Proceedings of the XIth International Congress and Exposition
June 2-5, 2008 Orlando, Florida USA

Digital image correlation for specimens with multiple growing cracks

Jeffrey D. Helm

Department of Mechanical Engineering, Lafayette College, Easton, PA 18042
HelmJ@Lafayette.edu

1. Abstract
Measuring the surface displacements of specimens having multiple, growing cracks is difficult with most implementations of the digital image correlation (DIC) method. This difficulty arises from the need to exclude the cracked area from the analysis, a process that oftentimes requires significant and time-consuming user input to achieve successful results. This work presents a set of modifications to the Newton-Raphson based DIC process that allows the method to automatically analyze specimens with multiple growing cracks. The modifications combine a relatively simple crack identification process that takes advantage of the consistency of quasi-regular speckle patterns with a method to reestablish the analysis in areas segregated by the crack growth. The use of a regular dot pattern does, however, introduce a greater chance for registration error in the correlation process. A method to minimize possible registration problems is also presented. Finally, the effectiveness of the method is demonstrated using images of concrete specimens with a complex and growing crack pattern.

2. Introduction
Performing digital image correlation on specimens with multiple growing discontinuities has traditionally been a labor intensive process because the user was required to specify the location of the discontinuities on the initial image to ensure that the correlation process would not fail while trying to cross a discontinuity. Without an automated method, the tedious process of excluding the crack often produced less then ideal results as the correlation process would either fail as it crossed a discontinuity the user failed to exclude or require the user would exclude an area significantly greater than needed around the crack. In addition, to exclude a growing crack path, the user would have to adjust the analysis area for individual images, or groups of images, as the discontinuity or settle for the exclusion of the entire crack path from the entire set of images. Réthoré [1] addresses this problem through the use of a DIC method based on a finite element like grid that refines its spatial resolution in the presence of discontinuities.

While it is possible for the user to make these modifications for a single or small number of cracks, the process becomes quit burdensome, or nearly impossible in the presence of a large number of cracks that grow in no pre-determined direction. Additional complications arise if the cracking produces segregated areas on the surface. In this case, additional analyses must be run with a new starting (seed) point within each segregated area. An example of this type of behavior is shown in Figure 1, an image of the cracks in a slab of reinforced concrete under a distributed pressure load.

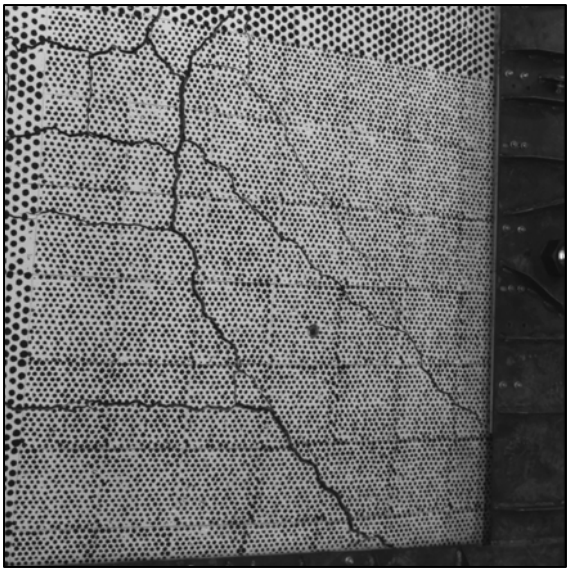


Figure 1. Concrete slab with cracking

To perform the analysis on this type of problem without the need for additional user input, a method was developed to automatically modify the analysis to exclude the growing cracks. The development of the system required the addition of two additional capabilities to the standard image correlation method. First, the modified method must be able to detect the discontinuities on the surface of the object and second, it must be able to establish a new starting point for the correlation in areas that get segregated from the general correlation. The first function capitalizes on the consistency of the correlation

value of quasi-regular pattern of dots on the surface. The second problem, growing discontinuities that segregate smaller subsections from the larger analysis, was addressed with a method to establish an initial guess for the smaller sub-region employing a combination of spatial initial guesses, derived from an adjoining subset, and temporal guesses, derived from the results of the previous image to reestablish the analysis. In addition, the use of a regular pattern requires the introduction of a system to significantly reduce or eliminate the possibility of registration errors.

3. Development and implementation

To achieve the desired goal of DIC analysis of specimens with growing crack patterns, three methods must be implemented; a crack identification method, a method to prevent pattern registration errors, and a method to reestablish the analysis in areas segregated by crack growth. The development and implementation of the methods is presented in the following sections.

3.1. Identifying cracks

The first difficulty in analyzing a system that has a randomly appearing and growing crack pattern is identifying the location of the cracks and determining when the influence of a crack is significant enough that the local area must be omitted from the analysis. The key to the accomplishing this task, was the use of a quasi-regular pattern of dots as a speckle pattern. The increased consistency in the correlation function that results when a dot pattern is used as the speckle pattern presents a significant advantage in the identification of cracks in the image.

