Automated 3-D Reconstruction Using a Scanning Electron Microscope

Nicolas Cornille†‡ Dorian García† Michael A. Sutton† Stephen R. McNeill†
Jean-José Orteu‡

† University of South Carolina
Department of Mechanical Engineering
300 Main St. — Columbia, SC 29308 (USA)
e-mail: sutton@sc.edu

‡ École des Mines d’Albi, CROMeP
Campus Jarlard
81013 Albi Cedex 09 (France)
e-mail: jean-jose.orteu@enstimac.fr

Abstract

Methods for both the accurate calibration and also the use of an Environmental Scanning Electron Microscope (ESEM) for accurate 3D reconstruction are described. Unlike previous work, the proposed methodology does not require either known motions of a target or use of calibration target with accurate fiducial marks. Experimental results from the calibration studies show the necessity for taking into account and correcting for distortions in the SEM imaging process. Moreover, the presence of high-frequencies components in the distortion field demonstrates that classic parametric distortion models are not sufficient to model distortions in a typical ESEM system. Results from preliminary 3-D shape measurements indicate that the calibration process removes measurement bias and reduces random errors to a range of ±0.05 pixel, which corresponded to ±43 nanometers for the objects being studied in this work.

Keywords: scanning electron microscope, accurate calibration, digital image correlation, stereo-vision, 3-D shape from motion.

1 Introduction

The increasing interest in micro- and nano-scale investigations of material behavior at reduced length scales requires the development of high-magnification, non-contacting, 3-D shape measuring systems. Optical methods have emerged as a de facto standard for 2-D and 3-D macro-scale measurements. Since its early development at the University of South Carolina [33], the digital image correlation (DIC) technique for measuring two and three-dimensional shapes and/or deformations has been shown to be a versatile and effective measurement method due to its high spatial resolution, its high sensitivity and its non-contacting nature. However, the extension of DIC to the micro- or nano-scale with an SEM system faces the challenges of calibrating the imaging system, as well as measuring 3-D shapes using the single imaging sensor usually available.

Many authors have successfully applied the DIC technique to a wide range of macro-scale problems [15, 22, 21, 11, 26, 34, 6, 32]. It has become commonplace to try to apply DIC at the micro- or nano-scale, using either optical microscopy or electron microscopy [31, 24, 12, 18, 25, 36]. However, few authors have investigated the problem of the accurate calibration of their micro- or nano-scale imaging systems, and specifically the determination and correction of the underlying distortions in the measurement process. One reason may be that the obvious complexity of high-magnification imaging systems weakens the common underlying assumptions in parametric distortion models (radial, decentering, prismatic, . . . ) commonly used to correct simple lens systems (e.g., digital cameras) [2, 7].

Recently, Schreier et al. proposed a new methodology to calibrate accurately any imaging sensor by correcting a priori for the distortion [30, 29] using a non-parametric model. The a priori correction of the distortion transforms the imaging sensor into a virtual distortion-free sensor plane using transformations of a grid-less planar target. The same target can then be used to calibrate this ideal virtual imaging sensor using unknown arbitrary motions. As opposed to classical calibration techniques relying on a dedicated target marked with fiducial points of some sort (ellipses, line crossings, . . . ), the approach of Schreier et al. can be applied on any randomly textured planar object (the so-called “speckle pattern”). Since it is rather difficult, if not impossible, to realize a proper classical calibration target aimed at being imaged at high magnification, this method appears well suited for calibrating a SEM, only requiring a suitable planar speckle pattern.

This paper presents our preliminary results for (a) the accurate calibration of a SEM system and (b) the application of the calibrated SEM system for 3-D shape measurement using a series of arbitrary specimen motions. The results indicate that two known translations of a randomly textured specimen are sufficient to transform a SEM into an accurate 3-D shape measurement system for high magnification studies.

