ABSTRACT

Digital microscopy is a field that is becoming increasingly popular for tissue imaging. Most digital microscopes have a limited field-of-view, requiring the acquisition of multiple tiles across the tissue that are then stitched together in software. During the acquisition, however, the microscope may intermittently fail to focus correctly, which will result in out-of-focus images that no longer provide diagnostic value. While many approaches have been proposed to address this using absolute measures of focus quality, we here introduce a novel approach that instead operates on the overlap regions between acquired images. This provides a relative measure that is independent of tissue type and staining protocol. These automatic measures can then be used to identify failed images that need to be re-acquired. Our quantitative and qualitative results on large datasets demonstrate the accuracy and robustness of the approach.

Index Terms—Microscopy, focus quality assessment, classification, image analysis

1. INTRODUCTION

Digital microscopy enables the digitization of large tissue slides, which can then be stored, transferred, and analyzed electronically. A critical component of such systems is accurate autofocus to improve workflow and speed up processing, but failure of the autofocus system can lead to images that are of no diagnostic value and are generally frustrating to the user. Therefore, there is a large demand for developing focus quality assessment tools to identify such focus errors.

To enable high-throughput digital microscopy studies, many focus algorithms have been developed to automatically predict the focus plane for tissue imaging [1, 2, 3, 4, 5, 6]. Such approaches generally calculate some features on the image as the microscope depth is shifted, and the global maximum (or minimum) of a curve fitted to the calculated feature values is chosen as the true focal depth [7, 8]. In practice, because the scanning time is critical for microscopy applications, autofocusing is often performed by using a small image region, instead of the entire field of view. However, various factors such as tissue topologies or stage tilt can cause these strategies to fail. For example, if the region used for autofocusing is at a different depth from the majority of the tissue, the scanned image will be out-of-focus. Another focusing solution is to measure the focus quality directly on a single image. However, in general, this is challenging because of the large variety of tissue types and structures. For example, the image features of an in-focus image of smooth tissue regions can look like an out-of-focus image of textured regions.

To overcome these limitations, we here present a novel solution that is fundamentally different from previous approaches. Instead of operating on a single image or stack, we operate directly on the overlap regions between neighboring images. Because the underlying tissue is the same in the overlap regions, the only differences of the two images in these regions are due to the acquisition. Thus, the measures are independent of tissue type, staining protocols, and other such variables. In the overlap regions, we calculate a large number of features that are combined in such a way as to overcome the limitations of the individual features, and we measure feature differences between image pairs. Using this approach, neighboring focused images are similar, whereas out-of-focus images display large differences from their neighbors. Such failed images can then be targeted for improved image acquisition or can be displayed as feedback to the user.

2. METHODS

Our overlap difference measurement approach consists of feature extraction, selection, and normalization followed by feature difference calculation and comparison. In our approach, we assess the focus quality of an image by measuring the differences in the regions overlapping with other images. The neighboring images are already registered by first using the coordinates given by the stage followed by a fast registration step using a correlation metric. Because the borders are registered, the overlap regions of adjacent images should ideally be identical, although there are slight differences due to variations in image acquisition such as camera and transmittal noise. Still, when neighboring images are both in focus, the differences are minor compared to the differences when one
of the neighbors is out-of-focus. We measure these differences using an extensive set of features extracted on the overlap regions. To illustrate this, Figure 1 presents two images of different quality: the top image is out-of-focus and the bottom is in-focus. The two bar plots present, respectively, the measured feature differences between overlap regions and a histogram of these differences. All features showed significant negative changes (with a median of \(-25\%\)) indicating that the upper image is out-of-focus. The same conclusion can be drawn from the histogram, which shows that over 95% of the features changes were greater than 10%. On the other hand, Figure 2 shows a pair of tiles that are both in-focus, and the plots show only small difference in feature responses. The histogram shows that no feature change was greater than 10%. The differences are calculated for all adjacent tiles, and the signs of the differences indicate which tile is out-of-focus. Therefore, if one of the first tiles in the scan were out-of-focus, this would not result in poor quality for the entire scan because it would be detected when it is compared with an adjacent in-focus neighbor. In the following subsections, we present the details of the features and methods used for image comparisons.