While one benefit of the image correlation method is its ability to use non-structured, random patterns on the surface of the object [2-4], there is no reason the same correlation process cannot be applied to a regular pattern on the surface. One reason random patterns are often preferred is that it is easier to produce a random pattern, through various painting techniques, on the surface of small specimens then to produce a regular pattern. However, as the observed surface increases in size it becomes increasingly difficult to produce a good natural random pattern but significantly easier to produce a quasi-regular pattern of dots on the surface. It is important to note that the dot pattern is used in the same manner as the random pattern in typical DIC analysis. Even though the pattern is composed of dots, the analysis uses square subsets from the initial image, irrespective of dot placement, for the analysis. Because of this, the size and relative position of the dots need not be highly controlled and hence the references to a quasi-regular pattern. While the early impetus for using the dot patterns was the ease of application, it was observed that the nature of the pattern also resulted in a more uniform correlation coefficient, particularly in two-dimensional analysis. As an example consider the two patterns shown in Figure 2. Both the random pattern and the quasi-regular pattern were subjected to a displacement in the x and y direction of approximately 8 pixels. The patterns were then analyzed with a DIC program that utilizes a cross-correlation error function. The error function provides a measure of the quality of the pattern matching, returning a value between 0 and 1. While both patterns produce accurate displacement values, the correlation function of the regular pattern is more consistent. The average, standard deviations, max and min values of the correlation function for the two patterns are shown in Table 1. As seen from the table, the correlation values for the dot pattern have less scatter, allowing variations in the correlation function due to growing cracks to be easily identified.

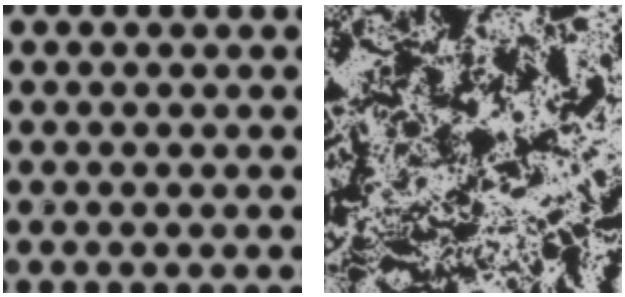


Figure 2. Regular and random speckle patterns

Pattern Type	Average	Standard Deviation	Minimum	Maximum
Dot	1.94×10^{-4}	1.89×10^{-5}	1.53×10^{-4}	2.59×10^{-4}
Speckle	7.26×10^{-4}	9.67×10^{-5}	4.75×10^{-5}	9.94×10^{-4}

Table 1: Correlation function values from dot and random speckle patterns

The foundation of image correlation is the mapping a subset from an initial image to a location in a later, loaded image through pattern matching. This is accomplished by spatially mapping the subset between images and comparing the gray levels of the original points in the subset to the gray levels at their mapped locations[5,6]. While this mapping may include rigid body motion, strains and/or perspective distortions, the subset is mapped as

a continuous surface. Because the surface pattern of a subset containing a crack deforms in a non-continuous manner, the changes in the pattern cannot be accurately modeled by the spatial mapping portion of the image correlation process, resulting in significantly higher correlation coefficients. This effect, combined with the greater uniformity of the correlation values from regular dot patterns, provides a method of detecting the presence of discontinuities in the subset. Unfortunately, the use of a regular dot pattern also increases the possibility of registration errors during the correlation process, a possibility that is heightened near discontinuities. A method of minimizing this effect is presented in the next section.

After adopting the use of quasi-regular dot patterns, the problem of crack identification becomes the solution of two problems; a) establishing suitable cutoff values for the correlation and b) dynamically altering the values to account for global changes in the pattern and the surrounding conditions. Two cutoff values for the correlation function are established, a warning factor and a failure factor. The two cutoff values are determined from a combination of user-set values for the warning and failure factor and the characteristics of the pattern as determined from the analysis of the first two images. The process assumes that the specimen will not exhibit any crack growth in the first two images and the user has seleced a valid analysis area for correlation. After the first analysis is completed the average and standard deviation of the correlation function is calculated. The warning and failure cutoff factors are then calculated. The warning factor is calculated from the average and standard deviation of the correlation function and a user supplied factor f_w, as shown in Equation 1. The failure factor is calculated from the average and maximum correlation value and a user supplied factor f_f, as shown in Equation 2. These base factors remain the same for the entire analysis. The cutoff values for a particular image are determined by multiplying the base warning and failure factors by the average correlation value from the previous image. The average correlation value for an image is calculated using all the correlation values below the warning threshold. In this way changes that effect the pattern as a whole, such as lighting changes, are incorporated into the cutoff factors for the next image. When applied to the three-dimensional DIC method used for this work [6], it was also observed that the presence of cracks was more consistently detected when comparing two images from the same camera, due to the lack of perspective distortions. For this reason the three-dimensional analysis was split into two passes, a first pass that correlates between the initial and deformed images from the same camera and a second pass that correlates the two deformed images to measure the three-dimensional displacements. During the analysis, data from subsets with correlation values above the failure cutoff value are discarded and data from subsets with values above the warning threshold are retained, but the results of these subsets are not used to develop initial guesses for adjacent subsets. While the resulting crack identification method is relatively simple, the use of regular patterns and dynamic calculation of cutoff values proved to be an effective method to locate cracks in the analysis.

$$Warning\ Factor = 1 + f_w * (\sigma(CoorFunction) / \overline{CoorFunction}) \qquad (1)$$

$$Failure\ Factor = f_f * (\max(CoorFunction) / \overline{CoorFunction}) \qquad (2)$$

