Scale factor in digital cameras

Anthony P. Badali, Yahui Zhang, Peter Carr, Paul J. Thomas, Richard I. Hornsey

Anthony P. Badali, Yahui Zhang, Peter Carr, Paul J. Thomas, Richard I. Hornsey, "Scale factor in digital cameras," Proc. SPIE 5969, Photonic Applications in Biosensing and Imaging, 59692B (13 October 2005); doi: 10.1117/12.632215

SPIE.

Event: Photonics North, 2005, Toronto, Canada

Scale Factor in Digital Cameras

Anthony P. Badali[a], Yahui Zhang[a], Peter Carr[a], Paul J. Thomas*[b], Richard I. Hornsey[a]
[a]Vision Sensor Laboratory, York University, 4700 Keele St., Toronto, ON, Canada, M3J 1P3
[b]Topaz Technology Inc., Toronto, ON, Canada

ABSTRACT

For objects on a plane, a "scale factor" relates the physical dimensions of the objects to the corresponding dimensions in a camera image. This scale factor may be the only calibration parameter of importance in many test applications. The scale factor depends on the angular size of a pixel of the camera, and also on the range to the object plane. A measurement procedure is presented for the determination of scale factor to high precision, based on the translation of a large-area target by a precision translator. A correlation analysis of the images of a translated target against a reference image is used to extract image shifts and the scale factor. The precision of the measurement is limited by the translator accuracy, camera noise and various other secondary factors. This measurement depends on the target being translated in a plane perpendicular to the optic axis of the camera, so that the scale factor is constant during the translation. The method can be extended to inward-looking 3D camera networks and can, under suitable constraints, yield both scale factor and transcription angle.

Keywords: Imager, Scale Factor, Correlation, Calibration, Camera

1. INTRODUCTION

A camera characterizes its environment as a two-dimensional image. Typically, the images from one or more cameras are used to analyze certain aspects of the scene, such as the size or position of objects. A camera must be calibrated to successfully infer spatial relations from its images. Using the pinhole approximation and homogeneous coordinates, the mapping between positions in the world, **X**, and their corresponding image plane locations, **x**, is governed by the projection equation[4] (1):

$$\mathbf{x} = \mathbf{PX} \tag{1}$$

The camera projection matrix, **P**, is determined by calibrating the camera. The merits of several techniques have already been investigated in previous work[6, 10, 11, 12]. In some circumstances, the camera may be further characterized by decomposing **P** into its intrinsic and extrinsic parameter matrices **K** and **M** respectively[4] (2):

$$\mathbf{P} = \mathbf{KM} \qquad \mathbf{K} = \begin{bmatrix} s_x & \gamma & x_0 \\ 0 & s_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{M} = \left[\mathbf{R}\middle|\mathbf{t}\right] \tag{2}$$

Here, s_x and s_y are the diagonal terms that give the scale factors for the x and y directions in the image, while γ describes the relative orthogonality of the image axes. In the present work, $\gamma = 0$ is assumed, which is normal for digital cameras[4]. The label "scale factor", *S*, is used to denote either s_x or s_y as determined by context, in order to emphasize the physical meaning of the parameter. The parameters of a camera's internal matrix are often found in a generalized least-squares fit of calibration measurements (from imagery) to the above parameters of the projection model. *A priori* knowledge of parameter values, such as $\gamma = 0$, can improve the convergence and accuracy of the fitting algorithm[4]. A precisely measured scale factor can be included in such a fitting algorithm as an additional constraint. The present work is

* thomas@cs.yorku.ca; phone 1 416 736-2100 x 33343; www.cs.yorku.ca/~visor

Photonic Applications in Biosensing and Imaging, edited by Warren C. W. Chan, Kui Yu, Ulrich J. Krull, Richard I. Hornsey, Brian C. Wilson, Robert A. Weersink, Proc. of SPIE Vol. 5969, 59692B, (2005) · 0277-786X/05/$15 · doi: 10.1117/12.632215

intended primarily for test/measurement applications where the scale factor is the only camera calibration parameter of interest.