2 SEM Imaging Considerations

The SEM imaging process is based upon the interaction between atoms of the observed specimen and the SEM electron beam: a heated filament generates electrons which are accelerated, concentrated and focused using a tray of electromagnetic lenses. When the electron beam strikes the specimen, a variety of signals are emitted which can be measured by dedicated detectors. Within all these emitted signals, two of them are of a particular interest: the secondary electrons (SE) and the back-scattered electrons (BSE). The SE originate close to the
surface of the specimen and are correlated to its topography: the SE signal amplitude is a function of the local orientation of the specimen surface with respect to the detector location. Relative to SE, the BSE signal is emitted from interactions that occur over a slightly larger depth under the specimen surface and is mainly correlated to the local atomic composition.

Fig. 1 — Difference between acquisition with SE (left) and BSE detector (right): The SE detector is strongly influenced by sample topography. The BSE is most strongly influenced by local specimen composition.

2.1 Operating Conditions

SEM operational parameters such as the detector type, accelerating voltage, working distance, etc. will affect the quality of the images. In this regard, the major requirements for image correlation in an SEM are (a) adequate image contrast, (b) random texture in the images, (c) appropriate spatial frequency in the random texture (d) temporal invariance in the images and (e) minimal image changes during the rigid-body motion of the specimen. Since SE detector imaging (which depends on topography of the sample) will violate (e) in most cases, BSE imaging is preferred. Moreover, the surface texture is improved when using the BSE detector (see Fig. 1).

Since BSE imaging is also slightly affected by topography, in this work we chose to acquire all BSE images at a low accelerating voltage. In this way, the primary beam electrons do not penetrate deeply into the specimen. Another advantage of a low voltage incident beam is that surface details are enhanced in comparison to high voltage [27].

2.2 Parallel and Perspective Projections

Fig. 2 — At low magnification, perspective projection is used. At high magnification, parallel projection is typically employed.

3 Calibration Procedure

Our calibration procedure for the SEM follows the methodology developed at the University of South Carolina [29]. It is a two-step process:

1. Determination of the distortion removal function (warping function) based on a series of in-plane translations of a flat, randomly textured, calibration target. This warping function transforms the SEM imaging system into an ideal, distortion-free, virtual imaging sensor.

2. Calibration of this virtual imaging sensor through a traditional bundle-adjustment calibration technique. This step requires another image sequence of the same target undergoing arbitrary rigid-body motions.

As will be shown below, the first step in the calibration process uses digital image correlation to precisely measure the displacements\(^1\) into the image sequence undergone by a set of arbitrary chosen points: the so-called “image matching” or “image registration” problem. This novel approach greatly simplifies the second step in the process by completely eliminating the need for a calibration target made up of fiducial markers such as ellipses or crossing lines. This is particularly important since, for obvious reasons, it is rather difficult, if not impossible, to realize a proper dedicated target which is suitable at high magnifications.

3.1 \textit{A priori} Distortion Removal

Fig. 3 — Motions of the calibration target made with the translation stage for the estimation of the distortion removal functions. The given numerical values are suitable for magnification 200×.

The determination of the warping function which transforms the physical imaging sensor into an ideal, virtual sensor

\(^1\)A displacement of a given object point into the imaging space is often called a “disparity”. A set of disparities related to the same two images is commonly referred as a “disparity map”.

plane is accomplished by acquiring a series of images of a randomly textured planar target undergoing in-plane translations [30, 29]. The theoretical background of this technique comes from the theoretical fact that any perspective transformation (and so any affine transformation) transforms a line into a line [5]. While only two known, non-parallel translations are required to determine the a priori distortion removal function, the accuracy of the warping functions can be greatly improved using additional (unknown) translations. A typical translation sequence is illustrated in Fig. 3 (AB and AC are usually chosen as the two known motions).

