

Measure: An Interactive Tool for Accurate Forensic Photo/Videogrammetry

Stephen C. Jensen and Leonid Rudin
Cognitech, Inc.

Abstract

The development and use of Cognitech's Measure package is described. The Measure program allows the user to enter the locations and dimensions of known points in an image, and through the established principles of photogrammetry, recover dimensions and positions of objects that are unknown. Several cases in which Measure was employed in the investigative process are discussed.

Keywords: Photogrammetry, Videogrammetry, Perspective, Measurements, Measure, Investigative Image Processing

1 Introduction

The importance of high quality image restoration tools in the field of investigative image processing can not be overstated. In many cases, however, a sharp, clear image is simply not enough. No amount of image restoration, for example, will determine the identity of a criminal who is wearing a mask. A clear image of an aircraft will shed little light on the behavior of that aircraft in the critical seconds before a crash. A sharp image of a construction site often will not, by itself, prove or disprove the existence of safety violations on the site. It was problems such as these which led to the development of Cognitech's Measure† package. Measure is a collection of tools which allows the user to perform an accurate perspective calibration of an image using positions and dimensions of known objects visible in the video image. Once this perspective calibration has been performed, the location and lengths of other objects in the image can be computed. These measurements further enhance the value of photographic or video evidence to investigators, often dramatically so.

2 Program Description

2.1 Photogrammetry

The Measure package performs two tasks. First, using the known two dimensional screen coordinates and three dimensional spatial coordinates of points in an image, it determines the unique perspective calibration of the image. Once this is done, the program uses this calibration to locate the unknown spatial coordinates of a point whose screen coordinates are known, or else locate the screen coordinates of a point whose spatial coordinates are known. The basic mechanisms for both these steps will be briefly described here. To begin the analysis, a point in space will be defined with the three dimensional, spatial coordinates X , Y , and Z . Using homogeneous coordinates, the coordinates of this point can be reexpressed as a four component vector \mathbf{V} , such as $[X, Y, Z, 1]^T$, where 1 is the homogeneous term, or any multiple of that vector, such as $[sX, sY, sZ, s]^T$ where s is arbitrarily chosen.

The general perspective projection can be represented in homogeneous coordinates as a four by four transformation matrix \mathbf{T} , such that

$$s\mathbf{V}^T \times \mathbf{T} = s'\mathbf{V}'^T \quad (1)$$

The 16 components of the \mathbf{T} matrix account for all the possible variations of perspective which may occur. There are two important properties of this matrix which should be noted. First, the perspective transformation actually maps one three dimensional space not into a two dimensional space as we might

expect, but into another three dimensional space. This third dimension can be considered the perspective depth of a point. This perspective depth is actually quite important in cameras, which need to be focused in order for the plane of the film to fall in the part of the three dimensional image we are most interested in. For our purposes, however, this third dimension will be removed. Secondly, one should note that, because this analysis uses homogeneous coordinates which may themselves be arbitrarily scaled, the perspective matrix itself can be scaled at will.

After dropping the perspective depth coordinate Z' , the projection matrix becomes

$$\mathbf{V}^T \times \mathbf{T} = s' [X', Y', 1] = s' [u, v, 1] \quad (2)$$

where we assume the scale factor s to be unity, and where u and v are the two dimensional screen coordinates of a point.

From (2), we can derive the following equations:

$$\mathbf{P} \cdot \mathbf{V}^T = 0 \quad (3)$$

where

$$\mathbf{P} = \begin{bmatrix} (T_{11} - T_{13}U') & (T_{21} - T_{23}U') & (T_{31} - T_{33}U') & (T_{41} - T_{43}U') \\ (T_{12} - T_{13}V') & (T_{22} - T_{23}V') & (T_{32} - T_{33}V') & (T_{42} - T_{43}V') \end{bmatrix}$$

