Imaging Buddhist Art with a Digital Large-Format Camera: A Field Study Report from the Dunhuang Caves

ZHENG LU and WEI LUO, Microsoft Research Asia
ZHIJUN SUN, Dunhuang Academy
MOSHE BEN-EZRA, Microsoft Research Asia
MICHAEL S. BROWN, National University of Singapore

This article describes recent field work undertaken by Microsoft Research Asia and the Dunhuang Academy to capture high-resolution images of Buddhist art at the UNESCO world heritage site, the *Mogao Caves*. This project is intended as a feasibility study examining the use of a digital large-format gigapixel camera to capture high-resolution images in a cultural heritage setting. In particular, we report on the current challenges faced by the Dunhuang Academy in their imaging efforts and how the use of a digital large-format camera can improve the quality of the imaging process while reducing time and effort. We also describe lessons learned from this field study as well as remaining challenges inherent to such projects.

Categories and Subject Descriptors: I.4.1 [**Image Processing and Computer Vision**]: Digitalization and Image Capture—*Scanning*; I.3.3 [**Computer Graphics**]: Picture/Image Generation—*Digitizing and Scanning*; J.5 [**Arts and Humanities**]: Fine arts

General Terms: Documentation, Experimentation

Additional Key Words and Phrases: Digitalization, 2D imaging, large-format camera, field study, Dunhuang, Mogao Caves

ACM Reference Format:

Lu, Z., Luo, W., Sun, Z., Ben-Ezra, M., and Brown, M.S. 2012. Imaging Buddhist art with a digital large-format camera: A field study report from the Dunhuang caves. ACM J. Comput. Cult. Herit. 5, 3, Article 9 (October 2012), 10 pages.
DOI = 10.1145/2362402.2362403 http://doi.acm.org/10.1145/2362402.2362403

1. INTRODUCTION AND MOTIVATION

The *Mogao Caves*, often referred to as the Dunhuang Caves, consist of a network of over 400 caves of various sizes housing some of China's oldest Buddhist artwork. The artwork, which spans a 1000-year period, includes statues, religious texts, and wall paintings that cover over 42,000m^2. The Mogao

Zheng Lu is currently affiliated with National University of Singapore and Wei Luo is currently affiliated with the Chinese University of Hong Kong.
Authors' addresses: Z. Lu and M. S. Brown, Computing 1, 13 Computing Drive, Singapore 117417; email: {luzheng, brown}@comp.nus.edu.sg; W. Luo, Multimedia Lab, Department of Information Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong; email: lw010@ie.cuhk.edu.hk; Z. Sun, Digitalization Center, Dunhuang Academy, Mogao Ku, Dunhuang, Gansu province, China 736200; email: sunzhijun@vip.sohu.com; M. Ben-Ezra, Microsoft Research Asia, Building 2, No. 5 Dan Ling Street, Haidian District, Beijing, P.R. China 100080; email: mosheb@microsoft.com.

