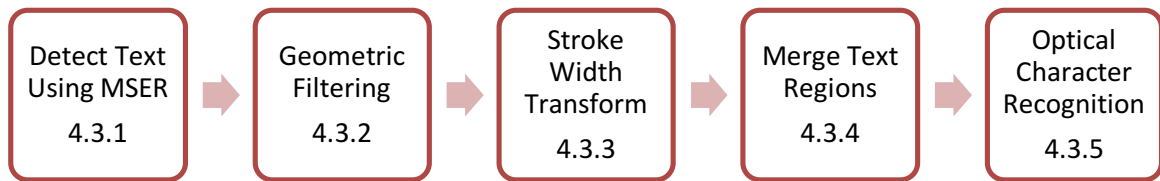


Figure 5: Text Detection and Recognition Workflow

To process over 12,000 images, a batch process was written in Matlab that took the process in Figure 5 and looped it over all images in a folder. Instead of writing out the text it detected from the OCR, it wrote the name of the file to a text file, along with a number. An output of zero (0) meant no text boxes were detected that contained text. An output of one (1) or greater meant text boxes were detected that contained text, and the resulting number represented the number of boxes detected in the image. The default text detection output variable assigns a confidence threshold to each text detection box. If the confidence value threshold was greater than 0.5 for a given word, then that word counted as a detected word.

4.3.1 Detecting Text using MSERs

The original input image (as shown in Figure 6) was converted to a grayscale image. The grayscale image is taken to be the input into the MSER detection algorithm. The minimum area required in order for a region to be selected is 200 pixels, and the maximum area allowed is 8000 pixels. The threshold delta, or the step size between intensity levels, is a value of 4 (out of possible values from 0 to 100). The threshold delta is the value at which the MSER detector steps through an image to detect stable regions.

The higher the value of the threshold, the lower the number of intensity increments it filters through to detect stable regions, and vice versa: the lower the value of the threshold, the higher the number of intensity increments the algorithm filters through to detect stable regions.



Figure 6: Original Input Image

The output of the detect MSER regions process is shown in Figure 7. It can be observed that more than just probable text regions are detected. Further processing is therefore needed to remove non-text regions.



Figure 7: Detect MSER's

4.3.2 Geometric Filtering

Geometric filtering was then performed on each image to remove non-text regions that were detected from the MSER algorithm. This process was completed with simple rule-based filtering; however, some researchers employ a machine learning approach to train a text vs. non text classifier (the hybrid methods mentioned previously in Section

2.4). Because of the size of the dataset considered in this research, the simple rule based filtering was employed, given its fast computation nature.

After calculating the statistics for each region, thresholds were employed to remove regions that didn't meet that threshold for that particular statistic. Regions with an aspect ratio greater than 3, an eccentricity greater than 0.995, a solidity less than 0.3, an extent less than 0.2 and greater than 0.9, and a Euler number less than -4 were all removed as possible text regions. The resulting image after geometric filtering is performed is shown in Figure 8.

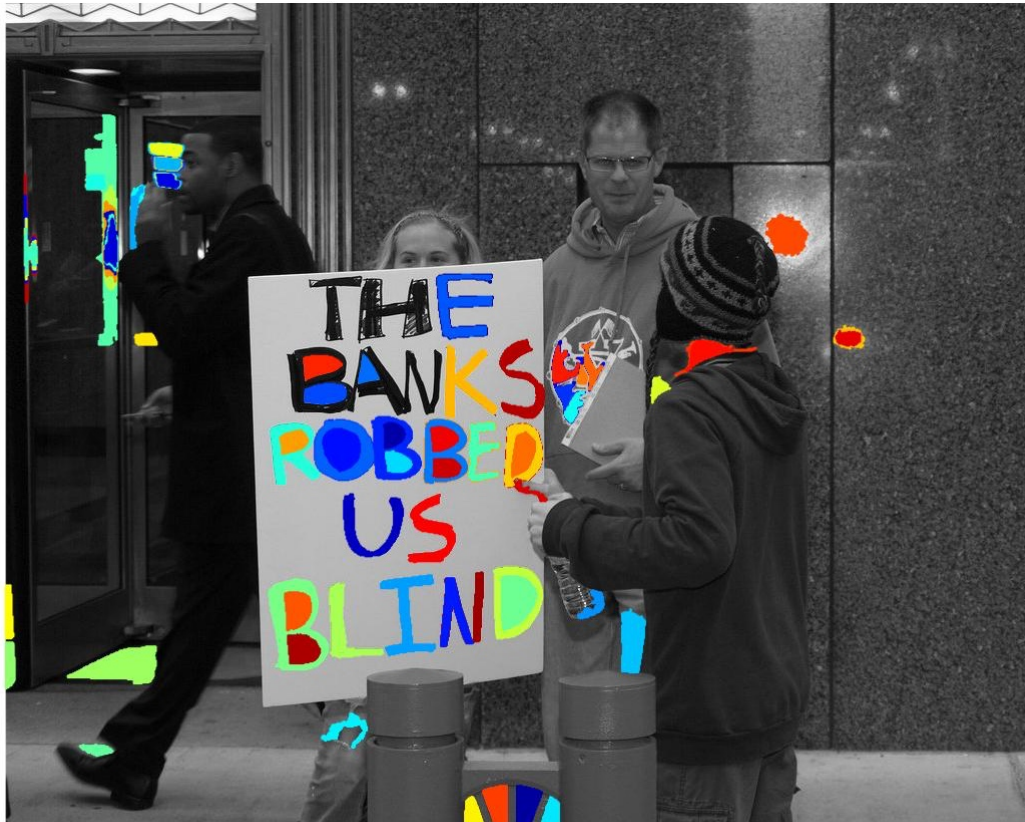


Figure 8: After Removing Non-Text Regions Based on Geometric Properties

4.3.3 Stroke Width Transform

As described in Epshtein et al. (2010), stroke width is the measure of the lines and curves that make up the characters. The length of a straight line from the text edge pixel to another along its gradient direction is the stroke width. The majority of text regions have little variation in stroke width, so it is easy to calculate stroke width and remove larger regions that have a higher variation. Regions with little variation are most likely text, because the lines and strokes making up the text character are usually similar and uniform.

The stroke width is estimated for one of the detected MSER regions by using a distance transform and binary thinning operation (Li & Lu, 2012). A binary region image is created, and then the Euclidean distance is calculated for every pixel on the stroke width regions to the nearest boundary of the corresponding MSER. A skeleton image is created by performing binary thinning. If an object does not have a hole in it (such as a text character), it shrinks the skeleton to the minimum connected strokes. Objects that do have a hole cause the skeleton to shrink to a connected ring halfway between the hole and the outer boundary. Figure 9 shows the region image (left) and the stroke width ‘skeleton’ image (right).

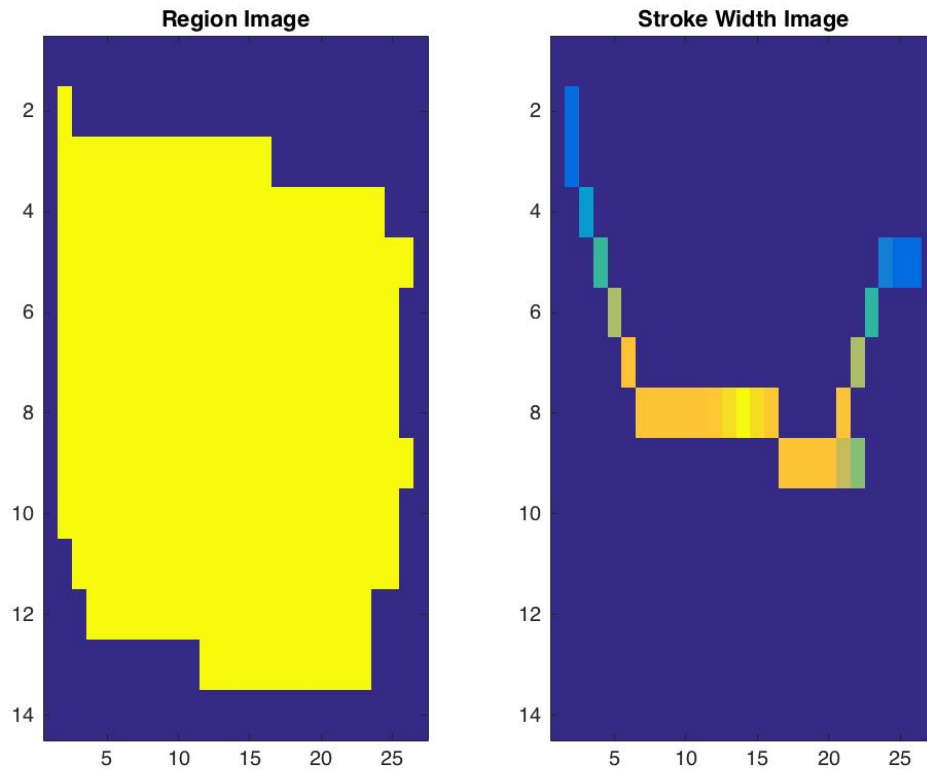


Figure 9: Region Image and Stroke Width Image

The stroke width variation must be calculated for the entire region to be used as a threshold value to remove non-text regions. The standard deviation of all the stroke width values for each region is divided by the mean of the stroke width values for that region. Regions with a stroke width variation metric of 0.4 or higher are determined to be non-text and are filtered out of the image. This value can be adjusted for different types of font, but in this case, the default was taken, as there is no a priori information about the font type used. A for-loop processes all of the regions in an image to determine if that region is text or non-text. The output of the stroke width and distant transform filtering is

show in Figure 10. Under ideal conditions, only text would be remaining at this step; however, each image is different, with different conditions.



Figure 10: After Removing Non-Text Regions Based on Stroke Width Variation

4.3.4 Merge Text Regions

In order to make sense of the text regions detected, the characters (which are detected individually up to this stage) need to be merged for further processing and recognition, such as in an OCR tool. For example, the characters that make up the word “BLIND” need to be properly ordered, because a random ordering, such as “L, N, D, I,

B,” does not have the proper meaning. Characters or text in the same word usually share similar properties, such as intensity, size, and stroke width, and text almost always appears in a straight or slightly curved line, thereby allowing the characters to be grouped together (Li & Lu, 2012; Chen, et al., 2011).

To merge text regions, boxes are drawn around the text characters, and the boxes are then expanded by a small amount (0.02) to find neighboring regions that overlap. Anything that exceeds the image’s boundary is clipped. The expanded bounding boxes are shown in Figure 11.



Figure 11: Expanded Bounding Boxes

Any boxes that are found to be overlapping with other neighboring bounding boxes are merged together around the words and character lines by computing the overlap ratio. The distance between all bounding box pairs is quantified, making it easier to find groups of neighboring text regions (overlap ratio of non-zero). A graph is then used to find all the connected (overlap ratio of non-zero) text regions, outputting indices for each connected text region. A single bounding box is created (merging the neighboring bounding boxes) by using the indices and taking the minimum and maximum of each individual bounding box that make up a connected component. Any

boxes that are just one text region are removed (which further refines false positives, as text is unlikely to be on its own). The resulting image of detected boxes is shown in Figure 12.



Figure 12: Detected Text

4.3.5 Optical Character Recognition

The resulting text regions are then run through the OCR program in Matlab, returning an ocrtext object, which gives the recognized text, the location of the text on the input image, and the confidence threshold of the results. The OCR program used in

Matlab is based on the popular OCR tool Tesseract (Smith, 2007). The output of the OCR text recognition for this example is shown in Figure 13.

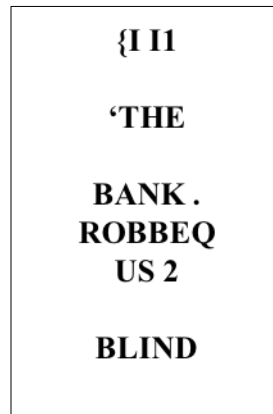


Figure 13: OCR Results for Example Image

4.4 Spatio-temporal analysis

To analyze the Flickr data that had text detected and its relation to physical space, a clustering algorithm was used to detect groups and clusters in physical space.