For each point with known two dimensional screen coordinates (u, v) and the three dimensional space coordinates (X, Y, Z) , expanding equation (3) provides two equations and 12 unknowns. Since T is arbitrarily scaled, we can set one of the elements to 1, thus giving us 11 unknowns. In order to find a unique solution (if it exists), we must have 11 equations, or 5 known points. Assuming one has N points, the matrix equation will look like the following:

$$\begin{bmatrix} [\mathbf{R}_1] & [1] & [0] \\ [0] & [\mathbf{R}_2] & [1] \end{bmatrix} [\mathbf{C}] [-\mathbf{b}] \times \tilde{\mathbf{T}} = 0 \quad (4)$$

where $[\mathbf{R}_1]$ and $[\mathbf{R}_2]$ are $(N \times 3)$ matrices, $[0]$ is the $(N \times 4)$ zero matrix, $[1]$ is the $(N \times 1)$ unit vector, $[\mathbf{C}]$ is a $(2N \times 3)$ matrix, $[\mathbf{b}]$ is a $(2N \times 1)$ matrix, consisting only of the screen coordinates (u, v) of the known points, and $\tilde{\mathbf{T}}$ is the reduced columnwise vector of \mathbf{T} . Setting the last element of $\tilde{\mathbf{T}}$ to 1, and moving $[\mathbf{b}]$ to the other side produces

$$\mathbf{m}\tilde{\mathbf{T}} = \mathbf{b} \quad (5)$$

where

$$\mathbf{m} = \begin{bmatrix} [\mathbf{R}_1] & [1] & [0] \\ [0] & [\mathbf{R}_2] & [1] \end{bmatrix} [\mathbf{C}]$$

It should be noted that \mathbf{m} and \mathbf{b} both consist only of the spatial and screen coordinates of the various points, and are thus known.

This equation can be solved using some type of best fit method, such as a least square fit, or the minimization of the L_1 of the error. L_1 fitting is employed if some measurements are statistical outliers compared to the rest of the data.

Once the elements of the matrix \mathbf{T} are known, one can easily solve for the screen coordinates (u, v) given the space coordinates (X, Y, Z) using the following relations:

$$\begin{aligned}
U &= \frac{F_{11}X + F_{21}Y + F_{31}Z + F_{41}}{F_{13}X + F_{23}Y + F_{33}Z + F_{43}} \\
V &= \frac{F_{12}X + F_{22}Y + F_{32}Z + F_{42}}{F_{13}X + F_{23}Y + F_{33}Z + F_{43}} \quad (6)
\end{aligned}$$

If both the elements of \mathbf{T} and the screen coordinates (u,v) of a point are known, equation (3) expands to

$$\begin{aligned}
a_1X + b_1Y + c_1Z + d_1 &= 0 \\
a_2X + b_2Y + c_2Z + d_2 &= 0 \quad (7)
\end{aligned}$$

This set of equations can be thought of as two planes, which intersect in a line passing through both the camera and the point. To uniquely determine the point, one more constraint must be provided. This is most commonly done by defining a third plane on which the point is known to lie. For example, if the point is sitting on the floor, and the floor has been defined as the plane $z = 0$, this provides three equations in three unknowns, and the exact point can be found. An alternative method would be to locate that same point in another image, taken at the same time but from a different viewpoint. Once the \mathbf{T} matrix of this new image is found, two more equations will be provided. This will give four equations in three unknowns, which can be solved using some type of best fit method.

2.2 Interface

Using the Measure interface to perform a photogrammetric analysis begins with using the mouse to click on points in the image whose spatial coordinates are known. After the points have been defined, Each point can then be selected in turn, and the spatial coordinates entered from the keyboard. Once this is completed, the perspective engine is engaged. If no unique perspective matrix can be found due to an insufficient number of known points, the user is notified and the returned to data entry mode. If enough points are available, the perspective matrix is computed. Since this matrix is a best fit of the available data, the program replots each point using the computed perspective matrix and the input spatial coordinates. This provides a very simple, yet powerful, feedback mechanism. If one or more of the points is incorrect, the replotted points will be scattered far from where they should be. As adjustments are made, either to a point's spatial coordinates, or to the point's position on the screen, the replotted points will converge to their correct positions. When all the points are correct, the replotted points will coincide exactly with their entered counterparts. This feedback method is one of the most powerful features of the Measure package, greatly increasing the accuracy of the final results. Figure 1 shows the results of inconsistent data. Of course, if all the points were measured and entered correctly, this will not occur.