DOI 10.1145/2362402.2362403 http://doi.acm.org/10.1145/2362402.2362403

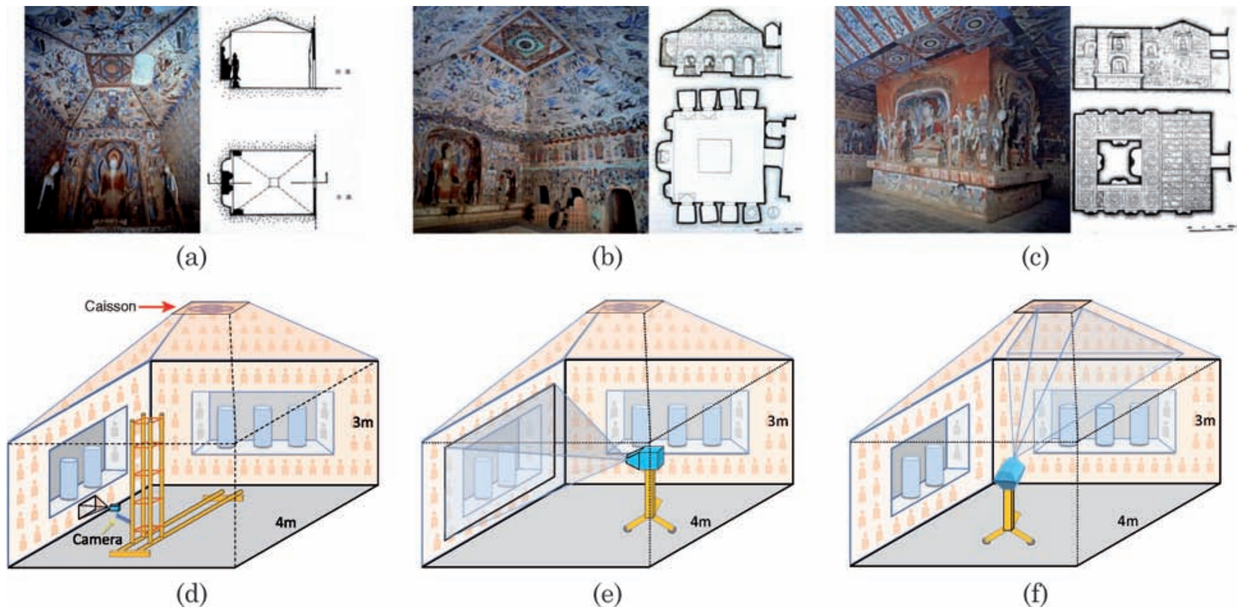


Fig. 1. Top Dunhuang Cave structure: (a) small cave example; (b) medium cave example with rooms; and (c) large cave example with central pillar. Bottom: (d) current imaging method, (e–f) proposed method for wall, niche and ceiling imaging using our digital large-format camera. (Image courtesy of the Dunhuang Academy.)

Caves have been a UNESCO world heritage site since 1987 [UNESCO 1987] and are managed by the Dunhuang Academy, which was established in 1943 by the Chinese government. As part of the many duties of the Dunhuang Academy, imaging the caves for purposes of preservation, studying, or virtual tourism [Lutz and Weintke 1999] has been on going since 1999. In the last two years, this imaging has turned its focus to high-resolution digital imaging, with the Dunhuang Academy's long term goal of capturing images of each cave's artwork at 300ppi resolution. To date, frescos in only 47 caves have been imaged at only 75ppi. This imaging task has proven to be so labor intensive that the Dunhuang Academy is now considering alternative options over its current procedure involving small-format cameras. The following discusses the physical environment, the current imaging procedure, and potential benefits of large-format imaging.

Environment. Photographing the caves at a minimum resolution of 300ppi presents a unique challenge due to large variations in cave size, structure, and sometimes difficult access. Figure 1(a–c) show three typical layouts of the Dunhuang Caves. A relatively small cave about $30m^3$), shown in (a), includes frescos on the walls as well as on the pyramidal ceiling. In addition, a few statues are located in niches in the wall. Special interest is given to the painted center of the ceiling known as the *caisson*. A medium size cave (about $100m^3$), shown in (b), can have attached small rooms that often house statues. Larger caves like the example shown in (c) often have one or more central pillars, with niches and statues. In most of the caves, all areas of the walls, ceilings, and niches are covered by frescos. The frescos and statues are still colorful, but the objects are covered with accumulated dust, which makes them very diffuse. The sizes of the statues in most caves range from less than 50cm to a few meters in height.[1] Given the complexity of imaging 3D objects at 300ppi, the Dunhuang Academy has focused

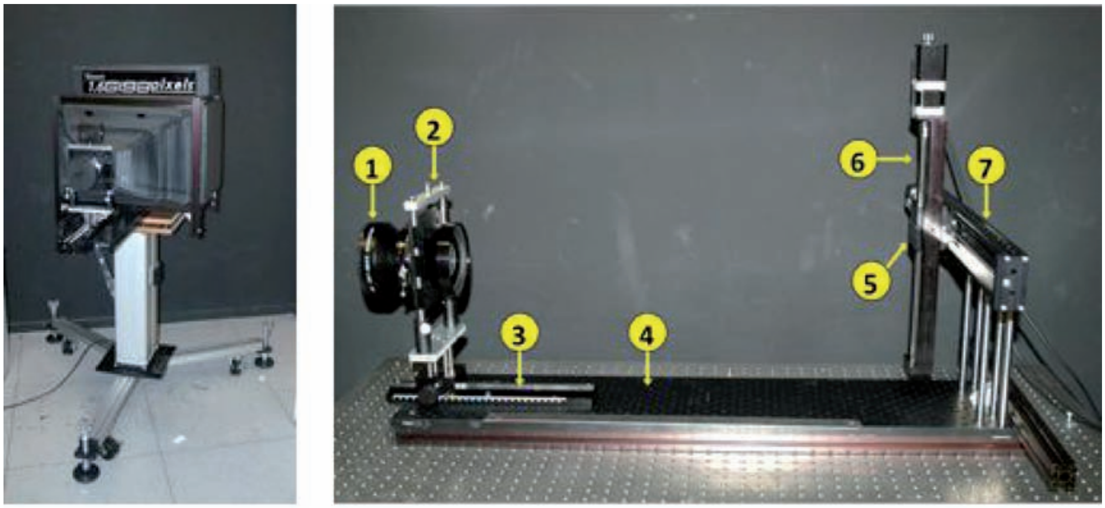
[1]There are exceptions, e.g., the Buddha statue in Cave #96 is 35m in height.

