ABSTRACT

There are numerous types of documents which are difficult to scan or capture in a single pass due to their physical size or the size of their content. One possible solution that has been proposed is mosaicing multiple overlapping images to capture the complete document. In this paper, we present a novel Graphcut-based document image mosaicing method which seeks to overcome the known limitations of the previous approaches. First, our method does not require any prior knowledge of the content of the given document images, making it more widely applicable and robust. Second, information regarding the geometrical disposition between the overlapping images is exploited to minimize the errors at the boundary regions. Third, our method incorporates a sharpness measure which induces cut generation in a way that results in the mosaic including the sharpest pixels. Our method is shown to outperform previous methods, both quantitatively and qualitatively.

Index Terms— document, image mosaicing, panorama, Graphcuts

1. INTRODUCTION

In the field of document image analysis, image mosaicing has received attention as mobile devices with low cost built-in cameras are used to image printed materials. The idea of acquiring a single, high-quality, digital copy of a document from multiple overlapping shots has become very attractive, especially for documents which are difficult to scan or capture in a single pass. Some examples of such documents are shown in Figure 1 including long receipts, posters on display, and framed documents.

Numerous approaches were introduced which address the general issue of image mosaicing which are now even built into popular commercial software [1]. Although they seem to perform well on natural scene images, they show unsatisfactory results on document images. Unlike scene image mosaics where discontinuities are less noticeable, document images show noticeable errors because most of the content is small and very high contrast.

Figure 2 depicts examples of document image mosaics using two state-of-the-art scene image mosaicing approaches: AutoStitch [1] and iPhone5s built-in panorama. Figure 2(a) shows regions where the same texts appear twice with a slight offset. This is typically referred to as “ghosting”. Figure 2(b) illustrates another erroneous result where contents are missing in the mosaic. Such artifacts are caused by two major components of a general image mosaicing process: image registration and image blending. The registration attempts to properly align the overlapping images, while image blending is responsible for compositing the images as naturally as possible.

Previous work can broadly be categorized into two groups based on which major component (registration or blending) they address. Most of the approaches [2–8] focus on enhancing the registration process. In early approaches [2, 3], registration between overlapping images were estimated using methods such as image pyramid, image correlation or Least Median of Squares. These approaches target planar registration, typical of scanned documents. In [4], a sliding window registration method was introduced, but is time consuming and only applicable for binary images. Kasar et. al.
[5, 6] began using feature descriptor-based registration methods. In [5], the Harris corner detector and the discrete cosine transform feature descriptors were exploited, while [6] employed angular radial transform for the description of each connected component for registration. As mobile devices became more popular, a mobile-based, user-interactive mosaicking scheme [9] was introduced which incorporated SIFT features and RANSAC-based homography estimation. Most recently, two methods [7, 8] were proposed which focus on compensating for perspective distortion of the overlapping portions of documents.

We note that, most of the approaches addressing the registration problem [2, 4–8], adhere to using the conventional alpha-blending (weighted averaging). Although it is not explicitly stated in [2, 4, 7, 8], we presume that they have used alpha-blending by carefully inspecting their experimental results.

Instead of focusing on the registration problem, Liang et al. [10, 11] addressed the blending problem by using "selective" image blending. The method was developed to handle text content, and thus performs binary morphology and word-level segmentation. It is likely to perform poorly when dealing with complex figures, tables or text with different sizes. Even if the given document image includes uniformly sized characters, words might appear jagged in the mosaicked image.

In this paper, we address the limitations introduced in the previous approaches by using a sharpness-aware document mosaicking based on Graphcuts performed at the pixel level. The contributions of our method are as follows. First, Graphcut-based blending method is a novel method which effectively stitches two overlapping images without requiring any prior knowledge of the document, thus being more robust and widely applicable. Second, boundary constraints are imposed which minimize discrepancy between overlapping and non-overlapping regions. Third, we incorporate a sharpness measure which promotes cuts which favor a mosaiced image with sharper pixels when blending the overlapping images.

2. OVERALL DOCUMENT MOSAICING APPROACH

Although the novelty of our method is primarily in the image blending step, we briefly summarize the overall framework for completeness, and additional detail can be found in [9].

The mosaicing process begins with the capture of a portion of the document with a user interactive approach. Motion of the mobile device is estimated in real-time and the user is notified when to move and when to stop while scanning. The result is a series of images suitable for mosaicing.

Once the images are captured, scale and rotation invariant SIFT [12] features are extracted and matched. Matched features are then used to estimate the homography, or perspective projection, between pairs of overlapping images. Since there may exist outliers in the feature correspondences, we employ RANSAC for completeness, and additional detail can be found in [9].

