High Resolution Large Format Tile-Scan Camera
Design, Calibration, and Extended Depth of Field

Moshe Ben-Ezra
Microsoft Research Asia
49 Zhichun Rd. Beijing 100190

`mosheb@microsoft.com`

Abstract

Emerging applications in virtual museums, cultural heritage, and digital art preservation require very high quality and high resolution imaging of objects with fine structure, shape, and texture. To this end we propose to use large format digital photography. We analyze and resolve some of the unique challenges that are presented by digital large format photography, in particular sensor-lens mismatch and extended depth of field.

Based on our analysis we have designed and built a digital tile-scan large format camera capable of acquiring high quality and high resolution images of static scenes. We also developed calibration techniques that are specific to our camera as well as a novel and simple algorithm for focal stack processing of very large images with significant magnification variations.

1. Introduction

Although the resolution of SLR and medium format digital cameras has been increasing in recent years, there are applications in cultural heritage preservation and computation photography that require even higher resolutions. For example, museums require a minimum density of 20 pixels per millimeter (on the object) for digital archiving of paintings [11]. Light field imaging [13, 16] trade spatial resolution with partial light field and even a small 3x3 matrix will reduce the number of pixels in a processed image by nine.

Unfortunately, cameras, like many physical systems, do not scale well without affecting their performance. The primary reason for this is the signal to noise ratio - reducing a pixel size by n will reduce the amount of collected photons by n^2 (for the same f-number and exposure time) and will noticeably affect the image quality [4].

Due to their large image planes (often as large as 10" $\times$ 8") large format cameras can achieve a very high resolution without compromising pixel size.
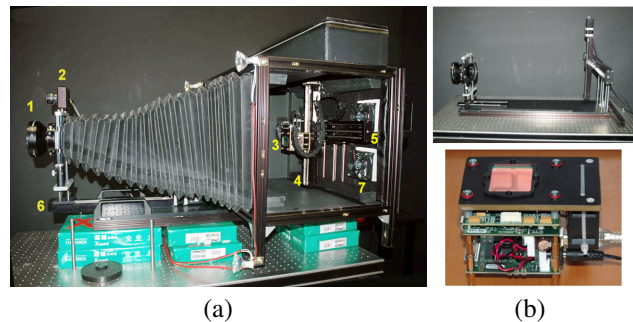

Figure 1. (a) Our camera with its door open. The following parts are shown: the primary lens (1), the video camera (2), and the focusing stage (6) are located at the front of the camera. The main sensor (3), vertical translation stage (4) and the horizontal translation stage (5) are located at the back of the camera as well as the thermoelectric cooling (7). (b)-top: The skeleton of the camera, shown here with a manual focusing stage, is made of optical table components, with few custom parts used in the lens holder. This makes the structure both rigid and accurate. (b)-bottom: The main sensor setup. A custom made front board was milled from hard plastic to hold the CCD, optical window and the electronics while permitting a wide FOV.

1.1. Goals

Our primary goal is to make a high resolution camera system for use by museums and cultural heritage sites for digital acquisition and archiving of art works. Our main challenge is to make a camera that is simple enough to be constructed by a small team and be reasonably cost effective for museums and cultural sites.

Our secondary goal is to make a high resolution camera that can serve as a platform for research that requires high resolution imaging and to enable the assembly of a database of high resolution images for research.

1.2. Related Work

A semi-digital high resolution large format camera was presented in 2001 by Graham Flint as part of his GIGAPXL PROJECT (www.gigapxl.org). Flint's camera uses a modified K-38 military aerial camera and a specially designed

high resolution wide angle lens to capture 9" × 18" frames on analog photographic film that are later scanned to produce gigapixel images. Being essentially an analog film camera, it is capable of taking stop-motion snapshots of dynamic scenes. This however is less important for imaging of static scenes as in museum and cultural heritage sites. The imaging cost of Flint's camera is approximately $50/exposure ($2K for a 100 frame film + processing and scanning costs) and the imaging cycle is relatively long compared to a digital imaging cycle due to the processing and scanning time.

Focal plane array technology uses an array of sensors to capture very high resolution images. For example, the Pan-STARRS telescope camera uses an array of 4096 CCDs to provide 1.4 gigapixel images [15]. Focal plane arrays can also capture snapshots. However, they are very expensive and currently used only in telescopes. Additionally, focal plane arrays have visible seams between individual detectors that are inconvenient for digital archiving purposes.

A low cost high resolution scan camera was introduced in 2004 by Wang et al. [17]. This camera combines a flatbed scanner with an 8" × 10" view camera to obtain up to 490 megapixel images. The camera is grayscale and requires multiple scans using a color filter wheel to obtain color images. Because it uses a flatbed scanner, the gain and exposure control are very limited and require strong illumination. Focusing requires removing the back or using trial scans. Additionally, images acquired by the scanner show scan-line artifacts and a significant amount of post processing is required to detect and remove these artifact by means of in-painting.