3.2. Reducing registration errors

As stated earlier, one downside to using a regular pattern of dots as the speckle pattern for DIC is the possibility of registration errors. The presence of cracks or discontinuities in the specimen greatly aggravates the problem. DIC is fundamentally based on the matching of a pattern from the first image or set of images to corresponding locations in later images or sets of images. This is typically achieved by minimizing an error function based on the comparison of gray levels from the initial image and the projected locations in the subsequent image. The error function is minimized using a non-linear optimization method to determine the parameters that best describe the mapping from one image to another. The optimization method typically employed in DIC is the Newton Raphson method or one of its variants[5, 6]. These methods attempt to optimize the system by driving the first partials of the correlation error function to zero by solving Equation 3 to obtain correction factors (δ) for each of the parameters from the gradient ($\nabla CoorFunction$) and Hessian of the correlation error function (H *CoorFunction*) for each iteration.

$$[H\ CoorFunction](\delta) = -\{\nabla CoorFunction\} \qquad (3)$$

The correction factors are applied to the parameters and the process continues in an iterative fashion until a minimum error value is achieved. The difficulty in using a regular pattern arises when the process over-corrects and dictates a significant shift in the location of the interrogated area. To overcome this problem many DIC formulations, including the one used in this work, employ the Levenberg-Marquardt variation of the Newton-

Raphson method (also called the Marquardt method)[7]. A schematic of the Marquart method is shown in Figure 3. This optimization method allows the optimization to smoothly transition from a Newton-Raphson type method to a steepest decent method. The shift is accomplished by varying the diagonal factor D. If D is close to zero, the process is nearly a strict Newton-Raphson method. As D increases the weight of the diagonal elements increase and the optimization smoothly transforms to a steepest descent method with steps that get smaller as D increases. The diagonal factor is changed based on the results of the previous iteration. If previous iteration produces a result with a decreased error function the diagonal factor is reduced by a factor of 10, the corrections are applied to the mapping parameters and the Hessian is recalculated using the new mapping parameters. If the error increased the diagonal factor is increased, the diagonal of the old Hessian is multiplied by the new diagonal factor the equation solved to produce another set of prospective corrections, and the new error value is evaluated using these corrections. This method works extremely well on random patterns, where almost any significant over-correction will produce an error higher than the current error. However, when using a regular pattern of dots it is possible that the over correction will move the interrogated area over by the period of the pattern. Not only does this move the solution space into the zone of an incorrect local minimum, but because the correct location may have a crack in the pattern, the shifted pattern may also produce a lower error than the value for the appropriate location on the image. This lower error would fail to trigger the needed increase in the diagonal factor, and the method would locate an incorrect local minimum. To correct for this possibility a check has been introduced into the process. A user selected factor, called the sanity factor (f_s), describes the maximum allowable distance the center of the subset may move without misregistration. This parameter is typically set to a value somewhat less then ½ the diamter, in pixels, of the dots in the pattern. The value describes the maximum distance the center of the subset may move on the image from the subset's starting, initial guess, position. To achieve the desired effect, the sanity factor must be applied at two different points in the analysis. The most obvious point to apply the sanity factor is at the end of the optimization. If the final location of the subset in the second image is further from the initial guess by more than the sanity factor the optimization is considered a failure. However, if this is the only place the sanity check is applied, valid subsets that experience an overcorrection during the optimization may be excluded. An

