

Resection of Long-Range Sensor Models for Mono and Stereo Exploitation of Non-Earth Imagery

Reuben Settergren

BAE Systems Information and Electronic Systems Integration Inc.

ABSTRACT

Measurement of Non-Earth Imagery (NEI) for Space Domain Characterization (SDC) is enabled by applying the principles of projective geometry to one or more images. Estimates of focal length, pixel size, and range are needed to establish scale. The analyst must decide on a target-body-fixed coordinate system, and measure the slopes of all three axes in image space, to establish rotation. Range determines the vertical component of camera translation, and choice of object-space origin on the body establishes horizontal translation, which is sufficient for a pinhole camera (camera matrix). The camera model supports transformation of object coordinates into pixel space, and projection of image coordinates onto a chosen plane in object space. Triangulation of image rays from two or more images allows estimation of 3-D coordinates of arbitrary points on the body.

1. INTRODUCTION

Measurement of 3-D objects observed from a distance in 2-D imagery is a fundamental photogrammetry task. Modern photogrammetry relies heavily on metric sensors that have a precise understanding of their position and orientation relative to a stationary scene. Scenes on the surface of the Earth may be observed from long-range (satellite) or medium-range (airborne) distances. Close-range imagery may be used for industrial part inspection (without reference to absolute geopositioning). All of these use cases rely on the 3-D scene content being stationary, and the camera moving to different positions and orientations relative to that fixed scene.

But if an object of interest has moved or rotated between collected images, the necessary conditions of stationary scene content do not hold. Existing capabilities [4], [5], [3] to measure 3-D objects in arbitrary poses are focused on close-range imagery (exhibiting vanishing points), and are not applicable for long-range cases such as Space Domain Characterization (SDC) of Non-Earth Imaging (NEI). For SDC/NEI contexts, an orbiting sensor which images another orbiting object will likely have an accurate understanding of its own position and pose, and a rough range to the target, but not a precise understanding of the natural axes of the observed body, and how they are moving and rotating over time.

We present a capability that creates synthetic long-range sensor models, which treat the axes of the observed rigid body as a fixed reference, and reposition the sensor to a pose in the object's coordinate system. A sensor model for a single image supports a variety of monoscopic measurements, such as lengths and angles (even for structures not parallel to the imaging focal plane, if it is known/assumed how they align with body XYZ axes). But having camera models for two or more images enables full stereo exploitation – determining the 3-D XYZ coordinates of every point that can be observed in multiple images. This opens the door to 3-D reconstruction of the object, stereo viewing for visual assessment, etc.

2. METHOD

The first step in synthesizing camera models is to select a coordinate system. A right-handed, orthogonal coordinate system should be chosen, for which object-space-parallel lines are observable in all images. The conventions of this

SOCET GXP is a registered trademark of BAE SYSTEMS. This document gives only a general description of the product(s) or service(s) offered by BAE Systems. From time to time, changes may be made in the products or conditions of supply.

Approved for public release as of 09/03/2024; This document consists of general information that is not defined as controlled technical data under ITAR Part 120.10 or EAR Part 772. 20240903-20.

work are illustrated in Fig. 1, with object axes chosen so that in the image perspective, X will be to the right, Y will be up, and Z will be most into the camera. (Without loss of generality, X can be in any direction in image space, as long as Y is counter-clockwise from it so that right-handed Z is toward the viewer.) The image axes x, y are also indicated in parallel in object space. The image plane is shown in positive, in front of the focal center (rather than behind C in mirror image). The perpendicular line segment from the focal plane's 'principal point' (c_x, c_y) to C is the focal length f , in units of pixels. Image z , out of the camera, completes a right-handed orthogonal system with x and y . Camera rotation matrix R describes the rotation between XYZ and xyz .

