Detecting Cloned Portions of Images

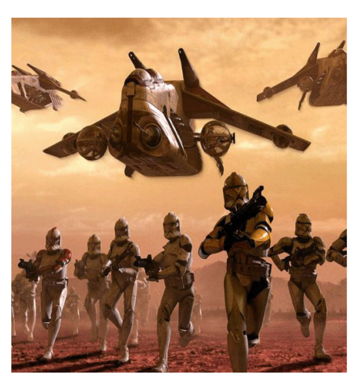
C1C Steven Glowacki, C1C Joshua Gminski
Computer Science 438 Information Warfare

INTRODUCTION

As digital photography becomes more common, the need to authenticate and protect the integrity of digital images has become a concern. Institutions ranging from law enforcement to journalism have stakes in the authenticity of images that are used to convey information.

Currently the ability to detect tampering in digital images is limited to detecting specific types of image tampering in specific types of images. Our project focuses on detecting images that have been maliciously tampered with using a cloning tool found in common commercial software. The goals of our project are to:

- Gain background understanding of tampering detection techniques that currently exist.
- Identify an area of image tampering detection that is currently lacking.
- Create a tool the fills a need by detecting a common type of image tampering.



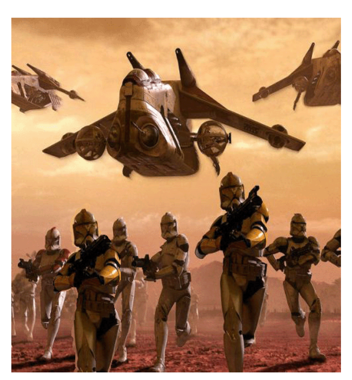


Figure 1: non-tampered image
Figure 2: tampered Image. Can you spot the clone?

BACKGROUND

In our background research we initially focused on techniques for detecting image tampering using analysis of expected color values due to the color filter array interpolation used in cameras. However, after our initial investigation, we chose a different approach and came up with the idea of analyzing color values of an image to detect portions of an image that have been cloned. Background research showed that an approach focusing on color matching to detect clones had not been documented. Most prior research focused on detecting anomalies left over by the cloning tools them selves.

METHODOLOGY

Algorithm:

- Pick a random pixel from the image.
- Iterate through the image and flag any pixels whose color values match.
- Check areas around a flagged pixel to determine if the area should be checked further.
- If it is concluded that an area needs to be checked further than an expanding search starts.
- Expanding search, seeks both left and right on the current rows of the corresponding pixels until a pixel pair is found that do not match.
- The search then move down a column and repeats the left and right seeking until a non-identical pixel pair is found.
- The search then moves upward on the column in the same manner.
- Continue to choose random pixels until the limit is matched

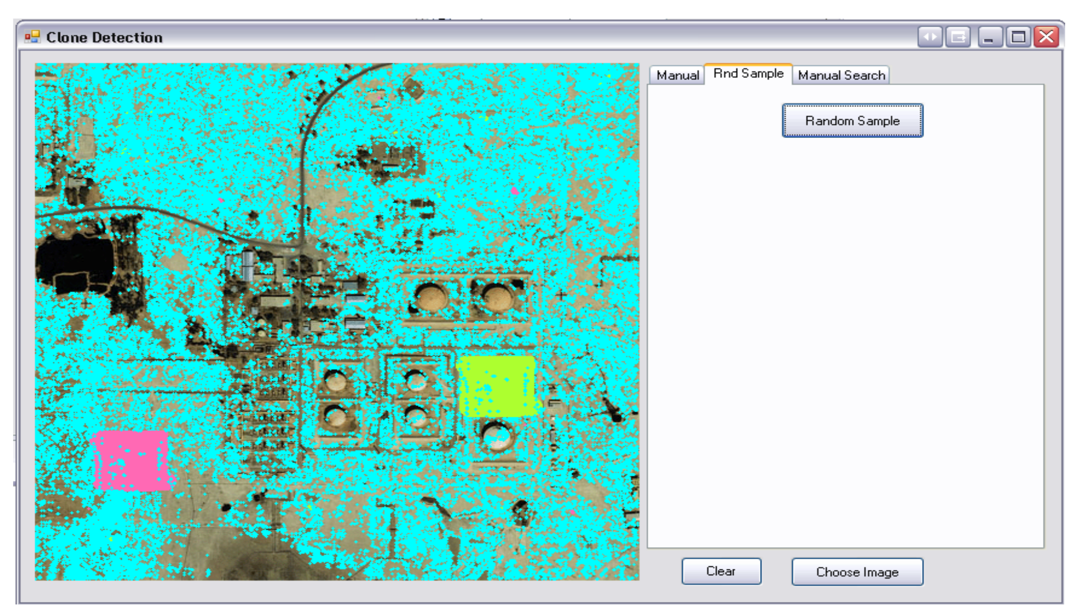


Figure 3(top): A screen shot of the tool being used to analyze a satellite image of Iraq. Figure 4(left): Original image of Rome, Italy. Figure 5(right): Tampered image of Rome, Italy. With the coliseum cloned over.