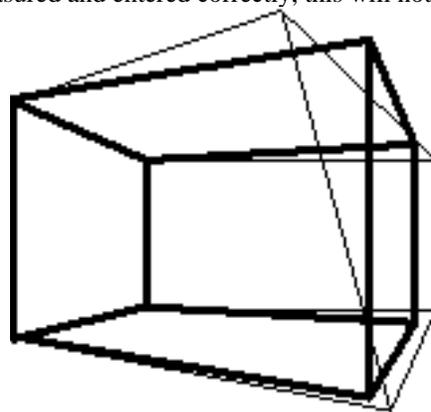


Figure 1

Thin lines represent the known points, as entered into the Measure program. The uppermost point is incorrect. The Heavy lines represent the replotted points, using the best-fit perspective transformation

Once the perspective matrix of the image has been calculated, unknown points can then be entered and solutions retrieved. If the screen coordinates are known, but the spatial coordinates are unknown, one first uses the mouse to select the point on the image. Then, the parameters of the plane on which the point is

known to lie are entered. When the perspective engine is activated, the spatial coordinates will be updated automatically. For points where the spatial coordinates are known but the screen coordinates are not, the user clicks anywhere on the screen to select a point, and enters the spatial coordinates of that point from the keyboard. When the perspective engine is activated, the point will be replotted in the correct location. All of these techniques are discussed in greater detail later on.

A videogrammetric version of the Measure package designed to take advantage of multiple frames/multiple views also has been developed, and has been used in a number of cases. Its interface is similar to that described above, except that the user is presented with multiple frames for data visualization and calibration.

3 Description of Use

As stated previously, Measure is a collection of tools which perform a perspective calibration of an image using known data, and then facilitate the calculation of unknown data from this calibration. The process can be summarized as the following series of steps. First, the image is clarified and restored using image restoration tools. The image is then studied for points which can be measured and used for the perspective calibration. These points are then measured, either by going back to the original scene or through the use of schematics and maps, and the position data is put into a form suitable for input into the measure program. Once the data is actually entered into the Measure program, it is checked for consistency. Points that are not consistent are then remeasured, and the data reentered into the program. This cycle continues until sufficient confidence in the accuracy of the input data is attained. At this point, the perspective calibration of the image is complete. Now, unknown points can be identified and their positions and/or dimensions automatically computed by the program. These results can then be output either for final presentation, or as the basis for some further calculation. This concludes the measurement process. Each of these steps will be discussed in more detail below.

3.1 Image Preprocessing

Image restoration and photogrammetric measurement should not be considered two entirely separate processes. Image restoration, for example, can have a dramatic effect on the quality of the resulting measurements. To obtain an accurate perspective calibration, for instance, the location of the known data points in the image is critical. If edges are blurred, correct placement of these points will be extremely difficult. In addition, if the image restoration algorithm is not feature accurate, important aspects of the image, such as edges and corners, might actually move as a result of the restoration process. These shifts will make accurate calibration impossible. These are all factors which must be taken into consideration when preprocessing an image for photogrammetric analysis.

3.2 Selection of Points

Possibly the most critical aspect of the photogrammetric process is the selection of known points. Errors in the calculation of unknown points increase with distance from the centroid of the group of known points, and decrease with increasing distribution of known points along the axis. The optimal configuration appears to be a "cloud" of known points, widely scattered in all of the x, y, and z directions. The known points should also be distributed around the area for which data is needed.

The estimation of accuracy at a point can be determined qualitatively by displacing known points a small amount, and noting the effects on the point in question. If small displacements of known points lead to large changes in the calculated point, the errors in the region surrounding this point are high. Conversely, if the calculated point remains stable, error in the vicinity of that point will be low. This is in fact a good method of determining if more known points are needed to provide an acceptable solution, and if so, where they should be placed. A quantitative estimation of accuracy is derived by calculating the **T** matrix fit residual, appropriately normalized. An L_1 fit assures that the least accurate points do not effect **T** matrix estimation (robust statistics).