Commercial solutions for large format imaging scan the image plane using a tri-linear sensor that does not require demosaicing. To the best of our knowledge, the highest resolution scanning back available today is the Anagramm David that can capture up to 340 megapixel images. Because these solutions use a linear sensor they can only capture a single column during each exposure. Therefore, capturing images using a long exposure time may not be practical and strong illumination is needed to shorten the exposure time.

An alternative approach is to stitch small images into a mosaic as in [14, 8]. However, panorama stitching works best for large, distant objects. Close objects often require adjustment of focus and/or viewpoint between images, both of which usually result in visible seams in the stitched image [17].

The rest of this paper is organized as follows: Sections 2 and 3 discuss the design consideration of the camera, in particular the selection of the lens and sensor. Section 4 briefly describes the implementation of the camera. Sections 5 and 6 describe calibration related issues, in particular the relation between focus and magnification. Sections 7 and 8 describe the image capture and focal stack process, and fi-
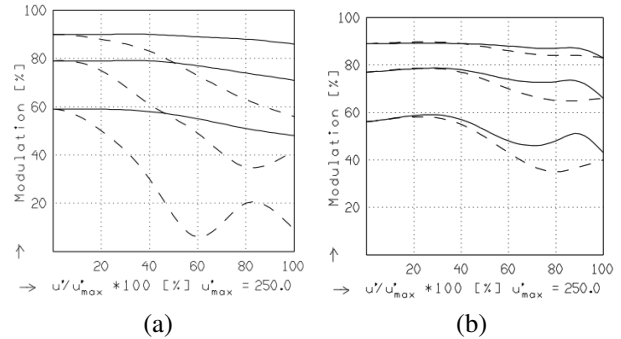

(a)   (b)

Figure 2. Modulation transfer function (MTF) of two different lenses at the same f-number (22), image circle (500mm), and magnification factor (10:1). X-axis is the distance from the optical center. Solid lines show the MTF for radial orientation at 5,10, and 20 lp/mm. Dashed lines show the MTF for tangential orientation for the same frequencies. (a) Schneider's Apo-Tele-Xenar 12/800 lens' MTF, (b) Schneider's Apo-Symmar 8.4/480 lens' MTF.

nally sections 9-12 provide experimental results and comparisons.

2. Lens selection

Roughly speaking, the effective image size produced by a given lens is equal to the resolution of the lens multiplied by the area of the projected image (image circle).

The size of the image circle is given by the lens manufacturer and usually fits common camera formats. Determining the resolution of the lens is more difficult because the lens resolution is subject to diffraction and aberration limits and changes with aperture, angular position, focal distance and orientation (radial or tangential). Lens manufacturers usually provide partial resolution information in the form of modulation transfer function (MTF) charts for a few spatial frequencies, apertures, and magnification factors. For example, Figure 2 shows the MTF of two different lenses having the same image circle size, magnification factor and f-number. We can see that while the lenses have very similar radial resolution, the Schneider's Apo-Tele-Xenar 12/800 lens has a noticeable degradation of the tangential resolution in part of its field of view, whereas the Apo-Symmar 8.4/480 lens has a nearly uniform resolution throughout its field of view (FOV).

When selecting the lens we also need to consider distortion and vignetting. We selected the Schneider's Apo-Symmar 8.4/480 for its large image circle of $500mm$, a standard FOV of 56°, low distortion and nearly uniform resolution throughout its FOV, as shown in Figure 2(b).

3. Sensor Selection

Most digital sensors are *not* suitable for work with large format lenses. Figure 3(a) shows a lens designed to work with a conventional digital sensor. We can see that the projection is telecentric at the sensor side and that all colors are
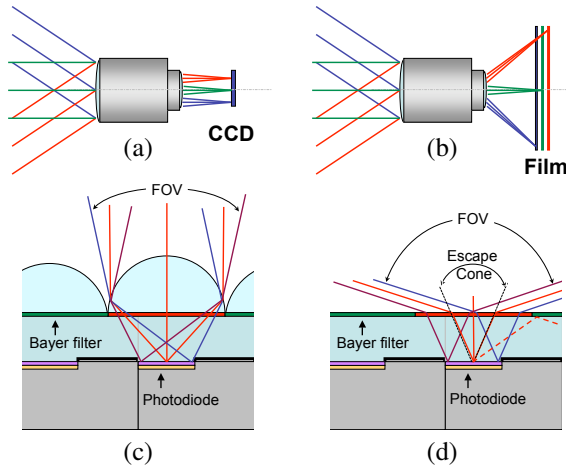
Figure 3. (a) A (good) "digital" lens is telecentric at the image side and all colors are focused on the same plane. (b) Large format film lenses cannot be telecentric at the image side due to size differences, and they may have slightly different focal planes for different colors to match the layered structure of photographic film. (c) A cross section through a nearly telecentric pixel with microlenses. (d) Same pixel without microlenses has half the fill factor (for interlined CCD) but a much wider FOV (up to nearly hemispherical FOV when the escape cone approaches the critical angle).

focused at the same plane. In contrast, a large format lens shown in Figure 3(b) has an image circle that is much larger than the lens and therefore the projection cannot be telecentric. Additionally, the lens has slightly different focal planes for each of the three primaries to match the film's layered structure.