2.1. Feature Extraction

The choice of features is an important step because many features have some extent of bias to the true focus [1]. To overcome such limitations, we utilize an extensive set of 78 features, many of which have been studied or suggested in previous microscope auto-focusing works [1, 2, 9]. These features include 24 local image contrast features [1], 20 features based on Canny edges [10], 15 image gradient and Laplacian features [11, 12, 1], 8 local image statistics features [1], and 11 wavelet features [9]. We implemented these features according to the referenced works.

2.2. Feature Selection

In order to improve classification ability, we implemented a method to select reliable features that are discriminative in distinguishing in-focus from out-of-focus images. Whereas the results of automatic feature selection methods vary depending on the training set, our method is adapted to the microscopy application by choosing features based on their behavior on a set of image stacks, each of which consists of microscope scans of the same tissue at different depths. Ideally, the responses of a discriminative feature at different depths should show a unique global peak (valley) at the focal plane and a smooth monotonic decrease (increase) on either side. Based on this requirement, the responses of each of the 78 features were inspected and classified into 3 categories: highly discriminative (19), moderately discriminative (25), and non-discriminative (34). Only the 19 highly discriminative features were then used for image comparison.

2.3. Feature Difference Measurement

While all of the features selected in Section 2.2 are effective for out-of-focus detection, they are not directly quantitatively comparable because their responses have different magnitudes and their responses change at different rates. We implemented a two-step normalization approach to normalize both the magnitude and rate of change of the features. This normalized feature difference value is described by

\[
\Delta f_i(I_1, I_2) = \frac{\min(k_i)}{k_i} \cdot \frac{f_i(I_1) - f_i(I_2)}{|f_i(I_1)| + |f_i(I_2)|}
\]  

(1)

where \(I_1\) and \(I_2\) are the overlapping regions in the two images, \(f_i\) is a feature value of feature vector \(\vec{f}\) with \(i\) as the index, and \(k_i\) is the slope of the feature difference relative to focus depth trained from image stacks.

The second fraction on the right implements the normalization of the magnitude. It converts the features from an absolute difference of feature responses to a relative measure by essentially normalizing relative to their magnitude. For a particular feature, as the feature values approach each other, this term goes to 0. At the other extreme, when the values are very different, the denominator term ensures that this term is bounded by \(-1\) or 1, depending on the sign of feature differences.

The first fraction on the right implements the normalization of the change rate. This is accomplished using a set of
stack images for training the change in feature differences versus distance in depth from the true focal plane. A linear function is fit to the response of each feature, and the slope \( k_i \) of each function is used to represent the sharpness of each feature response (all features pass through the origin at distance 0 by definition). The numerator of this fraction is set to the minimum of these slopes to ensure that the fraction is bounded between 0 and 1. Thus, with the combination of these normalization terms, the normalized feature differences are bounded between \(-1\) and \(1\). This normalized relative feature difference (referred to as “features difference” hereafter) can be directly compared among different types of features.

The feature differences are combined by taking the median of all included features differences, leading to the overall feature difference score, which can be compared with a threshold. Moreover, the sign of a feature difference indicates the relative focus quality of the compared images. For example, a positive sign means that image \( I_1 \) is better than image \( I_2 \). Thus, the magnitude and sign of the difference provides the desired measure of focus quality.

3. RESULTS

In this section, we demonstrate quantitatively and qualitatively the accuracy of the algorithms on various image stacks and whole tissue scans. We do not directly compare with methods that attempt to measure absolute image quality because our overlap region approach measures quality in a fundamentally different way.

We begin with a study of how the feature difference changes as a function of the distance in depth from the true focus plane. Ideally, the focus score would increase monotonically with an increase in depth from focus. For this experiment, we used a set of 38 image stacks, each of which contained 101 images scanned on both sides of the focus plane from \(-5 \mu m \) to \(5 \mu m \) with a step of \(0.1 \mu m \). Each image was of size \(2048 \times 2048 \) and was evenly divided into 64 regions of size \(256 \times 256 \). Features were extracted from each region and compared against those extracted from the same location in the focused image. For each stack, the feature differences measured at the same distances were averaged and are plotted in Figure 3 as blue curves. The average of all of these curves is plotted as the red line in the same figure.