Traditionally, the Density Based Spatial Clustering of Application with Noise (DBSCAN; Ester et al., 1996) is used to detect clusters by builds on density as a measure for defining and detecting clusters. While DBSCAN has a number of advantages (distinguishing noise in the data, the ability to form arbitrary cluster shapes, and performing clustering with no prior knowledge or assumptions), it does not have a temporal component. To account for time as a variable in the cluster, the DenStream method was used, which is a stream-based clustering algorithm, first proposed by Cao et

al., (2006). The use of DenStream in recent literature, as well as an in depth methodology, is described in Croitoru et al. (2015).

4.5 Spatial Data Visualization and Analysis

The images harvested from Flickr, which had metadata, such as tags, coordinates, and date taken associated with them, were parsed into a comma separated value (CSV) file and read into a spreadsheet. The table was then opened in ArcMap, and a point shapefile was created from the X and Y (longitude and latitude) values of the location of where the image was taken. The text file created from the process in Figure 5 was then joined using the Join by Attribute tool to create a new column in the attribute table. Each image now contained a value of zero (0) if text wasn't detected, or a one (1) or greater if text was detected. The data was projected to the WGS84 Web Mercator (Auxiliary Sphere) coordinate system, since the points span the entire United States (with a few in Canada).

The Twitter data were projected to WGS 84 UTM Zone 18N (since the dataset is located in New York), and then the point density tool was run in ArcGIS to visualize the density of points over the region per square kilometer. The point density tool allows a better visual representation of the 14,000 points (tweets), rather than points stacked on top of each other. In this case, the tool calculated how many points were calculated for each square kilometer. The resulting raster is then symbolized with a color ramp, which allows the human eye to better understand the amount of points in an area. The resulting Flickr images that had text detected were then spatially overlaid with the Twitter point

density results. Results of the visualization of Flickr data and Twitter data can be seen in Chapter Five.

CHAPTER FIVE

5.1 Text Detection Accuracy Assessment

In order to evaluate the text detection methodology, a subset of 200 images was taken from the larger dataset and run through the process. An equal number of images with text and without text were randomly selected, with the author determining if an image had text or not. These images act as the ground truth for determining if the text detection process was successful or not. Text detection was deemed a success if the output of the text recognition (OCR) identified any text, regardless of the accuracy of the text detection. For each image, a number was written to a text file; this was the number of text boxes detected from the OCR. This number was written if the confidence threshold of the OCR in Matlab was above 0.5 AND the text boxes detected were not empty spaces. This could be one text box detected, or several. The actual text that was recognized through OCR was not compared to the text in the image. OCR was only used as a tool to evaluate the text detection. Images with no text detected were given a value of zero; images with text detected were given a value of text boxes detected, starting at one.

A confusion matrix was built to represent the amount of true positives, true negatives, false positives, and false negatives that were a result of the process. An image

is a true positive if it was correctly classified as having text (output a number of 1 or greater). An image is a true negative if it was correctly classified as not having text (output a number of 0). An image is a false positive if it was incorrectly classified as having text, but the image did not actually contain text (output a number of 1 or greater but should have been a 0). An image is a false negative if it was incorrectly classified as not having text but the image did have text (output number of 0 but should have been 1 or greater). The resulting confusion matrix from the 200 ground truth images is located in Table 1. By adding up the number of true positives and true negatives and dividing by the total number of images tested, the overall accuracy of the text detection methodology is 78%.

Table 1: Confusion Matrix on subset of images for text detection and OCR process

True Positive: 74	False Positive: 18
False Negative: 26	True Negative: 82

Examples of images that were classified as true positives are shown in Figure 14. On the left is the result of the text detection image with bounding boxes drawn around where the algorithm determined text could be located. The potential text regions were then read through OCR. To the right of the image is the result of the OCR. While most

of the true positives capture at least one complete word, some, such as Figure 14d, did not. However, because it text was still detected, it was classified as a True Positive.



Figure 14: True Positive Examples

Examples of images that were classified as true negatives are shown in Figure 15. Because these are true negatives, no text was recognized through OCR. Figures that have

yellow boxes in them show the locations of potential text regions. However OCR did not recognize any text, therefore the images were still classified as true negatives.



Figure 15: True Negative Examples

Examples of images that were classified as false positive are shown in Figure 16. The image on the left is the result of the text detection/OCR, and the text it detected is on the right.



Figure 16: False Positive Examples

Examples of images that were classified as false negatives are shown in Figure 17. Because these are false negatives, no text was recognized through OCR, though there is text in the images. Figures that have yellow boxes in them show the locations of potential text regions. However, OCR did not recognize any text, therefore the images were classified as false negatives.

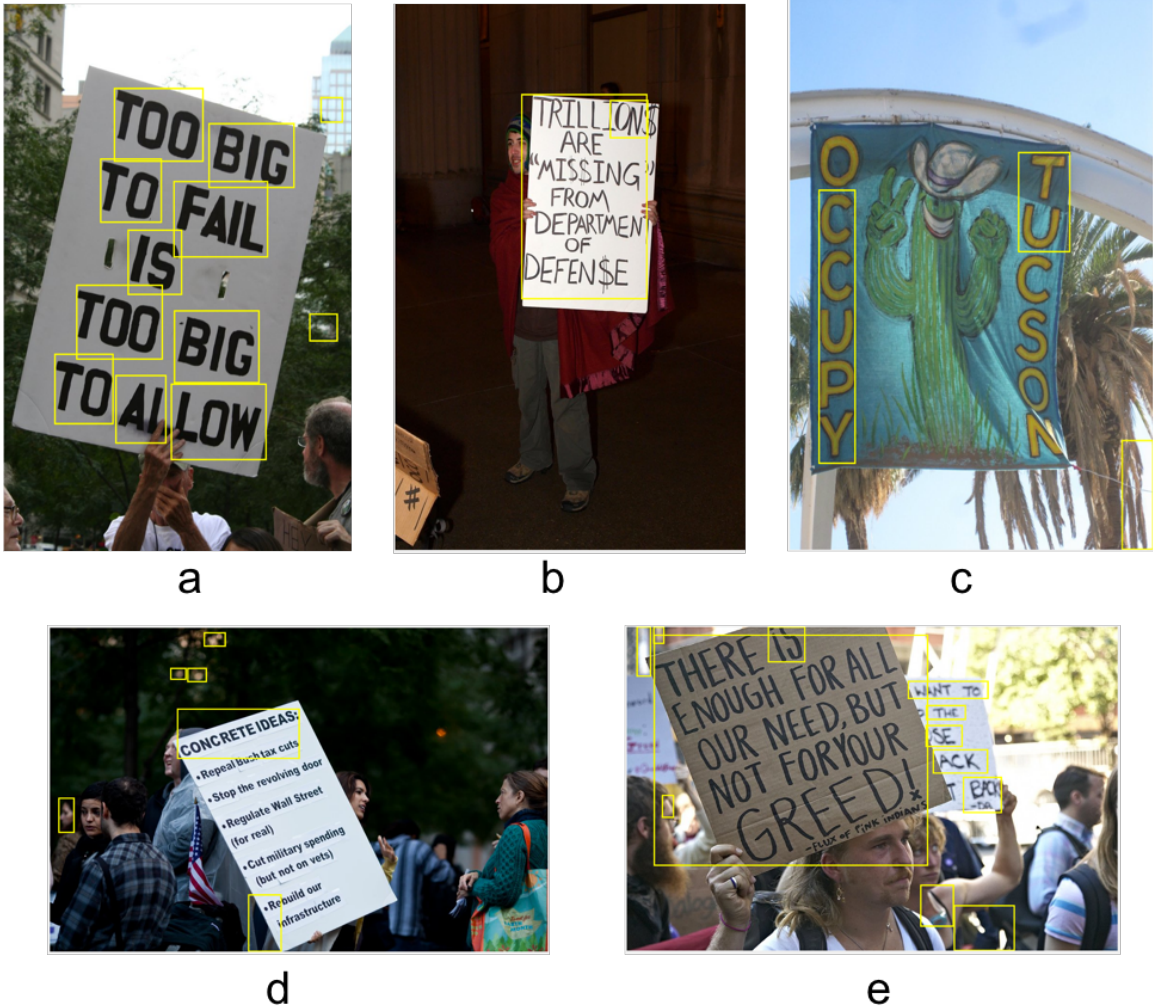


Figure 17: False Negative Examples

As mentioned in Chapter 4, taking a picture of a protest sign in an urban location most likely will have some form of text in the background, such as a street sign, a storefront name, a car driving by, etc. In the case of the OWS images, other signs can also serve as incidental text, as there were many protest signs in use. For the subset of

images used in the accuracy assessment study (200), only a small number of images had incidental text, and an even smaller number had detected and recognized incidental text. Examples of these are shown in Figure 18. For Figure 18a, text was detected (as shown by the yellow bounding boxes), but the OCR did not recognize any text in the image at all (including the focused text). In Figure 18b, the OPEN sign from the hot dog stand was detected, and the OCR detected the text.



a



b

Figure 18: Examples of Incidental Text

5.2 Spatial Data Visualization and Analysis Results

The geolocated images harvested from Flickr were added to a map with the latitude and longitude defined for X and Y using ESRI ArcGIS Desktop software. After the DenStream process was run (Section 4.4 Spatio-temporal analysis), the results produced 32 clusters of Flickr images where text was detected. The results of the clusters are shown in Figure 19.

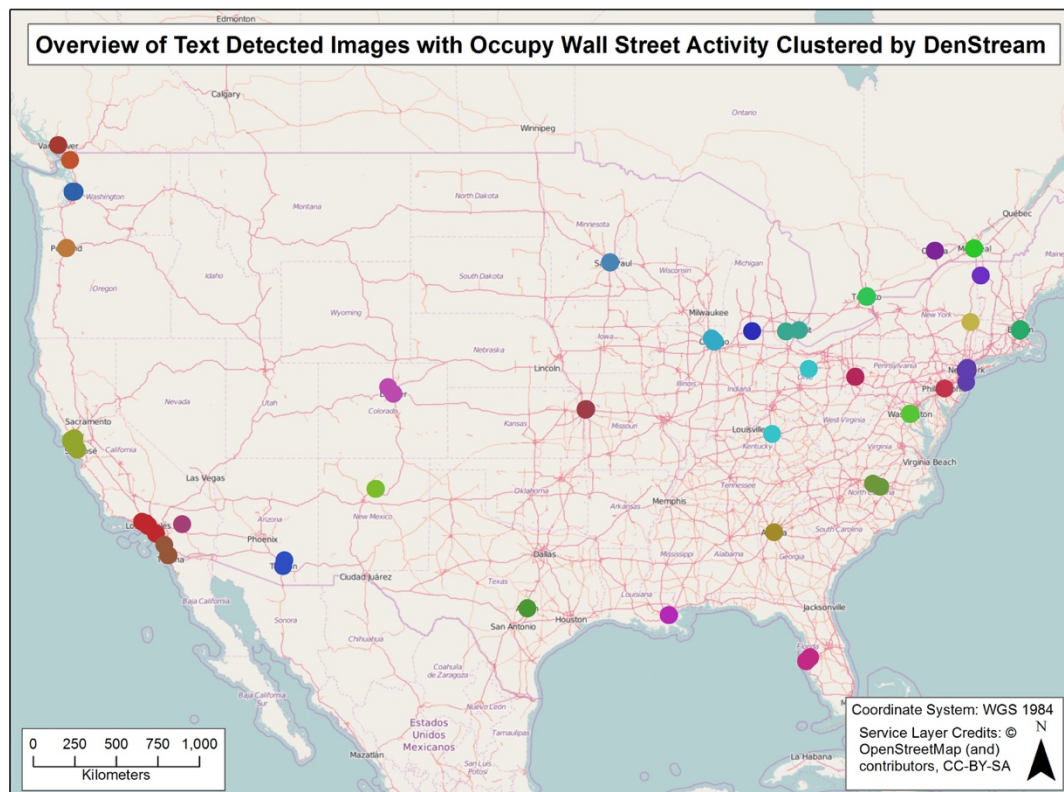


Figure 19: Overview Map of all Images Harvested from Flickr relating to Occupy Wall Street in Fall 2011

The cluster IDs, corresponding locations, and points per cluster are found in Table 2, ordered by points per cluster in descending order.