its efforts instead on imaging the frescos and other paintings in the grottos. However, even for these near-planar surfaces, acquiring high-resolution images is challenging.

Current Approach and Limitations. The current approach is a result of the Mellon International Dunhuang Archive project [MIDA], a recently completed collaboration between the Dunhuang Academy and Northwestern University. Figure 1(d) shows the current procedure used, where a rail and tower structure is erected in the cave and a small-format camera, typically a digital single lens reflex (DSLR) camera, is attached to the tower and placed close to the wall (closer than in the illustration). A sequence of images is then taken while manually moving the camera. To obtain the final mosaiced image over an area of several square meters at 75ppi resolution requires taking an enormous number of overlapping images using the small-format camera. To put this in perspective, the Dunhuang Academy reported that it requires 812 images using a 10 megapixel DSLR camera to obtain a 75ppi resolution image of a 5148 $\times$ 3329mm fresco located on the north wall of Mogao Cave #220. The total capture time for the wall is about four to five days excluding time for hardware setup. In addition, this camera needs to be carefully translated to facilitate the imaging process. We note that similar approaches have been used in other digital heritage projects, such as the Veridical Imaging of Transmissive and Reflective Artifacts (VITRA) project. The VITRA project used a robotic platform to assist the image capture [MacDonald 2006].

After image capture, the images are stitched together to form a single mosaic image. While image mosaicing is often considered a solved problem in computer vision and computer graphics, it is not necessarily the case when imaging artwork. Current image mosaicing techniques [Milgram 1975, 1977; Brown and Lowe 2003, 2007; Capel and Zisserman 1998; Szelishi and Shum 1997; Szeliski 2006] use homography-based image alignment, which only works in the ideal case of a purely planar surface or very distant scenes. Cave walls, however, are not perfectly planar and this causes errors in the final mosaic. The strategy used by almost all image mosaicing techniques is to hide these errors through image blending and/or seam-cutting. Perceptually masking errors is not an acceptable solution for the Dunhuang Academy. Instead, they use image-editing software, that is, Photoshop, to get an initial alignment. This is done using the mosaicing feature in Photoshop. Photoshop, however, uses seam-cutting and color adjustments in it's processing (applied as masks to the image layers). The Dunhuang Academy turns off these features and uses only the estimated planar warps as a starting point. The images are then manually adjusted using the local warping functionality in Photoshop to compensate for the nonplanar misalignments. In some case, Photoshop is unable to provide even the initial alignment. In these situations, the Dunhuang Academy is forced to align all images by hand, using Photoshop local warping tools. As in the previous example for Mogao Cave #220, it took approximately 54 man-days to stitch all 812 images. In fact, this mosaicing procedure is currently the major bottleneck in the Dunhuang Academy's imaging pipeline, with 50 caves' worth of images backlogged for processing.

Large-Format Camera Imaging. The impetus for our digital large-format camera project was to design a camera that could capture a perspective image of a much larger area with high resolution. In contrast to the previously described procedure, a well designed large-format camera can capture a $10m^2$ area at 300ppi resolution without moving the camera or lens. This allows capturing of statues and reliefs in a single image, as shown in Figure 1(e). The camera can also help capture the ceiling by just tilting it as shown in Figure 1(f). The Dunhuang Academy was interested in exploring the use of a digital large-format camera as it could significantly reduce overall time and labor as well as increase imaging resolution. A digital large-format camera cannot completely solve the imaging problems in a place as diverse as the Dunhuang Caves, as some caves and statues are simply too large, and in some, the space is too restricted. This would still require manual editing to stitch together the large-format images. However, its benefits were enticing enough to perform a initial study in the Dunhuang Caves.



(a) (b)

Fig. 2. Close look at our digital large-format camera. (a) the camera on the stand, (b) skeleton of our camera. (1) lens, (2) lens holder, (3) focusing stage, (4) breadboard, (5) mount for the sensor, (6) vertical stage of the image plane, (7) horizontal stage of the image plane.