In practice, we did not find the second issue to be a problem, due to the focus tolerance of large format lenses and large pixels. The first issue, however, is much more significant. This is demonstrated in Figure 4. We can see that a sensor without microlenses exhibits only a slight degradation at the edge of the FOV compared to the center of the FOV (near the optical axis), whereas a sensor with microlenses shows a very sharp degradation in image quality at the edge of the FOV.

This is explained in Figure 3(c),(d). Figure 3(c) shows a vertical cross section through a pixel as well as an illustration of the light rays' path. We can see that the optical configuration of the microlens significantly limits the FOV of the pixel, which is the reason a sensor-side telecentric lens is required. Figure 3(d) shows the same pixel without microlenses. We can see that field of view becomes significantly wider, however the fill factor is reduced by half.

A possible solution to the problem is to use a full frame or a frame transfer CCD. Unfortunately, full frame CCDs require a physical shutter to prevent smearing during readout. A ferroelectric shutter would reduce the light significantly when opened and would not fully block the light when closed, and can also effect the optical quality of the
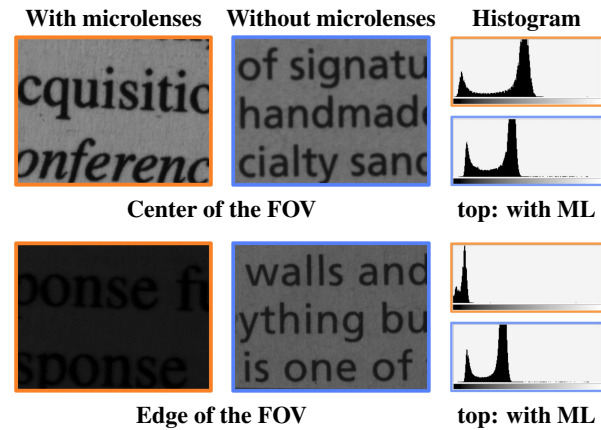
Figure 4. Comparison between two sensors tested by the same camera frame and lens. As seen in the left column, the sensitivity of the sensor with micro-lenses (Sony ICX424 7.4μ) drops sharply at the edge of the FOV whereas sensitivity drops only slightly for the sensor without microlenses (Kodak KAI-11002 $9\mu m$).

image, whereas many mechanical shutters have a limited cycle span and will only last for several hundred images (each requires over a hundred operations of the shutter). Though there are research grade mechanical shutters that can last for millions of cycles, finding one that is large enough and thin enough not to occlude the sensor at the edge of the image plane proved difficult. Additionally, a full frame CCD will not allow video streaming that is useful for focusing through the lens using the main sensor.

Frame transfer CCDs work best when the sensor is small or when the exposure time is very long (as in telescopes). Using a large frame transfer CCD with short exposure time will result in noticeable smear. Additionally, frame transfer CCDs are not common, and it is very difficult to find a camera that has a frame transfer CCD and fits our needs.

We therefore selected an interline CCD with no microlenses as shown in Figure 3(d). Interline CCDs without microlenses will reduce the amount of light less than a ferroelectric shutter without its additional complexity and also allow video streaming for easy focusing. To compensate for the low fill factor we selected a sensor with a large pixel size. We use the 11 megapixel Kodak KAI-11002 sensor, which has $9\mu m$ pixels. Even without microlenses, the sensitive area is equivalent to a full frame sensor with a pixel size of $6.3 \times 6.3\mu m$.

4. Implementation

In this section we provide a concise description of the core implementation of our camera. A detailed description is given in [2]. To get an accurate mechanical frame, which is essential for a focused and non distorted image, we used optical table components as building blocks. Figure 1 shows the mechanical setup of the camera. The "spine" of the camera is a 190×20 cm double density optical board from Thorlabs, to which we attached two long travel (300 and
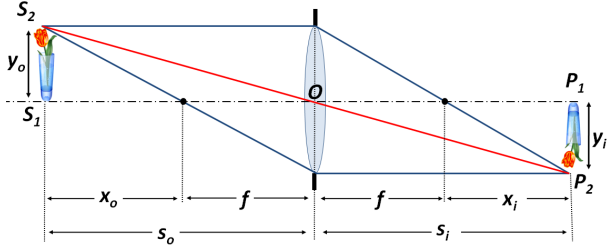
Figure 5. The thin lens model.
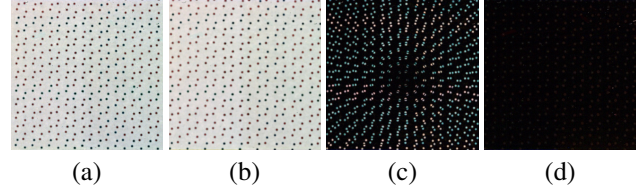

(a)   (b)   (c)   (d)

Figure 6. Calibration of the focusing stage: (a),(b) two (defocused) images taken at both sides of the focal plane at the estimated center of the image, (c) their difference. The change in magnification is clearly seen. (d) difference after calibration.