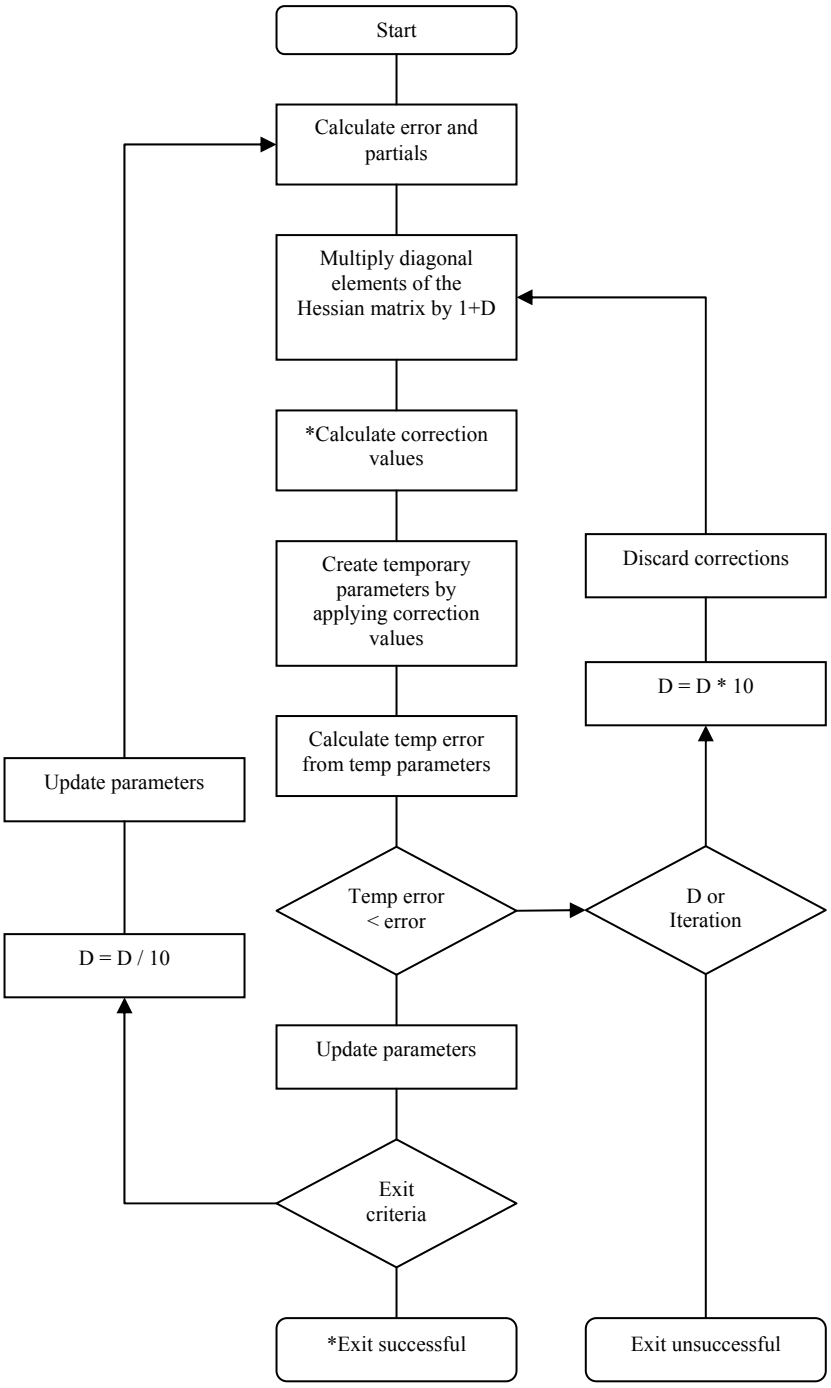


Figure 3. Diagram of the Levenberg-Marquardt optimization method

overcorrection, which would typically result in an increased error function and trigger a correction by the Marquardt method if applied to a random pattern, may result in a lower error value and defeat the inherent self regulation of the Marquardt method if the overcorrection results in a false registration. It should be noted that the check at the end would have detected the registration error, but the valid data from the subset would be lost. To overcome this limitation, the sanity check is also applied to the subset after the points have been mapped to the second image, prior to calculating the interpolated grayscale values and error. These points in the optimization process are indicated by an asterisk in Figure 3. If the movement of the subset exceeds the sanity factor the current error value is set to a value greater than 1. Because the cross-correlation error value is restricted to values from 0 to 1 the optimization will reject the correction factors, increase the diagonal factor and proceed with the correlation.

The combination of regular patterns with more consistent correlation values, the use of warning and failure cutoff values for crack detection and the inclusion of a sanity factor to insure proper registration, allow the analysis to effectively locate significant cracks on the surface without relying on prior user identification.

3.3. Reestablishing the analysis in segregated areas

One result of identifying and removing subsets that are part of a crack line is that sub-areas may be segregated from the analysis due to a crack line forming a boundary around an area or by clipping off a corner of the analysis. Automated analysis is only possible if there is a mechanism to re-establish the analysis in segregated areas. To reestablish the analysis a valid initial guess within the segregated area must be obtained. Two terms will be introduced for the purpose of this discussion. The first term, spatial initial guess, refers to an initial guess that was derived from information obtained from the successful correlation of an adjacent subset on the current image. The second term, temporal initial guess, refers to an initial guess derived from the same subset from the previous image set. Typically DIC analysis uses spatial guesses for all but the first seed point where temporal initial guesses are used. To ensure the temporal guess will be satisfactory, the seed point is chosen at an area of low displacement, typically near a fixed grip. The analysis then spreads out from the seed point using spatial guesses for all other points. This is done using some variation of a flood fill algorithm [8] that processes subsets that are neighbors of successful subsets until all contiguous subsets are correlated. Spatial initial guesses, being constrained to the physical movement of the surface, tend to be more robust than temporal initial guesses, particularly in areas of high strain. Provided the first temporal initial guess is successful, propagating the initial guesses in a spatial fashion is the most reliable. Unfortunately, when an area is segregated there is no path for the initial guess to propagate.

The spatial flood-fill algorithm was modified to produce a new starting point for the analysis in the segregated region. As is typical for a flood fill algorithm, after a successful correlation, the results from one subset are used to generate initial guesses for any neighboring subsets that have not been analyzed. This continues in a recursive fashion until all contiguous points have been analyzed. To obtain a starting point for areas that may have been segregated the algorithm then searches for any subsets that are in the analysis area but have not been attempted. The program then uses a temporal initial guess to try to establish a new starting location. If the correlation of the previously untied subset succeeds using the temporal guess, the point is used as a new seed point for the flood-fill algorithm. Successful correltion results if the subset has not moved off the image, has not been rejected by the sanity check, and results in a correlation factor lower than the current rejection threshold for the image. This sequence; flood-fill, search for a valid temporal guess, flood-fill, continues until no untried subsets remain. In this manner, spatial initial guesses are used to propagate the analysis while temporal initial guesses are used to seed segregated areas.

Combining the use of regular patterns, cutoff values derived from the correlation error statistics, sanity checks to reduce registration errors and a hybrid spatial/temporal initial guess scheme to handle areas that get segregated from the main analysis, correlation of specimens with randomly growing discontinuities may be analyzed.

4. Experimental Application

The techniques described in the previous sections were applied to the analysis of a slab of reinforced concrete subjected to a uniformly distributed pressure load. The 2.1 m x 2.1 m slabs were constrained in the out-of-plane direction by the steel structure around the periphery, as shown in Figure 4. When subjected to a uniform pressure load on the bottom surface of the slab, the slab bulges in the center and tension in the top surface causes crack patterns to form. Because the crack pattern in the surface of the concrete grows with the applied load is

widespread, non-deterministic, and significant enough to interfere with traditional digital image correlation methods, this specimen is an ideal candidate for the techniques presented in this work.