The remainder of this article presents an overview of the basic design of our digital large-format camera in Section 2, followed by the results of our initial field deployment in Section 3. A discussion and summary conclude the article in Section 4.

2. DIGITAL LARGE-FORMAT CAMERA

We have designed a digital tile-scan camera capable of acquiring high-quality and high-resolution images of static scenes. Our camera can capture images with a resolution of over one gigapixel.

A close at our camera is shown in Figure 2. The "spine" of the camera is a 190 $\times$ 20cm double density optical board to which is attached two long travel (300 and 450mm) motorized translation stages to move the sensor. A third translation stage is added for camera focusing. The lens holder is made of optical table posts and custom made aluminum bars. A custom made Neoprene coated Nylon bellows connects the lens to the main frame. The lens holder also contains a mounting point for a low-resolution video camera that is firmly attached to the main lens to act as a viewfinder. Schneider Optics Apo Symmar L 480/8.4, is used as the main lens of our camera. We select this large-format lens for its large image circle of 500mm, a standard field-of-view of 56°, low distortion, and nearly uniform resolution throughout its field-of-view. The main sensor attached to the translating stage is a Kodak KAI11002 CCD taken from a Lumenera USB camera. The sensor has 11 mega pixels at 9μm per pixel. The camera's dynamic range is estimated at 65db and supports exposure bracketing. Other details pertaining to the camera construction and characterization, such as the effective optical resolution in terms of the camera's modulation transfer function (MTF), can be found in Ben-Ezra [2011].

Our camera is connected to a PC with custom software that allows the user to perform tasks such as viewing the current scene, selecting parts of the scene for capture, moving the lens along the translation stage for focusing, and adjusting exposure parameters. After the user sets up all the parameters and starts capturing a high-resolution image, the camera sensor moves on the stages along the horizontal and vertical directions and takes several shots with a predetermined overlap between images (25% overlap is used in this work). The intermediate images are then automatically stitched to form

Fig. 3. Images from our field deployment in the Mogao Caves. (a) Our high-resolution digital large-format camera in Mogao Cave #46. The frescos and statues in the cave are from the Tang Dynasty (618–907 AD). (Image courtesy of Ben-Ezra [2011], photographed by Zhijun Sun). (b) Our camera with multiple lights placed around it to ensure uniform illumination. (Image courtesy of the Dunhuang Academy, photographed by Zhijun Sun.)

the final output image. Since our camera design is a single lens system, we do not have the same stitching problems as traditional mosaicing translating cameras. Instead, our resulting stitched image is performed on a true perspective image seen through the lens with more than 300ppi resolution and therefore does not suffer when the imaged scene is not perfectly planar. For further details see Ben-Ezra [2011] and the camera's Web site.[2]

3. FIRST FIELD DEPLOYMENT

Our first field deployment (see Figure 3) in the Mogao Caves was in March 2010. This section describes the image capture and postprocessing of this field work. We also show several results obtained from our field deployment.

Among the hundreds of caves, Mogao Caves #46, #418, and #420, were selected for our field work taking into account the following considerations: the size of our camera, the size and complexity of the cave, and the artistic value of the artifacts inside the cave. Cave #46 is a small cave whose walls contain both statues and frescos. Cave #420 contains frescos that have significant 3D texture, while Cave #418 has frescos that consist of several layers of paint from different dynasties.

In the three caves, high-resolution 2D images of the frescos together with statues were captured using our camera. All the images were captured in 16bits raw format under proper exposure. We used two 5000K fluorescent diffused lights to illuminate the scene, positioned at both sides of the camera when possible (see Figure 3(b)). No polarization is used because the diffuse nature of the scene. We found that in almost all cases, even for frescos, the required depth-of-field of the scene exceeded the lens depth-of-field. As a result focal stacking was used to extend the depth-of-field. There are various approaches to focal stacking (e.g. Pieper and Korpel [1983], Adelson et al. [1984], Nayar and Nakagawa [1994], Agarwala et al. [2004], Hasinoff et al. [2009], however, the basic idea for all methods is to capture the scene multiple times, while varying the focus of the camera. The multiple images are then composited to produce an all-in-focus image. During the field work, the focal stack parameters were estimated by hand, as we were not expecting to perform focal stacking. Our latest version of the control

[2]http://www.dgcam.org.

software is now able to automatically calculate the focal stacking parameters based on the object's distance to the camera and it's aperture setting.

