

The Preliminary Attempts to Quantify the Three-dimensional Details of Document Surfaces with Reflectance Transformation Imaging

Ning Liu¹, Lichao Zhang²

¹Forensic Science Department, Jiangsu Police Institute, Jiangsu, China 210031

²Department of Computer Information and Cyber Security, Jiangsu Police Institute, Jiangsu, China 210031

Reflectance Transformation Imaging (RTI) is a computational photographic method that captures a subject's surface shape details and enables the interactive re-lighting of the subject from any light direction. Multiple rendering modes of this computational photographic method could be very helpful for the analysis of morphological characteristics, among which the normal visualization mode provides the three-dimensional details of the surfaces of documents. However, since there might be visual illusions among different examiners, analyzing three-dimensional details just by observing the normal maps seems incomplete. With the method developed by the authors, the pen pressures of signature samples were experimented on by quantifying the three-dimensional surfaces with the normal data derived from the normal maps. By comparison, the 3D data derived from the normal maps of the signatures were similar to those generated with Confocal Laser Scanning Microscopy. Normal errors were discussed via the quantification method as well. For the inherent error of the RTI technique, the methods of error correction were developed by the authors.

Keywords: Reflectance Transformation Imaging, Quantification of Normal Maps, Error Correction, Pen Pressure

Introduction

Reflectance Transformation Imaging (RTI) was first developed for use in association with Polynomial Texture Mapping (PTM) by Malzbender et al in 2001.^{1,2,3} It is a computational photographic method that captures a subject's surface shape details and enables the interactive re-lighting of the subject from any light direction.⁴ RTI images are created from the information derived from multiple digital photographs of a subject shot from a fixed camera position while under different illumination conditions. By some computational ways, per-pixel surface normal of the subject (three-directional information) are extracted from the two-directional photographs.

After having conducted several experiments on RTI applications in signature examination, tampered document examination, printer iden-

tification, paper examination, and so forth, RTI was found as an effective method for questioned document examiners. The rendering modes of this computational photographic method could be very helpful for the analysis of morphological characteristics, among which the normal visualization mode provides the three-dimensional (3D) details of document surfaces that are not disclosed under direct empirical examination of physical objects. RTI offers a significant method for exhibiting pen pressure dynamic (see Figure1). However, empirically analyzing 3D details just by observing normal maps seems incomplete, because there might be visual illusions among different examiners. Therefore, the authors have done a preliminary study to quantify and measure 3D data of the normal maps, conducting experiments on the pen pressure of signature samples.

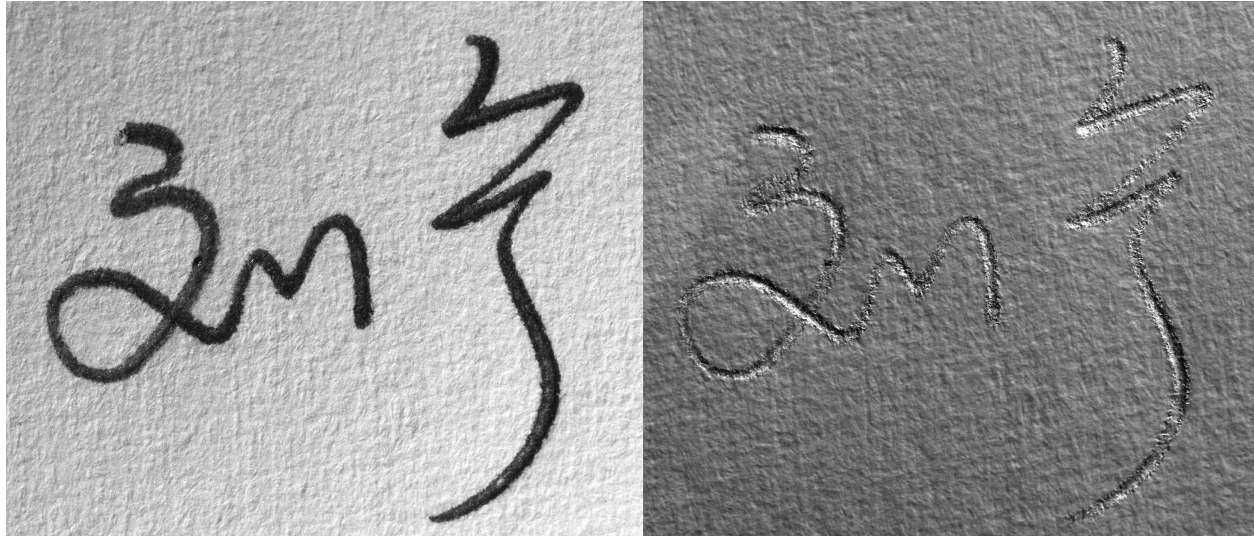


Figure 1. One of the signature sample in this study. The left is a common digital photograph; the right is the RTI normal visualization image (converted into grayscale and enhanced in contrast) of the left signature.