450 mm) motorized translation stages (for sensor motion) from Zaber Technologies. A third translation stage (either manual or motorized) was added for camera focusing. Aluminum rails were used to support the optical board and to build the enclosure frame.

The lens holder was made of optical table posts and custom made aluminum bars. A custom made Neoprene coated Nylon bellows from Gortite connects the lens to the main frame. The lens holder also contains a mounting point for a video camera that is firmly attached to the main lens and moves with it.

The main sensor is a Lumenera USB camera with a Kodak KAI-11002 CCD. The camera was stripped of its original housing to reduce its mass and to allow a wide field of view. A low-weight custom made front plate shown in Figure 1(b) protects the sensor and allows the attachment of optical windows or filters.

5. Focus and Magnification

When a conventional camera is set to a different focal distance, the magnification of the camera (and the effective focal length) also changes (with the exception of cameras that are telecentric at the image side [18]). In most cases, where the object is relatively distant with respect to the focal length and the image is relatively small, the change in magnification results in small motion and is often ignored. Having a digital lens with (nearly) telecentric projection also helps in this respect. However, as shown below, a large format camera operating in a relatively close range to the object exhibits a very significant magnification change that cannot be ignored. Given that a large format camera cannot easily be made telecentric due to the difference between the image size and lens size, the change in magnification becomes a significant problem when trying to focus at a point and when trying to extend the depth of field (DOF) using a focal stack. In this section, we compute the magnification factor change due to focus shifts. Focal stack computation is addressed in section 8.

The image or transverse magnification of an object using the thin lens model is defined as $M_T = \frac{y_i}{y_o}$, where y_i and y_o, shown in Figure 5, are image size and object size respectively. From the triangular similarity, $M_T = -\frac{s_i}{s_o}$ (negative M_T indicates an inverted image). From the thin lens equation

$$\frac{1}{f} = \frac{1}{s_o} + \frac{1}{s_i} \qquad (1)$$

we get:

$$s_o = \frac{f s_i}{(s_i - f)} \qquad (2)$$

$$M_T = -\frac{s_i - f}{f} = -\frac{x_i}{f}. \qquad (3)$$

Equation 2 is most commonly used for depth from focus. Equation 3, known as the Newtonian expression for magnification [7], provides the magnification factor as a function of internal parameters only (for in-focus objects).

The ratio $\psi(\delta)$ of the magnification factor as a function of lens movement by δ is given by:

$$\psi(\delta) = \frac{M_T(x_i + \delta)}{M_T(x_i)} = \frac{(x_i + \delta)}{f} \frac{f}{x_i} = \frac{x_i + \delta}{x_i}. \qquad (4)$$

For example, pixels at the corner of a 512×512 image focused one meter away and taken by a $50\ mm$ lens will be displaced by 2.6 pixels when the lens is focused $10mm$ further (in the object side). In contrast, pixels at the corner of a $20k \times 20k$ image taken with a $500\ mm$ lens under the same conditions will be displaced by 196 pixels. Clearly, this displacement cannot be ignored.

Since the motion vector of point (i, j) for an image centered at the focus of expansion (FOE) is simply given by:

$$(i, j) \rightarrow \psi(\delta)(i, j), \qquad (5)$$

we can now move the horizontal and vertical translation stages in synchronization with the focusing stage. This will keep the image tile centered through the focusing process, which is useful also for focal stack processing. Note however, that unlike [18] this will not correct the magnification change within a tile.

6. Calibration

This section describes the calibration steps that are specific to our camera. We also apply standard calibration (not described here) to the main camera and to the video camera.

6.1. Calibrating the focusing stage

In the previous section we computed the change in magnification due to focus change. However, this assumes that the motion of the lens is perfectly aligned with the optical axis, (or that the focus of expansion (FOE) is known) and that we know the accurate value of x_i.

We therefore seek to find the FOE due to magnification change, and the value of x_i. There is, however, a problem caused by the fact that the image is defocused when the lens is moved, making it more difficult to estimate the exact motion.

To solve this problem we use the fact that the defocus of a symmetrical feature (such as a point) is symmetrical on both sides of the image plane, whereas the magnification is not. We therefore take a pair of similarly *defocused* images of a set of points at both sides of the focal plane. Then we register both images using the Lucas-Kanade method [10] restricted to a translation plus scale motion model. Figure 6 shows two such images, as well as the difference before and after calibration. The change is clearly visible.