Before each scene was captured, we first imaged a dark current,[3] a white image[4] and a Macbeth color chart. The postprocessing steps include: (1) *dark current processing* to remove the fixed pattern noise due to bad pixels on the sensor; (2) *demosaicing* to convert the raw format images to RGB format with gamma correction; (3) *white image processing* to correct dark regions in the input image due to sensor/lens dust using white images. When images were captured with focal stacking, the method described in Ben-Ezra [2011] was used to extend the depth-of-field. We did not perform any radial distortion or light falloff correction in our field work. This is because only the center part of the image circle is used with our camera and the lens of the camera has very minimal distortion (see Ben-Ezra [2011] for more details). In addition, during actual image-capturing, we placed multiple lights around the camera and used light meters to ensure uniform illumination. In total, two complete walls containing both fresco and statues, and portions of frescos from the three caves were captured.[5] The color chart was not used by our software, however, it was used as a reference by the Dunhuang Academy to adjust colors with other software.

3.1 Result

Images shown in this article, were produced with HDView [Kopf et al. 2007] which is one of the few programs able to handle images at our resolutions and sizes. Images shown have at least 300ppi on the object's surface. In Figure 4 and Figure 5, we show two examples from images captured in our field study. Figure 4 captures part of the north wall in Cave #46. The pictured walls and statues are from the Tang Dynasty (618–907 AD). This example required focal stacking with five images at varying focal distances. The size of the image is 45320 × 32645 (1.4 gigapixels). The size of the scene is about 2.6 × 1.7m. The total capture time was 124minutes including the time to capture the focal stacking. Figures 4(a,b) show the overall image and zoomed region of the image and Figures 4(c–f) show further zoomed regions. Note that Figure 4(e) shows a close look at the slanted wall, which requires extended depth-of-field.

Figure 5 shows part of the fresco on the east wall in Cave #418. The size of the image is 37392×31718 (1.1 gigapixels). The size of the scene is about 1.3 × 1m. Although the fresco appears to be flat, focal stacking was still required. The total capture time was 77minutes including the time for focal stacking. Figure 5(a) shows the overall image. Figures 5(b–h) show regions at different zoom levels. Note that this image shows frescos with two layers. While Figure 5(b) shows a Buddha from the Western Xia Dynasty (1038–1227 AD), Figures 5(c–h) not only show fresco fragments from the Western Xia Dynasty, but also reveal a Buddha from the Sui Dynasty (518–618 AD).

4. DISCUSSION AND SUMMARY

4.1 Lessons Learned

Several valuable lessons learned from this field work, from both a technical and cultural heritage perspective, are summarized in the following.

Dust. Unlike a lab setting, imaging inside a cave involves in a great deal of dust. Bad sensor pixels and dust on the lens/sensor become artifacts in the final image. Since our digital large-format camera

[3]Raw images that are captured when the aperture is completely closed.
[4]Images that are captured when white card is in front of the camera and defocused. White image correction is also known as *shading correction*.
[5]The post processing are performed using an ordinary PC with Intel Pentium IV CPU.