Table 2: Cluster ID, Location, and Number of Points Per Cluster

ClusterID	City, State	Points Per Cluster
1	New York City, NY	3452
4	Chicago, IL	534
2	San Francisco/Oakland/Bay Area, CA	352
3	Los Angeles, CA	145
9	Washington, DC	139
5	Boston, MA	92
8	Seattle, WA	85
7	Portland, OR	54
12	Detroit and Ann Arbor, MI	50
25	Toronto, Ontario, Canada	49
13	Minneapolis, MN	48
30	Montreal, Quebec, Canada	37
26	Vancouver, British Columbia, Canada	36
6	Tampa/St. Petersburg, FL	34
21	Yucca Valley Airport, CA	33
23	Santa Fe, NM	30
27	Ottawa, Ontario, Canada	27
10	New Orleans, LA	25
29	Raleigh and Chapel Hill, NC	21
31	Albany, NY	21
18	Denver and Boulder, CO	20
15	Bellingham, WA	18
14	Austin, TX	14
16	Atlanta, GA	12
32	Lexington, KY and Mansfield, OH	12
22	Kalamazoo, MI	11
20	Burlington, VT	10
24	Pittsburgh, PA	10
11	Kansas City, MO	9
17	San Diego and Oceanside CA	8
28	Philadelphia, PA	8
19	Tucson, AZ	6

Based on Table 2, the top five cities (with a high concentration of Flickr images relating to OWS) were subset to be studied in depth. Table 3 shows the location of the subset, the number of records included in the subset, and a percentage of how many records were classified as Text Detected vs. Text Not Detected. With the exception of Los Angeles, CA, all the cities had a lower percentage of text detected than text not detected. Maps showcasing the location of the images as points are shown in Figures 20 through 24.

Table 3: Number of Records Per City and Percent Detected vs Not Detected

	Number of Records	% Text Detected	% Text Not Detected
<i>New York City, NY</i>	7659	3539 = 46.21%	4120 = 53.79%
<i>San Francisco/ Oakland, CA</i>	777	350 = 45.05%	427 = 54.95%
<i>Chicago, IL</i>	1318	533 = 40.44%	785 = 59.56%
<i>Los Angeles, CA</i>	258	134= 51.94%	124= 48.06%
<i>Washington, D.C.</i>	335	139 = 41.49%	196 = 58.51%
<i>Entire Dataset</i>	12599	5558= 44.11%	7041=55.89

Figure 20 shows a map of the Flickr images and the results of the text detection process for Washington, D.C. The red dots signify that text was detected, and the blue dots signify text was not detected. The two inset maps show a close up of McPherson Square and Freedom Plaza, locations of heavy protest activity.



Figure 20: Washington, D.C. Inset Map

Figure 21 shows a map of the Flickr images and the results of the text detection process for Los Angeles, CA. The red dots signify text was detected, and the blue dots signify text was not detected. The inset map shows Los Angeles City Hall, a location of heavy protest activity.

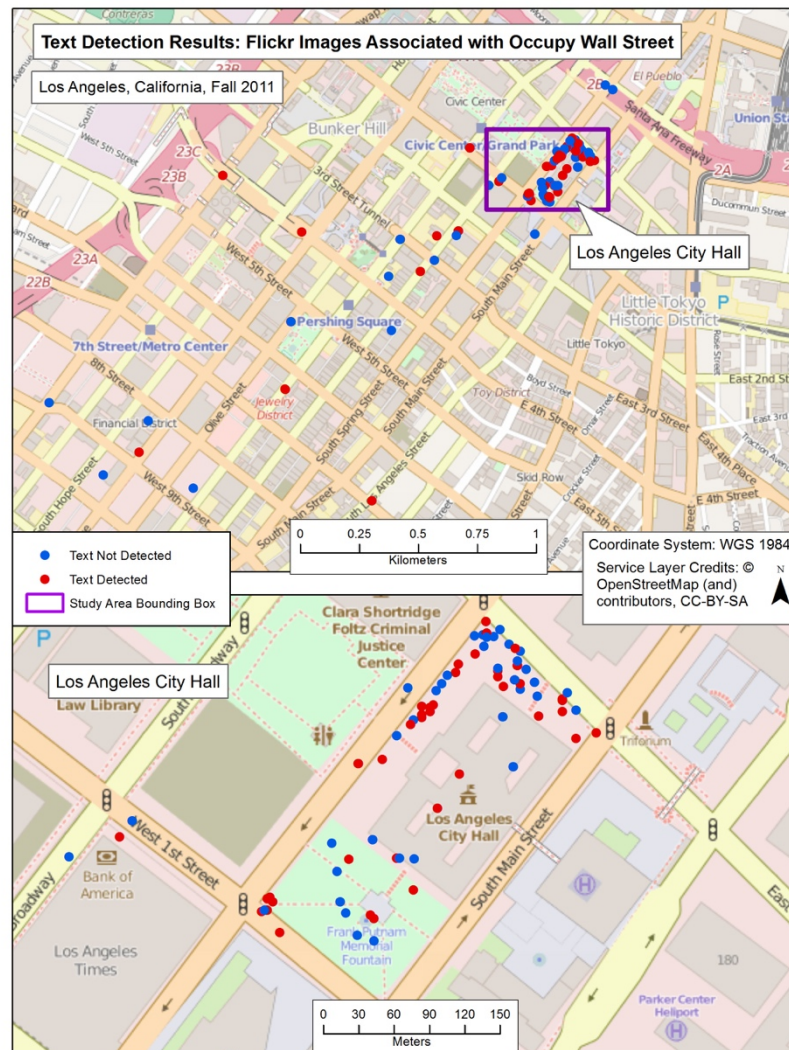


Figure 21: Los Angeles, California Inset Map

Figure 22 shows a map of the Flickr images and the results of the text detection process for San Francisco and Oakland, CA. The red dots signify text was detected, and the blue dots signify text was not detected. The two inset maps show San Francisco, CA and Oakland, CA, locations of heavy protest activity.

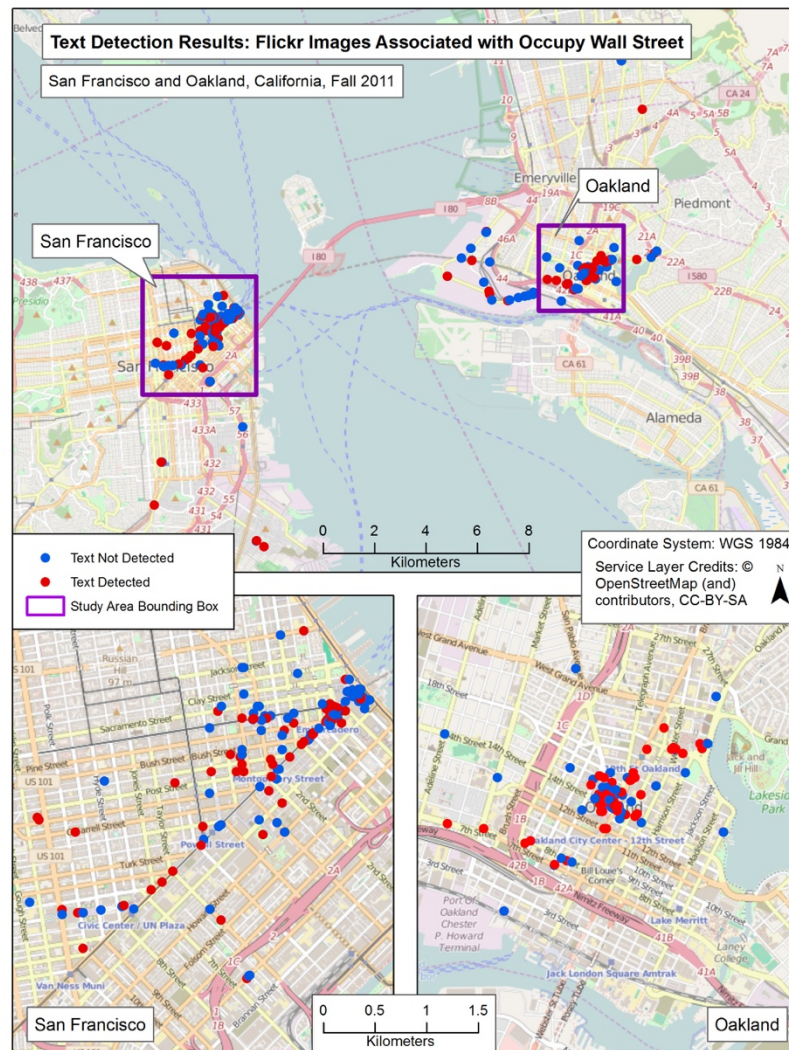


Figure 22: San Francisco and Oakland, California Inset Map

Figure 23 shows a map of the Flickr images and the results of the text detection process for Chicago, IL. The red dots signify text was detected, and the blue dots signify text was not detected. The two inset maps show the Federal Reserve Bank of Chicago and Congress Plaza Garden, locations of heavy protest activity.

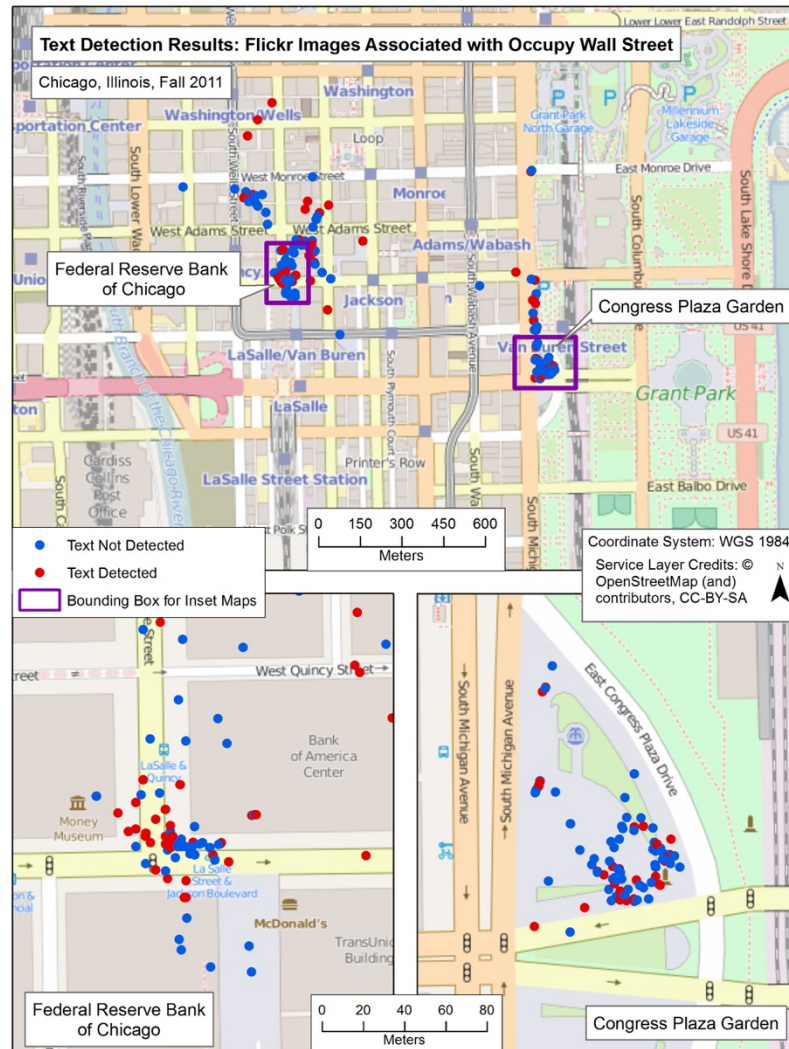


Figure 23: Chicago, Illinois Inset Map

Figure 24 shows a map of the Flickr images and the results of the text detection process for New York City, NY. The red dots signify text was detected, and the blue dots signify text was not detected. The inset map shows Zuccotti Park, one of the main locations of protest activity.