The translation vector is the correction to the true FOE, whereas the scale is $\psi(\delta)$ from Equation 4. Since δ is accurately known (the focusing stage maximal error is no more that $45\mu m$) we can compute x_i from Equation 4.

6.2. Calibrating the image plane stages

The image plane stages have a nominal error of no more than $23\mu m$ (2.5 pixels) and a repeatability error of no more than $3\mu m$ ($\frac{1}{3}$ pixel). However, these numbers refer to the 1D motion of each stage. When the stages are placed in an XY configuration they are subject to additional errors, mainly due to imprecise alignment of the two stages as well as imprecise alignment of the sensor. Additionally there could be a small error in the lateral direction that affects the other stage. The resulting error is a constant bias (caused by the angular errors) plus some perturbations. To calibrate for these errors, we took many images of a textured test target and then registered the images using a translation only motion model and measured the bias as well as the average residual error and variance at each position. In most cases the average error was sub pixels (or zero when the error was too small to detect), but in some locations the residual error was a few pixels long and is corrected during stitching.

7. Image capture

When capturing the image there are several properties that we wish to obtain: (i) we want the scan to be as fast as possible. (ii) we must avoid motion blur. (iii) for certain multi-exposure images such as HDR images and photometric stereo, we would like to keep the images very well aligned.

To achieve these goals, we scan the image in a step-scan manner, in which the sensor stops completely before each image is taken. Stopping the sensor completely allows for long exposure times as well as very good align-ment in mutli-exposure conditions (the scan is done only once, where potentially several images are taken at each location). The image is scanned column by column in a zig-zag order. The captured tiles overlap with each other. The overlapping regions are used later for better alignment and are necessary for tiled focal stack (see section 8).

Due to the low moving mass and the zig-zag motion, we are able to capture a $40k \times 30k$ image in less than 5 minutes (for $1/10$ sec exposure time). This is significantly faster than other techniques including mosaics and an 1D scan by a linear sensor).

Multi-exposure techniques such as HDR and photometric stereo will increase this time by only the additional exposure and image readout time as no mechanical motion is needed. Focal stack will require additional time for mechanical motion, but only at the areas designated by the user.

8. Focal Stack

Scaling up a camera greatly improves the image quality and the range of non-diffraction limited apertures. However, somewhat surprisingly, it does not significantly improve the DOF. The reason is that for the same object distance the optical magnification of the large camera is significantly higher - similar to macro lenses (an intuitive way to see this is to consider that while we "scale the camera" we do not "scale the physical world"). Therefore, in order to keep the same high level of details for 3D objects that have depth variations larger than the DOF we must extend the DOF of the camera.

Several computational methods for extending the depth of field using coded exposure or aperture have been proposed in recent years [12, 20, 19, 16, 9]. In this paper we use the traditional focal stack method, which is simple, performs well and is time efficient [6].

However, as mentioned before, unlike conventional cameras and in particular cameras that are telecentric at the image side [18], large format cameras are subject to a significant magnification change with the change of focus, which makes the focal stack non trivial because of the large magnification change between the edges of the stack. Additionally, a conventional focal stack algorithm performs poorly when applied to very large images due to extensive memory consumption. We therefore suggest a simple tile based focal stack algorithm. Our algorithm processes only two (or optionally three) tiles at each step, which minimizes its memory footprint. Our algorithm also uses the minimal scale change possible at each step. To address possible inconsistencies at the frames' edges due to scaling, we process each tile (except the edges of the focal stack) in both directions to maximize the robustness of the algorithm. Our algorithm is described by Algorithm-1 and is illustrated in Figure 7. The principle of operation is to first create local focal stacks, merge them, and then move up one level and merge these using new local centers until all tiles are merged.

Algorithm 1

Input: A set of n (n odd number) images taken at different focal distances where the DOFs of adjacent images are overlapping.

Output: Extended DOF image

1. Using the known scale and FOE, divide each image into overlapping tiles.

2. For each tile location (x, y) arrange the n focal stack tiles into triplets: $(1, 2, 3), (2, 3, 4) \ldots (n-2, n-1, 2)$

3. For each triplet register (and warp) the two edge tiles to the center one, which is the 'local focal stack center'.

4. For each triplet (i,j,k) merge the pairs (i,j), (j,k), using a Laplacian pyramid [1]

5. For each triplet (i,j,k) merge the newly created pairs (i,j), (j,k) into a single tile. Resulting in a total of $n-2$ tiles

6. If $(n > 2)$ then set $n \leftarrow (n-2)$ and repeat from step number 2 (using different local focal stack center(s)).

7. For each location (x, y) remove the overlapping edges (that are corrupted by the warp operator), and stitch the clean central region to obtain an extended DOF image located at the global center of the focal stack.

Notes: (1) The Overlapping DOF are needed to help the registration - at least on one side. As we merge more images the overlapping region increases. (2) The merging in step 4 is a non linear selection, therefore (a * b) *a = (a * b) where '*' is the blending operator. (3) Each image (except the images at the edge of the focal stack) is warped toward both its neighbors . This provides some redundancy in case one side fails to register.