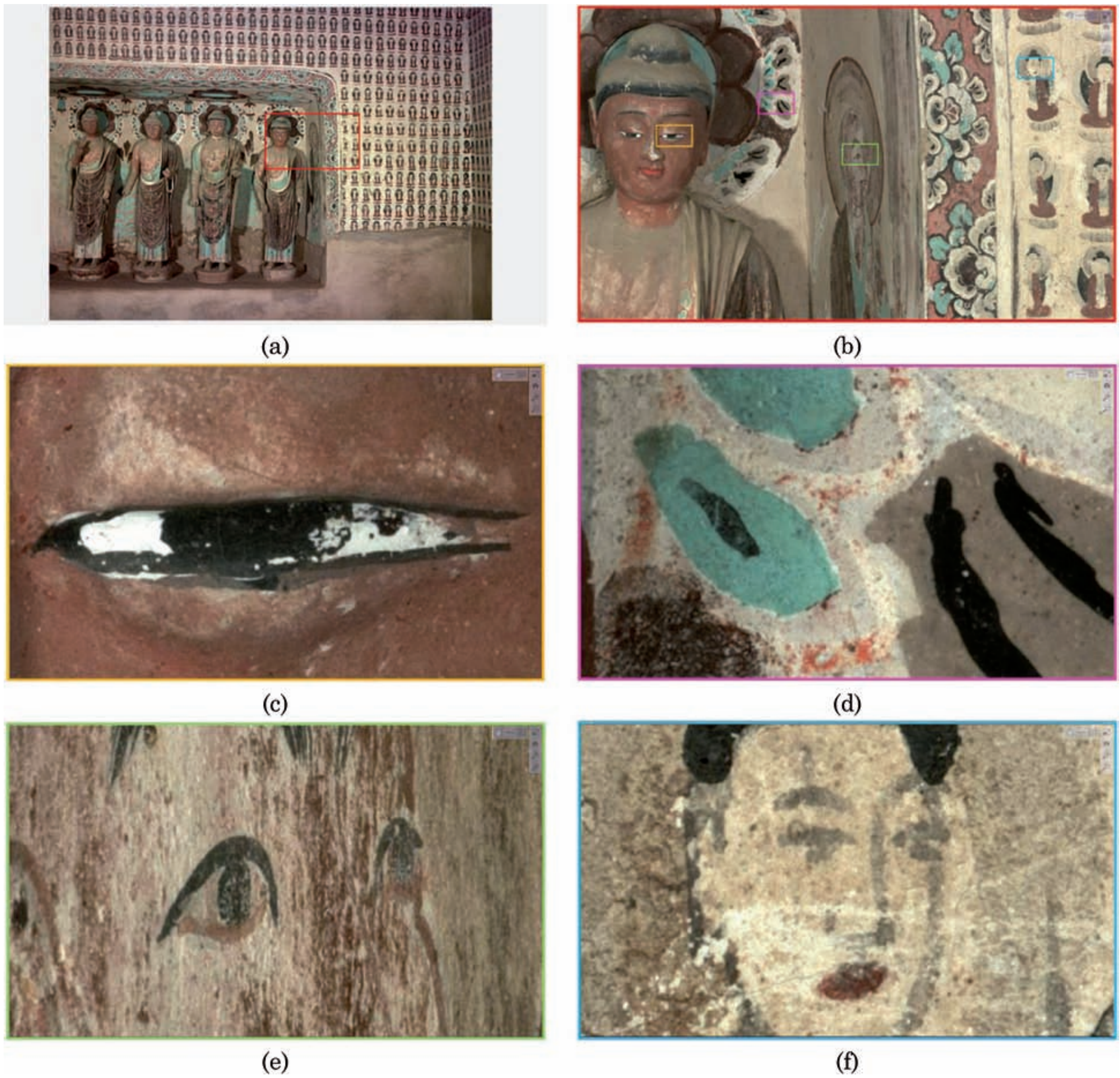


Fig. 4. An image of the north wall of Mogao Cave #46. (a) the final image, (b) zoomed region of the image, (c–f) further zoomed regions. Note that (e) shows a close look, at a wall in the niche with large depth-of-field. The walls and statues captured are from the Tang Dynasty (618–907 AD). (Image courtesy of the Dunhuang Academy.)

design uses a translating sensor, dust on the sensor creates artifacts that form repeated patterns in the final stitched images. To deal with this problem, we needed to capture a dark current and white image, before each imaging session, to help detect both bad sensor pixels and dust.

Focal Stack. While we have already mentioned the use of focal stacking to extend the depth-of-field of our camera, we were surprised at how often this was necessary. Focal stacking was even necessary