RESULTS

Evaluation of the technique

In order to evaluate the success of our technique we ran our program on an image bank of twenty images. These images consisted of 10 images that had no cloned regions and 10 images that had a cloned region of some type.

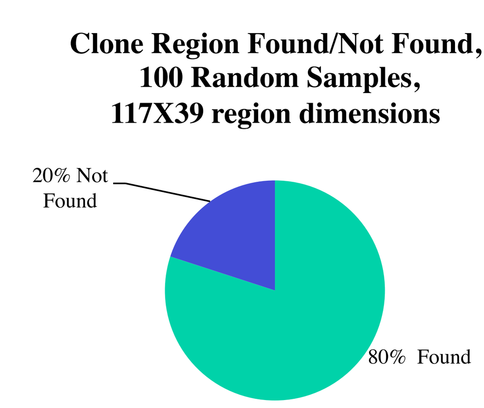
The images were chosen to provide a variety of backgrounds for clones and a variety of cloned objects of different sizes.

False Positives

Because the technique we use does not fully eliminate a person from the determination of whether or not a cloned region exists, false positives were counted if a reasonable person would assume with little or no working knowledge of the tool's algorithm that the results they were looking at could be a cloned region when in fact there wasn't any such tampering done to the image in that region.

Run Number	Clone Region Flagged	Clone Region Present at flagged location	Number of pixels sampled	Sized of cloned region
1	N	N/A	10	117X39
2	N	N/A	10	117X39
3	Y	Y	10	117X39
4	Y	Y	10	117X39
5	N	N/A	10	117X39
6	N	N/A	10	117X39
7	Y	N/A	10	117X39
8	N	N/A	10	117X39
9	Y	N/A	10	117X39
10	Y	N/A	10	117X39

Table 1: Example of data collected during testing. Taken from image Bejing2 which contained a cloned region.



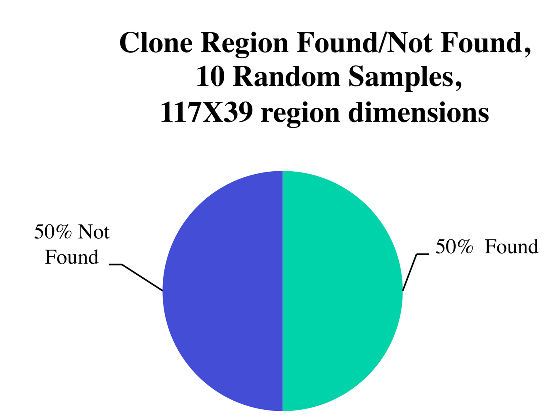


Figure 6(left) and 7 (right): Shows the results of the our technique when run on the same picture with a different number of samples taken.

CONCLUSIONS

Results are promising as the cloned regions of the images were found more than eighty percent of the time using a sample size of one-hundred randomly generated pixels.

Major factors that influenced results were:

1. The size of the cloned region
2. Number of pixels sampled
3. Presence of any regions in the image consisting of a consistent hue.

All of these results were expected. It makes sense from a probabilistic standpoint that the larger the cloned region is the more likely a randomly generated pixel is to lie within its borders. Similarly the more chances we give ourselves to find a cloned region, by choosing more pixels, the more likely it is that we find the region. Finally because we examine the color values of each pixel it is not surprising that areas which have a large area of consistent color, for instance bodies of water, tend to throw off our approach.

Suggested Refinements to our methods

1. Through further collection of data optimize the number of pixels searched as a function of the image size.
2. Create a method for declaring portions of the image off limits to cut down on false positives generated by water or other similar image textures.

Figure 8: Beijing with a cloned region corresponds to the data to the left.

ACKNOWLEDGEMENTS

Dr Dino Schweitzer, our mentoring professor, for providing invaluable advice, direction, and encouragement during the course of our research.