In the present context, the "scale factor", *S*, is defined as the ratio of the physical dimension of an object (in microns), d_{object}, to its size (in pixels), d_{image}, as determined from an image of the object. As illustrated in Figure 1, the scale factor in the thin-lens limit is approximately:

$$S = \frac{d_{object}}{d_{image}} = \frac{R}{i_{eff}} \quad [\mu m/pixel] \tag{3}$$

Here, d_{object} is the distance translated by the object in the object plane (in μm), d_{image} is the change in position of the corresponding image (in pixels in the image), R is the distance from the lens plane to the object (in μm), and i_{eff} is the effective distance from the lens plane to the imager (in pixels). The distance (in μm) between the ideal lens plane and image plane is recovered by scaling i_{eff} by the edge length of a pixel, d_{pixel} (in μm/pixel).

When the scale factor, Equation (3), is normalized to unit range, it is simply the inverse of the image plane distance, expressed in pixels. The Gaussian-optics approximation for the image distance is shown in Equation (4)

$$S_{norm} = \frac{S}{R} = \frac{1}{i_{eff}} = \frac{1}{f} - \frac{d_{pixel}}{R} \tag{4}$$

Here, the f represents the focal length of the lens (in pixels). At large range, $i_{eff} \approx f$, the focal length of the camera lens and S_{norm} become consistent with the diagonal term s_x (or s_y) in the camera internal matrix of the projection model[4].

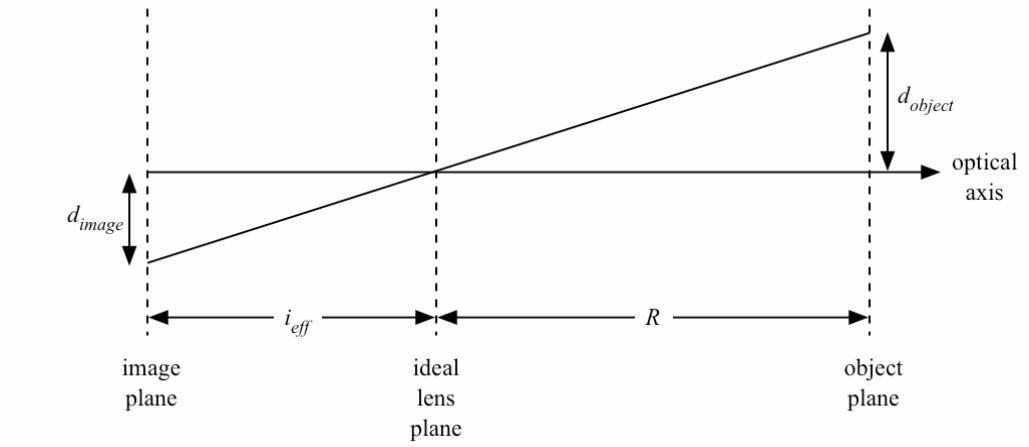


Figure 1: Scale Factor for Pinhole Cameras

For asymmetrical camera pixels, d_{pixel} depends on the orientation of the camera with respect to the target-translation axis. This definition of scale factor can be identical to either the s_x or s_y parameter of the camera internal matrix, depending on the orientation of the camera with respect to the translation axis of the target. In the following 2D work, unless otherwise indicated, the camera and target translation are oriented so that the scale factor being measured is equivalent to the s_x parameter. That is, the target translation direction is aligned to be parallel to the camera x-axis.

In many applications, multiple-lens optical systems are used and the "lens plane" is not clearly defined or accessible. In these cases, the scale factor may best be determined by the analysis of imagery rather than by the direct measurements of range and image distance.

For a single camera, the scale factor can be measured by the determination of the length in pixels of the image of a scene feature for which the physical dimension is known. Alternatively, a recognizable point-like scene feature can be translated by a known distance in the scene and the corresponding image positions measured. Scene translation can be considered to be the transfer of the calibration of a precision translator from the object plane to the image plane with the aid of a lens. The latter technique is used here, but with an extended target in order to improve measurement accuracy.

With a multiple-camera network, each camera of the network can be independently calibrated for scale factor in the method just described. However, when viewing a common target, the constraint on the direction of translation being perpendicular to the optic axis can generally be met for only one camera at a time, as the cameras may not be aligned in a parallel configuration[3]. Alternatively, the scale factor of each camera can be measured in a common test system before being added to the network. This procedure is expected to enhance the accuracy of the calibration of the multi-camera network.

