The VISAR Process

David Hathaway
Paul Meyer
Gary F. Templeton

The Video Image Stabilization And Registration (VISAR) process is an award winning video image processing software developed at NASA's Marshall Space Flight Center. VISAR has a wide variety of application areas where the refinement of digital video is needed. It is used to correct jitter, rotation, and zoom effects by registering and processing on individual image captures that are a part of normal video capturing. Its most prominent uses were the 1996 Olympic Bombing case and in identifying Saddam Hussein during the Iraq war.

Based on first-hand knowledge, this paper describes the VISAR process, which consists of several steps designed to refine digital video using VISAR software. The process determines the differences between two video images so that one, or both, of the images can be changed in ways that make them match as well as possible. Corrections include changes in position (horizontal and vertical image shifts), changes in orientation (image rotation), and changes in magnification (image zoom). While much of the VISAR process is automated, in its current embodiment it requires the user to initially identify the area of interest and to reset a threshold parameter if the default gives unacceptable results.

The basic process that is used is an old tried and true method that determines how well the two images match. This process is called *cross-correlation.* It gives a single number, the correlation coefficient, that is equal to 1.0 if the images are perfectly matched, is equal to 0.0 if the images have nothing in common, and is equal to –1.0 if one image is the negative of the other. This basic process is used by many image stabilization methods. With VISAR we use it in a manner that provides statistical information needed to best determine orientation and magnification.

To understand VISAR, we need to first understand this cross-correlation process. Consider two black-and-white images taken moments apart. Each image consists of a rectangular array of picture elements, *pixels*, with the brightness at each pixel represented by a number (typically 0 for black, 255 for white, and values in between for the shades of gray). Note that video images usually have either 640 or 720 pixels from left-to-right and 480 pixels from top-to-bottom and usually have three brightness values at each pixel, one each for red, green, and blue brightness.

Each image is first rescaled so that the brightness levels are negative for the dark pixels and positive for the bright pixels. This rescaling process is implemented by first finding the average brightness of all the pixels in the image and then subtracting this average from each pixel. This gives positive and negative values. The scale of the variations about this average is given by the *standard deviation*, the square root of the average of these values squared. The value at each pixel is divided by the standard deviation to complete the rescaling process. An example of this process is shown in Figure 1.

The correlation coefficient is then calculated by multiplying the two rescaled images together, pixel-by-pixel, adding up the results of this multiplication and then dividing by the total number of pixels. If the two images are well correlated then the positive values in the bright areas will multiply together to give positive contributions and the negative values in the dark areas will multiply together to give positive contributions as well. The net result will be a correlation coefficient near +1. If the two images are uncorrelated then there will be some positive values in one image that multiply with negative values in the other image to give negative contributions that cancel out the positive contributions where they happen to match. The net result will then be a correlation coefficient near 0. If one image is the negative of the other then positive values will always multiply negative values to give negative numbers. The net result will be a correlation coefficient near –1. An example of this process is shown in Figure 2.