The Technologies of 3D Surface Reconstruction

The normal maps are two-dimensional pseudo-color pictures with three-dimensional data: per-pixel surface normal. In a normal map, the R, G, and B components of a pixel represent the X, Y, and Z coordinates, respectively, of the surface normal on this pixel. X and Y are in the range $[-1, 1]$ mapped to $[0, 255]$, while Z is in the range $[0, 1]$ mapped to $[128, 255]$, assigned into three color channels. Therefore, a RTI normal map depicts a real 3D surface, different from those fake 3D renderings based on grayscale values of a single photograph.

Many different technologies can be used to preform 3D reconstruction; each technology comes with its own limitations and advantages. A 3D laser scanner captures 3D information by collecting the height/depth data. Hammer et al (2002) found that PTM representations gave the better results than laser scanning for the specimens with very low surface reliefs. They noted that spatial resolution was compromised by computation of geometric surface normal from laser point cloud data, because of the convolution with a kernel having a spatial extent, whereas for PTM the normal estimation for each pixel is performed independently.⁵ Furthermore, 3D laser scanning systems are not suitable for describing the surface of a document because of their accuracies. For example, 3D laser scanners with triangulation mechanism which range finders are on the order of tens of micrometers. However, the depths of

handwriting indentations on a paper document are around $30\mu\text{m}$. This requires much higher accuracy on Z-axis of the surface reference frame to capture and measure 3D details of handwriting.

Metallographic microscopy can reach this requirement. With certain software, successive 2D slices captured at different heights can be 'stacked together' to produce a 3D representation. Confocal Laser Scanning Microscopy (CLSM) is an optical imaging technique for improving optical resolution and contrast of a micrograph by means of adding a spatial pinhole placed at the confocal plane of the lens to eliminate the out-of-focus light. With CLSM, one can obtain a more accurate 'z-stack' to create a 3D image.

3D reconstruction from multiple images is mainly classified into two categories: stereo-photogrammetry and photometric stereo. Stereo-photogrammetry captures surface data of an object by estimating the spatial coordinates of points on the object from a number of photographs taken from different positions, whereas, photometric stereo is a technique for estimating the surface normals by picturing the fixed subject under different lighting conditions with a fixed camera.

Photometric stereo technique is based on the fact that the amount of light reflected from a surface is dependent on the orientation of the surface in relation to the light and the observer, and is also dependent on the reflectance property and the condition of the surface. Lucia Docuscan System (the product of Laboratory Imaging, Czech)

was designed to display “texture free topography”, adopting a photometric stereo technique.⁶ It is equipped with LED lights in a circle around its lens. As a matter of fact, with photometric stereo, three successive images from three different lighting directions are adequate to estimate the surface normals. However, the more lighting positions and more photos in a stack, the more reliable normals are estimated.

RTI is one of photometric stereo techniques, but with dozens of different lighting positions distributed on a hemisphere. In this study, Highlight-based RTI was adopted. For Highlight RTI, lighting information from the images is mathematically synthesized by recovering the lighting direction from the specular highlights on a black sphere that is included in the field of view.⁴ Then, Hemispherical Harmonics (HSH) fitter or PTM fitter is applied to determine the coefficients of the reflectance function on each pixel, from which per-pixel surface normal is extracted.^{1, 7} Although the albedos of materials are important and can be quantified as well, in this paper, the authors only tried to discuss the quantification of 3D data related to surface normal. Elfarargy et al (2013) presented a method to derive a 3D point cloud of an object from the height map generated from a RTI normal map.⁸ In this paper, the authors developed a quantification method and attempted to investigate the errors by quantifying surface normals of handwriting indentations on documents, since error control is a prerequisite for quantification.

Methods and Materials

1. Capturing and Processing for RTI

One of the authors signed her autograph several times on a sheet of copier paper as the tested samples. In this study, highlight-based method was used to capture image stacks. Canon EOS 700D (1800-megapixel resolution) with Canon EF-S 60mm f/2.8 Macro Lens and Canon EOS 5Ds (5000-megapixel resolution) with Canon MP-E 65mm f/2.8 Macro Lens were used to capture the images. For each image stack, the signatures on the sheet of paper were photographed with a 2mm diameter black specular sphere placed at the center of the field of view, and a scale was included.