2. EXPERIMENTAL SET-UP & DATA ACQUISITION

The experimental determination of scale factor used the test system shown in Figure 2. The camera field of view is filled by a planar calibration target that is oriented so that the target normal and camera optic axis are parallel. A laser beam coincident with the optic axis facilitates the alignment and allows the convenient replacement of the target and the rotation of the camera while maintaining the alignment to better than 1 mrad. A spatially-extended white-light source illuminated the target in reflection. The spatial variation of light intensity on the target was not important, as long as there was negligible drift during the time of acquisition of the images used in the analysis (a few minutes).

A PixeLINK PL-A781 monochrome CMOS camera (with 3000 x 2208 pixels and a FireWire connection) was used for capturing images[9]. The pixels of this camera are square, with an edge length of 3.5 μm.

The target was mounted on a precision linear translator whose translation axis was known relative to the plane of the target to ± 1 mrad. Use of a Newport ESP300 controller and ILS200 translator[8] allowed the target to be shifted with a position precision of ± 0.5 μm and a repeatability of ± 1 μm. The total travel of the translator can be greater than 10 cm (10^5 μm). Combined with the above specification for precision, the fractional error in the translation distance can be as little as $1x10^{-5}$, assuming that temperature drift and other effects do not interfere. This fractional error of the translation represents a lower-bound to the uncertainty in the scale factor determination. Translator systematic errors are expected to limit the uncertainty as well.

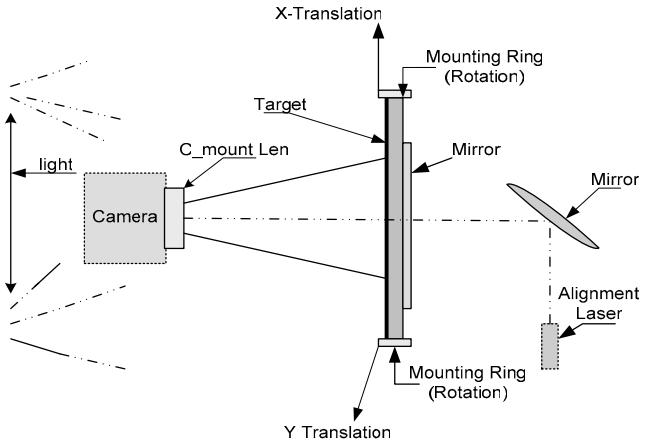


Figure 2(a) Setup for two Dimensional Scale Factor Measurements

Figure 2: Set-up for Two Dimensional Scale Factor Measurements

The alignment-laser beam is adjusted to be coaxial with the optic axis of the laser and is used to centre the target about the optic axis. With the aid of the mirror on the back of the target, the target normal can be oriented to be parallel to the optic axis of the camera. When appropriate, the target could be rotated accurately and reproducibly by monitoring the laser retro-reflection spot. It was assumed that the optic axis of the camera lens was normal to the detector array. Deviations from this assumption could affect the accuracy of the measurements, but were not addressed.

Steps in the determination of scale factor are indicated schematically in Figure 3. The camera, target and translator were aligned as described above. Afterwards, a sequence of images was acquired at various translation distances. A set of translations was always performed sequentially and in the same direction, in order to reduce mechanical backlash. Each image was analyzed by a correlation method to determine the image shift with respect to a reference image, which could be any convenient image in the sequence.

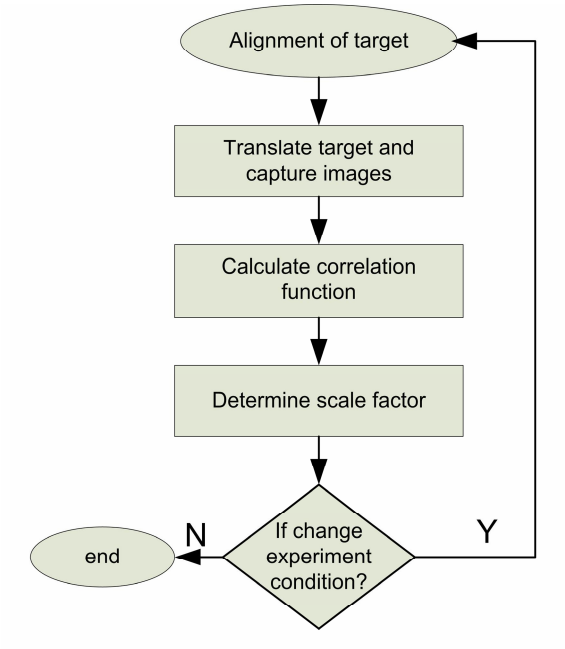


Figure 3: Block Diagram of Algorithm

In a measurement sequence, the target was generally moved to between 10 and 300 positions in translation steps that could be as small as 1 μm. At each position, one or more images were acquired. Multiple images allowed camera temporal noise to be reduced by averaging.