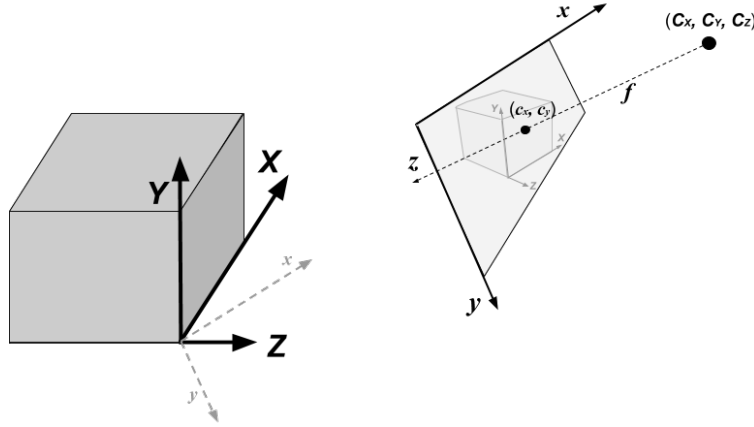


Fig. 1: The scene has right-handed axes X, Y, Z , with Z most into the image, and X and Y projecting into image space as nearly right and up. The image axes x and y are right and down from the upper-left of the image, and can be seen paralleled in object space. Object-axis Z is the axis most toward the camera; Image z is out of the camera.

3-D points in the scene will be indicated with capital coordinates like (X, Y, Z) , and 2-D image coordinates (pixels from the image upper-left corner) will be indicated like (x, y) . It is convenient to present coordinates as row-vectors in inline text, but context may dictate they are actually used as column-vectors.

It is possible to observe the origin and lines parallel to the X, Y , and Z object axes, as they appear in perspective in 2-D image coordinates. For normal handheld camera perspective (short focal length, small range), groups of object-space-parallel lines will converge to vanishing points (the number of vanishing points depends on how many object axes the image plane is parallel/perpendicular to). But as the focal length increases, object-space-parallel lines appear closer to parallel in perspective, and converge to vanishing points further and further outside the image. At telescopic focal lengths, vanishing points are so far outside the image that, within the bounds of the image, object-space-parallel lines are practically parallel (that is, parallel to within a small fraction of a pixel). This work assumes that condition: that all sets of lines which are parallel in object space, are measurably parallel in image space.

2.1 Camera Matrix

The standard [3] matrix representation of a projective camera (equivalent to the photogrammetric collinearity equations [1]) uses a calibration matrix:

$$K = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

On the diagonal of K , f represents the focal ratio, in units of pixels (the ratio between the physical focal length and the size of an individual pixel on the focal plane). $PP=(c_x, c_y)$ are the image coordinates of the principal point (where the image ray is perpendicular to the focal plane). K summarizes the calibration or 'interior orientation' of the camera.

The pose or 'exterior orientation' is captured in a rotation-translation matrix

$$Rt = [R|t],$$

where R is a 3×3 rotation matrix, and t is the 3×1 focal center, in the camera's xyz frame of reference (note how the

image axes in Fig. 1 are paralleled in object space); if the camera position in object-space coordinates is (C_X, C_Y, C_Z) , then $t = -RC$ and $C = -R't$.

Multiplied together, $K \cdot Rt$ is the 'camera matrix.' The projection of object coordinates into image space is

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} = K \cdot Rt \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

The 3-dimensional coordinates of an object point (X, Y, Z) are augmented with a fourth homographic scale coordinate of 1, and after application of the 3×4 camera matrix, the output is a 3-dimensional image coordinate, also augmented with a homographic scale. The actual 2-D image coordinate to which the object coordinate projects is then $(x/w, y/w)$. This homographic unscaling is well-defined everywhere except for camera center (C_X, C_Y, C_Z) , which is the only point that yields $w = 0$. Points behind the camera yield $w < 0$.

2.2 Rotation

The orientation of the camera can be uniquely determined by the slopes of the 3 parallel object-axis directions, as observed in the perspective image.

For close-range imagery, the Principal Point $PP = (c_x, c_y)$ is the unique point on the focal plane for which the imaging ray is perpendicular; all other imaging rays spread out around the PP toward the scene. But for long-range imagery, not only are all object-space-parallel lines practically parallel in perspective, all outgoing image rays are practically perpendicular to the focal plane. So the PP can be chosen as $(0, 0)$ or the center of the image (or any other point), as convenient.

In addition, long range necessarily means that $\text{Range} \sim t_z \gg t_x, t_y$, and (t_x, t_y) captures whatever shift perpendicular to the range direction is necessary to capture the distance of object (X, Y, Z) from the PP. Thus (t_x, t_y) can also be assumed to be $(0, 0)$.