The experimental setup for the concrete testing is shown in Figure 5. The specimen was observed by two stereoscopic camera systems, a global system that imaged the entire surface and a local system that imaged one quarter of the slab to achieve greater spatial resolution. The results and images presented in this section come from the local camera system. In preparation for the testing, the specimen was painted with a quasi-regular pattern of dots. The dots were applied with a stamp made from a steel plate and covered with a pattern of small felt disks, similar to those used to cushion the bottom of furniture legs. The dots in the local image area were approximately 10 mm in diameter which corresponded to a dot diameter of approximately 8 pixels in the images. The specimens were loaded by a pair of bladders that were located underneath the concrete slab and pressurized with water. As the specimen was loaded images were taken every 30 seconds, resulting in a sequence of 400 images for the test. Figure 1 shows the crack pattern at the maximum load.

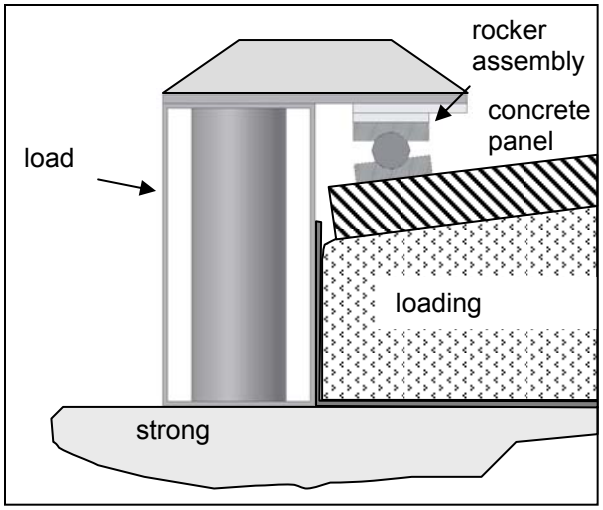


Figure 4. Load frame details

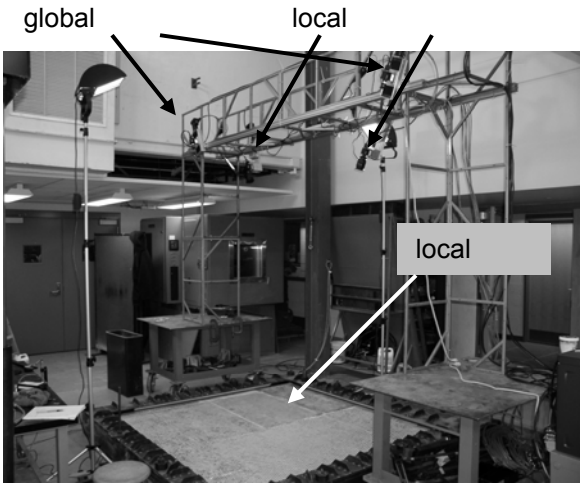


Figure 5. Experimental setup

4.1. Analysis results

The following section discusses the results obtained from the analysis of the concrete slab images with and without the crack detection, sanity and spatial/temporal guess systems. The initial image from the concrete slab sequence is shown in Figure 6. The area used for the analysis is indicated by the dashed lines and the location of the seed point is indicated by the white X. For all analyses the subset size was 25x25 pixels.

The analysis of the panel was first run with the crack detection, sanity factor, and spatial/temporal initial guess systems disabled. Therefore, no rejection limits were placed on either the correlation value or the displacement of the subset. The correlation failed shortly after crossing over one of the cracks and the analysis degraded to the extent that less than 60% of the points were even attempted. More importantly, the results, suffering from poor initial guesses and registration errors, were not reasonable with displacement values up to 1 m in magnitude.

To continue with the analysis user factors were determined for each of the there crack detection systems. The sanity factor was determined by selecting a value equal to ½ the average diameter of the dots in the pattern. It is assumed that the initial guess from the previous subset should be within this radius. Running the analysis on images without cracks demonstrated that the sanity factor was not rejecting subsets in uncracked areas. The failure factor was determined by adjusting the factor until subsets in uncracked areas were not being rejected but the system did reject subsets that ran over significant

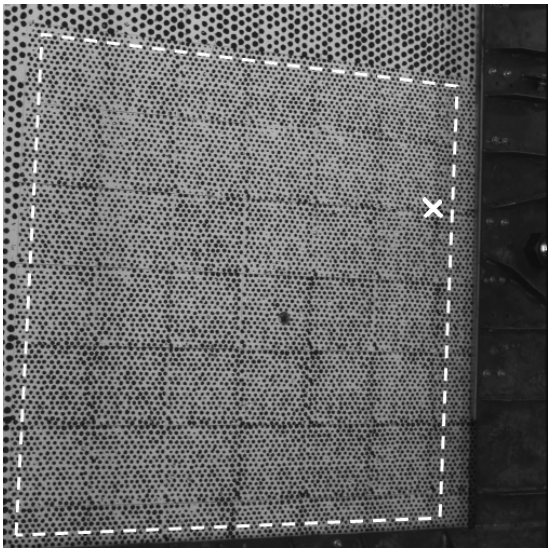


Figure 6. Initial image with correlation area and seed point locations

cracks in the analysis. Finally the warning factor was adjusted so that most subsets adjacent to the crack, that most likely overlap the crack at the edges of the subset would be indicated by the warning factor, once again without indicating false positives in non-cracked regions.