9. Level of Detail Evaluation

In this section we try to evaluate the level of detail *at the surface of the object* that our camera can obtain using its standard lens. The minimum level of detail required by museums for archiving purposes is 10lp/mm at the object surface [11].

For this test we used the Kodak TL-5003 test chart. This is a paper reflective test chart with maximum resolution of 18lp/mm. We imaged the chart from a distance of 120cm at f-22. Figure 8(a),(b) shows the vertical and horizontal resolution at the center and far corner of the FOV respectively. We can see that the camera can resolve at least 18lp/mm at the object's surface. This translates to an angular resolution of 0.159 minutes of arc. Figure 8(c) shows a star test target and demonstrates the camera resolution uniformity for different orientations.

Figure 8(e) shows a subjective test using a picture of a US\$20 note taken with our camera and with a Nikon-D70 at the same distance and FOV. Figure 8(f) shows 3D details of
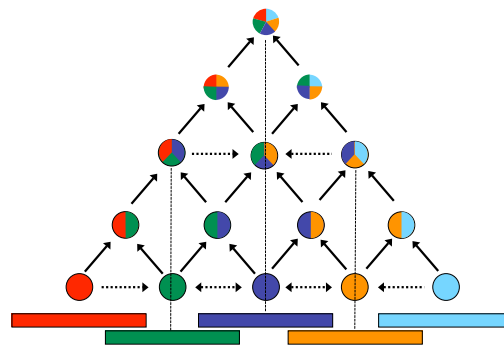

Figure 7. Single location focal stack merging: The colored bars represent the overlapping DOF of each tile. Circles represent images, where multiple colored circles are DOF composite tiles. The horizontal dashed arrows show the direction of image registration and warping. The dashed vertical lines show the different local focal stack centers, (with the middle one also being the global focal stack center). Finally, the solid diagonal arrows show the blending operators that are intentionally redundant (to simplify and increase robustness). See Algorithm-1 for details.
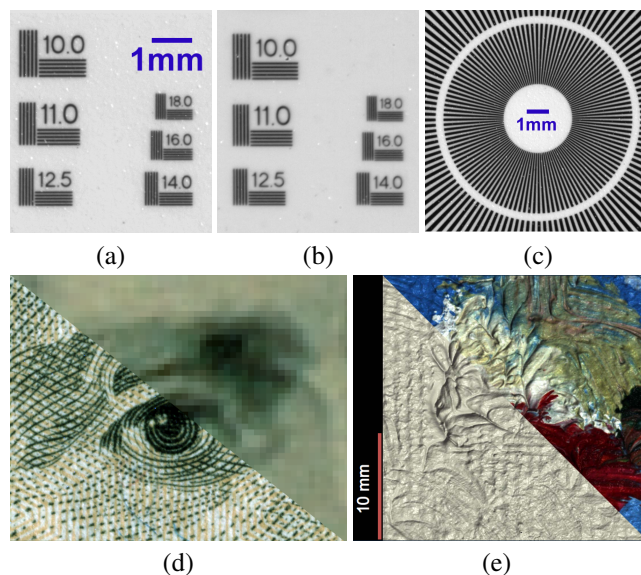

(a)   (b)   (c)

(d)   (e)

Figure 8. Resolution on the object results, (zoom (x6) in to see details). All images were taken at a distance of 120cm between the lens and the target. (a) Kodak test chart TL-5003, Microcopy resolution test pattern taken at f22. Numerals indicate line pairs per mm. Blue bar shows 1 mm world size. We can see that a pattern of 18 lp/mm is clearly resolved. (b) Same as (a) but imaged at the far corner of the field of view. We can see 18 lp/mm are still resolved. (c) Star-test target shows the uniformity of the resolution across different orientations. (d) Part of a US\$20 note. The bottom left part is taken with our camera while the top right half is taken with a Nikon D70 using the same focal length and distance. (e) 3D details of a small patch of an oil painting obtained by photometric stereo and a texture mapped color.

a small patch of an oil painting and a color textured region obtained by photometric stereo using our camera.

10. Resolution Evaluation

The maximal number of pixels of a camera can be easily determined. For our camera it is determined by the scanned area ($450 \times 300mm$), the sensor size (36×24 mm), and the pixel pitch ($9\mu m$) or 1944 mega pixels. However, the resolution of the camera changes with aperture, focal distance, location at the image plane and spatial orientation. Moreover, the color of the imaged object can significantly affect the resolution due to differences in the diffraction limit for different wavelengths, but mostly due to sampling frequencies of the color filter array of the sensor - if one is used.