Choosing $(c_x, c_y) = (t_x, t_y) = (0, 0)$, and labelling the individual elements of Rt from (1), the object origin projects to:

$$\begin{aligned} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_X^X & R_X^Y & R_X^Z & 0 \\ R_Y^X & R_Y^Y & R_Y^Z & 0 \\ R_Z^X & R_Z^Y & R_Z^Z & t_z \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} &= \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ t_z \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 0 \\ t_z \end{bmatrix} \\ &\equiv \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{aligned}$$

We can project any object X-axis coordinate $(X, 0, 0)$ into image space as follows:

$$\begin{aligned} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_X^X & R_X^Y & R_X^Z & 0 \\ R_Y^X & R_Y^Y & R_Y^Z & 0 \\ R_Z^X & R_Z^Y & R_Z^Z & t_z \end{bmatrix} \begin{bmatrix} X \\ 0 \\ 0 \\ 1 \end{bmatrix} &= \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} XR_X^X \\ XR_Y^X \\ t_X \end{bmatrix} \\ &= \begin{bmatrix} f(XR_X^X) \\ f(XR_Y^X) \\ XR_Z^X + t_z \end{bmatrix} \\ &\equiv \frac{f}{R_Z^X + t_z/X} \begin{bmatrix} R_X^X \\ R_Y^X \end{bmatrix} \end{aligned}$$

Thus, the object X-axis projects into the image as a line from the origin with slope $\frac{R_Y^X}{R_X^X}$, and because of the parallelism of long-range perspective, all object-space lines in the direction of the X-axis have that same slope in the image.

Reversing that logic, if we observe the slope of any/all parallel X -axis lines in the image, that provides a constraint on the ratio of rotation matrix terms R_X^X, R_Y^Y . Similarly, the apparent slopes of Y - and Z -lines in perspective, constrain the ratios of R_X^Y, R_Y^Y and R_X^Z, R_Y^Z .

Being a rotation matrix, R must be orthogonal. So if the slopes of projected X, Y , and Z lines are (respectively) $\frac{\delta y_X}{\delta x_X}, \frac{\delta y_Y}{\delta x_Y}, \frac{\delta y_Z}{\delta x_Z}$, then for some scalars s_X, s_Y, s_Z , and values a, b, c ,

$$R = \begin{bmatrix} R_X \\ R_Y \\ R_Z \end{bmatrix} = \begin{bmatrix} s_X \delta x_X & s_Y \delta x_Y & s_Z \delta x_Z \\ s_X \delta y_X & s_Y \delta y_Y & s_Z \delta y_Z \\ a & b & c \end{bmatrix}$$

Orthogonality constraints gives us constraints:

$$R_X \cdot R_X = 1$$

$$R_Y \cdot R_Y = 1$$

$$R_X \cdot R_Y = 0$$

Which allow us to solve for the (squares of the) scalars as follows:

$$\begin{bmatrix} \delta x_X^2 & \delta x_Y^2 & \delta x_Z^2 \\ \delta y_X^2 & \delta y_Y^2 & \delta y_Z^2 \\ \delta x_X \delta y_X & \delta x_Y \delta y_Y & \delta x_Z \delta y_Z \end{bmatrix} \begin{bmatrix} s_X^2 \\ s_Y^2 \\ s_Z^2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad (2)$$

The solved scalars determine the first two rows of R . The third row a, b, c can be solved for as the cross-product of the first two rows.

2.3 Rotation Insolubility/Ambiguity

Not every possible combination of slopes $\frac{\delta y_X}{\delta x_X}, \frac{\delta y_Y}{\delta x_Y}, \frac{\delta y_Z}{\delta x_Z}$ can arise from real imaging conditions. Namely, slopes for which the solution of (2) determines that any $s_*^2 < 0$ do not result in a real-valued orthogonal matrix R .

Fig. 2(a) shows four legal near-nadir configurations. The X - and Y -axes are red and green, and the foreshortened Z -axis is blue. The perspective XY angle being slightly greater or less than 90 degrees, is compatible with the Z vector being in quadrants I,III or II,IV, respectively. The other parity options lead to unsolvable R .