As noted previously, the disparity maps for each image of the translation sequence with respect to a common reference image are computed using digital image correlation. To establish the warping function, software has been developed [14] to accurately determine the two B-spline vector functions relating image coordinates to coordinates on a virtual sensor plane; here, the calibration object is the virtual sensor plane.

3.2 Calibration

After determining the warping function that establish the ideal, virtual sensor plane, the virtual imaging system can be calibrated using bundle adjustment techniques. The bundle adjustment technique was originally developed for photogrammetry applications [3, 16, 17], i.e., the shape measurement using multiple views of an object. The technique has since been used with great success for various single camera calibration applications [4, 19]. It is clear that a reliable shape measurement using different views of a rigid object can only be made if sufficiently large triangulation angles are realized between the different orientations of the imaging system. Equivalently, sufficient triangulation angles can be realized using a stationary imaging system by rotating the calibration object in front of it. This principle is used to achieve reliable calibration of the virtual imaging system by moving the calibration target with the SEM rotation stage.

4 Automated 3-D Reconstruction

The most common approach for calibration and use of a single-sensor imaging system in 3-D surface reconstruction is to require accurate knowledge of the rigid-body object motions within a sequence of images (see Fig. 4). To mitigate the requirement for accurate motions, a novel approach is proposed that does not require any information regarding the required rigid-body motions.

Indeed, a relationship exists between any image pair in the scene: the epipolar geometry [5, 23]. This geometry depends upon the relative positions of the two imaging sensor locations. If this geometry can be estimated from a pair of images in the sequence, then the rigid-body transformation between the two respective views can be estimated. By repeating this process for each image pair, the motion between each view and then the general motion of the scene is obtained.

The main problem in implementing this approach is the estimation of the epipolar geometry from a pair of images. This is a rather delicate process and will not be fully developed here; the reader can refer to [1, 35, 37] for details. The first stage in this process is the extraction of some points (called feature points) from each image in order to obtain appropriate corresponding matches throughout the image sequence.

Feature Points Extraction

Point detectors are most often compared with respect to their localization accuracy criterion. For this work, where the precise location of a feature point in the image (e.g., a corner) is not required but the ability to repeatedly identify corresponding points in the detection process is paramount, the work of Schmid et al. [28] indicates that the best detector evaluated with this criterion is the Harris detector [9].

Feature Points Matching

In order to estimate the epipolar geometry between each pair of images, feature points are optimally matched using a Zero mean Normalized Cross Correlation (ZNCC) function. To prevent false matching, a robust method (Least Median of Squares) is used to detect and remove outliers.

Epipolar Geometry and Motion Estimation

A minimum of seven point correspondences is required to estimate the epipolar geometry with an appropriate parameterization. In practice, an 8-point algorithm [20] is implemented which uses eight or more point pairs. To improve the stability of the process, isotropic scaling is applied to the coordinates [10] before carrying out the 8-point algorithm. Finally, an
optimization approach is used to minimize a physically meaningful quantity such as the distance from each point to the appropriate epipolar line. Then, using the intrinsic parameters given by the calibration, the motion is recovered [13, 38] (note that specific scene information is required to recover the scale factor for the magnitude of the translation).

Dense Matching and Refinement of the Estimated Motion

After the epipolar geometry is estimated, one approach to complete the matching process for each feature point restricts the search area to its epipolar line and the surrounding area, oftentimes referred to as the “epipolar band” (thick epipolar line). In this work, a different approach is employed. Using the feature points where corresponding matching points have been identified, an estimation of the disparity map is determined in order to make the dense correlation easier: a triangular mesh is generated [8] from the matched points and then the disparity map is estimated for each point by interpolation into the triangle which contains it. Once the dense disparity map is computed, the epipolar geometry is improved as well as the corresponding motion.

Triangulation

As shown in Fig. 5, triangulation is used to compute the 3-D coordinates of a point that is common to at least two images. The 3-D information for each point is recovered by intersecting the different rays defined by the optical center of the sensor and the projected point in the image. Since the rays may not intersect, an over determined equation system is solved using a bundle-adjustment process to minimize the distance between the 2-D image point and the reprojection of the computed 3-D point.