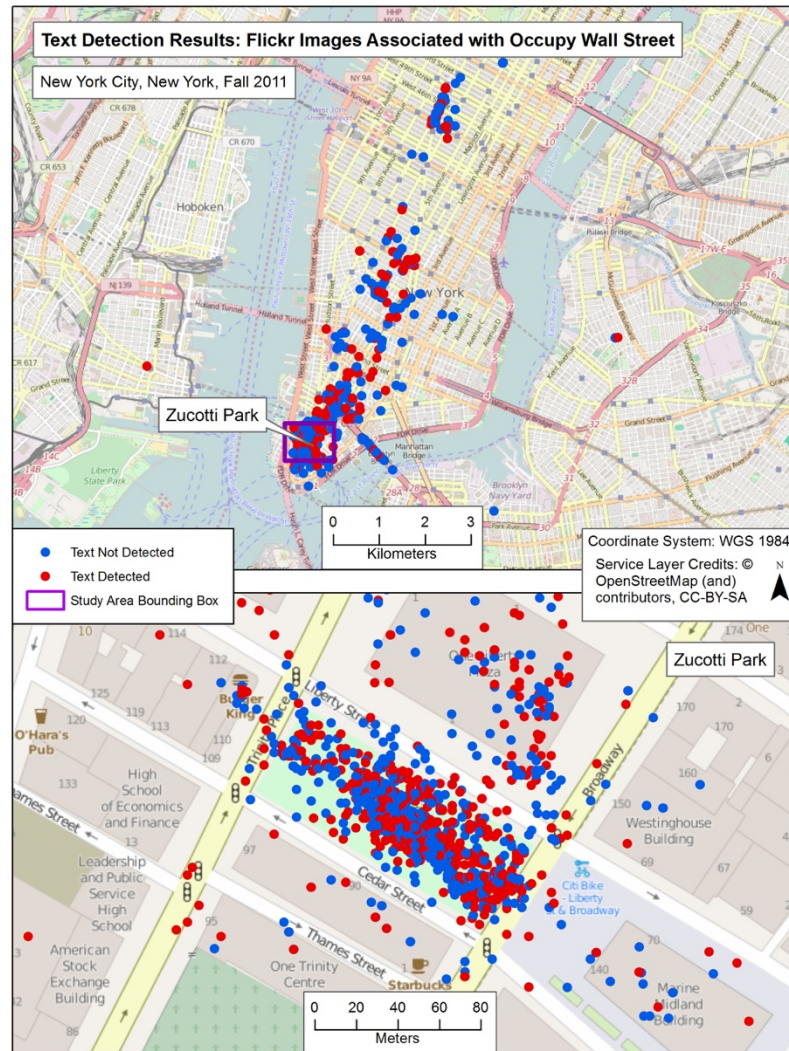


Figure 24: New York City, New York Inset Map

Subsets of clustered Flickr images with text detected are shown in Figure 25.

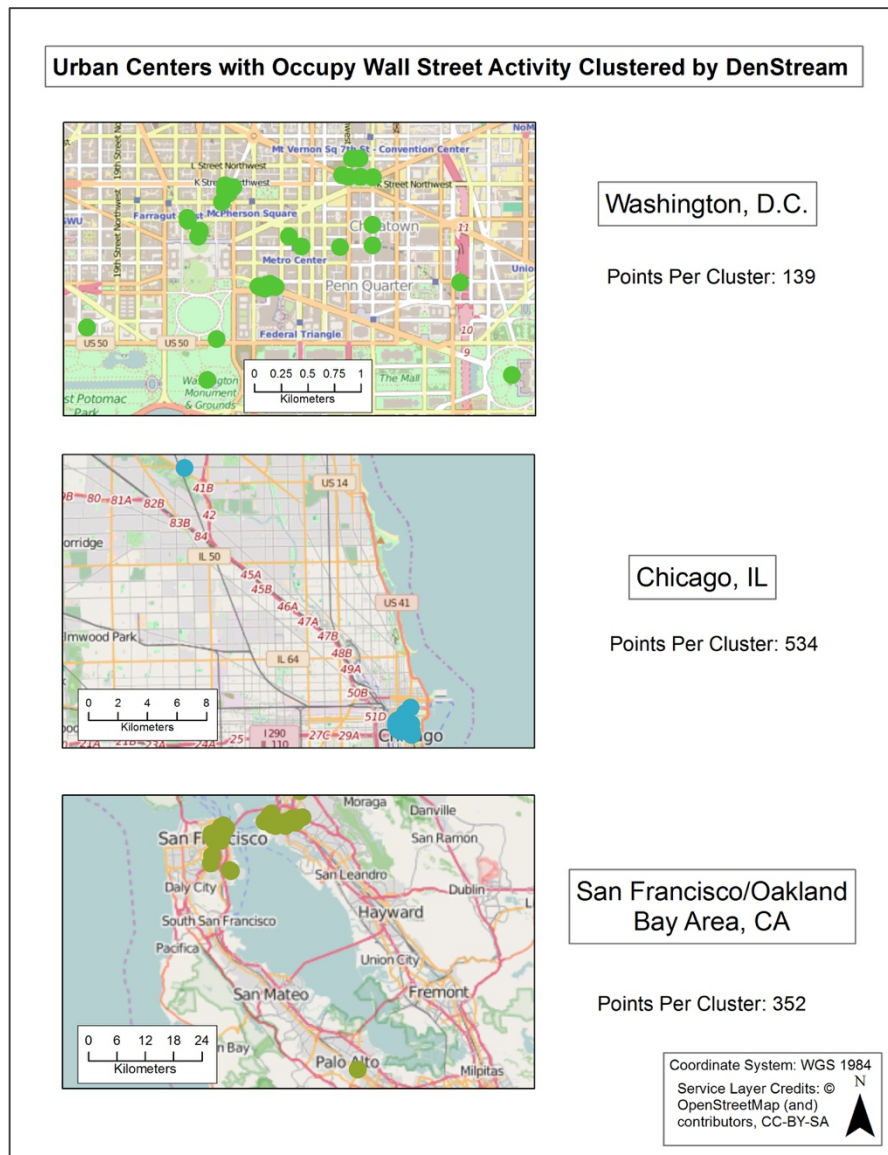


Figure 25: Urban Centers with OWS Activity Clustered by DenStream

The Twitter dataset, consisting of over 20,000 geolocated tweets relating to OWS during October 2011 to April 2012 around the greater New York City area, was subset to be a smaller region consisting of the five main New York City boroughs (Manhattan, Brooklyn, Queens, Staten Island, and the Bronx) containing 14,083 precisely geolocated tweets (5,946 individual users). The Twitter data are used as a proxy for the overall

protest activity (Wayant et al., 2012). Figure 26 and Figure 27 show the Twitter point density overlaid with the results of the text detection process of the Flickr images. Figure 26 shows the Twitter density and the results of the Flickr images that did have text detected, and Figure 27 shows Twitter density and the results of the Flickr images that did not have any text detected.

The point density shows that, in the lighter green areas, there are fewer points per square kilometer than in the yellow and red areas, which have a higher point density per square kilometer. The densest areas of Twitter activity relating to the protest activity occur in midtown Manhattan and Zuccotti Park. The Flickr activity relating to Occupy Wall Street has the majority of images in or around Zuccotti Park.