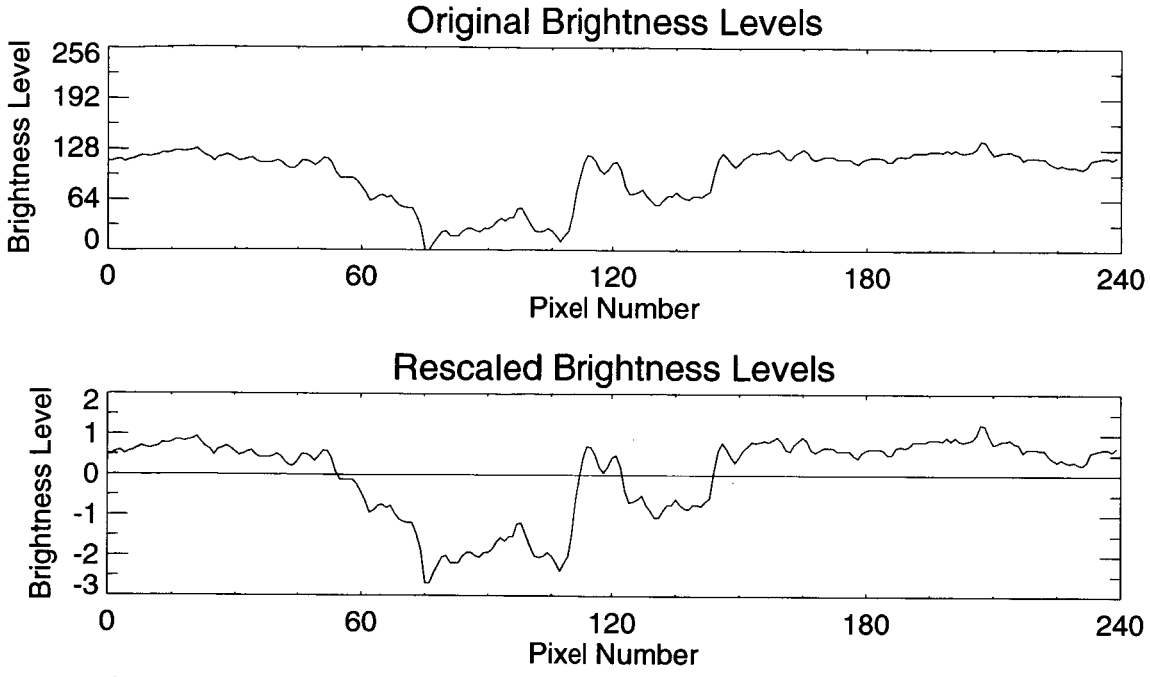

Figure 1. An example of the rescaling process with a cut through a sample image. The original brightness values are shown in the top panel. The rescaled values are shown in the bottom panel.