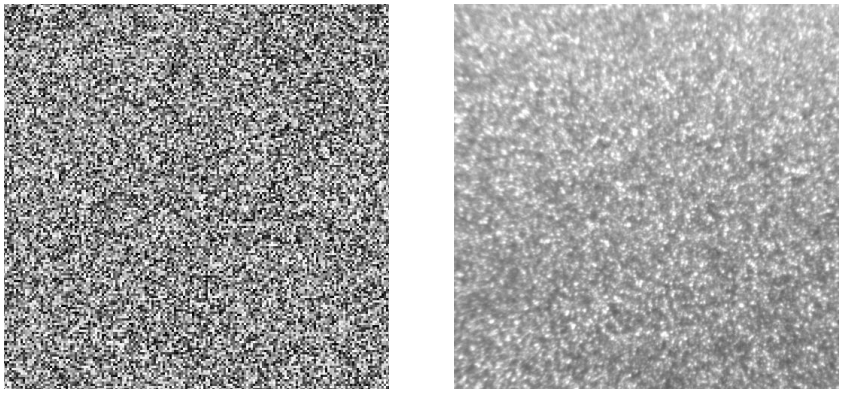
Data acquisition and analysis used standard MATLAB[7] functions. In particular, the image shift due to target translation was determined by a correlation technique described in more detail below. Image correlation allowed the full scene to be used in the comparison of images, so that the comparison was less likely to be influenced by camera fixed pattern noise, spatial variations of the intensity of the light source, flaws in the geometry of the target and other uncontrolled factors of the test system.

The calibration target was selected to complement the correlation analysis, by using a complex scene with a high degree of spatial variability. To improve the precision of the correlation value, a large number of pixels (up to 10^6) were used in the correlation. The correlation function was also selected to be monotonically decreasing about its peak with no significant secondary maxima due to scene periodicity. These requirements of the target were achieved by the creation of a matrix (up to 2048x2048 elements) in MATLAB, each cell of which contained a uniform random value from a common distribution. The matrix was formatted as an image with up to 8 bits of dynamic range, and printed using an inkjet printer. The resolution at which the target was printed determined the spatial correlation length. The printout was

cropped to a circular shape 87mm in diameter, attached to a flat aluminum disc, mounted in a precision rotation micrometer and centred on the optic axis of the camera. The back of the aluminum disc had a suitable mirror for alignment of orientation, and the target sub-assembly was mounted on a linear translator with sufficient travel to move the target out of the path of the alignment laser during set-up. In this work the target parameters (distance from camera, lateral dimension, correlation length, gray scale, flatness) were not optimized.

Figure 4(a) shows a representative calibration target pattern, and Figure 4(b) is the corresponding image. The target used for illustration has coarser spatial features than the optimal target, but shows that the camera did not significantly distort the image over the region of the camera field of view used for correlation (less than 20% of the total field of view). Figure 4(b) exhibits evidence of non-uniform illumination and lens point-spread effects.

During image acquisition, the target and camera were stationary. All movements of camera or target included a settling time to ensure that the objects (target or camera) had a known location before the next image set was acquired. For translations of less than 100 μm, a settling time of 1 second (or more) was used, while larger translations and camera rotations used an even longer settling time. These times were established qualitatively and were found to be satisfactory for meeting the "stationary" constraint.