3.3 Measurement of Points

Obviously, the more accurate the measurements made in the field, the more accurate the final results. It should be noted that, more often than not, users will be supplied with a series of measurements taken by someone else, instead of actually going to the scene and taking the measurements themselves. Because of this, a determination of which points are necessary should be made before any measurements are taken, and the importance of measuring all these points as accurately as possible should be emphasized. If this is not done, the resulting measurements will usually be clustered around two or three easily accessible objects (a counter top and a doorway, for instance), while other equally important but harder to reach areas are almost completely ignored. In addition, measurements may be taken of objects which appear on first glance to be

immobile, but may in fact have been moved since the time when the photograph or video image was taken. Cash registers and display racks are good examples of this.

3.4 Data Entry

Once a good set of measurements has been obtained, the data must be put into a form acceptable by the Measure tool. Typically, this means that dimensions (height, width and length of a counter, for instance) must be transformed into locations of individual points (x , y , and z coordinates of the corner points, for example). A convenient axis should be chosen which simplifies this task. It is often convenient to place the origin somewhere on the floor, with the x and y axes being chosen so that the floor sits in the plane $z = 0$. This is certainly not always possible, but it simplifies analysis later on.

Once all the points are entered, the perspective calibration engine in Measure is enabled. If there are a sufficient number of unique data points (i.e. the calculated perspective matrix is not singular), the perspective calibration will be calculated, and the feedback points will be plotted. From these feedback points, it can quickly be determined if the entered points do in fact all contribute to one unique perspective matrix. If they do not, one or more of the points must be incorrect. By looking at where the feedback points fall, and by observing the effects of eliminating suspect points from the calibration calculation, incorrect points can quickly be isolated. These points can then be checked, and remeasured if necessary. This new data is reentered into the program, and the process starts over again. Eventually, a consistent and accurate set of data will be obtained.

3.5 Measurement of Unknown Points

Once the perspective calibration of the image is completed, the operator can then proceed to determine the location of unknown points visible in the image. At this point, there are two classes of problems which can be solved. The first class of problem occurs when the spatial (x , y , z) coordinates of a point are known, and the screen coordinates (u, v) are unknown. The second class of problem occurs when the screen coordinates of a point are known, and one wishes to compute the location of that point in space.

For the first class of problem, one knows the location of a point in space, and wishes to see where that point would appear in the image. If one wants to show that a certain object was not present at the time the image was taken, for instance, the position of where that object should be can be plotted and zoomed to show there is nothing there. Conversely, if the object is not visible due to the occlusion of some other object, or if the predicted location of the object actually falls outside the region visible in the image, this can also be shown using this method. It can also show the location of points which are not normally visible, and whose positions can therefore only be inferred from the locations of other points. The center of gravity of an aircraft is one such point. Finally, if one has two or more images of an area, taken at approximately the same time but from different locations, this method can be used to find spatial points of correspondence between the images. If an object is visible in both images, for instance, and one wishes to show they are in fact different views of the same object, one can show that they both coexist in the same spatial location. Conversely, if the location of the objects in space is different, one can show that the objects themselves must be different.

A more common situation occurs when one wishes to determine the exact location in space of a point visible in the image. In this case, the screen coordinates are known, and the spatial coordinates are required. This case is slightly more complicated than the previous case, since a point in screen coordinates uniquely describes not a unique point, but a line in space which passes through both the camera and the point. As a result, some additional piece of information is required to specify the point uniquely. This additional information is usually provided by specifying a plane on which this point is known to lie. For instance, if the point in question sits on the floor of the scene, one can click on the point (to specify the screen coordinates), and then fix the point on the $z = 0$ plane (which, in this example, describes the floor). The intersection of the line and the plane is the point in question. If two views of the same scene exist, each taken from a different angle, one can select the same point on the screen in each view, thus providing two lines which must intersect at that point in space. It should be noted that, if the views of the scene are not sufficiently different, the lines will be almost parallel, and the error in computing their point of intersection will be quite large.