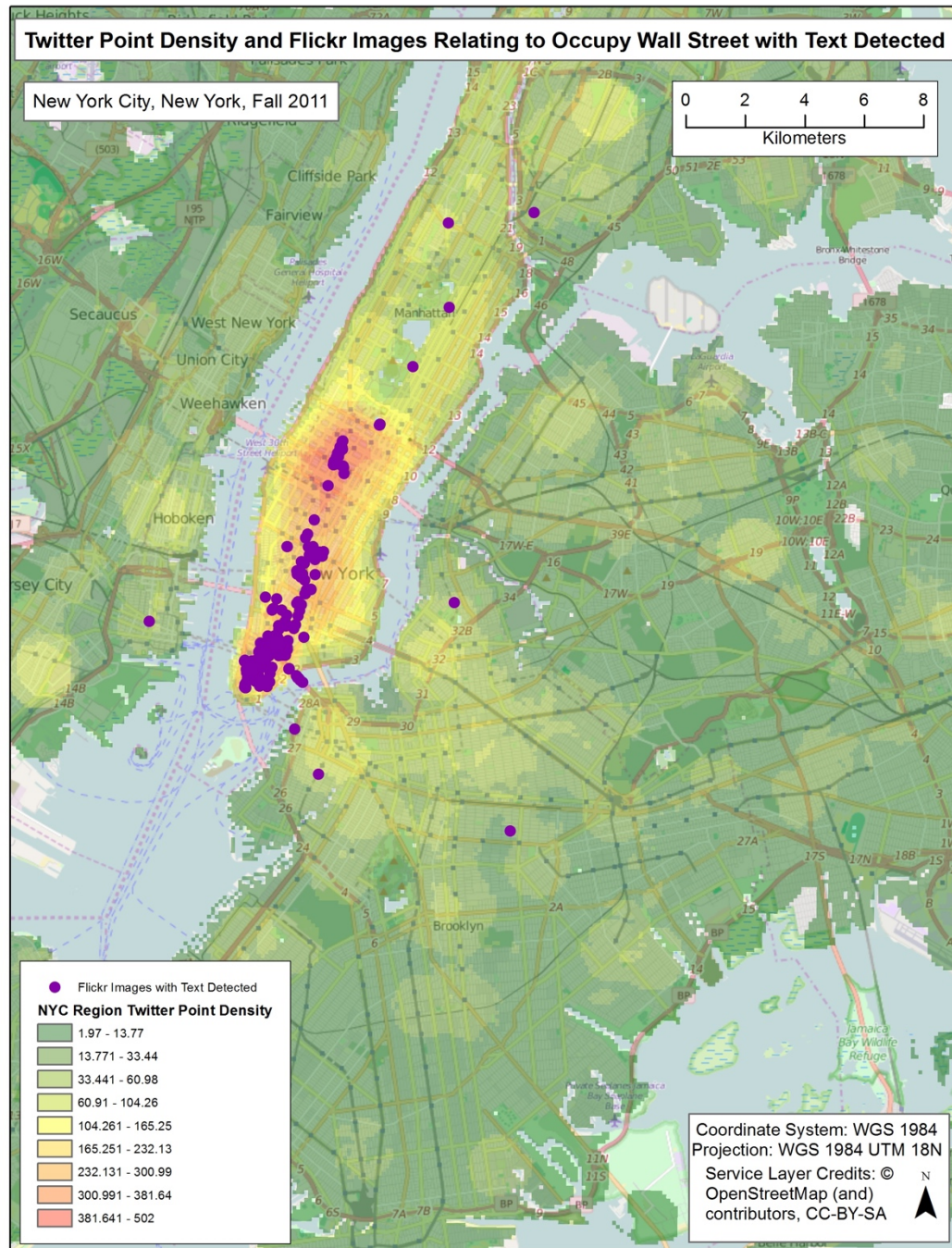


Figure 26: Flickr Images with Text Detected and Twitter Point Density

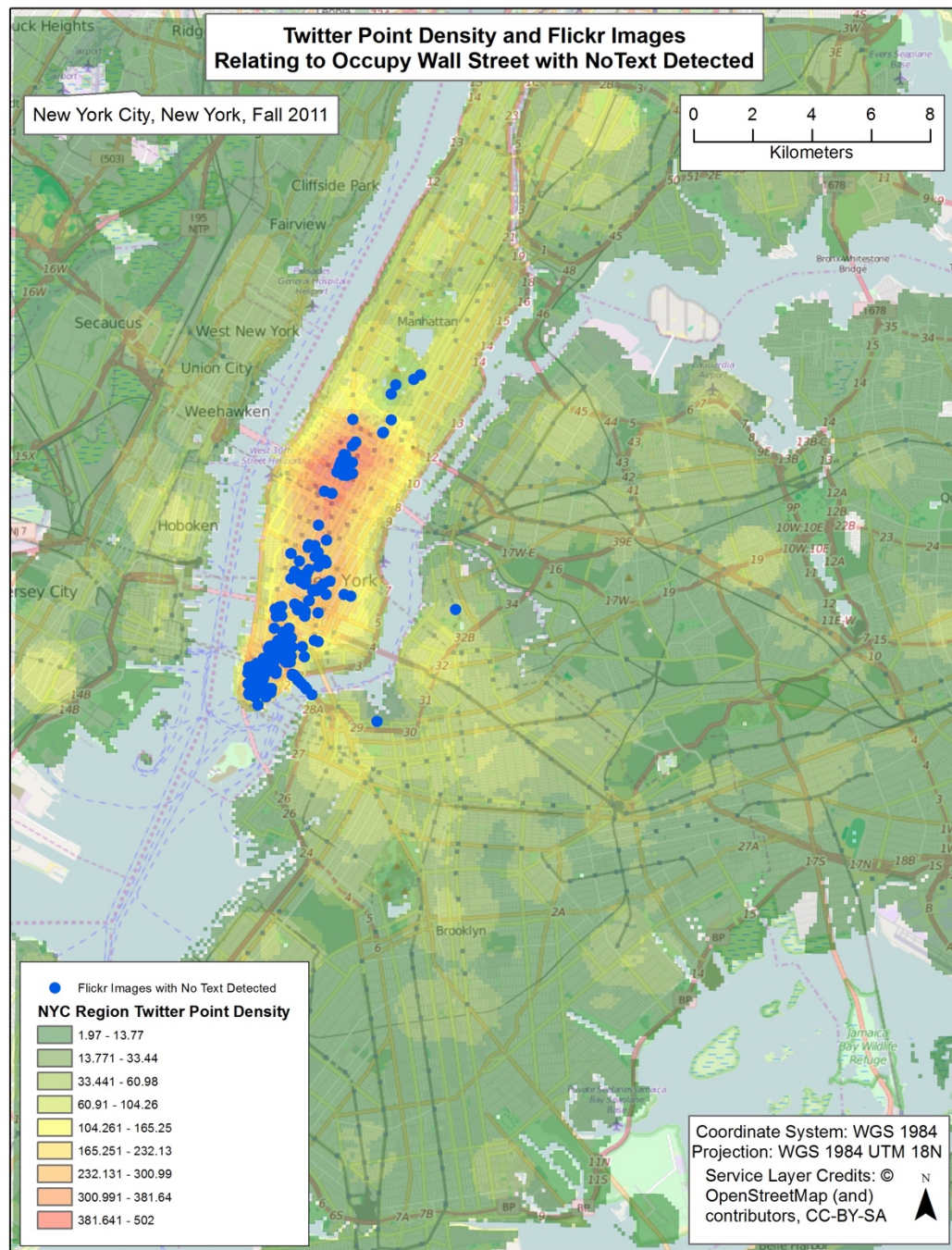


Figure 27: Flickr Images with No Text Detected and Twitter Point Density

Figure 28 and Figure 29 are close-ups of Manhattan, where the majority of Flickr images were located, along with the highest point density of Tweets. The locations with text detected and text not detected are the same.

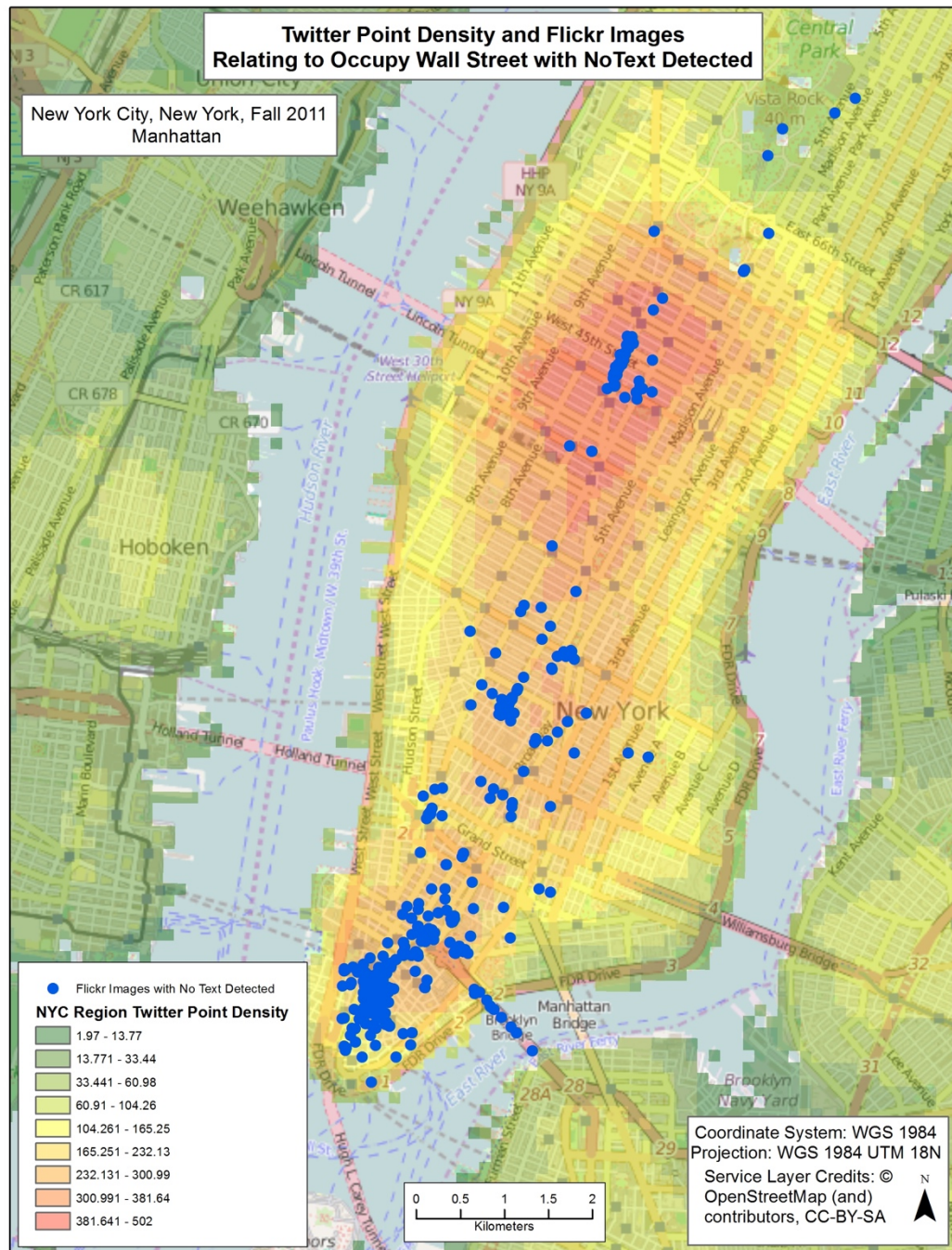


Figure 28: Manhattan Subset of Twitter Density and Flickr Images with no Text Detected

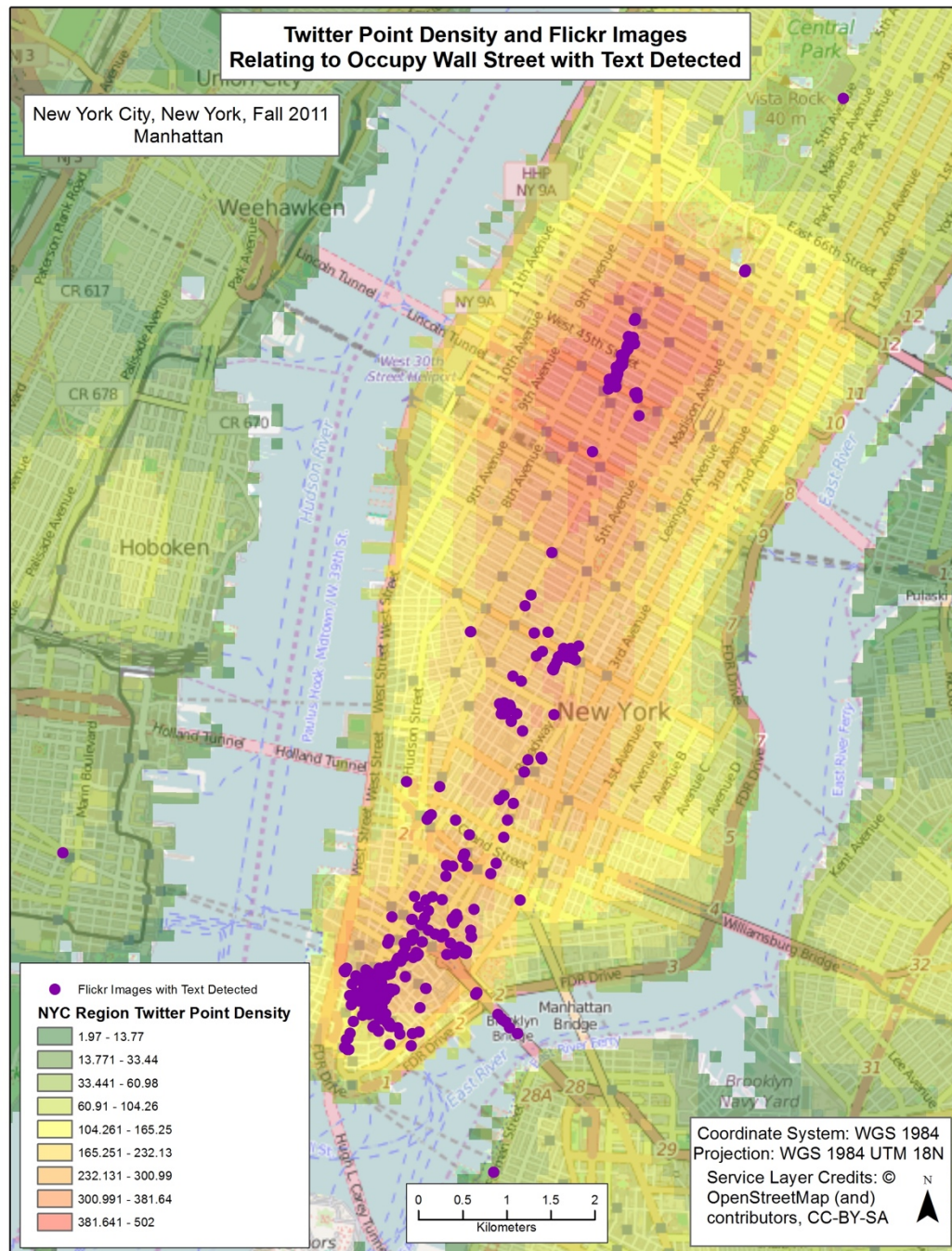
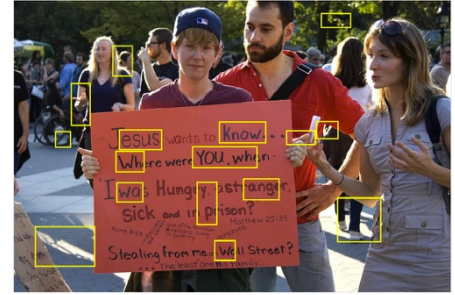


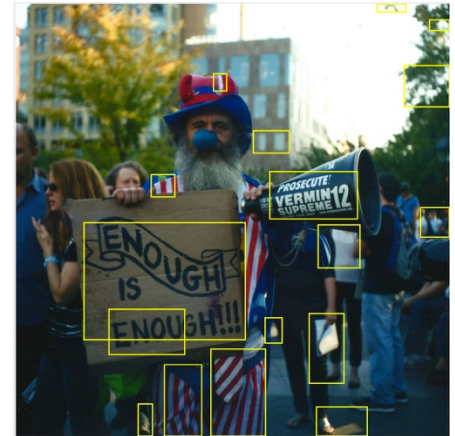
Figure 29: Manhattan Subset of Twitter Density and Flickr Images with Text Detected



(a)



(b)



(c)



(d)

Figure 30 (a) is an inset map on Washington Square Park showing 67 Flickr images that had text detected relating to OWS. Figure 30 (b), (c), and (d) are examples of the text detected images called out to their location in the park. The image is the output from the overlapping merged text regions to create bounding boxes around groups of text (similar to Figure 12).



Figure 30: Washington Square Park Inset Map with 67 Text Images Detected and Examples

Figure 31(a) is an inset map on Brooklyn Bridge showing nine Flickr images that had text detected relating to OWS. Figure 31 (b), (c), and (d) are examples of the text detected images, called out to their location on the bridge. The image is the output from

the overlapping merged text regions to create bounding boxes around groups of text (similar to Figure 12).