(a) (b)

Figure 4: Calibration Target (a) Matlab design and (b) image under experimental conditions

A normalized correlation was calculated, for each shifted position of the target, using MATLAB's "corr2" function[7] (5):

$$r = \frac{\sum_m \sum_n \left(A_{mn} - \overline{A}\right)\left(B_{mn} - \overline{B}\right)}{\sqrt{\left(\sum_m \sum_n \left(A_{mn} - \overline{A}\right)^2\right)\left(\sum_m \sum_n \left(B_{mn} - \overline{B}\right)^2\right)}} \tag{5}$$

The test image and reference image pixel values, A_{ij} and B_{ij} respectively, can be integers or real numbers. The correlation coefficient, r, is normalized to be 1 for a perfect correlation and 0 for no correlation. Any of the other target positions could be used as the reference image for the correlation.

3. ANALYSIS OF RESULTS

3.1 Correlation Calculation

The effects of camera noise and varying experimental conditions (such as illumination and temperature) on the correlation coefficient, r, were examined by comparing images of a stationary target. The middle image of the sequence was generally used as the reference image (B in Equation (5)) for the other trials. The correlation calculation was performed over a sub-region of the camera's frame, typically 512x512 pixels in size. The location of the reference window did not change between trials. As the target is translated, the "reference region" of the target moves in the

image plane. It is important for the correlation calculation that the image of the entire reference region be on the detector array.

The correlation coefficient is a precise number. As shown in Figure 5, the histogram of the correlation function over repeated data sets has an uncertainty of about 1×10^{-4}. This precision is a reflection of the large number of points (10^5 – 10^6) that contribute to the correlation.

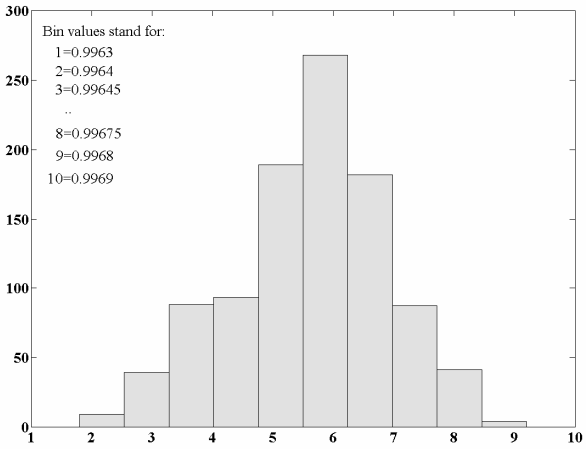


Figure 5: Variation in the correlation coefficient, *r*, for a stationary target

3.2 Correlation Function

A "correlation function" for the system was found by computing the correlation coefficients for several target positions. The target was translated in one direction, and the reference image for the correlation process, *B*, was taken from the middle of the sequence. Figure 6 shows the variation in correlation coefficient values over a small translation range (< 100 μm). As expected, the correlation is close to unity for the self-correlation trial and drops monotonically and symmetrically as the target is shifted in either direction. The uncertainty of each data point is ± 0.01%, as shown in Figure 5, and the small linear component indicates the correlation function is independent of direction. Over the range of the peak, a quadratic relation (solid line in Figure 6) is a good fit to the correlation function. In any case, this fit was only used for interpolating between the known data points of the correlation function.