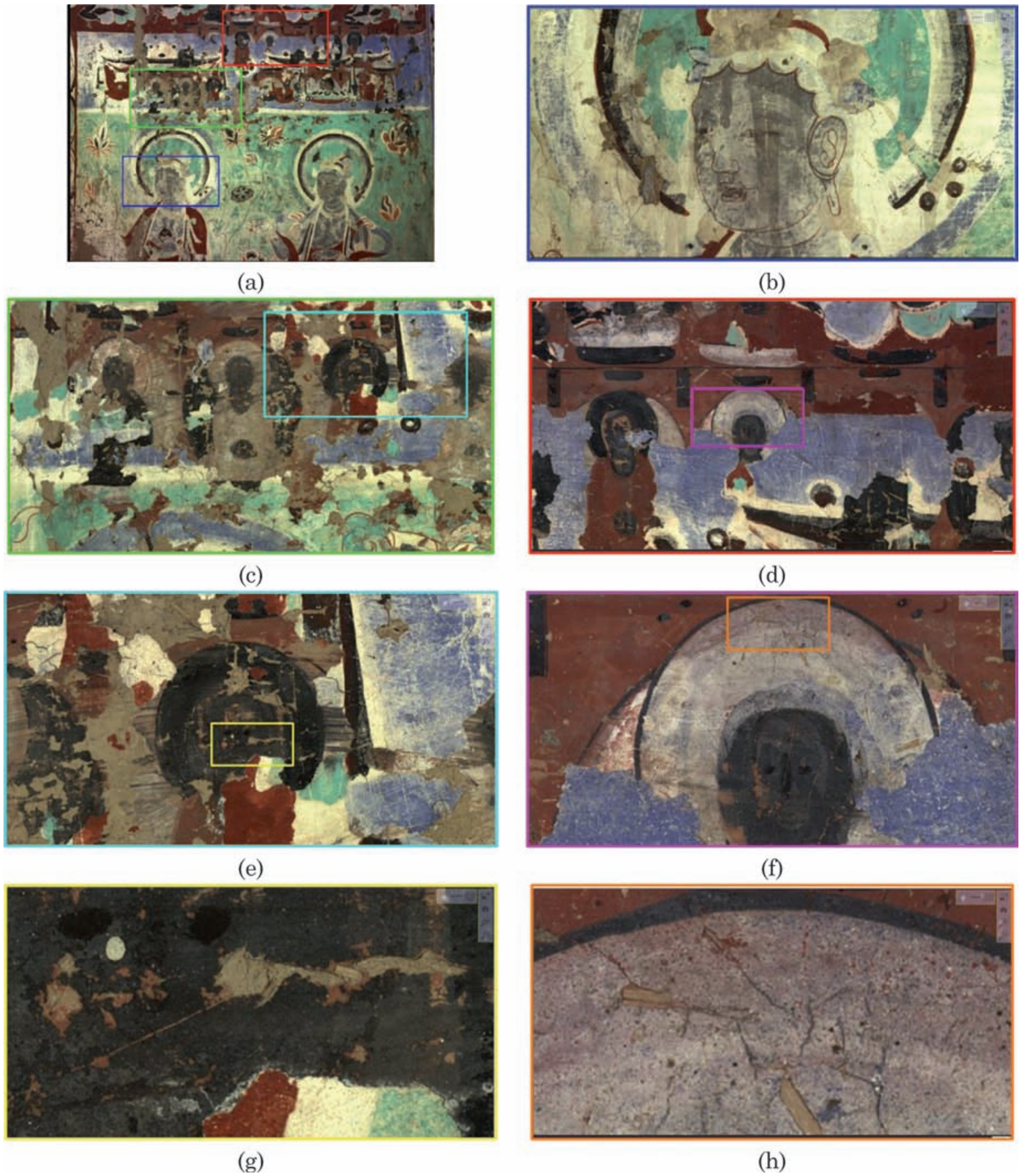


Fig. 5. An image from part of the east wall in Mogao Cave #418. (a) the final image, (b–d) zoomed regions of the image, (e–f) further zoomed regions of (c–d), (g–h) further zoomed regions of (e–f). This image captures frescos with two layers. While (b) shows a Buddha from the Western Xia Dynasty (1038–1227 AD), (c–h) not only show fresco fragments from the Western Xia Dynasty, but also reveal a Buddha from the Sui Dynasty (518–618 AD). (Image courtesy of the Dunhuang Academy.)

for capturing frescos that appear flat but had depth variation. Apparently this is also an issue for the small-format cameras. From the feedback of the Dunhuang Academy, our camera's ability to automatically capture and process the focal stack is very valuable for their work.

Color Management. Proper handling of color is absolutely essential for imaging heritage sites such as the Mogao Caves. Color needs to be carefully managed before and after capture. For a large-format camera, one capture covers a large area of the scene. Hence we need to place several high-quality lights in the cave to ensure proper illumination. Devices such as hand-held light meters are used to ensure the scene is as uniformly illuminated as possible. For each capture, we imaged a Macbeth chart under the same illumination for color calibration in the postprocessing step. Furthermore, all the output devices need to be calibrated also (e.g., monitors, printers, etc.). Both technical staff and professional archivists/artists need to work together to fine-tune the color if required.

Remaining Challenges for Mosaicing. Our camera does not completely remove the need for image mosaicing. For a large scene, multiple images from our camera will be required. The problem with non planar scenes remains, however our digital large-format camera does require less manual stitching, since large portions of the image are correct. This points to the need for more effective mosaicing algorithms that are capable of overcoming 3D surface variations in a proper manner, versus hiding artifacts using seam cutting and blending.