Since the translation component of the rigid-body motion linking two views is estimated up to an unknown scale factor, the 3-D shape is also reconstructed up to the same scale factor. The scale factor can be determined (a) knowing a distance between two points on the specimen and/or (b) knowing the magnitude of one translation component of the rigid-body motion of a given view in the image sequence.

5 Experiment and Results

For all experiments, a FEI ESEM Quanta 200 is operated in high-vacuum mode (SEM mode). The operational parameters for these studies are as follows:

- BSE detector
- 200× magnification
- non-dimensional spot-size of seven
- 10mm working distance
- accelerating voltage of 8kV
- seven-bit gray scale for all images (1024 × 884 size)
- 1.3mm by 1.1mm field of view

Figure 6 shows three fields of view for a US penny. The 200× image detail shows a speckle texture that is adequate for selective use of digital image correlation to identify and match corresponding subsets during rigid-body motion. Note that the 200× image has a small depth range (< 50μm) does not introduce significant perspective distortion to the disparity maps between the images in a translation sequence. Thus, the randomly textured regions of the coin can be considered “flat” and be used for the computation of the a priori distortion removal function.

5.1 Calibration

![Fig. 7 — Overview of calibration procedure: this first image of the translation sequence is correlated with the all other images of both the sequences.](image)

![Fig. 8 — BSE detector images of the planar aluminum calibration target covered with a thin speckle pattern layer of gold realized by microlithography.](image)

![Fig. 9 — Setup of the experiment: the coin to measure is stuck using a thin adhesive on an aluminum wafer covered with a gold speckle pattern deposited by microlithography.](image)
Procedure

As shown in Fig. 7, the FEI ESEM is calibrated (a) using the a priori distortion removal technique based on a 29 image sequence of a translated target and (b) using bundle adjustment calibration of the perspective model based on a 24 image sequence of the same target undergoing arbitrary rigid-body motions.

Since the a priori distortion removal technique requires that a planar calibration target contains an appropriate speckle pattern texture, an aluminum plate covered with a gold speckle pattern realized by a microlithography process (see Figs. 8 and 9) is employed for one set of experiments. An- other set of experiments used the random texture on the US penny. Note that the relative “flatness” of the coin detail should not produce significant heterogeneities in the disparity maps when the target is slightly translated at magnification 200×.

Results Using the Micro-lithography Target

Figure 10 shows the distortion field of the SEM imaging system when using the target shown in Fig. 8. The magnitude of the distortion is up to 4 pixels in the corners of the images (size 1024 × 884). Careful inspection of the data in Fig. 10 indicates that the distortion field contains high-frequency distortion components. To investigate these features, a high-pass filter is applied to the distortion field shown in Fig. 10. Fig. 11 displays the two-dimensional form for the high frequency components. The data in Fig. 11 clearly shows a periodic structure in the distortion along the x- and y-axis of the images, with an amplitude of up to 0.1 pixel. This high-frequency distortion field is probably related to the scan control system for the electron beam.

After performing the calibration process using bundle-adjustment, the rotation sequence gives a standard deviation of the re-projection errors in the virtual sensor plane of 43 nanometers (error is less than 0.05 pixel in image space).

Results Using Coin Texture

Figure 12 shows the distortion field of the SEM imaging system when using the random texture on the coin as the target for the a priori distortion removal technique. The results are very similar to those obtained by using the “ideal” planar target.

The high frequency components are shown in Fig. 13. These components are slightly disturbed compared to those computed with the planar target sequence.

After performing the calibration process by bundle-adjustment, the rotation sequence gives a standard deviation of the re-projection errors in the virtual sensor plane of 91 nanometers, which represents an error of less than 0.1 pixel in image space. Note that these values also include the correlation errors which are higher with the coin texture than the “ideal” speckle pattern used in the previous section.