To acquire the image stacks more efficiently, the authors developed an automatic acquisition device that formed a hemispherical dome of light encircling the target signature. The camera was fixed onto a copy stand, facing down with the light axis of camera lens vertical to the document plane (see Figure2). The auto-acquisition device has three arc-shaped light frames with three different sizes, which are exchangeable for different sizes of object distances. On each of the light frames, six LED lights were mounded toward the center of the dome, which was also the center of the field of view. The acquisition device fired a different LED each time and synchronized with the camera shutter. When the device operated, it rotated the light frame step by step, and

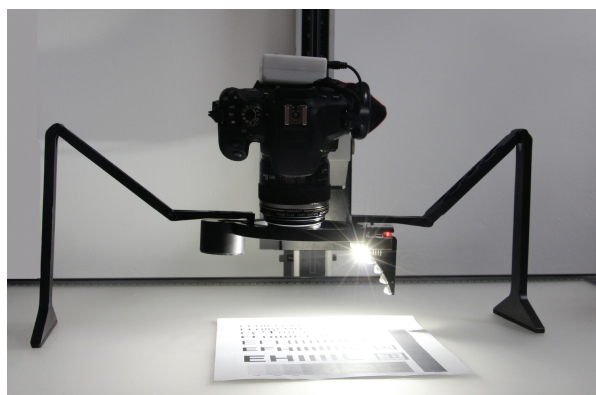


Figure 2. The automatic acquisition system in this study.

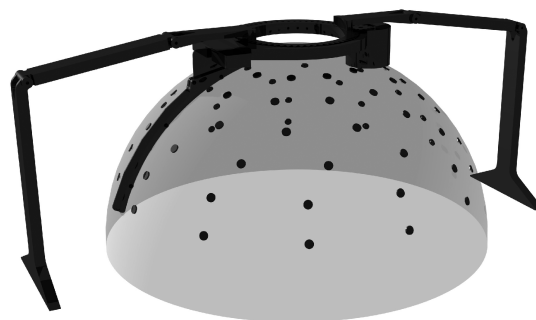


Figure 3. The schematic layout of the light dome formed by the acquisition device developed by the authors.

the LED lights turned on and off in sequence to form a light dome (see Figure 3) with the radius of 10cm, 20cm, and 25cm, respectively. In addition, a flashlight (920 Lumen, ONE THE ROAD M3) was manually handheld to distribute lights on a dome of 80cm radius. Each image stack was set to record 48 images.

All photographed images were stored both in RAW format and largest JPEG format. If needed, the images in RAW format were converted into DNG (embed original Raw file) and then into JPEG (highest quality) with Adobe Camera Raw 9.0 to obtain higher image quality. Each captured image stack was input into RTIbuilder v2.0.2 and fitted to both a 2nd order HSH file and a LRGB PTM file. The normal maps of the signature samples were generated through the Normal Visualization mode of RTIviewer v1.1. The software RTIbuilder v2.0.2 and RTIviewer v1.1 can be freely downloaded on the website of the Cultural Heritage Imaging.²

2. The Self-Developed Quantification Method

Assuming each pixel is a small square plane, the image of a signature is composed by these small planes that have different orientations respectively. RTI technique applies normal mapping to visualize the surface normal of a subject. The upper left picture in Figure 4, Figure 5, and the upper picture of Figure 11 are examples of normal map. To process normal mapping, the coordinate values (x , y , and z) of the normal vector on a pixel plane were represented by the intensity values of the three-color channels (R, G, and B). Therefore, a normal map carries the per-pixel normal data, which can be retrieved. The spatial coordinates (x , y , and z) of normal on each pixel were recovered by reversing the process of normal mapping.

Using the program developed by the authors with Eclipse Integrated Development Environment, surface profiling was conducted as a pen pressure quantification method. Generally speaking, we first put a section plane orthogonal