We note that, most of the approaches addressing the registration problem [2, 4–8], adhere to using the conventional alpha-blending (weighted averaging). Although it is not explicitly stated in [2, 4, 7, 8], we presume that they have used alpha-blending by carefully inspecting their experimental results.

Instead of focusing on the registration problem, Liang et al. [10, 11] addressed the blending problem by using "selective" image blending. The method was developed to handle text content, and thus performs binary morphology and word-level segmentation. It is likely to perform poorly when dealing with complex figures, tables or text with different sizes. Even if the given document image includes uniformly sized characters, words might appear jagged in the mosaicked image.

In this paper, we address the limitations introduced in the previous approaches by using a sharpness-aware document mosaicking based on Graphcuts performed at the pixel level. The contributions of our method are as follows. First, Graphcut-based blending method is a novel method which effectively stitches two overlapping images without requiring any prior knowledge of the document, thus being more robust and widely applicable. Second, boundary constraints are imposed which minimize discrepancy between overlapping and non-overlapping regions. Third, we incorporate a sharpness measure which promotes cuts which favor a mosaiced image with sharper pixels when blending the overlapping images.

3. GRAPHCUT-BASED BLENDING

As mentioned previously, the phenomenon of the same content appearing twice with a slight offset, referred to as the ghosting artifact, is caused by alpha-blending (weighted averaging) the two overlapping images when the homography estimation contains errors. It is very difficult to have zero error in the homography estimation throughout the overlapping region. Thus, the proposed method seeks to eliminate such ghosting artifacts by using a Graphcut-based blending scheme which performs well even when slight registration error exists. The proposed method also has advantages over the selective image blending [10, 11] in that it does not require any segmentation.

Our method is capable of acquiring a cut line where two overlapping images can be stitched together like two pieces of puzzle. Since documents tend to have empty space where text or other contents like figures or tables are not present, it is desirable for the cut to be generated in those empty regions as shown in Figure 3.

3.1. Boundary Constraints

This problem can be viewed as a 2-class labeling problem where each of the labels indicates which of the two images, pixels are being copied from. The energy function $E$ which is being minimized in solving this labeling problem is represented by the sum of two terms: a smoothness term $\sum_{p,q} V_{p,q} (f_p, f_q)$ and a data term $\sum_{p} D_p (f_p)$, as shown in (1). The objective is to find a labeling $f$ that labels each pixel $p \in P$ as $f_p$.

$$E(f) = \sum_{(p,q) \in N} V_{p,q}(f_p, f_q) + \sum_{p \in P} D_p(f_p) \quad (1)$$

Fig. 3: dynamic programming based horizontal cut blending
The smoothness term is the sum of the penalty $V_{p,q}$ for all the pairs $(p,q)$ included in $N$, where $N$, $f_p$, $f_q$, indicate the set of neighboring pairs of pixels, label of pixel $p$, and label of pixel $q$, respectively. It can be described as the penalty imposed on the edge between pixel $p$ and $q$, whenever a cut is being made. The data term is the sum of the penalty $D_p$ for all the pixels in $F$. $D_p$ measures the penalty imposed on pixel $p$ when $p$ is labeled as $f_p$. A detailed explanation of the energy function and the Graphcut algorithm can be found in [14–16].

In our method, we use the following equation [17] as the smoothness term, which is defined for edges between every pair of neighboring pixels in the overlapping region

$$V_{p,q}(p, q, A, B) = |A(p) - B(p)| + |A(q) - B(q)|,$$

where $p$ and $q$ are the neighboring pixel locations in images $A$ and $B$.

For the data term in (1), we have incorporated two different terms, a boundary constraint and a sharpness measure, to guide the cut to minimize the discrepancy where the overlapped regions meet with the non-overlapping regions, and to favor the sharper image.

Our initial idea was to simply acquire a horizontal cut by using a method in [18]. This approach performs well in generating a seamless mosaic near the cut. However, a considerable number of discrepancies appear as shown in the dotted circular region of Figure 3.

In order to mosaic the two images with minimum discrepancies where the overlapped and non-overlapped regions meet, we have employed hard-constraints to constrain which image the boundary pixels are copied from. We have adaptively applied one of six different hard-constraints determined by the geometrical disposition of the two overlapping images shown in Figure 4. This can also be viewed as designating the locations of the two end points, $X$ and $Y$, of the cut being made within the overlapping region.

All the pixels located on the red dashed boundary line in Figure 4, are copied from image A, by setting $D_p(A) = 0$, $D_p(B) = \infty$. In the same way, pixels on the blue dotted boundary will be set with the data terms of $D_p(A) = \infty$, $D_p(B) = 0$.