When only one image is available, and the plane in which the point falls is unknown, locating that point in space is a little more difficult. Usually, however, it is possible to determine the location of these points using some knowledge of the scene, and the relationships between known and unknown points. One could determine the height of a vertical pole whose position is unknown by determining the locations of both the top and bottom of the pole, and then using Measure to calculate the distance between them. At first, this seems impossible, because nothing is known about the location of the top point. Since the base of the pole



sits on the ground, however, its position can be computed. Since the pole is vertical, the top must be directly above the bottom. This additional constraint is sufficient to calculate the position of the top of the pole, and the height of the pole is obtained. The location of this new point can then be used in the same manner to calculate the position of another unknown point, which can be used to calculate yet another point, and so on. With this method, and with careful placement of the coordinate axis, it is often possible to calculate a number of points in an image for which, at least on first

glance, nothing is known. It should be noted, however, that the further one gets from the known points, the less the accuracy that may be obtained.

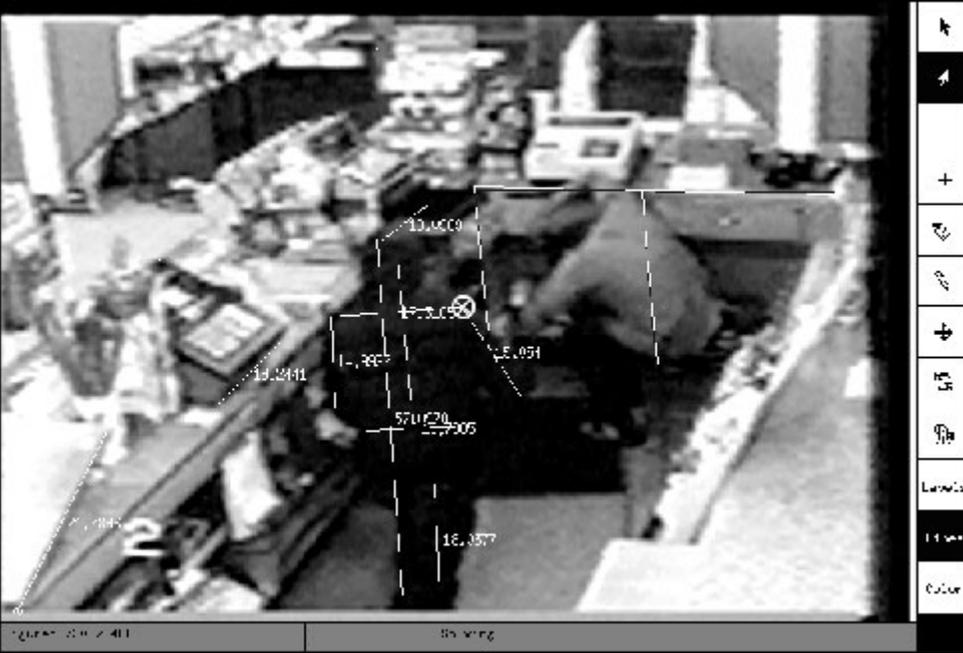
4 Applications

4.1 Civil Liability—Construction Safety Video Analysis

Cognitech's Measure tool was employed in a civil complaint filed by a homeowner against a large home service company. The plaintiff alleged that, during the course of a service procedure, a hole was drilled in his floor which cut the sewer line, causing extensive damage. The company argued that the hole drilled too far away from the pipe to have caused the damage. The only evidence was a video tape, made by the client, showing him digging up the bathroom floor, eventually uncovering the broken sewer pipe. Using the bathroom tiles as a grid, Cognitech was able to accurately calibrate the scene from the video, and measure both the location of the drilled hole and the broken sewer pipe. This information showed that the hole drilled by the service company was in fact offset several inches to the side of the sewer pipe, and therefore could not have been the cause of the damage. Figure 2 illustrates those results.