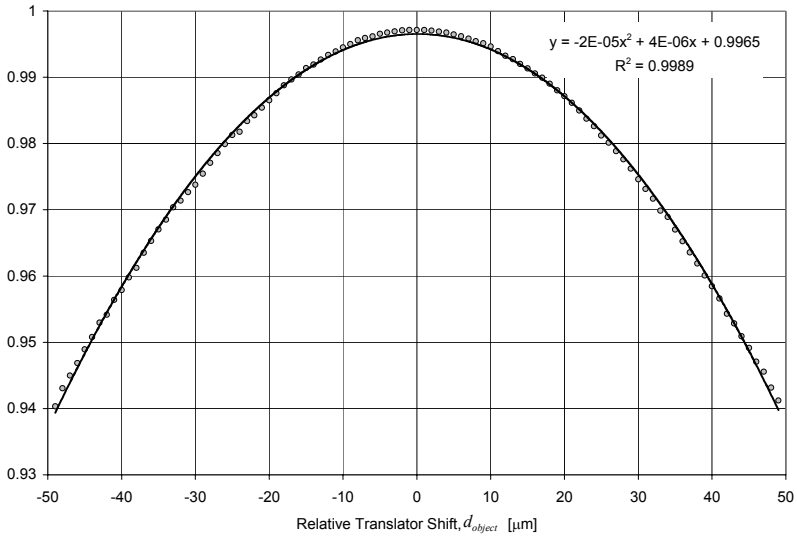


Figure 6: Variation in the correlation coefficient, *r*, with target position

The variation in correlation values over larger translation ranges was carried out using a slightly different technique. For each test trial, the position of the 512x512 sub-region was shifted relative to its previous location (either left or right) to find the displacement at which a maximum correlation occurred. The sub-window location was changed in increments of 1 pixel only, since this was the minimum shift possible without re-sampling the image. In Figure 7, the amount by which the sub-window was shifted to produce a maximum correlation value is plotted as a function of the horizontal translation of the target. The abscissa of the graph gives the translation of the image whose maximum correlation occurs at the value of shift indicated on the ordinate axis. The translation needed to shift the peak of the correlation from one step to the next is the desired scale factor. The scale factor can be read from the figure and is listed in Table 1. This procedure assumes that the shape of the correlation function near the peak has, at most, a weak dependence on the translation position.

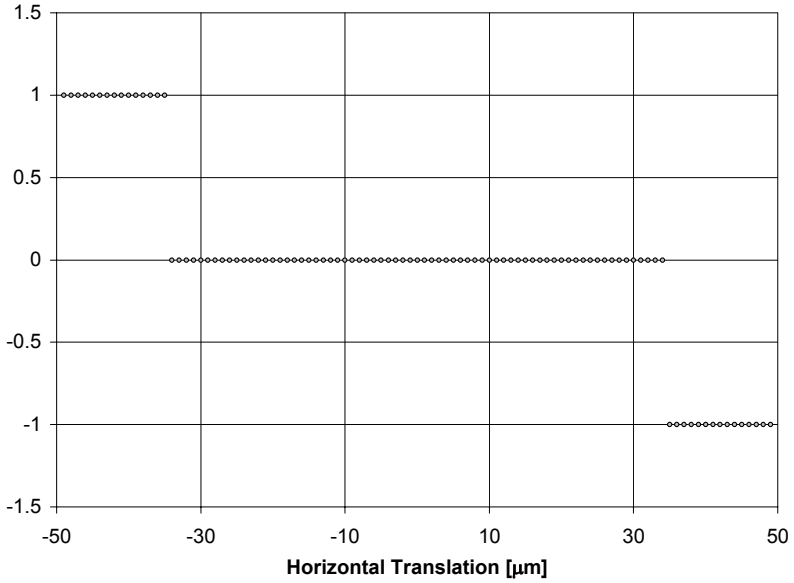


Figure 7: Required Shift of Sub-Window Location to Maximize Correlation

The uncertainty in the scale factor determined in this way is primarily due to the size of the translation step. The translator had a minimum step size of 1 um and a repeatability of ± 1 μm. When the length of the translation is increased, the uncertainty remains the same but the number of pixels in the shift increases, so that the precision of the determination of scale factor increases proportionately. Measurements for longer translations, shown in Table 1, take advantage of the large travel available with the precision translator. Potential improvement to precision by averaging are not included in Table 1.

Table 1: Scale Factor Versus Length of Translation

Length of Translation d_{object} **[μm]**	**Sub-Window Shift** d_{image} **[pixels]**	**Change in Object Distance** dR **[μm]**	**Scale Factor** S **[μm/pixel]**	$\frac{\delta d_{image}}{d_{image}}$ **[%]**	$\frac{\delta d_{object}}{d_{object}}$ **[%]**
65	1	0	65	1.54	1.538
859	13	0	66.1	0.11	0.11
1718	26	0	66.08	0.06	0.06
1784	27	0	66.08	0.06	0.06
1002	14	10000	71.57	0.10	0.10
2007	28	10000	71.68	0.05	0.05
3009	42	10000	71.64	0.03	0.03

As the number of pixels is increased, camera noise, illumination fluctuations, temperature drift and other factors are expected to limit the precision of the scale factor measurement.