The second analysis was then run with crack detection parameters of $f_w = 4$ for the warning factor, $f_f = 5$ for the failure factor and $f_s = 4$ pixels for the sanity factor. Figure 7 shows the subsets that successfully correlated with the crack detection system active. With the rejection of the crack from the analysis the correlation no longer had the tendency to get lost due to the influence of the crack on the surface pattern. However, the area of successful correlation was reduced to the segregated area that contained the seed point for the analysis. To acquire additional data using only the crack detection methods, the analysis would need to be run multiple times with seed points located in each segregated area. It should be noted that a small bridge, where the effect of the crack was diminished due to a bifurcation, allowed the analysis to spread into the second area.

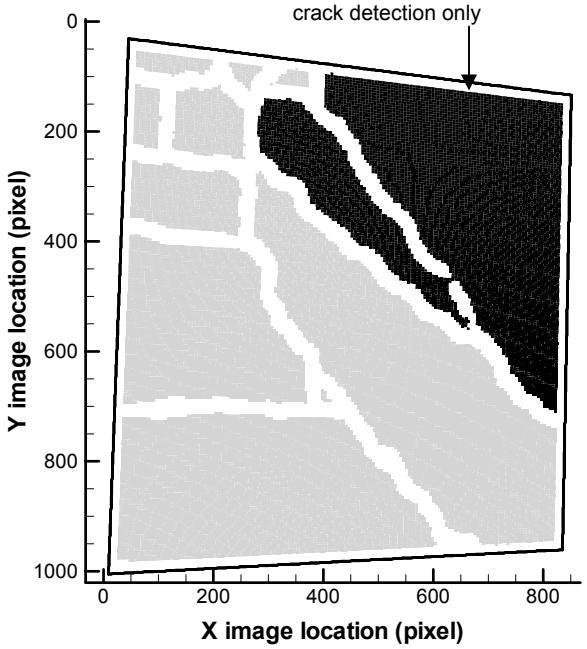
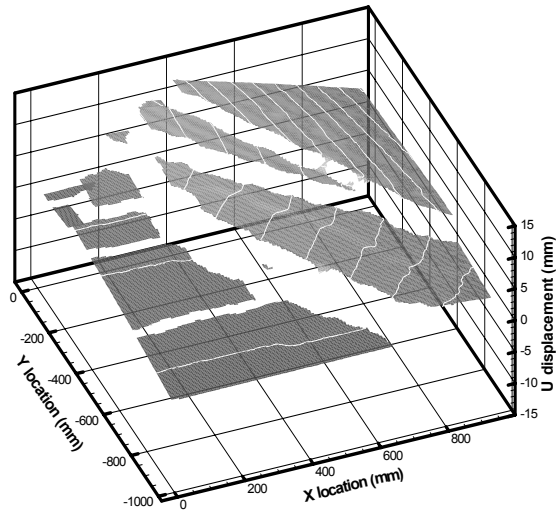


Figure 7 Successful analysis points with crack detection and crack detection and reestablished areas

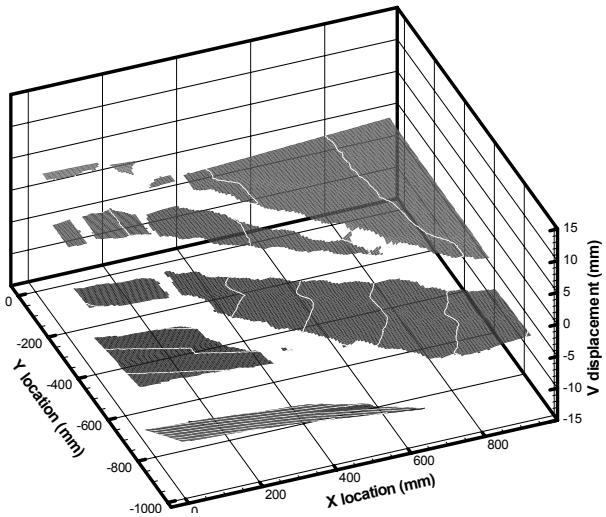
Combining the crack detection method and spatial/temporal guesses for segregated areas resulted in the analysis of the entire surface of the specimen, except for those areas directly affected by the growing cracks. Figure 7 shows the points where the analysis was successful. The ten areas, segregated by the growing cracks are clearly visible. Figure 8 presents plots of the three surface displacements from the analysis. The magnitude of the discontinuities between each of the segregated areas is visible, particularly in the plots of the in-plane displacements. The propagation of the crack pattern due to increasing load is illustrated in the plots for images 112, 124, 148, 176, 300 and 400 in Figure 9.

The effect of the crack detection system on the analysis can be shown by comparing the results from the first horizontal line of data from image 400 using different aspects of the crack detection system. Figure 10 shows the combined results for the horizontal, in-plane displacement, U analyzed; a) with no sanity checks, no crack detection and no method to reestablish the analysis, b) with sanity and crack detection systems employed but no method to reestablish the analysis and c) with sanity checks, crack detection and a method to reestablish the analysis. As shown in the graph, all three analyses produced the same results from the location of the seed point, x = 906 mm, to the edge of the search area and back to the location x = 400 mm. Crack locations, indicated as gaps in the data, are present at x = 225 mm, x = 400 mm and x = 600mm. While the non-constrained correlation (a) was able to bridge the gap at x = 600 mm, it was unable to bridge the gap at 400 mm. Because the analysis continued by assuming initial guesses derived from the results of previous subsets, the correlation process was effectively lost and could no longer provide valid data. Because analysis (b) had no means of reestablishing itself, it was forced to stop when it reached the crack located at x = 400 mm. The combination of all three methods (c) allowed the analysis to avoid the crack locations and to continue across segregated cracked areas.