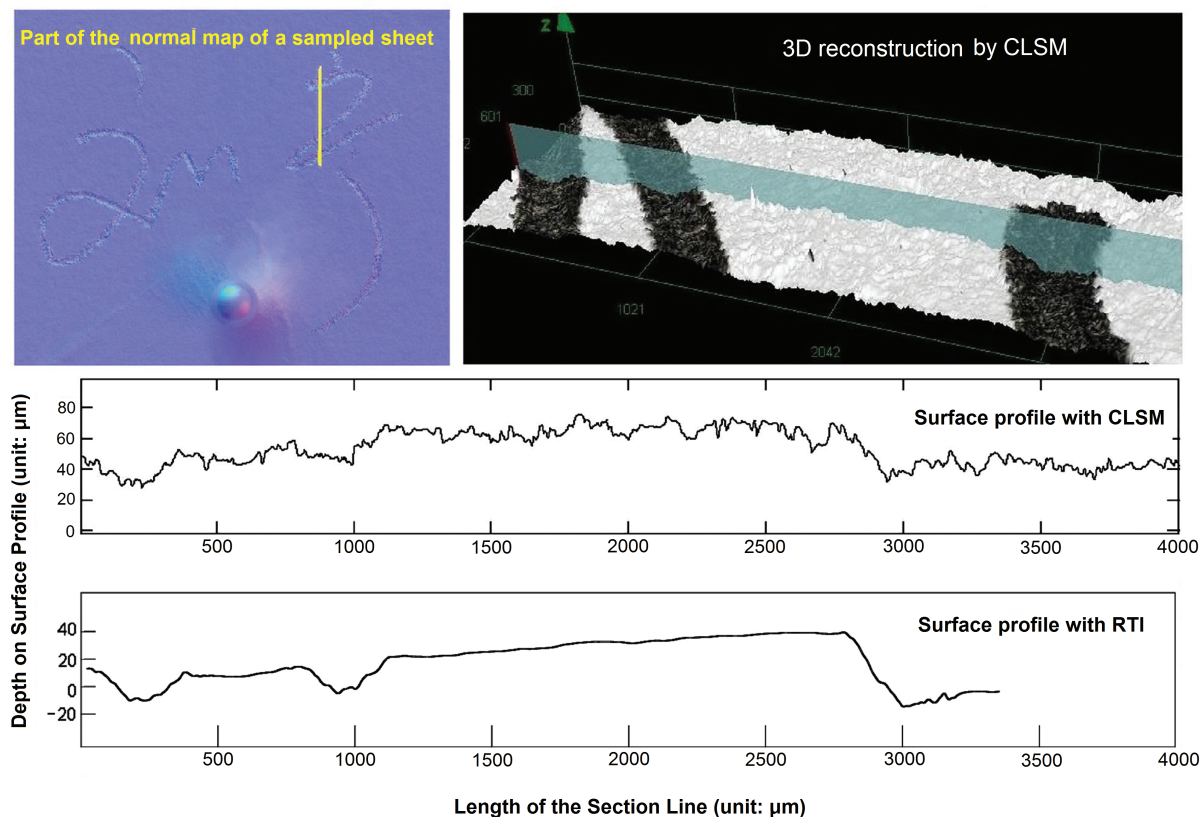


Figure 4. The comparison between one of the surface profiles derived from RTI and that generated from CLSM. The upper left picture is the normal map of the RTI representation captured with 80cm light distance, in which the yellow line is the section line profiled in the two graphs below. The upper right picture displays the 3D reconstruction of the section line with the CLSM. The ratio of the XY axis of the graphs is 1:5 for enhancing the visual effect of depth.

to the image plane (normally set a document as the horizontal plane) and formed a section line that crossed the selected strokes of the signature; next, every normal vector on the section line (1 pixel width) was projected onto the inserted perpendicular plane; then the straight lines perpendicular to the projected vectors were made and connected successively; finally, the surface profile was plotted, which shows the depths of the interested strokes.

To confirm the validity of the quantification via RTI, the same parts of the signature were also measured with an Olympus LEXT OLS4100 confocal laser scanning microscope with the 20X objective lens. In comparison, the 3D data derived from the normal maps were similar to those generated by CLSM (see Figure 4). The captured images with higher resolution produced more detailed surface topography, therefore, created more detailed profiles. It is well accepted that confocal laser scanning microscopic images depict 3D surface texture with high resolution and accurate measurements. However, due to the limited field, it took about 40 minutes to capture a single stroke even though the CLSM device was equipped with a software to preform automatic image stitching. In the authors' opinion, quantifying surface of a document via RTI is promising for the Questioned Documents field.

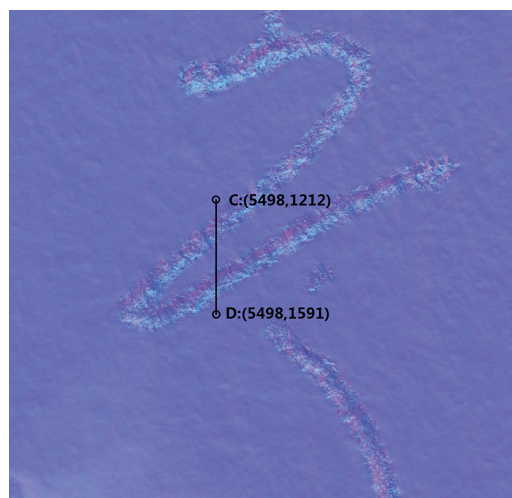


Figure 5. The part of the normal map of a signature sample. The line CD was the same section line profiled in Figure 6, 7, and 8.

Results

1. Normal Error Control

Since RTI mathematically estimates surface normals from a set of digital photographs, error control must be considered before quantification. The quantification program developed by the authors was also useful for quantification of normal errors by the comparison between the normal data from the different RTI representations of the same subject.

There are various causes of normal bias during every phase in the entire course of generating a RTI normal map. The errors introduced by camera lens distortions are well known. These errors can be minimized by placing the subject in the middle of the field of view as much as possible.