Fig. 2(b) demonstrates generally which XYZ -axis combinations are R -solveable. According to the coordinate system convention of section 2, the angle from the X -axis to the Y -axis must be between 0 and 180 degrees (beyond 180 degrees would imply a left-handed coordinate system). The angle between X - and Z -axes is not constrained. The combinations highlighted by the green triangles are R -solveable; while combinations outside the legal zones are unsolvable due to causing at least one $s_*^2 < 0$ in equation (2)

Fig. 3 demonstrates an ambiguous configuration. If any of the two projected object axes are parallel in the image, there are multiple camera orientations that can project object axes into that perspective. In equation (2), two columns of the matrix are identical or parallel, so the system is singular. In order to unambiguously solve for the rotation matrix R with which the camera observed the scene, additional information about scale or aspect ratio is necessary; such as, are the windows in Figure 3 square? Or a known aspect ratio in portrait or landscape orientation?

2.4 Translation

As mentioned previously, “long-range” means $t_Z \gg t_X, t_Y$. For range r , a camera (1) with $t = (0; 0; r)$ and PP= (c_x, c_y) (center of the image), projects object $(0,0,0)$ to image (c_x, c_y) . If a specific point on the body of the observed object is desired to be the object origin, then t_x, t_y can be adjusted to shift the camera parallel to the focal plane, without otherwise affecting the perspective. Since vector $t = (t_x, t_y, t_z = r)$ is then slightly larger than desired range r , t can be slightly scaled down to exactly $\|t\| = r$.

2.5 Scale

Each of the quantities: range, focal length, pixel size; linearly affect the scale of measurement using camera matrix (1). For a long-range calculation, typically the details of the sensing camera (focal length and pixel size) are known precisely, and the trajectory of the sensing platform is known somewhat better than the trajectory of the target object

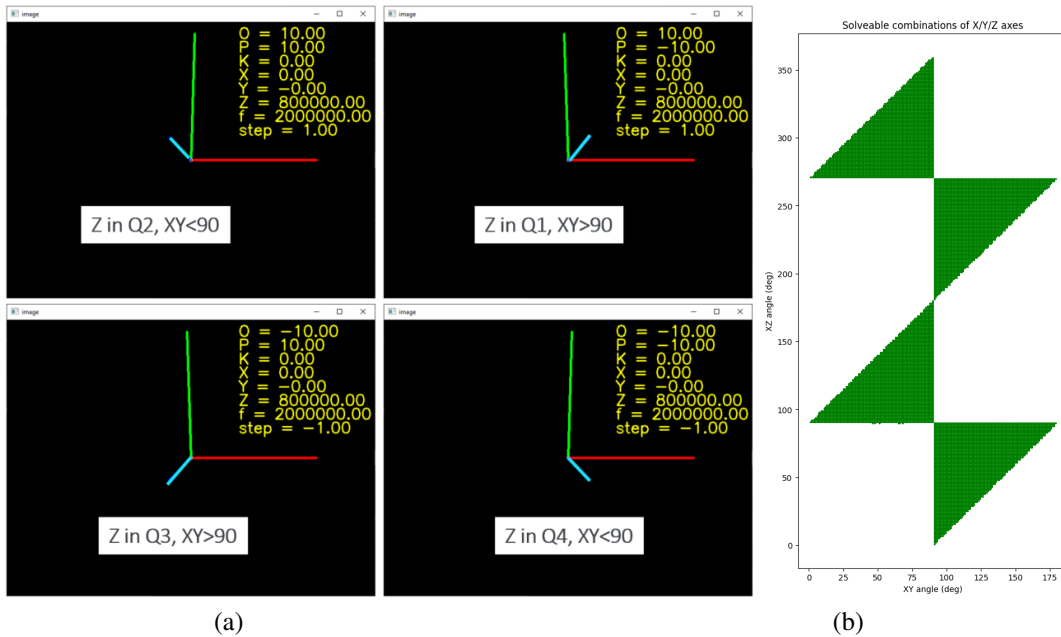


Fig. 2: (a) Near nadir views demonstrating the relationship between XY angle $\lessgtr 90$ and Z quadrant for four feasible configurations. (b) R -solveable zones for XZ angle, given XY angle.