Assuming that the uncertainty in the translation distance, d_{object}, is independent of the uncertainty in the measured shift on the image, the errors combine in quadrature (Equation (6)).

$$\frac{\delta S}{S} \approx \sqrt{\left(\frac{\delta d_{object}}{d_{object}}\right)^2 + \left(\frac{\delta d_{image}}{d_{image}}\right)^2} \tag{6}$$

Based on an uncertainty of ± 1 μm for the determination of the object distance, d_{object}, the fractional error, $\delta d_{object}/d_{object}$, is tabulated in Table 1 along with estimates of $\delta d_{image}/d_{image}$ from the same measurement sequence. Larger values of d_{object} and d_{image} will reduce the uncertainties. As long as the desired region of the target remains in the field of view of the camera and the correlation is not significantly degraded by image distortion, camera gain variations, illumination changes or other factors, d_{object} can be increased. These considerations are illustrated in Figure 8, where the image correlation procedure is depicted in terms of the shift in location on the detector array of the pixel region that gives the maximum correlation.

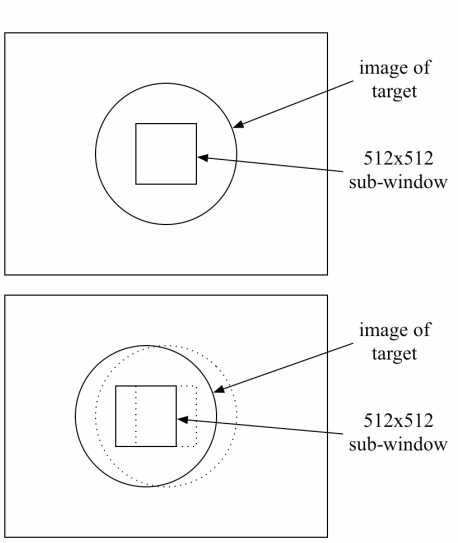


Figure 8: An example reference image (top) and test image (bottom) for the correlation calculation. The sub-region of the image target used in the reference image is demarcated with a solid square. The region in the test image producing the largest correlation value is also shown as a solid square. The dotted outlines show the relative positions of the target and sub-window in the reference frame. The required pixel shift has been exaggerated for clarification in this example.

The methodology of shifting the sub-window to maximize the correlation between two images could be extended to sub-pixel shifts, which should reduce the uncertainty in δd_{image}. The scale factor, S, could be determined directly, as the sub-window shift represents the corresponding change in d_{image} from any variation in d_{object}. However, the precision and accuracy of any sub-pixel shift measurement will depend on the image interpolation method used – i.e., nearest neighbour, bilinear or bicubic. In either case, re-sampling the image is computationally intensive, and similar results for S may be obtained through faster methods. In the measurements summarized in Table 1, the uncertainty of the translator dominates the error, so that this added processing step was not required.

3.3 Agreement with Alternative Metrics

Differentiating Equation (3) shows the derivative of scale factor with respect to range in the Gaussian-optics limit to be:

$$\frac{dS}{dR} = \frac{d_{pixel}}{f} \tag{7}$$

When the scale factor is measured for various ranges, this derivative can be estimated. A measurement of dS with dR = 100 μm gave $f = 6.3 \pm 0.3$ mm, which is consistent with the manufacturer's nominal focal length of 6 mm.

Besides the determination of focal length and image distance, it may be possible to use the scale factor test setup for the measurement of other camera properties. Pixel aspect ratio is readily determined from the scale factors in the row and column directions. Axis orthogonality can be verified by measuring the scale factor at different angles of rotation of the camera about its optic axis. The correlation-function procedure is sensitive to random noise, but not to fixed-pattern noise, and so has potential for the measurement of camera random noise. Likewise, the blurring of a point source due to the camera point spread function may have a convenient characterization using the correlation function. These applications are under investigation.

The scale factor measurement is sensitive to the orientation of the plane of the object. Conversely, the high precision of the scale factor means that it can be used to determine the angle between the object plane normal and the camera's optic axis by measuring S for targets at different points on the plane. In addition, the high precision of S generally allows the range of the target to be positioned more accurately (albeit less conveniently) by monitoring S than by using, for example, the depth of focus of the image.

4. APPLICATIONS TO CAMERA NETWORKS

As mentioned previously in Section 1, a camera's calibration parameters may be determined directly if specific constraints are imposed, such as square pixels (which requires $s_x = s_y$). Further, if some parameters are known to a high precision, these can be treated as constraints. For instance, the method for scale factor determination described here is capable of achieving a precision of 0.1% or better. Therefore, the effective focal length of each camera could be calculated using correlation, and the remaining parameters (such as principal point) could be found through other established techniques[6, 10, 11, 12].