Some causes of normal errors, which have been found by the RTI developers, are indicated in the "Guide to Highlight Image Capture", including variable intensity of the lights, variations of camera settings during an image stack capturing, fewer images in a stack, false-registration of the image stack, out-of-focus, highlight and shadow, improper exposure, uneven light source distribution, non-uniform illumination distribution across image area, compression errors from the JPEG image encoding, the algorithm of the light position detection, and the algorithm for nor-

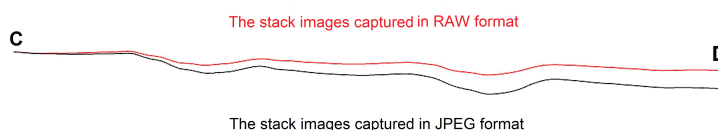


Figure 6. An example of the quantitative comparison between the RTI representations of the same signature captured (with the 80cm light distance) in different formats.

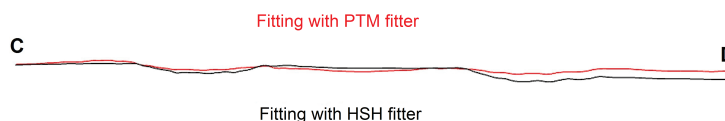


Figure 7. An example of the quantitative comparison between the RTI representations of the same image stack (with the 80cm light distance) fitted with different fitters.

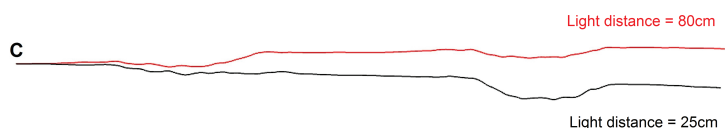


Figure 8. An example of the quantitative comparison between the RTI representations of the same signature captured with different light distances.

mal calculation, and so forth.⁹ For example, with the same acquisition settings, different qualities (more specifically, storage format and resolution) of the images resulted in variations in normal errors. In Figure 6, we can see the normal biases of the same signature captured at the same time but stored in different formats. The RTI user's guide recommends that storing the captured images in RAW format will help to minimize the compression errors caused by the JPEG compression encoding. The errors caused by the algorithms, however, cannot be controlled by average users. Figure 7 depicted the comparison of the differences between outcomes of HSH fitter and PTM fitter.

The normal biases were presented not only in the measurements for the depth of indentation, but also in the deformation of the cross section of paper. Figure 8 displays the deviation in the surface profile from the images captured with 25cm light distance, comparing with the representation of 80cm light distance. Figure 5 shows the section line on the signature sample, which is the same one profiled in Figure 6, Figure 7, and Figure 8.

Some of the factors during the capturing can be easily avoided. The auto-acquisition device and proper settings of camera helped to avoid or minimize many errors introduced during capturing. However, there are inherent normal errors of the RTI method, which are not mentioned in the RTI user's guide. One of them could significantly affect the result of normal calculation, in relation to the light distance or the radius of the light dome.

2. The Inherent Normal Error of RTI

Normal estimation via RTI lies on the per-pixel reflectance functions modeled from captured data, including the luminance of pixels under varying light source directions, the spatial coordinates, and the parameters encoding the directions of incident illumination.¹⁰ To construct the reflectance function for a particular pixel, the light source directions corresponding to the known intensities of the pixel have to be known as the inputs for fitting. With the Highlight-RTI method, the unknown direction of the incident illumination of each image is determined as the light vector of the specular sphere based on its highlight and other captured data. The light vector of the specular sphere will be considered as the lighting direction at each pixel within the whole image for fitting. Even with the method of capture us-

ing pre-known light positions, all the pixels in a same image share only one light vector.

Assuming the light vector detection is accurate for each image in a stack, the pixels within an image, have their own locations, away from the sphere in longer or shorter distances, thereby errors are introduced into the normal calculations. Figure 9 shows that the actual light directions on most pixels are not the same one on the sphere, but at certain angles with it. The authors call the normal error yielded by this kind of deviation as the inherent error of RTI method. The pixels farther away from the sphere will be calculated with more deviate light directions, therefore yield more biased normal results. Because of this normal deviation, every normal map of the documents in this study demonstrated a slightly convex cambered surface, rather than a flat paper (see Figure 11). If the light distance is long enough, the deviate angle approximately equals to zero and can be neglected. In Figure 9, it is demonstrated that the deviated angle ($\angle FAD$) caused by the shorter light distance is bigger than that ($\angle FAE$) caused by the longer light distance.

In addition, since the light source for RTI should be a point source, the shorter light distances result in the more variations of irradiance intensity for the same pixel due to the different light distances (the difference between the light distances of DA and GA is an example demonstrated in Figure 9), and consequently more errors of normal estimation. In our future study, the light intensity fluctuations due to short light-subject distances will be managed to reduce the resultant errors, especially for RTI representations captured with shorter light distances.