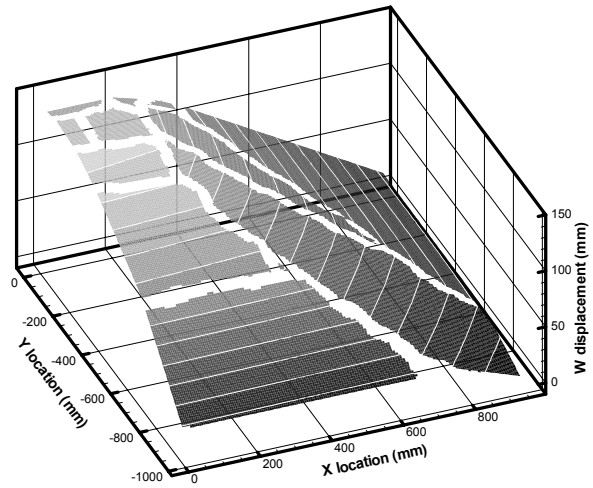
While the experimental data presented above is from a single concrete panel test, the test program consisted of a total of 15 tests, each with two camera systems. The initial user parameters were determined using the data from the global DIC system for the first panel test and the parameters were applied, without modification, to the local



U displacements (1 mm contours)



V displacements (1 mm contours)



W displacements (10 mm contours)

Figure 8. Surface displacements at maximum load

and global analysis for all subsequent tests. The crack detection system was robust enough so that the same

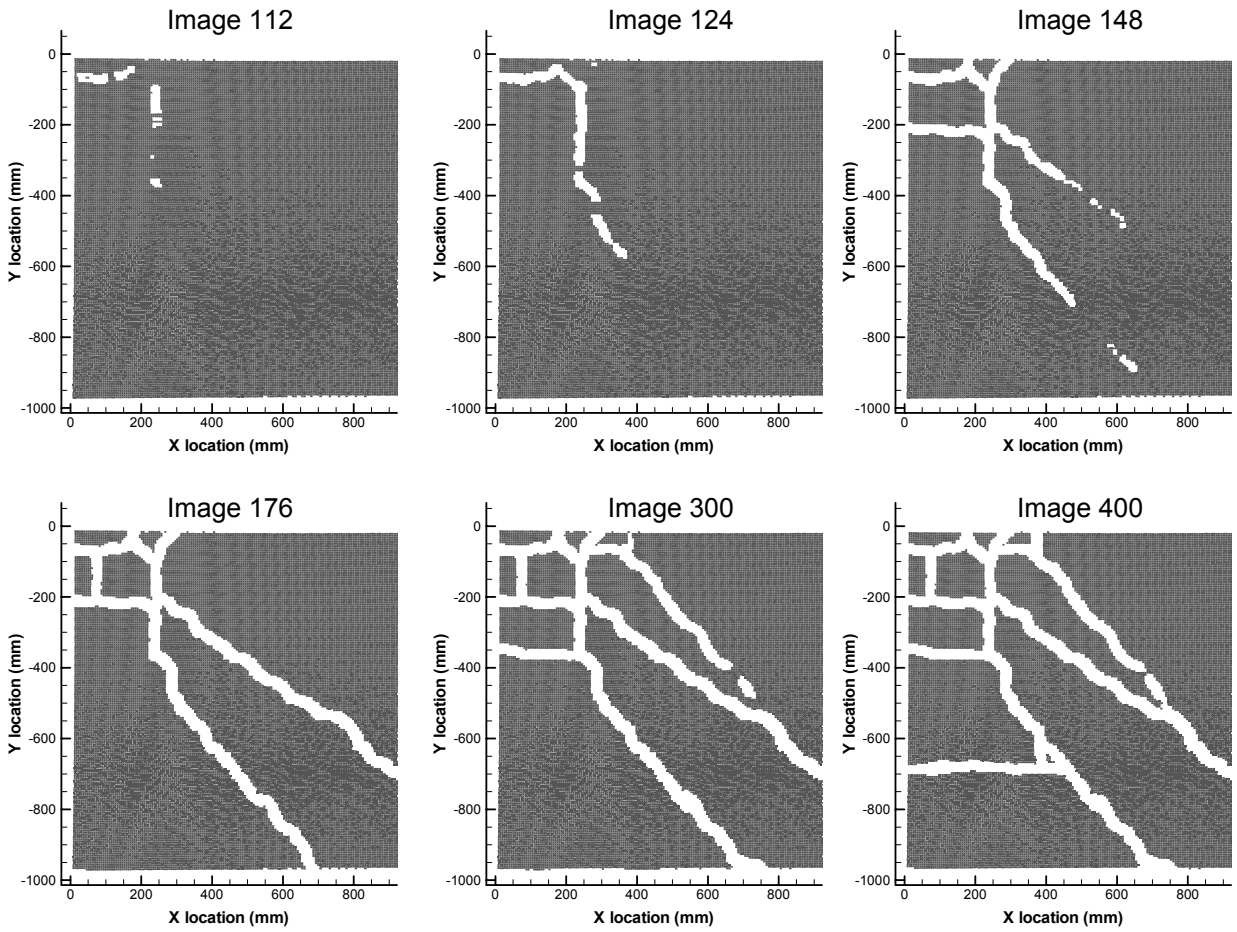


Figure 9. Crack growth patterns with increasing loads

parameters could be applied to all 29 sets of images (one test did not have the standard local system) containing a total of approximately 8000 image pairs. The factors that will most significantly affect the value of the parameters are the imaged pattern size and period and the contrast of the cracked areas. Because the pattern was scaled to appear roughly the same in both the global and local systems and both systems saw the same contrast changes from cracking it was not unexpected that the same user parameters could be applied to all the other data sets.