Figure 31: Brooklyn Bridge Inset Map with 9 Text Images Detected and Examples

Figure 32(a) is an inset map on Union Square Park showing 48 Flickr images that had text detected relating to OWS. Figure 32 (b), (c), and (d) are examples of the text detected images called out to their location in the park. The image is the output from the

overlapping merged text regions to create bounding boxes around groups of text (similar to Figure 12).



Figure 32: Union Square Park Inset Map with 48 Text Images Detected and Examples

Figure 33(a) is an inset map on Times Square showing 128 Flickr images that had text detected relating to OWS. Figure 33 (b), (c), and (d) are examples of the text detected images called out to their location in the park. The image is the output from the overlapping merged text regions to create bounding boxes around groups of text (similar to Figure 12).



Figure 33: Times Square Inset Map with 128 Text Images Detected and Examples

Figure 34 (a) is an inset map on Zuccotti Park showing 2587 Flickr images that had text detected relating to OWS. Figure 34 (b), (c), and (d) are examples of the text detected images called out to their location in the park. The image is the output from the

overlapping merged text regions to create bounding boxes around groups of text (similar to Figure 12).

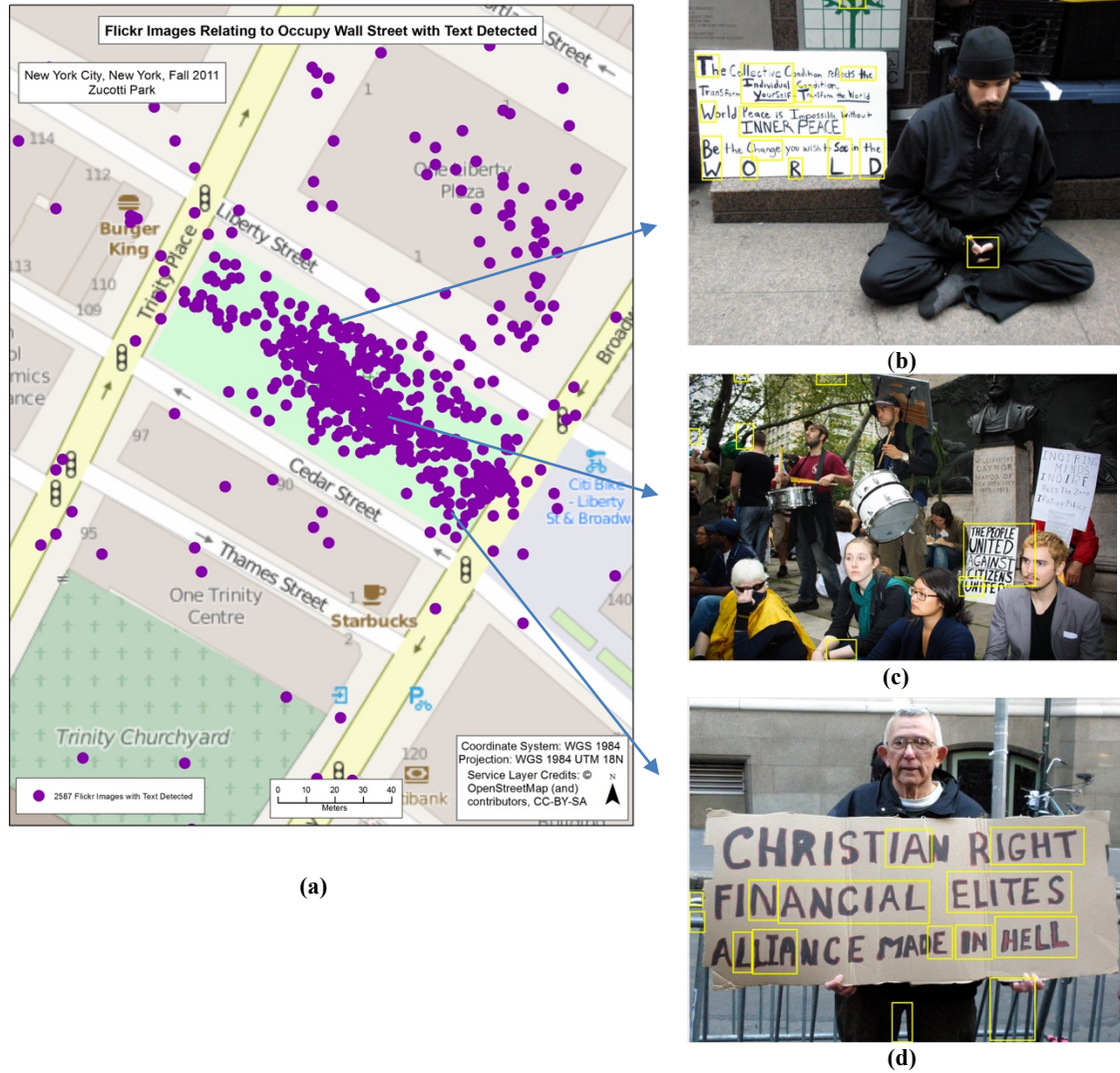


Figure 34: Zuccotti Park Inset Map with 2587 Text Images Detected and Examples

5.3 Discussion

The accuracy assessment of the subset of 200 images shows that 78% of the time, the images are correctly identified as having text or not. Looking at samples of the true positives, true negatives, false positives, and false negatives shows interesting results. In Figure 14, showcasing examples of true positives (images that were correctly identified as containing text) shows images with text that is evenly distributed, printed (vs. handwritten), has similar font size throughout, and a vertical orientation of text (as opposed to slanted). The text detection process correctly identified areas in the image that contain text in each image, and for four out of the five examples, the text recognition process (OCR) correctly identified the text in the image. In Figure 14d, there is variation in font size, non-vertical text orientation, and variation in font color. However, some text was still detected (vs. no text at all).

Examples of true negatives (images that were correctly classified as not having text) in Figure 15 show small yellow boxes in 15b and 15e, signifying potential text from the text detection process, however, when run through the OCR, no text was found. The patterns found in those two images from a person's legs (15b), or the bag the dog is being carried in and the container with the straight line (15e) are the potential text regions. It is also important to note that in 15b and 15c, incidental text (text in the background of an image such as road signs, storefronts, or other construction not relating to signs) is not detected, leading to a positive evaluation of the text detection process overall. Because there was no text in the foreground, the algorithm didn't immediately look for text in the background, which is not relevant to the protest activity.

In Figure 16, which showcases examples of false positives (images incorrectly classified as having text), 16e has a building with windows that are mistakenly identified as text. The even distribution, size, and color of these windows and their contrast from the background make it an easy mistake for the text detection and recognition algorithms.

The false negative (images incorrectly classified as not having text) examples in Figure 17 showcase how the variations in orientation specifically cause issues with automatic text extraction (Yin et al., 2014; Li & Lu, 2012; Jung et al., 2004), even though font and size are consistent. The text detection process outputting the yellow boxes around potential text regions does accurately highlight the text, which is in line with claims made by Neumann and Matas (2012) and Su and Xu (2015) that connected components are not sensitive to orientation issues. This suggests that the reason these images were classified as false negatives were due to the OCR text recognition process and not the connected component process. If that is the case, many images that were not classified as having text in them could have been included, expanding the locations where text was detected in the spatio-temporal clustering and comparisons with the Twitter density maps.

The clustered map showing text detected Flickr images in Figure 19 and the inset maps from the five cities (New York, Chicago, Los Angeles, Washington, D.C, and San Francisco and Oakland) show that the protest activities in physical space primarily occurred in urban centers around the country (and Canada). The subsets of New York, Chicago, Los Angeles, Washington D.C., and San Francisco/Oakland were chosen

because they were the top five areas with the most points per cluster as highlighted in Table 2.

Figure 20: Washington, D.C. Inset Map shows that McPherson Square appears to have more images with text not detected vs. text detected. Freedom Plaza appears to have a mixture of images with text detected and without, with no pattern emerging on the location of the signs, other than inside the public park. Washington D.C. is the home to the National Mall, a space known for social movements and actions (Endres & Senda-Cook, 2011); however, very few images are located on the mall, with the majority in McPherson Square and Freedom Plaza. McPherson Square is on K Street, which is where lobbyists (people who influence policy makers) have offices. Freedom Plaza, another popular location of Occupy protests in D.C., is a few blocks from the White House. Additionally, the actual name of the plaza, “Freedom,” suggests that the activists are using that space intentionally as a base to share their message.

Figure 21: Los Angeles, California Inset Map has the majority of images with and without text located just outside the Los Angeles City Hall, the local government office of the region. Figure 23: Chicago, Illinois Inset Map has a cluster of images outside the Federal Reserve Bank of Chicago, a symbol of the federal government and a place of power in the finance industry. These subset maps (in addition to Figure 19, the cluster map of the entire United States and Canada) answer the questions of where the protest activity was occurring, confirming the work from Sewell (2001), who claimed that the majority of protests occur in urban areas because of the access to public spaces and a wider audience. All of the Flickr images were geolocated to be in parks, on streets, or

near government centers, all areas of high visibility and audience. The symbolic significance of the protest can better be characterized because people put signs there, as opposed to the other locations around the city where activity occurred. In New York City, the location of over 2,000 signs in Zuccotti Park has immense meaning. This park is located in the financial district of lower Manhattan, right outside Wall Street, which controls the United States stock market and other big banks, believed by the activists to be the culprit in the growing wealth inequality and the 2008 financial downturn in America. By protesting outside Wall Street, protestors wanted the people who work on Wall Street to know they weren't happy. Having this same level of protestors in Central Park would not have the same effect, because of the distance from Wall Street. To better characterize the overall protest activity, Twitter data were used as a proxy (Wayant et al., 2012).

The Twitter point density maps in Figure 26 and Figure 27, which show the overall New York City Region, and the maps that focus on Manhattan in Figure 28 and Figure 28 show that the majority of the protest activity occurred in Manhattan. The color ramp is stretched so that the light green areas have less activity, progressing through yellow, orange, and red areas, with the red areas having the highest level of tweets per square kilometer. The red areas signify the highest levels of protest activity because more tweets were geolocated to those locations. The Twitter maps show more points per square kilometer in lower Manhattan near Zuccotti Park, and in Midtown, near Times Square. Both locations had almost all of the Flickr images relating to OWS. The Flickr images are concentrated around two locations of high protest activity (red areas), answering the

question that the Flickr images with text detected spatially overlay with the majority of the tweets, or the locations of the overall protest activity.

Subset images of New York City protest locations are shown in Figure 30 through Figure 34 (Washington Square Park, Brooklyn Bridge, Union Square Park, Times Square, and Zuccotti Park, respectively). By highlighting some of the specific images that were determined by the text detection and text recognition process to contain text, this shows and confirms that those signs relate to the OWS movement and that the signs were located in public places throughout the city that have high visibility, in line with the claims made by Sewell (2001), Endres & Senda-Cook (2011) that space in contentious politics is relevant.

The Twitter point density maps show that the location of the most tweets is in line with the location of the Flickr images. As shown in the literature, social media (and Twitter in particular) are very effective and entwined with social protests, especially the OWS movement. The location of the protest activities in OWS in public spaces, urban areas, parks, proximity to government centers and buildings, policy makers and policy influencers, and the financial institutions confirm the theory that space in contentious politics is important to consider when examining the overall protest activity. These spaces could have already had meaning or they were transformed to be meaningful during the protest (areas such as Times Square, the Brooklyn Bridge, etc.). In the case of Zuccotti Park, the protest was so significant that it has been permanently linked to the OWS activities and any future protest activity in that space, whether related or unrelated to the OWS activity; this is a byproduct of the Fall 2011 activities that occurred. These

results suggest that, overall, signs are immersed in the protest activity, but tend to concentrate in specific locations that are likely to have a more central role in context of the protest.

CHAPTER SIX

6.1 Conclusions

Social media are offering a new lens through which researchers can study complex human systems, and the amount of data is increasing as time goes on. By integrating a social network analysis with a geographic analysis, a new understanding of the location of a space can be defined and interpreted (Croitoru et al., 2014). Twitter in particular is one way for users to document, share, coordinate, live-report, and provide commentary on a phenomenon or event, such as a protest activity. Signs and posters, which have been used in protest activities for many decades, are a version of an analog tweet. Short messages are shared with the wider audience. Rhetoricians have claimed that the messages shown on those signs are important to study, and geographers have claimed that the location of those signs is important.