Interestingly, the bundle adjustment technique re-estimates the shape of the calibration target (the coin) while also computing the perspective model parameters of the imaging system. Due to the a priori distortion removal technique constraint, the target is first assumed to be perfectly flat. Assuming a flat object as the initial guess for the calibration by bundle-adjustment, Fig. 14 shows that the shape of the target is properly re-estimated.

5.2 3-D Reconstruction

The process described in Section 4 is used to complete the calibration and 3-D reconstruction process for a specific example. For this application, four images of the US penny were acquired at 200× (see Fig. 6). The four images were obtained after undergoing unknown rigid-body motions.
Since the US penny texture is used to obtain feature points and perform image correlation in this application, then the distortions shown in Figs. 12 and 13 will be similar to those present in these images. Hence, the distortion results for this application are not reported.

**Feature Point Extraction**

A set of feature points is first extracted using Harris detector. This set is then processed to keep only the best feature points in a given circular neighborhood (typically 5 pixel radius) with respect to their Harris "cornerness" function response. Using this approach, good feature points are regularly scattered throughout the image and the epipolar geometry to estimate in the next steps is likely to be better [38]. Depending upon the image being used between 9200 and 9800 points are extracted.

**Robust Matching**

For each pair of images, feature points are matched using:

- ZNCC criterion
- 15 x 15 pixels correlation window
- Correlation threshold of 70%
- Least Median of Squares method to detect and remove outliers (false matches) while estimating the epipolar geometry

An average of 3100 pairs of points are robustly matched from each set containing initially 9500 points (about 100 initial matches were removed as outliers).

**Motion Estimation**

As the feature points are only extracted with pixel accuracy, the computation of the epipolar geometry is a first estimation for the following stage (at this point, the mean distance of points to their epipolar line is 0.26 pixel). An estimation of the motion is recovered and a first approximation of the 3-D shape is computed. Figure 15 shows the reconstructed shape based on 3100 matched feature points.

**Dense Matching**

From the matched feature points obtained previously, an estimation of the dense disparity map is computed and used as an initial guess for an optimization approach (maximization of the ZNCC score with affine transformation of the correlation window). Then, the epipolar geometry is refined (mean distance of points to their epipolar line is now 5.10^{-3} pixel) improving the motion estimates.

**Triangulation**

Figure 16 shows the 3-D shape of the coin detail reconstructed by triangulation of the dense sub-pixel disparity map. The results in Figure 16 confirm that the calibration and 3-D reconstruction processes can be performed with arbitrary rigid-
body motions using a single sensor imaging system such as the FEI ESEM.

6 Conclusion

Using a priori distortion removal methods, a FEI ESEM system has been calibrated and used to accurately measure the 3-D shape of an object. The experimental results from the calibration studies show the necessity for taking into account and correcting for distortions in the SEM imaging process. Moreover, the presence of high-frequencies components in the distortion field demonstrates that classic parametric distortion models are not sufficient in the case of an SEM. Furthermore, the results from preliminary 3-D shape measurements are promising; calibration accuracy is estimated to be ±0.05 pixel, which corresponded to ±43 nanometers for the objects being studied in this work.

7 Future Work

Work is on-going to utilize the a priori distortion removal process and extend the work to the accurate measurement of strains in SEM systems. Issues such as temporal stability, presence of an unknown scale factor in the 3-D reconstruction process and accuracy of the measurements will be primary areas of emphasis.

Acknowledgment

The authors wish to thank Dr. Oscar Dillon, former director of the CMS division at NSF, and Dr. Julius Dasch at NASA HQ for their support of this work through grants NSF-CMS-0201345 and NASA NCC5-174. In addition, the support of Dr. Dana Dunkelberger, Director of the USC Microscopy Center, by providing unlimited access to the FEI ESEM for our baseline studies is gratefully acknowledged.

References