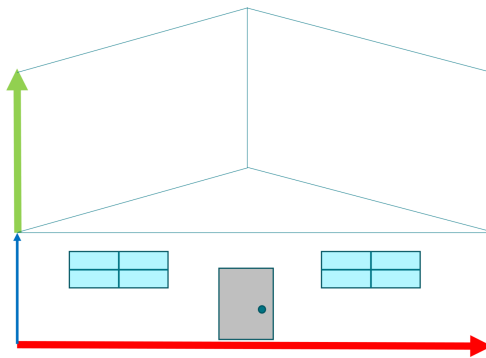


Fig. 3: Ambiguous collection perspective, with observed Y (horizontal roofline/green) and Z (vertical building edge/blue) axes parallel in image-space. This building could have windows that are tall, or wide, or square.

(although the target trajectory must necessarily be known well enough to orient the camera to be able to catch the target in frame). Thus the principal source of error is the uncertainty of the range to target.

If the size of the observed body is known (for instance the length of a feature from the chosen object origin to a point along the X -axis) is s meters, then the solved camera matrix can be used to project object point $(s, 0, 0)$ into image space. If the projected image coordinate is exactly on the point which is known to be s meters from the origin, then the scale is correct. But if the desired point is p pixels from the origin, and the projected point is p' pixels from the origin, then the scale is incorrect by a factor of $\frac{p}{p'}$. Correcting the range by a factor of $\frac{p}{p'}$ and re-solving for (1) will cause the object coordinate $(s, 0, 0)$ to project exactly where it should.

If the size of the observed body is not known, and two or more images are to be exploited in stereo, it is necessary to ensure their scales are same. For a feature observable in all images, a fixed scale length can be imposed to correct all ranges as described above. Absent any information of any one range being known more accurately than another, the common range/scale can be from one image imposed on other(s), or an intermediate correction. (If the size of the

observed body is known, that information can be used to impose common scale on all images.)

3. EXPLOITATION

Once a camera matrix of the form (1) is determined, standard photogrammetric/computer vision techniques can be used to collect measurements of the object. For a fuller treatment, see reference works such as [3], [1].

3.1 Monoscopic

Camera matrix (1) transforms 3-D coordinates into 2-D (after accounting for the homographic coordinates), thus must necessarily be a surjection (each 2-D output is projected to by multiple 3-D inputs). The set of 3-D points which project into a specific 2-D pixel coordinate, comprise the *locus* of that pixel. The locus of a pixel is the linear ray that can be visualized as emanating from the focal plane at that pixel (although physically it is the path of photons entering the camera).

If $P = K[R|t]$ is a 3×4 camera matrix, and P' is its 4×3 SVD pseudo-inverse, then

$$C = -R't$$

is the position of the camera center in object-space, and

$$\begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} = P' \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

is an arbitrary homogenous point on the image ray out of pixel (x, y) . The line through C and $(X/W, Y/W, Z/W)$ is the locus of that pixel, and can be intersected with any plane in object space (most usually, a horizontal plane at a specific value of object-space Z).

This entire analysis depends critically on an assumption of orthogonality, and that certain features of the object can safely be assumed to be linear/planar as they appear, and perpendicular. So for planar surfaces that are (assumed to be) in the principal axes, pixel loci can be intersected with those principal planes to determine object-space coordinates. But planes that have an unknown orientation in object-space cannot be exploited in this way.

For example, for an extent which lies in an XY plane, the pixel coordinates of the observed endpoints can be projected out of the camera to image rays. If the Z coordinate of the plane is known, the image rays can be intersected with the specific plane to determine the 3-D coordinates of the endpoints on the object. But if the Z coordinate is not known, the image coordinates can at least be projected onto the $Z = 0$ plane, and the distance between them calculated; because of the long range, image rays within the image are nearly parallel, the distance resulting from intersecting with a plane a little nearer or further than $Z = 0$ is insignificant.

The previous type of measurement works for any line observed to be in a principal plane, regardless of the orientation of the line. For lines that are observed to be parallel to a principal axis, lengths along that axis can be determined.

For instance, if two points on a 'vertical' edge (parallel to the Z axis) are observed, then the 'height' of the feature can be measured. The 3-D location of one endpoint can be (without loss of generality) be situated on the $Z = 0$ plane, say at object-space position $(X_0, Y_0, 0)$. Then the image ray projecting from the other endpoint, can be intersected¹ with the vertical ray

$$(X_0, Y_0, 0) \rightarrow (X_0, Y_0, 1)$$

to determine the location of the other end point in object-space.