The experiments showed that the shorter light-to-subject distance contributed to the larger deviation of normal. However, a long light distance is impractical for capturing the subjects with small sizes (such as signatures), which need a macro lens with shorter object distance. One of the reasons is because the higher light positions cannot be used as the camera or the Lens cast shade on the subject. By comparison with CLSM, RTI representations derived from images captured under light distance at 80cm have fewer normal errors, see Figure 10.

3. Error Correction for the Inherent Error of RTI

Despite that the inherent error of RTI technique is not avoidable, they can be corrected with some

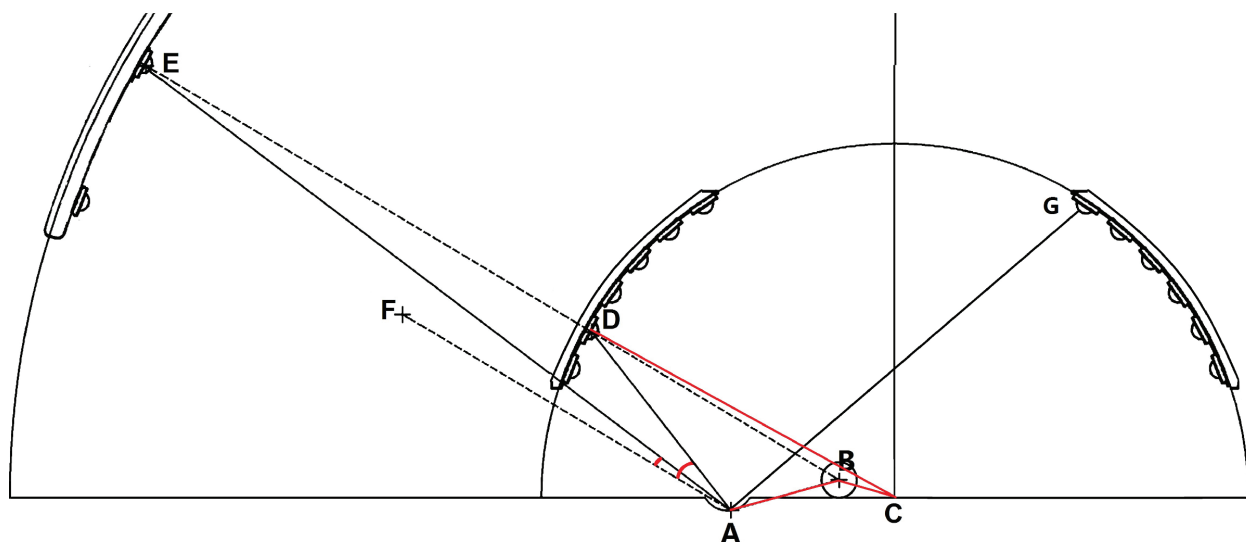


Figure 9. The schematic diagram of an inherent error of RTI. Point C is the center of the field. Point B is the center of the specular sphere. Point A is an arbitrary pixel of the subject. The actual light directions on the pixel A could be on the lines of DA, EA, and GA. When light position D or E is fired, the light vector estimated for the pixel A is parallel to the line DB/EB on which the light vector on the sphere is located, namely, on the line FA. Comparing between the different light distances, the deviated angle ($\angle FAD$) caused by the shorter one is bigger than that ($\angle FAE$) caused by the longer one.

mathematical methods. The authors developed two ways for the attempt of normal error correction: 1. Pre-correction refers to correcting the normal errors by modifying the light direction for each pixel within every image of the stack before the fitting, according to the distance between the

pixel and the center of the specular sphere. 2. Post-correction refers to correcting the normal errors by adjusting the angle of the normal vector of each pixel in the normal map according to the distance between the pixel and the center of the specular sphere.