4.2 Summary

This article described recent field work for imaging Buddhist art at the UNESCO world heritage site, the Mogao Caves in Dunhuang, China. Our article outlined current challenges faced by the Dunhuang Academy using conventional mosaicing techniques with a small-format camera. The need for human intervention in the mosaicing of hundreds of images each covering a small area was the major bottleneck in the Dunhuang Academy imaging. Our preliminary results showed that using a digital large-format camera not only improves the resolution of the imaging, but can significantly reduce time and effort by capturing a vastly larger area (equivalent to around 800 small format images) in a single photograph. We have also outlined several lessons learned from our experience.

Following the success of our field study, the Dunhuang Academy has decided to pursue a long term imaging effort using a digital large-format camera. We have designed and assembled a second generation digital large-format camera. The new camera has tilt/swing gears, for easily capturing frescos on the ceiling or adjacent to the floor. While still over a gigapixel, the new camera is slightly smaller and better protected from dust. The new design also uses a 300mm lens instead of 480mm to cover a wider area with slightly lower resolution. The smaller design will also allow operation in tighter spaces. In August 2011, the second generation camera, officially called *Apsara*, after the Mogao Caves' famous "flying fairy" paintings, has been handed over to the Dunhuang Academy [Apsara 2011].

ACKNOWLEDGMENTS

The authors thank the Dunhuang Academy Vice Managing Director Xudong Wang. We also thank related staff, Hongcai Sun, Zhaofu Qiao, and Tianxiu Yu for their assistance in capturing the images and for their kind permission to use the relevant images shown in this article.

REFERENCES

ADELSON, E. H., ANDERSON, C. H., BERGEN, J. R., BURT, P. J., AND OGDEN, J. M. 1984. Pyramid methods in image processing. *RCA Engineer 29*, 6, 33–41.

AGARWALA, A., DONTCHEVA, M., AGRAWALA, M., DRUCKER, S. M., COLBURN, A., CURLESS, B., SALESIN, D., AND COHEN, M. F. 2004. Interactive digital photomontage. *ACM Trans. Graph. 23,* 3, 294–302.

APSARA. 2011. Apsara. http://news.xinhuanet.com/english2010/culture/2011-09/30/c131169403.htm.

BEN-EZRA, M. 2011. BEN-EZRA, M.2011. A digital gigapixel large-format tile-scan camera. *IEEE Comput. Graph. Appl. 31,* 1, 49–61.

BROWN, M. AND LOWE, D. 2003. Recognising panoramas. In *Proceedings of ICCV.*

BROWN, M. AND LOWE, D. 2007. Automatic panoramic image stitching using invariant features. *Intern. J. Comput. Vis. 74,* 1, 59–73.

CAPEL, D. AND ZISSERMAN, A. 1998. Automated mosaicing with super-resolution zoom. In *Proceedings of CVPR*.

HASINOFF, S., KUTULAKOS, K., DURAND, F., AND FREEMAN, W. 2009. Time-constrained photography. In *Proceedings of ICCV*.

KOPF, J., UYTTENDAELE, M., DEUSSEN, O., AND COHEN, M. F. 2007. Capturing and viewing gigapixel images. *ACM Trans. Graph. 26,* 3, 93.

LUTZ, B. AND WEINTKE, M.1999. Virtual Dunhuang art cave: A cave within a CAVE. *Comput. Graph. Forum 18, 3,* 257–264.

MACDONALD, L. W. 2006. A robotic system for digital photography. In *Proceedings of SPIE*.

MIDA. Mellon Interational Dunhuang Archive. http://www.artstor.org/what-is-artstor/w-html/col-mellon-dunhuang.shtml.

MILGRAM, D. L. 1975. Computer methods for creating photomosaics. *IEEE Trans. Comput. 24*, 11, 1113–1119.

MILGRAM, D. L. 1977. Adaptive techniques for photomosaicking. *IEEE Trans. Comput. 26*, 11, 1175–1180.

NAYAR, S. K. AND NAKAGAWA, Y. 1994. Shape from focus. *IEEE Trans. Pattern Anal. Mach. Intell. 16,* 8, 824–831.

PIEPER, R. J. AND KORPEL, A. 1983. Image processing for extended depth of field. *Appl. Opt. 22,* 10, 1449–1453.

SZELISKI, R. 2006. Image alignment and stitching: A tutorial. *Found. Trends Comput. Graph. Vis. 2,* 1, 1–104.

SZELISKI, R. AND SHUM, H. Y.1997. Creating full view panoramic image mosaics and environment maps. In *Proceedings of SIGGRAPH*.

UNESCO. 1987. Mogao Caves. http://whc.unesco.org/en/list/440.

Received March 2011; revised June 2011, November 2011; accepted December 2011