4.2 Insurance Investigations—Workers Compensation Fraud

In another case, a civil complaint was filed by an insurance company against an individual for Workers Compensation fraud. This individual was receiving benefits for a disabling stroke he allegedly suffered as a result of work-related stress. During the insurance company's investigation of the case, several covert video tapes of the man were taken over a period of a few weeks. Some of these tapes showed the man entering and exiting his doctor's office, and he made no attempt to contest the fact that he was the one being taped. One tape, however, showed what appeared to be the same man playing football in front of his house, as well as juggling his keys in his allegedly paralyzed hand and using that hand to hold a heavy duffel bag. After viewing this tape, the man claimed the person pictured was actually his brother, who was in town at that time and now could not be found. The goal of Cognitech's investigation was to show that the man seen playing football on the second tape was the same man as shown in the first tape coming out of the doctor's office. Since the man and his brother might look very similar, merely showing the two men looked alike was not sufficient. Instead, Cognitech employed photogrammetry to accurately measure the two individuals. Since the man's van was visible in both videos, it was used as the source of known measurements. From these, the two men's shoulders, waists, legs, arms, feet, and heads were measured and compared. The results showed that the measurements of the men in the two videos were consistent within 2%, greatly increasing the probability that they were indeed the same person.



4.3 Criminal Investigations

During one summer evening in a Northern California town, a man entered a small liquor store robbed the owner of the store's receipts, and, before leaving, shot and killed him. Fortunately, the store had a surveillance video camera, which was operating at the time.

Unfortunately, the robber was wearing a mask, and was careful never to stand up straight the entire time he was inside the store. This

made the video almost useless for standard identification purposes. A few weeks after the crime occurred, police received a tip, and a suspect was arrested. Still, the only evidence in the case, beside the word of the informant, was the video taken the night of the robbery. Cognitech was asked to analyze the video tape, taking as many measurements of the perpetrator as possible. The security system recorded images on tape in a "Quad" configuration. Quality and resolution of the video were thus very poor, but it did provide four simultaneous views of the scene from different viewpoints. Several points on the suspect were located in these multiple views, and a singular value decomposition (SVD) solver estimated the optimal 3-D coordinates of these points using this multi-view data. Since the assailant never stood up straight, these additional points allowed us to determine his height piecemeal, starting with the length of lower leg, his upper leg, his torso, and the distance from his shoulders to the top of his head. In addition, measurements of his feet and arms were also calculated (see Figure 3). All of these measurements fell within 2% of the measurements taken of the suspect arrested in the case.

4.4 Accident Investigations—Air Crash

An entirely different application of the Measure tool was in the investigation of the crash of a B-52H bomber at Fairchild Air Force Base, in June of 1994. The aircraft was practicing for an airshow which was to take place the following weekend and was performing a low speed pass over the runway. After passing the runway threshold, the pilot attempted to bank the airplane steeply around the control tower. Midway through this turn, the aircraft rolled completely on to its side and plunged into the ground. All aboard were killed. Since this was a military aircraft, no flight data recorder was on board. Without the flight parameter data usually provided by these recorders, determining with any degree of certainty whether the accident was caused by hardware malfunction, pilot error, or reckless flying becomes extremely difficult. Cognitech was provided with a video tape of the accident and asked to retrieve as much information as possible regarding the flight parameters of the aircraft during its final turn. To this end, a multiframe, videogrammetric version of Measure was employed to determine the exact position of the aircraft's center of gravity in 6 degrees of freedom throughout the last 30 seconds of flight (see Figure 4). From this position information, the relevant flight parameters (ground speed, angle of attack, bank angle, etc.) could be determined. It should be noted that these parameters are calculated using derivatives of the position

information. Since there is some degree of noise in these position measurements, denoising was performed on the data before these parameters were calculated.

Cognitech's final results confirmed the investigation board's theory that the accident was caused by the pilot's failure to stay within the safe flying limits of the B-52. Cognitech's final report was included in the mishap investigation board's final report to the US Senate subcommittee.