### 3.2. Incorporating sharpness

The sharpness measure is also incorporated into the data term of the energy function. This, in turn, penalizes the blurred pixels with higher cost and the sharper pixels with lower cost when computing the energy function in (1).

The sharpness measure is computed for every pixel location within the overlapping region of the two images using a method introduced in [19] which is designed to estimate sharpness for documents or scenes. For the proposed method, the penalty value, $D_p(A)$ or $D_p(B)$ for each of the pixels in the overlapping region is controlled by the difference of the sharpness of the two images as shown in (3). $\gamma_{pA}$ and $\gamma_{pB}$ are the sharpness value of image A and B, respectively, computed at the pixel location $p$.

Let $\delta = |\gamma_{pA} - \gamma_{pB}|$

If $\gamma_{pA} \geq \gamma_{pB}$, then $D_p(A) = -\delta$ and $D_p(B) = \delta$ (3)

Else, $D_p(A) = \delta$ and $D_p(B) = -\delta$

Thus, the Graphcut favors the pixels with higher sharpness, which guides the cut so that regions with sharper pixels are included in the final mosaiced image.

### 4. EXPERIMENTS

#### 4.1. Dataset

To the best of our knowledge, there are no publicly available datasets for document image mosaicing. Thus, we have constructed a dataset where each session is comprised of two partially overlapping shots of a document using the camera on the iPhone5s. The images were captured with the resolution of $3264(w) \times 2448(h)$ in a reasonably lit, indoor environment.

Ten different documents were selected so that the method could be tested on not only the text lines but also on other types of frequently appearing contents such as equations, graphs, pictures, and tables. For each document, 6 sessions were captured, for a total of 60 sessions. The images in a session may have no blur or blur in one of the two images. Note that the blur is added to the dataset for the purpose of verifying the performance of sharpness-aware approach.

#### 4.2. Performance Comparison

Our experiments compare alpha-blending and selective blending to our method using our dataset. As the target objects for the mosaicing are documents which typically include text contents, OCR performance was used as a measure for quantitative performance comparison. Character and word level OCR accuracy were obtained using the OCR Frontiers Toolkit [20]. Table 1 shows that our method significantly outperforms the previous methods in both character and word-level OCR accuracy.
Table 1: OCR Performance Comparison

<table>
<thead>
<tr>
<th></th>
<th>Alpha-blend</th>
<th>Selective blend</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>character</td>
<td>72.31%</td>
<td>80.90%</td>
<td>83.70%</td>
</tr>
<tr>
<td>word</td>
<td>62.25%</td>
<td>71.98%</td>
<td>77.25%</td>
</tr>
</tbody>
</table>

Figure 5 shows the resulting mosaics of two documents generated by three different blending approaches. The gray regions indicate the overlaps between each pair of partially overlapping images. Observe that the ghosting artifacts clearly occur when using the alpha-blending as depicted in Figure 5(a) and (d). Meanwhile, the selective blending approach generates several different types of artifacts due to its binary morphology based procedures which incorporate dilation, thresholding and connected component labeling. In result, the mosaic shows unwanted fragments of contents as seen in Figure 5(b).

Moreover, neither of the previous approaches demonstrate the smooth transition between the overlapping and the non-overlapping regions, thus generating text or figures on the boundary with improper alignment. Such phenomenon can also be seen in Figure 5(b). The selective blending may even lose some of the contents which reside on the boundary. Notice that almost an entire text line is missing in Figure 5(e), while the same portion is properly recovered in Figure 5(f).

4.3. Limitations

Although the results show that the proposed method outperforms the previous approaches, Graphcut-based blending does not address registration errors. In other words, if the registration error is considerably large, the resulting mosaicked image may contain duplicate contents as shown in Figure 6. The red crosses and blue dots indicate the corresponding feature points. Note that in an ideal case, only one of the two matching features should appear.

We can see that such problems arise when the cut runs between the corresponding feature points. The relative positions of the cut and the matching features could be used do an automatic check on the mosaicking quality.

5. CONCLUSION

In this paper, we have proposed a novel method for document image mosaicing based on Graphcuts. We have focused on comparing the proposed blending approach with the two blending approaches used in previous literature. We have verified that our approach outperforms other approaches both qualitatively and quantitatively, showing its advantage in eliminating the ghosting effects and being capable of handling various types contents other than text. In the future, it will be worthwhile to devise a method which incorporates the registration error along with the matching correspondence information when running the Graphcut so that the result could effectively avoid having duplicate contents in possible erroneous cases.
6. ACKNOWLEDGEMENT

The partial support of this research by the US Government through NSF Awards IIS-0812111 and IIS-1262122, the Xerox Foundation is gratefully acknowledged.

7. REFERENCES