This plot shows that the average of the feature differences increases monotonically as a function of the distance, which indicates that this metric is highly correlated with the focal depth of the tissue section.

The next experiment provides an intuitive understanding of how our feature differences correspond to human visual perception. Figure 4, which shows three image pairs with their feature differences, illustrates that the quality difference between the 25% pair can be easily perceived, whereas the quality difference becomes more subtle for smaller feature difference. Although a feature difference of 8% or less may be hardly perceptible to a non-expert, this could still be unacceptable to a pathologist. In this sense, the sensitivity of the proposed measure is very valuable for high-throughput digital microscopy.

In another experiment, we applied the proposed measure to the entire scans of various tissue slides. Here we show how the user can define an acceptable threshold for the feature difference so that the algorithm will report images that are beyond this threshold. Table 1 presents a summary of the reports for 8 different tissue scans representing nearly 1,100 tiles. For each of the scans, the threshold of the feature difference score from Section 2.3 was set at 5%; tiles with feature differences greater than this threshold are classified as out-of-focus. The number of tiles given in the second column represents only the tiles that include tissue regions; background tiles were automatically removed. The third column shows the percentage of all tiles that are considered out-of-focus. The last four columns show the assignments of the out-of-focus tiles into four categories based on the direction: since each tile is compared to its four neighbors, this results in a north (N), south (S), west (W), and east (E) score. From this summary, the user can decide whether the overall out-of-focus rate is acceptable or whether a re-acquisition is needed. Also, a trend in the directionality of the error from the table may reveal a systematic scan error. For example, for scans 7 and 8, the error in the “south” direction (marked red in the table) is significantly larger than other directions. This suggests that the slides might be tilted during the loading process, or there might be some optical problems in the microscope. Such information can be very useful to both the user and the microscope designer.

In the final experiment, we visualize the features differences as a color heat-map overlaid on the original scan. The colors of the heat-map represents the magnitude of the feature difference from the four directions. Since only tile borders with negative changes are of interest for out-of-focus analysis, the positive changes are discarded in computing the over-
Feature difference of 25%.

Feature difference of 12%.

Feature difference of 8%.

**Fig. 4.** Feature correspondence with visual perception. In each box, the top region is always poorer than the bottom, and the scores correspond closely with visual perception. While a change of 25% is obvious, 8% is barely perceptible.

**Table 1.** Quantitative out-of-focus detection results. For each of these datasets, the threshold for maximally allowed negative feature changes was set at 5%.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Tiles</th>
<th>Out-of-Focus Tiles (%)</th>
<th>N</th>
<th>S</th>
<th>W</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>62</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>52</td>
<td>17.3</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>226</td>
<td>3.54</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>224</td>
<td>1.34</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>57</td>
<td>10.5</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>52</td>
<td>1.92</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>209</td>
<td>21.1</td>
<td>5</td>
<td>35</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>209</td>
<td>8.13</td>
<td>2</td>
<td>13</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

**Fig. 5.** Whole tissue scan with quality score overlaid. The color bar indicates the value of the magnitude.

all magnitude. Figure 5 shows a visualization example for a whole slide, where each square in the image represents a tile of size $2048 \times 2048$. The heat-map overlay provides a convenient way to interact with the data. For example, a user can zoom in to the most out-of-focus tile and decide whether a reacquisition is required. In Figure 5, the maximum feature difference measured in the scan was about 5%, which suggests that all tiles are of good quality. A pathologist independently confirmed that all of the tiles in this image are in focus.

4. CONCLUSIONS

While many approaches have been designed for image focus detection, our approach introduces the new paradigm of measuring features in the overlap regions of acquired tiles rather than measuring absolute features on the images. This provides a relative measure that is robust to a large set of variables including tissue type and staining protocol. The results demonstrate the quantitative accuracy of the algorithms on a large variety of datasets, the sensitivity of the feature difference measure to the focus quality changes and its correlation with human visual perception, and an interactive manner of visualizing the results.

5. REFERENCES