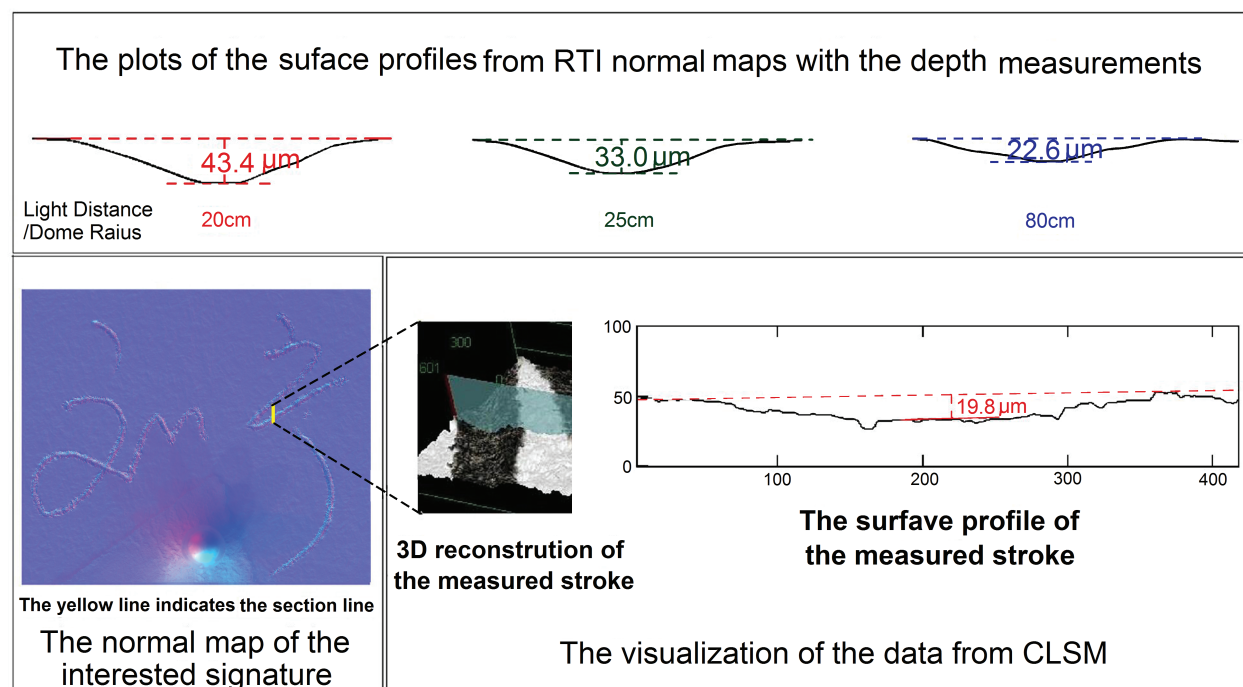


Figure 10. The comparisons of the measurements obtained from CLAM and the RTI normal maps.