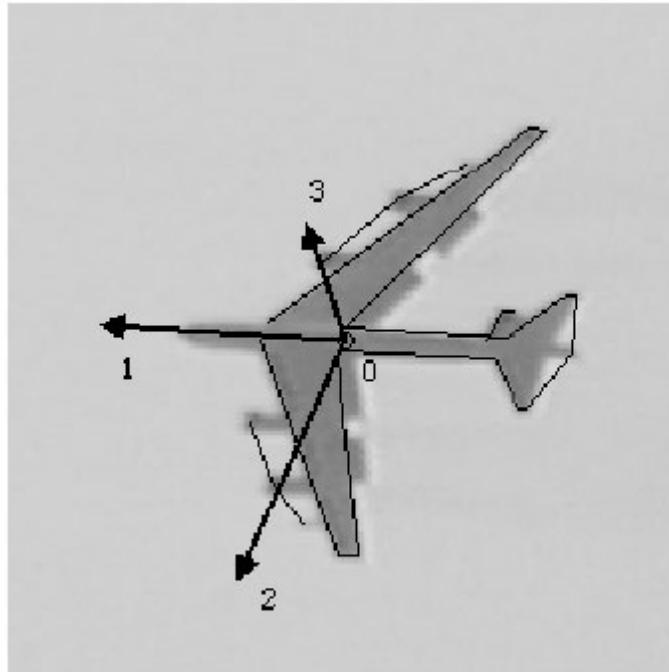


Figure 4

Example of known (light) and calculated (dark) points used in the analysis of the B-52 crash in Spokane, Washington. Point 0 is the calculated center of gravity of the aircraft

4.5 Crime Scene Analysis from Photo and Video Data

In yet another case, two individuals were brutally murdered outside their apartment. After a brief investigation, a suspect was identified and arrested. During the ensuing trial, the defense claimed that one important piece of evidence presented against their client had become contaminated during the investigation, and should therefore be thrown out. Specifically, photographs taken by the police show this item marked and undisturbed in its original location at the crime scene. A video tape, taken at around the same time by a citizen across the street, appears to show the object in a different location, where some contamination could have occurred. Because the photograph was taken from a position almost directly above, whereas the video was taken almost parallel with the ground, it was extremely difficult to reconcile the two views by eye.

After taking careful measurements of the crime scene, the photograph and video images were calibrated, and the position of the object in each image was measured. The location of the object in the photograph was plotted on the video, and the location of the object in the video was plotted on the photograph. In addition, the height and width of the object in the video were determined. From this, it was shown that the object seen in the photograph was not visible in the video, due to landscaping and other obstructions. Furthermore, it was shown that the location of the object in the video corresponded exactly with the location of a fold in a blanket placed on the ground of the crime scene, visible in the photograph. Further analysis showed that the fold was indeed the same width and height, and stained the same color, as the object in the video. Thus, the conclusion was reached that the object in the video was not an object at all, but instead just a stained fold in a blanket laid down on the ground of the crime scene. Therefore, there was no indication the evidence had been moved during that time.

5 Conclusion

The Measure package was developed in response to a set of problems unsolvable by image restoration alone. By using established photogrammetric methods, a great deal of information about the content of an image can be recovered. This tool complements image restoration by allowing the retrieval of quantitative

data from a restored image. This data can be used to determine the speed of an aircraft, the slope of a hillside, the height of a criminal, the location of evidence, and many other quantities useful to investigators in many fields. To the criminal, the Measure package provides law enforcement with an identification tool which can not be defeated by simply wearing gloves or a mask. We believe that photogrammetry and videogrammetry will become indispensable tools in the forensic image processing field.

6 References

- [1] Faugeras, Olivier D. "*Three Dimensional Computer Vision: A Geometric Viewpoint*", MIT Press, 1993.
- [2] *SPIE Photogrammetry Handbook*, SPIE

† Measure Package is a Trademark of Cognitech, Inc.