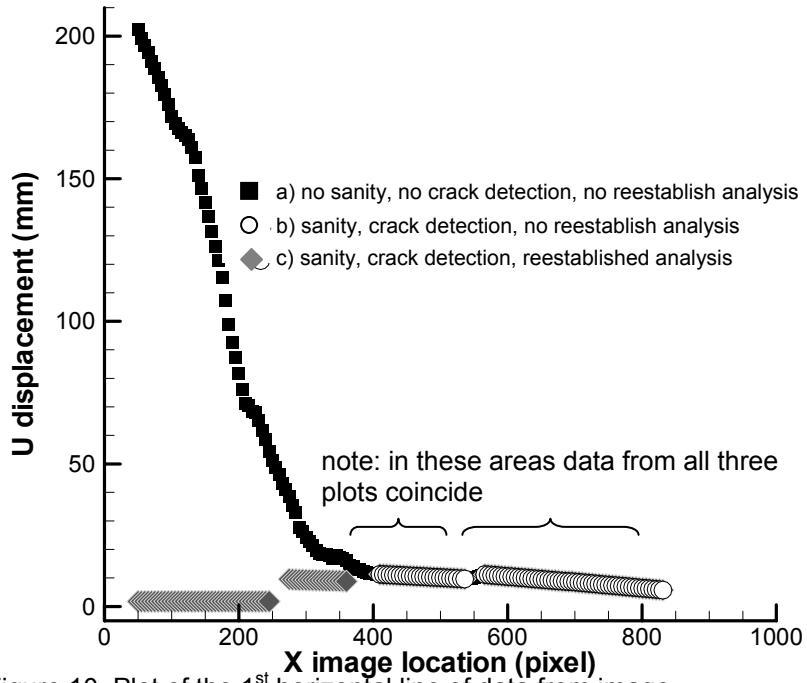


Figure 10. Plot of the 1st horizontal line of data from image

5. Conclusions

The problem of analyzing specimens with growing cracks or discontinuities is problematic and user-intensive with traditional implementations of digital image correlation. Combining a simple, but effective crack

identification process and a spatial/temporal initial guess method to reestablish the analysis in areas that get segregated by the growing cracks allows the analysis of specimens with randomly growing cracks with little or no additional user interaction. The crack identification process utilizes the greater consistency in the correlation function offered by quasi-regular surface pattern, the sensitivity of the two-dimensional DIC process to the change in patterns due to the presence of the crack and a threshold value that is established from the statistical analysis of the correlation error and dynamically updated for each image's correlation. In addition, the incorporation of a sanity factor into the optimization process greatly reduces the chance of registration errors, while rejecting the fewest number of subsets possible. A combined spatial/temporal initial guess system is effective in restarting the analysis in areas that get segregated due to crack growth.

The effectiveness of the combined processes is demonstrated on the analysis of a pressure loaded concrete slab. Even though the slab experienced widespread crack growth, separating the area into ten distinct areas, the process was able to complete the entire analysis without additional user interference. The inclusion of these systems into DIC programs will allow the analysis of a wider range of problems, where the inclusion of growing cracks makes traditional implementations too burdensome on the user.

6. References

1. Rethore J, Hild F, Roux S (2007) Extended digital image correlation with crack shape optimization, International Journal for Numerical Methods in Engineering, DOI:10.1002/nme.2070
2. Sutton M A, McNeill S R, Helm J D, Chao Y J, (2000) Computer vision applied to shape and deformation measurement, Trends in Optical Non-Destructive Testing and Inspection, Rastogi P K, Inaudi D Eds. Springer Verlag, New York, p. 323-368.
3. Zhang D, Eggleton D C, Arola D D, (2002) Evaluating the Mechanical Behavior of Arterial Tissue using Digital Image Correlation. Experimental Mechanics 42:409-416.
4. LeBlanc M M, Florando J N, Lassila D H, Schmidt T, Tyson J, (2006) Image Correlation applied to Single Crystal Plasticity Experiments and Comparison to Strain Gage Data, Experimental Techniques 30:33-37.
5. Bruck H A, McNeill S R, Sutton M A, and Peters W H, (1989) Digital Image Correlation Using Newton-Raphson Method of Partial Differential Corrections, Experimental Mechanics 29:261-267.
6. Helm J D, McNeill S R, Sutton M A, (1996) Improved three-dimensional image correlation for surface displacement measurements, Optical Engineering 35:1911-1920.
7. Press W H, Teukolsky S A, Vetterling W T, Flannery B P, (1992) Numerical Recipes in Fortran (2nd ed.), Cambridge University Press, New York, NY, p. 372.
8. Foley, vanDam, Feiner, Hughes (1990)Computer Graphics: Principles and Practice (2nd ed), Addison-Wesley, Reading, MA, p. 979-982