To evaluate the resolution of the camera we use the ISO12233 slant edge resolution test [3, 17]. This test measures the MTF of the camera for two perpendicular step edges (a slanted vertical edge and a slanted horizontal edge). The spatial frequency for an MTF of 10% is considered the resolution limit by the Rayleigh criterion. According to the ISO12233 procedure we used grayscale (demosaiced) images of a B/W test target. We emphasize that this test does not reflect the worst case scenario of a monochromatic red or blue target that is aligned with the sensor grid - we address the worst case scenario at the end of this section.

The test was conducted at a distance of three meters from the lens, which is close to the peak performance focal distance of the lens, and an aperture of f-27, which is just below the diffraction limit for the pixel pitch. We used a commercial software (www.imatest.com) to process the images and compute the MTF graphs. The computed MTF graphs for the center and corner of the image plane at two perpendicular orientations are shown in Figure 9.

We can see that in all cases, the MTF at the Nyquist frequency of the sensor was above 10%, which indicates that for the test conditions the resolution of our camera is limited by the sensor that has lower resolving power than the lens in these conditions (note that MTF values above the Nyquist spatial frequency of the sensor appear in the chart since the ISO 12233 test procedure super-samples the image before computing the MTF. However, only the values up to the Nyquist frequency are considered for resolution estimation).

Since the image circle of our lens is $500mm$, we only use an area of $400 \times 300mm$, which results with maximal resolution of: $400 * 300 * (2 * 55.56)^2 = 1481$ mega pixels. We repeated the test using lower contrast images (the differences between the average grey levels of the bright and dark sides of the test target images were 190 and 165 for center and edge of the FOV respectively). The results showed that the camera is still sensor limited at the center of the FOV (10% MTF above 55.5lp/mm), and at the edge the 10% MTF was obtained at 46.6lp/mm. This puts the resolution of the camera for the test conditions above 1042 megapixels (the value for 46.6lp/mm).

In the worst case scenario - a monochromatic target that is aligned with the Bayer sampling grid, the sensor's cutoff frequency (at each axis) will drop by $\sqrt{2}$ for the green chan-
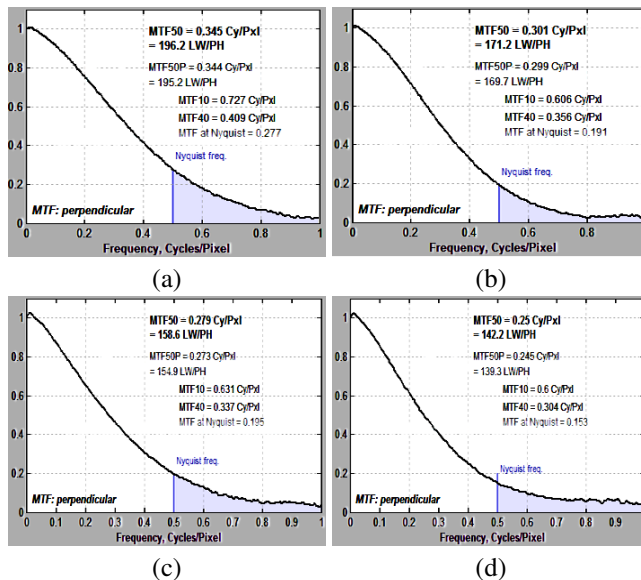

Figure 9. ISO12233 slated edge MTF test results. (a),(b) were computed from a slanted vertical and horizontal edges respectively taken at the center of the FOV. (c),(d) same for the edge of the FOV.
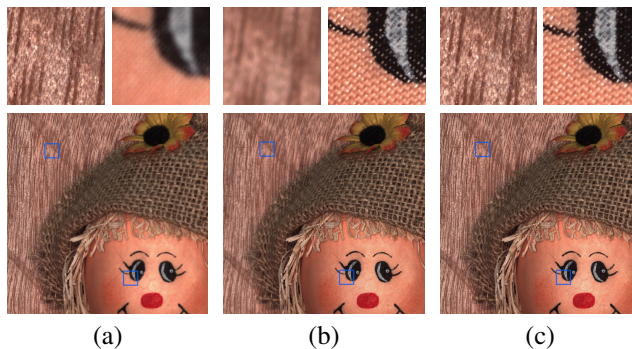

Figure 10. Focal stack example: (a) farthest focus, (b) closest focus (c) focal stack result.

nel and by two for the blue and red channels. This problem can be addressed either by using multi-frame demosaicing [5], or by taking multi-frames using a grayscale sensor with a tricolor strip filter instead of the Bayer sensor.

11. Focal Stack Example

Figure 10 shows a focal stack result using one triplet. The far focal edges are shown in (a) and (b) and the results are shown in (c). Magnified rectangles show the difference in focus in each example.