Typically, monoscopic mensuration depends on assumptions such as these, that observed points lie on a common line or plane comprised of principal axes. Thus, points or surfaces that are arbitrarily diagonal or curved, cannot be measured with confidence.

¹Due to inevitable measurement errors, the locus and the vertical ray will not actually intersect in 3-D space. Their 'intersection' will be a point close to both, an equivalent problem to stereo-/multi-sopic intersection as in section 3.2.

3.2 Stereoscopic

If two or more images are available of the same object, and camera matrices can be determined for both/all of them, then stereo-/multi-scopical exploitation can be used to determine 3-D coordinates of any point observable in multiple images. In order to assure proper alignment between camera matrices, sections 2.4 and 2.5 should be used to ensure a common origin and scale.

Then if a point is observed in two images, the 3-D location of the point is the intersection of the image rays projecting out of the two cameras from the observed pixels. Since the camera matrices will not be perfect, the image rays will not actually intersect, so the 3-D location can be taken to be the least-squares solution for the point closest to all the image rays. (When there are two images, it is generally sufficient to solve for the midpoint of the closest approach of the two image rays.)

By these means, the object-space coordinates of point observable in multiple images can be determined, which enables 3-D reconstruction of the object (as much as is observable, potentially augmented with assumptions of symmetry).

3.3 Software

The solution to camera matrix (1) can be converted to equivalent camera position, and euler angles ω , ϕ , κ (as well as known focal length and pixel size, and center-of-image principal point). Such data can be exploited in any software package that is designed to handle blocks of Airborne frame imagery. If an unanchored coordinate system is not supported, the camera positions can be considered offsets to an arbitrary geolocation.

SOCET GXP[®] [2] is one such software package. Once position/orientation data is imported as frame camera(s), SOCET GXP serves as an “electronic light table” (ELT) for convenient exploitation in mono or stereo; supporting for instance marking the coordinates of points, measuring lengths and angles, scribing 3-D subfeatures such as boxes or cylinders of arbitrary aspect, etc.

But more importantly, with appropriate stereo viewing hardware, an ELT can present a stereopair of images directly to the left and right eyes, giving the viewer the subjective experience of seeing the object in 3-D. When the capabilities of the human vision system are thus engaged, many details on the object may become apparent to the viewer – details that don’t stand out when viewing two separate 2-D images.

4. CONCLUSION

A method is presented for synthesizing camera models for long-range NEI imagery. The necessary inputs are focal length, pixel size, range, and image markings of a chosen object-space origin, and lines on the body of the object which are observed/assumed to be in the perpendicular X, Y, Z axes of the body.

Long-range perspective causes projected object-parallel lines not to converge to vanishing points, but to also be parallel in the imagery. The trio of slopes of X -, Y -, and Z -axes in the imagery are explained only by a unique rotation matrix of the camera. The position of the camera is determined by backing the rotated camera up by the known range, and translating the camera parallel to the focal plane in order to align object coordinate $(0,0,0)$ with the marked origin. Scale bars can be imposed to refine the range.

After camera(s) are established for one or more images, traditional photogrammetric techniques can be used to visualize and measure the object monoscopically or stereoscopically.

REFERENCES

- [1] ASPRS. *Manual of Photogrammetry, 4th ed.* ASPRS, Falls Church, VA, 1980.
- [2] BAE Systems Inc. SOCET GXP Software, version 4.5.1.4. www.GeospatialExploitationProducts.com, 2024.
- [3] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, Cambridge, 2 edition, 2003.
- [4] Reuben Settergren. Resection and monte-carlo covariance from vanishing points for images of unknown origin. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2020:487–494, 2020.

- [5] James R. Williamson and Michael H. Brill. *Dimensional Analysis from Perspective: A Reference Manual*. ASPRS, Falls Church, VA, 1990.

SOCET GXP is a registered trademark of BAE SYSTEMS. This document gives only a general description of the product(s) or service(s) offered by BAE Systems. From time to time, changes may be made in the products or conditions of supply.

Approved for public release as of 09/03/2024; This document consists of general information that is not defined as controlled technical data under ITAR Part 120.10 or EAR Part 772. 20240903-20.