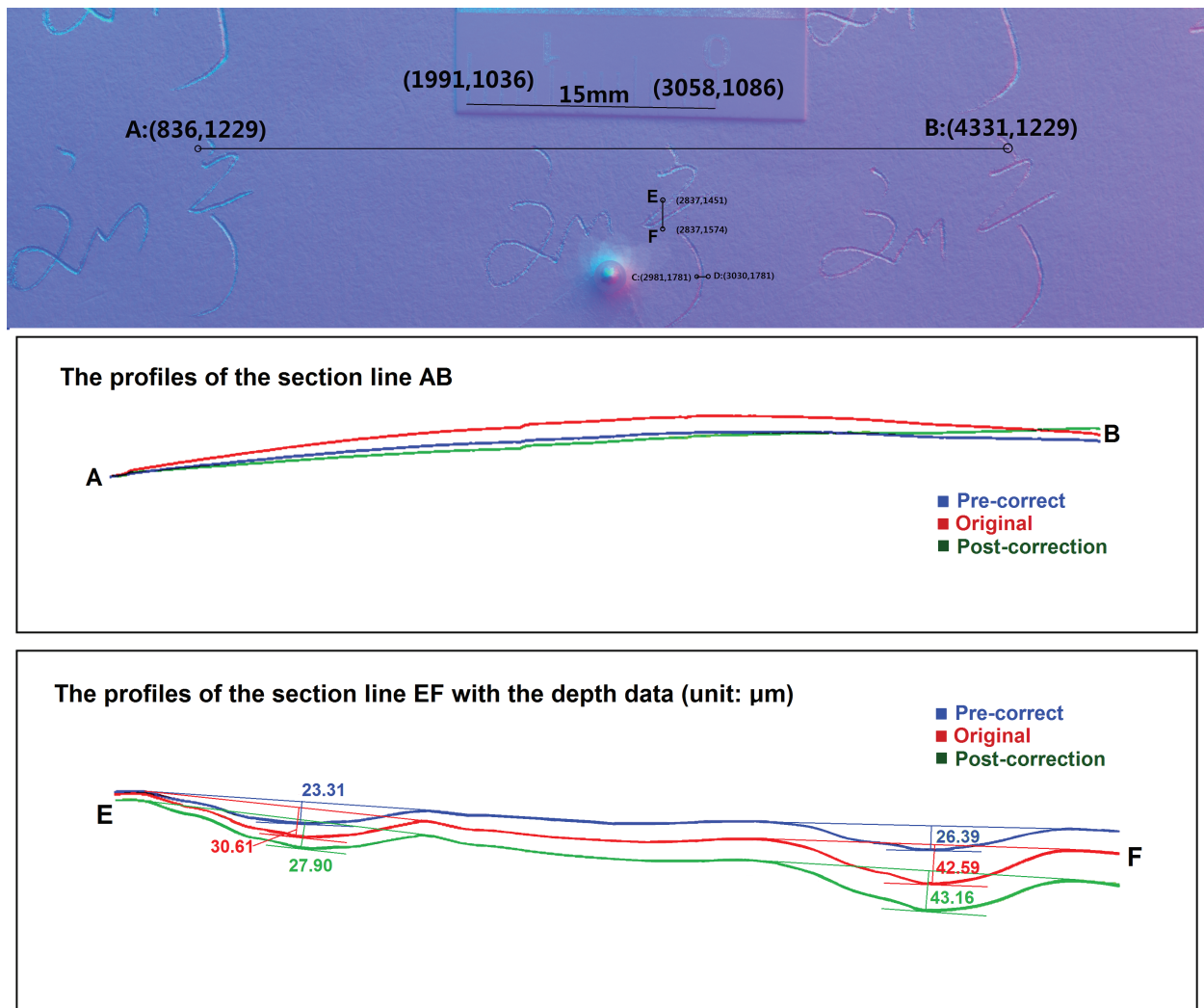


Figure 11. The quantitative comparison of pre- and post-correction ways. The upper picture is a normal map that was generated from the images taken with 25cm light distance and stored in RAW format.

As Figure 9 illustrated, the lengths of the light distances (the line CD and CE) and the radius of the sphere were known, the distances from the pixels to the center of the specular sphere (the line AB) were converted into the true distances of the subject based on the scale put inside the images, therefore, applying pre-correction could first obtain the real light vectors (along the line DA/EA) using geometrical computation, and then replace the light vectors on the sphere as the input data for fitting. The illustration of pre-correction can explain post-correction as well, however, the point D, E, and G in Figure 9, which represent the real light positions for pre-correction, indicate the virtual light positions for post-correction. When fitting without pre-correction, every pixel

of a subject was regarded as if it was on the point of the specular sphere, because the inputs of light direction were the same as those on the sphere. With post-correction, the coordinates of the corrected normal vectors were computed based on the known data and geometry. The new normal vectors were then normalized and replaced original ones in the normal maps.

Theoretically, pre-correction is supposed to reduce the inherent error more than post-correction does. Because the deviated light vectors might combine with other error sources during a RTI constructing, there might be more uncontrollable deviation of the estimated surface normals. Whereas, the pre-correction modifies the light directions before the fitting thus effectively avoids

the inherent errors. It was shown that the normal errors were corrected at certain extent with these two methods. The post-correction was better at flattening the surface (see the middle graph in Figure 11) but less correction on the details in the indentations, while the depth data were obviously modified by the pre-correction and closer to the data from CLAM (see the bottom graph in Figure 11).

Conclusion

In summary, a RTI normal map effectively reveals the whole range of pen pressure dynamic of handwriting that is significant and previously illegible. The experiments have shown that a quantification of document surfaces via RTI technique is promising. To achieve quantification of 3D features on a document, further studies need to be done to specify the normal errors in every aspect. This study has addressed that some situations could cause bigger normal errors than others. It was also shown that error control can be established if the error has a rule to follow. With further research, better error correction methods could be developed.

References

1. Malzbender T., Gelb D., Wolters H. (2001) Polynomial Texture Maps, Proceedings of ACM Siggraph. <http://www.hpl.hp.com/ptm/papers/PTM.pdf>
2. <http://culturalheritageimaging.org/Technologies/RTI>
3. http://en.wikipedia.org/wiki/Polynomial_texture_mapping
4. Mudge M., Malzbender T., Schroer C., Lum M. (2006) New Transformation Imaging Methods for Rock Art and Multiple-Viewpoint Display. The 7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage, VAST.
5. Hammer Ø., Bengtson S., Malzbender T., Gelb D. (2002) Imaging Fossils Using Reflectance Transformation and Interactive Manipulation of Virtual Light Sources, *Palaeontologia Electronica*, 5(4).
6. <http://www.forensic.cz/en/products/toolscan>
7. Wang O., Gunawardane P., Scher S., Davis J. (2009) Material Classification using BRDF Slices. 2009 IEEE Computer Vision and Pattern Recognition, pp. 2805-2811.
8. Elfargy M., Rizq A., & Rahswan M. (2013) 3D Surface Reconstruction Using Polynomial Texture Mapping. *Advances in Visual Computing. Lectures Notes in Computer Science*, 8033, pp. 353-362.
9. Guide to highlight image capture (Document version 2.0), Guide to highlight image processing (Document version 1.4 for RTIBuilder v. 2.0.2beta). http://culturalheritageimaging.org/What_We_Offer/Downloads
10. MacDonal L., Robson S. (2010) Polynomial Texture Mapping and 3D Representations. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII, Part 5, Commission V Symposium, Newcastle upon Tyne, UK, pp.422-427.