In order to determine which signs from the Occupy Wall Street movement contained text, a text detection process was used. Text detection in computer vision is a very active and hard problem. Every year, there are hundreds of new methods being produced and papers being published that try to improve the text detection and recognition processes. This research showcased one text detection method in which

images with natural scenes (with varying orientations, fonts, sizes, colors, and blurriness) can be enhanced before running that image through an optical character recognition tool. That process was chosen based on the complexity and type of data that were being studied. There were assumptions made that if text was detected, it was relating to the Occupy Wall Street movement, as Flickr tags were used to harvest the images.

Once text was detected, those images were clustered using a spatio-temporal clustering method that showed that images in urban areas were related to each other in space and time, solidifying the work that studies the theory of space and place in contentious politics. Comparing that work with precisely geolocated Twitter data (as a proxy of the overall protest) shows that the majority of images were in areas of high protest activity, such as urban squares, parks, plazas, and financial and government institutions.

The quantitative work shown in this research confirms the qualitative claims made by the researchers regarding the need and validity of studying space and contentious politics, and that studying the signs of protests is important to studying contentious politics rhetoric. Signs and posters are immersed in protest activity and tend to be located in areas with a central role and meaning.

6.2 Future Work

Although the text detection process was evaluated using a subset of images, it is important to remember that the text detection process was not used to evaluate the accuracy of the text detection methodology, but as a catalyst for studying the relationship

between social media images and physical space. Assumptions are that if text was detected, it actually was the text from the image. As referenced in the examples of true positives, true negatives, false positives, and false negatives (Figure 14 through Figure 17), the text detection process with the detected bounding boxes around areas of text does an adequate job (78% of the time, it is correct) finding text. However, there seem to be issues with the resulting text recognition piece - the optical character recognition tool. Even though the text detection method clearly outlines boxes of text, there are still issues with the OCR interpretations. As mentioned in the literature, OCR works very well for documents with evenly formatted, distributed, and vertical text. However, when it comes to natural scenes with varying degrees of orientation, illumination, fonts, sizes, colors, and blur, OCR all but fails to accurately output the proper text. Handwritten posters also are not as easily detected as a printed poster image. The connected components, stroke width, and geometric filtering recognize text in the images, as determined by the yellow bounding boxes; however, the OCR process needs improvements. One possible method of improving the OCR results could be rotating the result of the text detection images to be completely vertical, so that the orientation of the text is straight up and down.

There are many text detection processes, all with their advantages and disadvantages. The process chosen for this work was used because of the ease of computational complexity and lack of a priori knowledge or lexicon to determine the words used in the images. Future work could enhance these methods by implementing a Canny edge detector to better separate out characters of potential text, such as in the research conducted by Chen et al. (2011), or by using a 'contrast enhanced' method (Li &

Lu, 2012) to overcome the issues that MSERs have with detected blurred text. With the process used in this research, all of the defaults were taken because of the large dataset size. Further refining could adjust those variables (aspect ratio, stroke width, MSER region size, etc.) for improved text detection accuracy.

The resulting images with detected text boxes around potential text regions generally gave good results. However, problems arose with the OCR process to determine human readable text from that image. Future improvements to optical character recognition could benefit from improved text recognition in natural scenes.

This study focused on Twitter data that were isolated to the New York City region. The reason for this is because that location was the epicenter and the catalyst for all other OWS movements. Further research should be done to examine the relationship of Twitter and Flickr images in other urban spaces, such as San Francisco and Oakland, two prominent areas of OWS activity.

The Flickr images and tweets used in this study have dates associated with them, and it is assumed that the date associated is the date the image was taken or the date the tweet was uploaded. However, further analysis should be done to explore the latency involved when images are taken and when they are uploaded to Flickr. Some users could upload images immediately, and others could wait a few days before uploading a batch of images all at once.

This research did not explore the density of signs in an image, but further work could be done to weight images based on how many signs are contained within the image. Many images include only one sign, but some, such as in a larger crowd, can

contain many signs, with as many as 10 to 20 signs per image. Exploring density of signs in one image could expand on the location of signs, since one image does not necessarily equal one sign.

To explore the level of coordination in the Occupy Wall Street movement, looking at the differences in types of signs could be examined. Coordination could be higher if there are many professionally produced posters or many copies of signs held by many people. Coordination could be lower if every sign was unique and made in a ‘do it yourself,’ or DIY, fashion. Similar to the work performed by Steinert-Threlkeld et al. (2015), which looked at the coordination of Tweets in protest, investigating coordination of Flickr images could show interesting results.

Finally, Flickr tags were explored here, but not actually used in this research. Flickr tags could be compared with the hashtags and content from Twitter messages to create tag clouds and compare the rhetoric of the protest signs and social media.

REFERENCES

- Ahern, S., Naaman, M., Nair, R., & Yang, J. H.-I. (2007). World Explorer: Visualizing Aggregate Data from Unstructured Text in Geo-referenced Collections. In *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 1–10). New York, NY, USA: ACM. <http://doi.org/10.1145/1255175.1255177>
- Al-Ali, N. (2012). Gendering the Arab Spring 1. *Middle East Journal of Culture and Communication*, 5(1), 26–31. <http://doi.org/10.1163/187398612X624346>
- Ames, M., & Naaman, M. (2007). Why We Tag: Motivations for Annotation in Mobile and Online Media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 971–980). New York, NY, USA: ACM. <http://doi.org/10.1145/1240624.1240772>
- Bailey, J., & McAtee, D. (2003a). The Politics and Poetics of Union Transgression: The Role of Visual Methods in Analyzing Union Protest Strategy. *Social Analysis: The International Journal of Social and Cultural Practice*, 47(3), 27–45.
- Bailey, J., & McAtee, D. (2003b). The Politics and Poetics of Union Transgression: The Role of Visual Methods in Analyzing Union Protest Strategy. *Social Analysis: The International Journal of Social and Cultural Practice*, 47(3), 27–45.

- Ballan, L., Bertini, M., Bimbo, A. D., & Serra, G. (2009). Semantic annotation of soccer videos by visual instance clustering and spatial/temporal reasoning in ontologies. *Multimedia Tools and Applications*, 48(2), 313–337. <http://doi.org/10.1007/s11042-009-0342-4>
- Bastos, M. T., Mercea, D., & Charpentier, A. (2015). Tents, Tweets, and Events: The Interplay Between Ongoing Protests and Social Media. *Journal of Communication*, 65(2), 320–350. <http://doi.org/10.1111/jcom.12145>
- Bekkers, V., Beunders, H., Edwards, A., & Moody, R. (2011). New Media, Micromobilization, and Political Agenda Setting: Crossover Effects in Political Mobilization and Media Usage. *The Information Society*, 27(4), 209–219. <http://doi.org/10.1080/01972243.2011.583812>
- Ben-Ami, I., Basha, T., & Avidan, S. (2012). Racing Bib Number Recognition. BMVC. Retrieved from <http://www.eng.tau.ac.il/~talib/papers/RBNR.pdf>
- Burridge, J. (2008). “Hunting is not just for blood-thirsty toffs”: The Countryside Alliance and the visual rhetoric of a poster campaign. *Text & Talk-An Interdisciplinary Journal of Language, Discourse Communication Studies*, 28(1), 31–53.
- Cao, F., Estert, M., Qian, W., & Zhou, A. (2006). Density-Based Clustering over an Evolving Data Stream with Noise. In *Proceedings of the 2006 SIAM International Conference on Data Mining* (Vols. 1–0, pp. 328–339). Society for Industrial and Applied Mathematics. Retrieved from <http://epubs.siam.org.mutex.gmu.edu/doi/abs/10.1137/1.9781611972764.29>

- Castañeda, E. (2012). The Indignados of Spain: A Precedent to Occupy Wall Street. *Social Movement Studies*, 11(3-4), 309–319. <http://doi.org/10.1080/14742837.2012.708830>
- Chen, H., Tsai, S. S., Schroth, G., Chen, D. M., Grzeszczuk, R., & Girod, B. (2011). Robust Text Detection in Natural Images with Edge-Enhanced Maximally Stable Extremal Regions. In *2011 18th IEEE International Conference on Image Processing (ICIP)* (pp. 2609–2612). <http://doi.org/10.1109/ICIP.2011.6116200>
- Croitoru, A., Crooks, A. T., Radzikowski, J., Stefanidis, A., Vatsavai, R. R., & Wayant, N. (2014). Geoinformatics and social media: A new big data challenge. *Big Data Techniques and Technologies in Geoinformatics*, CRC Press, Boca Raton, FL, 207–232.
- Croitoru, A., Wayant, N., Crooks, A., Radzikowski, J., & Stefanidis, A. (2015). Linking cyber and physical spaces through community detection and clustering in social media feeds. *Computers, Environment and Urban Systems*, 53, 47–64.
- DeLuca, K. M., Lawson, S., & Sun, Y. (2012). Occupy Wall Street on the Public Screens of Social Media: The Many Framings of the Birth of a Protest Movement. *Communication, Culture & Critique*, 5(4), 483–509. <http://doi.org/10.1111/j.1753-9137.2012.01141.x>
- Earl, J., & Kimport, K. (2011). *Digitally Enabled Social Change: Activism in the Internet Age*. MIT Press.
- Epshtein, B., Ofek, E., & Wexler, Y. (2010). Detecting text in natural scenes with stroke width transform. In *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2963–2970). <http://doi.org/10.1109/CVPR.2010.5540041>
- Eskilson, S. (2012). *Graphic Design: A New History, Second Edition*. Yale University Press.

- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, pp. 226–231). Retrieved from <http://www.aaai.org/Papers/KDD/1996/KDD96-037>
- Fabrizio, J., Marcotegui, B., & Cord, M. (2013). Text detection in street level images. *Pattern Analysis and Applications*, 16(4), 519–533. <http://doi.org/10.1007/s10044-013-0329-7>
- FAQs about trends on Twitter. (n.d.). Retrieved April 12, 2016, from <https://support.twitter.com/articles/101125>
- Find every photo with Flickr's new unified search experience | Flickr Blog. (2015). Retrieved from <http://blog.flickr.net/en/2015/05/07/flickr-unified-search/>
- Flickr | tags - Search Results. (n.d.). Retrieved April 9, 2016, from https://help.yahoo.com/kb/index?page=answers&startover=y&y=PROD_FLICKR&source=search.landing_search&locale=en_US&question_box=tags#37
- Flickr API. (2007). Retrieved April 9, 2016, from <https://www.flickr.com/groups/api/discuss/72157594497877875>
- Flickr: The Help Forum: [Official Thread] Updates on tags. (2015). Retrieved April 9, 2016, from <https://www.flickr.com/help/forum/en-us/72157652019487118/>
- Gerbaudo, P. (2012). *Tweets and the Streets: Social Media and Contemporary Activism*. Pluto Press.
- Gillham, P. F., Edwards, B., & Noakes, J. A. (2013). Strategic incapacitation and the policing of Occupy Wall Street protests in New York City, 2011. *Policing and Society*, 23(1), 81–102. <http://doi.org/10.1080/10439463.2012.727607>

- Gleason, B. (2013). #Occupy Wall Street: Exploring Informal Learning About a Social Movement on Twitter. *American Behavioral Scientist*, 0002764213479372.
<http://doi.org/10.1177/0002764213479372>
- González, Á., Bergasa, L. M., & Yebes, J. J. (2014). Text Detection and Recognition on Traffic Panels From Street-Level Imagery Using Visual Appearance. *IEEE Transactions on Intelligent Transportation Systems*, 15(1), 228–238.
<http://doi.org/10.1109/TITS.2013.2277662>
- González, Á., Bergasa, L. M., Yebes, J. J., & Bronte, S. (2012). Text location in complex images. In *2012 21st International Conference on Pattern Recognition (ICPR)* (pp. 617–620).
- Gould, R. V. (1991). Multiple Networks and Mobilization in the Paris Commune, 1871. *American Sociological Review*, 56(6), 716–729. <http://doi.org/10.2307/2096251>
- Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 78(6), 1360–1380.
- Grothe, C., & Schaab, J. (2009). Automated Footprint Generation from Geotags with Kernel Density Estimation and Support Vector Machines. *Spatial Cognition & Computation*, 9(3), 195–211. <http://doi.org/10.1080/13875860903118307>
- Hartsell, T. (n.d.). In the privacy of our homes | code.flickr.com. Retrieved from <https://code.flickr.net/2011/08/30/in-the-privacy-of-our-homes/>
- IEEE Xplore Abstract Record. (n.d.). Retrieved from http://ieeexplore.ieee.org.mutex.gmu.edu/xpl/login.jsp?tp=&arnumber=6116200&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D6116200

IEEE Xplore Full Text PDF. (n.d.). Retrieved from

[http://ieeexplore.ieee.org.mutex.gmu.edu/ielx5/6094293/6115588/06116200.pdf?tp=&ar
number=6116200&isnumber=6115588](http://ieeexplore.ieee.org.mutex.gmu.edu/ielx5/6094293/6115588/06116200.pdf?tp=&arnumber=6116200&isnumber=6115588)

Ikica, A., & Peer, P. (2014). Influence of image quality on SWT voting-based color reduction method for detecting text in natural scene images. In *2014 International Work Conference on Bio-inspired Intelligence (IWOBI)* (pp. 142–148).