For convenience, the assessment of scale factor as a preliminary calibration technique for camera network was investigated primarily by modeling. Issues of spatial resolution, obscuration and illumination were not addressed. Further, only the specific configuration of an inward-looking camera network was considered[3]. As with any camera network, enough of the target must be visible to each camera to ensure calibration of intrinsic and extrinsic parameters[1]. Since the scale factor requires a target which is translated perpendicular to a camera's optical axis, the effective focal length (and pixel aspect ratio) of each camera could be calculated before the camera is situated on the underlying supporting structure of the network. Afterwards, the remaining parameters could be found by observing a translation that is not necessarily perpendicular to a camera's optical axis.

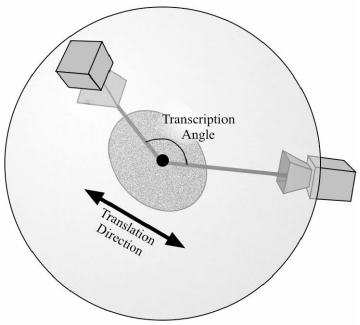


Figure 9: Transcription Angles for a Spherically Symmetric Inward-Looking Camera Network

If the scale factors for two cameras are already known by single-camera 2D measurements, then the configuration of Figure 9 allows the direction cosine between the target normal and each camera optic axis to be calculated from the ratio of the two measurements of scale factor. The direction cosine is in the plane that includes the normal to the target and the line defined by the translation. Target motions in two orthogonal directions are needed to fully define the transcription angle[5] without additional constraints. The analysis is convenient if the target is close to the centre of the camera network (for the spherical example, at any rate), since the measured angle between cameras is then close to the "transcription angle". The locations of the optic axes of the cameras can be defined in terms of these transcription angles.

5. SUMMARY

An accuracy of better than 0.1% in scale factor was measured with an object distance of approximately 100 mm. This level of accuracy could be enhanced by a longer travel distance of the translator or a smaller translator positioning error, both of which are readily achieved with common laboratory equipment. Likewise, the effect of displacement along the optic axis was measured to a comparable precision, thus providing a method of orienting the object plane with respect to the camera optic axis using imagery alone. Finally, a measure of the scale factor and its range derivative provides an accurate determination of the range to the target plane from the ideal-lens plane of the camera and the equivalent focal length of the camera.

ACKNOWLEDGEMENTS

The financial support of NSERC Canada is greatly appreciated.

REFERENCES

1. P. Baker and Y. Aloimonos, "Complete Calibration of a Multi-Camera Network", *Proc. IEEE Workshop on Omnidirectional Vision,* 134-141, Hilton Head Island, SC, 2000.
2. J.-Y. Bouguet, "Camera Calibration Toolbox for MATLAB", Computer Vision Research Group, California Institute of Technology. *http://www.vision.caltech.edu/bouguetj/calib_doc/*
3. P. Carr, "Design Considerations for Efficient Camera Networks", Master thesis, York University, Toronto, ON, Canada, 2005.
4. R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision (Second Edition), Cambridge University Press, 2003.
5. T. Kanade, P. Rander and P.J. Narayanan, "Virtualized Reality: Constructing Virtual Worlds from Real Scenes", *IEEE Multimedia*, **4**(1), 34-37, 1997.
6. R. K. Lenz and R. Y. Tsai, "Techniques for Calibration of the Scale Factor and Image Center for High Accuracy 3-D Machine Vision Metrology", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **10**(5), 713-720, 1988.
7. The MathWorks Inc., Natick, MA, USA. *http://www.mathworks.com*
8. Newport Corp., Mountain View, CA, USA. *http://www.newport.com*
9. PixeLINK Inc., Ottawa, ON, Canada. *http://www.pixelink.com*
10. R. Y. Tsai, "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", *IEEE Journal of Robotics and Automation*, **3**(4), 323-344, 1987.
11. Z. Zhang, "A Flexible New Technique for Camera Calibration", Technical Report MSR-TR-98-71, Microsoft Corp., 1998.
12. Z. Zhang, "Camera Calibration with One-Dimensional Objects", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **26**(7), 892-899, 2004.