12. Comparisons

We compare our camera to the Wang et al. low cost camera [17] and to the state of the art (to the best of our knowl-

Table 1. Comparison to scanning back cameras with a linear sensor

	Resolution	Focusing	Color
Wang et al.	122/490MP	focus screen	fltr. wheel
Anagramm	340MP[1]	not specified[2]	tri-linear
Our camera	>1042MP[3]	sensor's video	Bayer[4]

	time @10ms	time @1s	Cost
Wang et al.	2 min / 4 min	n/a	$1.5K[5]
Anagramm	4 min	6.5 hours	$62K[6]
Our camera	5 min	8 min	$25K

1. Requires 100lp/mm over 139mm image circle lens.
2. Most likely it uses a ground glass focus screen like the Linof M 679cs camera.
3. As of ISO 12233 resolution test procedure.
4. Can be replaced with tricolor strip filter to avoid demosaicing.
5. As of writing time, neither the lens nor the view camera kit are manufactured anymore - this cost may not be applicable to alternatives.
6. Scanning back 56.4K, camera 3.6K, lens 2.2K.

edge) professional camera[1]. We compare cost, camera resolution, focusing method, color method, image capture time for $10\ ms$ exposure time and one second exposure time. The capture time is not normalized by the image size and reflects the capture time of a full size image. The comparison is shown in Table 1. The better choice at each category is in green. Red text indicates clear disadvantages. As seen in this table, using an area sensor results in a significantly faster capture time, particularly in low light conditions.

13. Conclusion

In this work we have focused on some of the main aspects related to large format digital imaging. We analyzed the requirement from the lens and sensor, built a large format digital camera and demonstrated its performance. We have also addressed the magnification change problem and calibration issues that are relevant to lagre format imaging and propose a new tile based algorithm for the focal stack of very large images. Our camera fills the gap between medium format digital camera and large focal plane arrays and opens new opportunities for cultural heritage applications and research.

References

[1] E. Adelson, C. Anderson, J. Bergen, P. Burt, and J. Ogden. Pyramid methods in image processing. *RCA engineer*, 29(6):33–41, 1984.

[2] M. Ben-Ezra. High resolution large format tile-scan camera - design and implementation. *MSR-TR-2010-20*, 2010.

[3] P. Burns. Slanted-edge MTF for digital camera and scanner analysis. In *IS AND TS PICS CONFERENCE*, pages 135–138. SOCIETY FOR IMAGING SCIENCE & TECHNOLOGY, 2000.

[4] R. N. Clark. Digital Cameras: Does Pixel Size Matter? In *http://clarkvision.com/imagedetail/index.html*, 2008.

[5] S. Farsiu, M. Elad, and P. Milanfar. Multi-frame demosaicing and super-resolution from under-sampled color images. *Computational Imaging II*, 5299:222–233.

[6] S. Hasinoff, K. Kutulakos, F. Durand, W. Freeman, and M. CSAIL. Time-Constrained Photography. In *ICCV 2009, Kyoto, Japan, Sep. 27- Oct. 4*, pages 333–340, 2009.

[7] E. Hecht. *Optics*. Addison-Wesley, 1998.

[8] J. Kopf, M. Uyttendaele, O. Deussen, and M. Cohen. Capturing and viewing gigapixel images. *ACM Transactions on Graphics*, 26(3):93, 2007.

[9] A. Levin, R. Fergus, F. Durand, and W. Freeman. Image and depth from a conventional camera with a coded aperture. In *International Conference on Computer Graphics and Interactive Techniques*. ACM New York, NY, USA, 2007.

[10] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International joint conference on artificial intelligence*, volume 3, page 3. Citeseer, 1981.

[11] K. Martinez, J. Cupitt, D. Saunders, and R. Pillay. Ten years of art imaging research. *Proceedings of the IEEE*, 90(1):28–41, 2002.

[12] H. Nagahara, S. Kuthirummal, C. Zhou, and S. Nayar. Flexible Depth of Field Photography. In *European Conference on Computer Vision (ECCV)*, Oct 2008.

[13] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2, 2005.

[14] H. S. R. Szeliski. Creating full view panoramic mosaics and environment maps. *Proc. of ACM SIGGRAPH 97*, pages 251–258, 1997.

[15] J. Tonry, P. Onaka, B. Burke, and G. Luppino. Pan-STARRS and Gigapixel Cameras. *ASTROPHYSICS AND SPACE SCIENCE LIBRARY*, 336:53, 2006.

[16] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Transactions on Graphics*, 26(3):69, 2007.

[17] S. Wang and W. Heidrich. The design of an inexpensive very high resolution scan camera system. In *Computer Graphics Forum*, volume 23, pages 441–450. Citeseer, 2004.

[18] M. Watanabe and S. Nayar. Telecentric Optics for Focus Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(12):1360–1365, Dec 1997.

[19] C. Zhou, S. Lin, and S. K. Nayar. Coded Aperture Pairs for Depth from Defocus. In *IEEE International Conference on Computer Vision (ICCV)*, Oct 2009.

[20] C. Zhou and S. K. Nayar. What are Good Apertures for Defocus Deblurring? In *IEEE International Conference on Computational Photography*, Apr 2009.

---

[1]Anagramm David scanning back, Linhof RD 1 camera and Rodenstock Apo-Sironar Digital lens