<http://doi.org/10.1109/IWOBI.2014.6913954>

Irwin, J. A. (2007). The Guilt in Mom, We're Home: The content, control, and critique of modern war protest posters. In *International Communication Association conference, San Francisco, CA*.

Jung, K., In Kim, K., & K. Jain, A. (2004). Text information extraction in images and video: a survey. *Pattern Recognition*, 37(5), 977–997. <http://doi.org/10.1016/j.patcog.2003.10.012>

Juris, J. S. (2012). Reflections on #Occupy Everywhere: Social media, public space, and emerging logics of aggregation. *American Ethnologist*, 39(2), 259–279.

<http://doi.org/10.1111/j.1548-1425.2012.01362.x>

Karatzas, D., Gomez-Bigorda, L., Nicolaou, A., Ghosh, S., Bagdanov, A., Iwamura, M., ...

Valveny, E. (2015). ICDAR 2015 competition on Robust Reading. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)* (pp. 1156–1160). <http://doi.org/10.1109/ICDAR.2015.7333942>

Keßler, C., Maué, P., Heuer, J. T., & Bartoschek, T. (2009). Bottom-Up Gazetteers: Learning from the Implicit Semantics of Geotags. In K. Janowicz, M. Raubal, & S. Levashkin

- (Eds.), *GeoSpatial Semantics* (pp. 83–102). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-10436-7_6
- Koo, H. I., & Kim, D. H. (2013). Scene Text Detection via Connected Component Clustering and Nontext Filtering. *IEEE Transactions on Image Processing*, 22(6), 2296–2305.
<http://doi.org/10.1109/TIP.2013.2249082>
- Leitner, H., Sheppard, E., & Sziarto, K. M. (2008). The spatialities of contentious politics. *Transactions of the Institute of British Geographers*, 33(2), 157–172.
<http://doi.org/10.1111/j.1475-5661.2008.00293.x>
- Li, Y., & Lu, H. (2012). Scene text detection via stroke width. In *2012 21st International Conference on Pattern Recognition (ICPR)* (pp. 681–684).
- Marlow, C., Naaman, M., Boyd, D., & Davis, M. (2006). HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, to Read. In *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia* (pp. 31–40). New York, NY, USA: ACM.
<http://doi.org/10.1145/1149941.1149949>
- Matas, J., Chum, O., Urban, M., & Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10), 761–767.
<http://doi.org/10.1016/j.imavis.2004.02.006>
- McAdam, D. (1986). Recruitment to High-Risk Activism: The Case of Freedom Summer. *American Journal of Sociology*, 92(1), 64–90.
- McAdam, D., & Paulsen, R. (1993). Specifying the Relationship Between Social Ties and Activism. *American Journal of Sociology*, 99(3), 640–667.

McCaughey, M., & Ayers, M. D. (2003). Cyberactivism. *Online Activism in Theory and Practice*. New York/London.

Miller, T. A. (2014, May 20). Computer vision at scale with Hadoop and Storm | code.flickr.com. Retrieved from <https://code.flickr.net/2014/05/20/computer-vision-at-scale-with-hadoop-and-storm/>

Neumann, L., & Matas, J. (2010). A Method for Text Localization and Recognition in Real-World Images. In R. Kimmel, R. Klette, & A. Sugimoto (Eds.), *Computer Vision – ACCV 2010* (pp. 770–783). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-19318-7_60

New-York Historical Society | Art as Activism: Graphic Art from the Merrill C. Berman Collection. (2015). Retrieved April 7, 2016, from <http://www.nyhistory.org/exhibitions/art-activism-graphic-art-merrill-c-berman-collection>

Nistér, D., & Stewénus, H. (2008). Linear Time Maximally Stable Extremal Regions. In D. Forsyth, P. Torr, & A. Zisserman (Eds.), *Computer Vision – ECCV 2008* (pp. 183–196). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-88688-4_14

NY General Assembly. (n.d.). About OWS. Retrieved April 23, 2016, from <http://occupywallstreet.net/learn>

#Occupy Wall Street | Declaration of the Occupation of New York City. (2011). Retrieved April 9, 2016, from <http://www.nycga.net/resources/documents/declaration/>

- Occupy Wall Street| Journal of the mental environment. (n.d.). Retrieved from <https://www.adbusters.org/occupywallstreet/>
- Pan, Y. F., Hou, X., & Liu, C. L. (2011). A Hybrid Approach to Detect and Localize Texts in Natural Scene Images. *IEEE Transactions on Image Processing*, 20(3), 800–813. <http://doi.org/10.1109/TIP.2010.2070803>
- Pearce, K. E., & Kendzior, S. (2012). Networked Authoritarianism and Social Media in Azerbaijan. *Journal of Communication*, 62(2), 283–298. <http://doi.org/10.1111/j.1460-2466.2012.01633.x>
- Penney, J., & Dadas, C. (2013). (Re)Tweeting in the service of protest: Digital composition and circulation in the Occupy Wall Street movement. *New Media & Society*, 1461444813479593. <http://doi.org/10.1177/1461444813479593>
- Pollock, J. (2011, August 23). Streetbook. Retrieved April 7, 2016, from <https://www.technologyreview.com/s/425137/streetbook/>
- Poynor, R. (2012). Why the Activist Poster is Here to Stay. Retrieved April 9, 2016, from <http://designobserver.com/feature/why-the-activist-poster-is-here-to-stay/36068>
- Quy Phan, T., Shivakumara, P., Tian, S., & Lim Tan, C. (2013). Recognizing Text with Perspective Distortion in Natural Scenes (pp. 569–576). Presented at the Proceedings of the IEEE International Conference on Computer Vision. Retrieved from http://www.cv-foundation.org/openaccess/content_iccv_2013/html/Phan_Recognizing_Text_with_2013_ICCV_paper.html
- Sewell Jr, W. H. (2001). Space in contentious politics. *Silence and Voice in the Study of Contentious Politics*, 78, 18.

- Sigurbjörnsson, B., & van Zwol, R. (2008). Flickr Tag Recommendation Based on Collective Knowledge. In *Proceedings of the 17th International Conference on World Wide Web* (pp. 327–336). New York, NY, USA: ACM. <http://doi.org/10.1145/1367497.1367542>
- Smith, R. (2007). An Overview of the Tesseract OCR Engine. In *2013 12th International Conference on Document Analysis and Recognition* (Vol. 2, pp. 629–633). Los Alamitos, CA, USA: IEEE Computer Society. <http://doi.org/10.1109/ICDAR.2007.56>
- Tan, E. S., & Tollenaar, A. (2007). Aesthetic Communication through Posters. *Empirical Studies of the Arts*, 25(1), 21–39. <http://doi.org/10.2190/N8K4-3482-0503-T763>
- Tang, G. (2016). Happy 12th Birthday, Flickr! | Flickr Blog. Retrieved from <http://blog.flickr.net/en/2016/02/10/happy-12th-birthday-flickr/>
- Tilly, C. (2008). *Contentious Performances*. Cambridge University Press.
- Twitter Q4 2015 Earnings Report. (2015). Twitter. Retrieved from http://files.shareholder.com/downloads/AMDA-2F526X/1594070831x0x874450/36429F80-3FCD-46FF-8AFF-232AE9B6CEDD/2015_Q4_Earnings_Slides.pdf
- Valenzuela, S. (2013). Unpacking the Use of Social Media for Protest Behavior The Roles of Information, Opinion Expression, and Activism. *American Behavioral Scientist*, 57(7), 920–942. <http://doi.org/10.1177/0002764213479375>
- Valenzuela, S., Arriagada, A., & Scherman, A. (2012). The Social Media Basis of Youth Protest Behavior: The Case of Chile. *Journal of Communication*, 62(2), 299–314. <http://doi.org/10.1111/j.1460-2466.2012.01635.x>

- Wang, K., Babenko, B., & Belongie, S. (2011). End-to-end scene text recognition. In *2011 IEEE International Conference on Computer Vision (ICCV)* (pp. 1457–1464).
<http://doi.org/10.1109/ICCV.2011.6126402>
- Wang, Q., Lu, Y., & Sun, S. (2015). Text detection in nature scene images using two-stage nontext filtering. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)* (pp. 106–110). <http://doi.org/10.1109/ICDAR.2015.7333735>
- Wang, T., Wu, D. J., Coates, A., & Ng, A. Y. (2012). End-to-end text recognition with convolutional neural networks. In *2012 21st International Conference on Pattern Recognition (ICPR)* (pp. 3304–3308).
- Wang, X., Song, Y., & Zhang, Y. (2013). Natural Scene Text Detection with Multi-channel Connected Component Segmentation. In *2013 12th International Conference on Document Analysis and Recognition (ICDAR)* (pp. 1375–1379).
<http://doi.org/10.1109/ICDAR.2013.278>
- Watt, S. S. (2012). A postfeminist apologia: Susan G. Komen for the Cure's evolving response to the Planned Parenthood controversy. *Journal of Contemporary Rhetoric*, 2(3/4), 65–79.
- Wayant, N., Crooks, A., Stefanidis, A., Croitoru, A., Radzikowski, J., Stahl, J., & Shine, J. (2012). Spatiotemporal clustering of twitter feeds for activity summarization. *GIScience (short Paper)*. Retrieved from
http://www.giscience.org/past/2012/proceedings/abstracts/giscience2012_paper_78.pdf
- Wildermuth, S., Davis, C. B., Frederick, E., & Bolton, J. (2014). Sign, Sign, Everywhere a Sign: An Analysis of the Argument, Target and Content of Protest Signs Displayed

- During the Wisconsin Budget Repair Bill/Act 10 Protests of 2011. *American Communication Journal*, 16(1). Retrieved from http://ac-journal.org/journal/2014-2015/Vol16/Iss1/ACJ_2014-019_Corey2.pdf
- Yao, C., Wu, J., Zhou, X., Zhang, C., Zhou, S., Cao, Z., & Yin, Q. (2015). Incidental Scene Text Understanding: Recent Progresses on ICDAR 2015 Robust Reading Competition Challenge 4. *arXiv:1511.09207 [cs]*. Retrieved from <http://arxiv.org/abs/1511.09207>
- Yin, X. C., Yin, X., Huang, K., & Hao, H. W. (2014). Robust Text Detection in Natural Scene Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(5), 970–983. <http://doi.org/10.1109/TPAMI.2013.182>
- Yun, S., & Chang, W.-Y. (2010). Political Participation of Teenagers in the Information Era. *Social Science Computer Review*. <http://doi.org/10.1177/0894439310363255>

BIOGRAPHY

Kathryn M. Kash graduated from Massaponax High School, Fredericksburg, Virginia, in 2005. She received her Bachelor of Science in Geographic Science with a focus on geospatial information systems from James Madison University in 2009. She is currently employed as a Geographer by the Army Geospatial Center (AGC) in Alexandria, VA. Prior to joining AGC, she was a Cartographer with the Topographic Engineering Center (now the Geospatial Research Laboratory), a division of the Engineering Research and Development Center (ERDC), part of the US Army Corps of Engineers in Alexandria, VA.