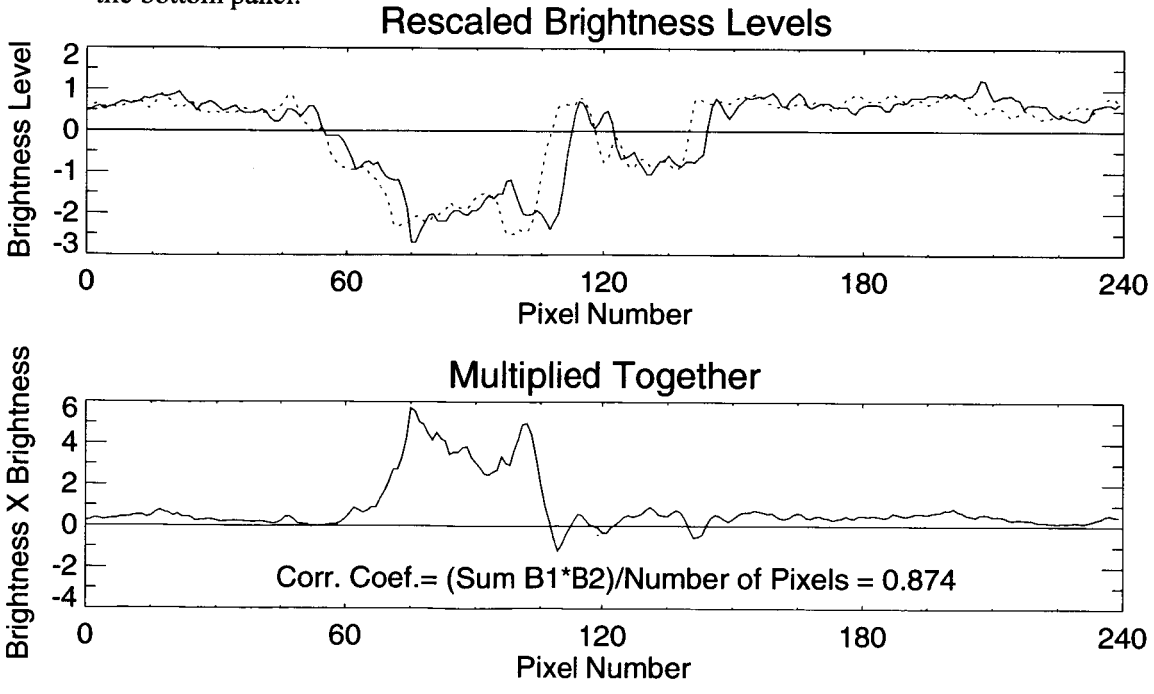

Figure 2. An example of the cross correlation process with cuts through two sample images. The rescaled brightness values are shown in the top panel with a solid line for one image and a dotted line for the other. The pixel-by-pixel multiplication of the two is shown in the bottom panel. The correlation coefficient is the sum of this product divided by the number of pixels.

The VISAR process determines how to shift, rotate, and magnify one image to get the best correlation (correlation coefficient closest to 1.0) between the two images. It does this in a series of steps designed to give the best estimate of the shift, rotation angle, and magnification. The results of this information can then be used to produce a stabilized video sequence (by shifting, de-rotating, and de-magnifying each image).

Step1:

The video *fields* must be extracted from each video frame. Video images are usually *interlaced* images consisting of two video fields. The first video field consists of the even-numbered lines from the top to the bottom of the image. The second video field consists of the odd-numbered lines from the top to the bottom. These two fields are obtained $1/60^{th}$ of a second apart and are interlaced to form a single video frame. The first step is to separate the video fields to produce two separate images. This involves interpolating between the lines. This process is not unique or original and several different interpolation methods may be used. The problem is that each video field represents an individual image but with alternating lines missing. The VISAR process, as a video processing technique, must start with these individual video fields extracted from each video frame. Each of these video fields is reduced to a gray-scale (black and white) image by averaging together the red, green, and blue brightness values at each pixel.

Step 2:

A *key field* and *key area* must be identified. The operator must chose a video field as the key field and outline a rectangular area within this image. All fields in the video sequence will be processed to obtain the best correlation between this key area and corresponding areas within those fields.

Step 3:

The key area is sub-divided into smaller blocks of pixels. The key area is adjusted so that it can be sub-divided into smaller and smaller blocks of pixels (or sub-areas). For example, if the key area was 358 pixels wide by 242 pixels high it would be adjusted to a size of 360-by-240 pixels. It could then be sub-divided into 24 60-by-60 blocks of pixels, each of these would be sub-divided into 4 30-by-30 blocks of pixels and these would, in turn, be sub-divided into 4 15-by-15 pixel blocks. VISAR obtains its information about the image shift, rotation, and magnification by looking at the horizontal and vertical shifts of each of these 384 15-by-15 blocks of pixels. This process is one of the innovations in VISAR and is illustrated in Figure 3.

Step 4.

Produce a *data mask* to block out information from poorly correlated or unwanted blocks of pixels. A data mask is constructed by producing a rectangular array of numbers with each element of the array corresponding to one of the 15-by-15 pixel blocks. The element is set equal to 1 if the corresponding block of pixels is to be included in the calculations. The element is set equal to 0 if the corresponding block of pixels is to be excluded from the calculations. The operator can use this data mask to exclude parts of the key area that are not of interest. The VISAR process automatically excludes pixel blocks that are featureless by setting the corresponding mask value to 0 when the scale of the variations in that pixel block become smaller than a predetermined value. The VISAR process also masks out pixel blocks when the correlation between a key area pixel block and the corresponding block in a new video field is too small. If the correlation coefficient is above some predetermined level (typically 0.7-0.8) then that block is assigned a 1 in the data mask. If the correlation coefficient is less than this level that block is assigned a 0 in the data mask. We use this mask to exclude information from pixel blocks that did not have a good match with the corresponding block in the key area. This data mask allows us to

get more reliable information about the movement of the image. This data mask is an innovation in the VISAR process and it requires the use of correlation coefficients. It helps to give more reliable determinations of the image motion and allows us to correct for parallax effects in which foreground objects appear to move at a different rate or in a different direction than background objects.

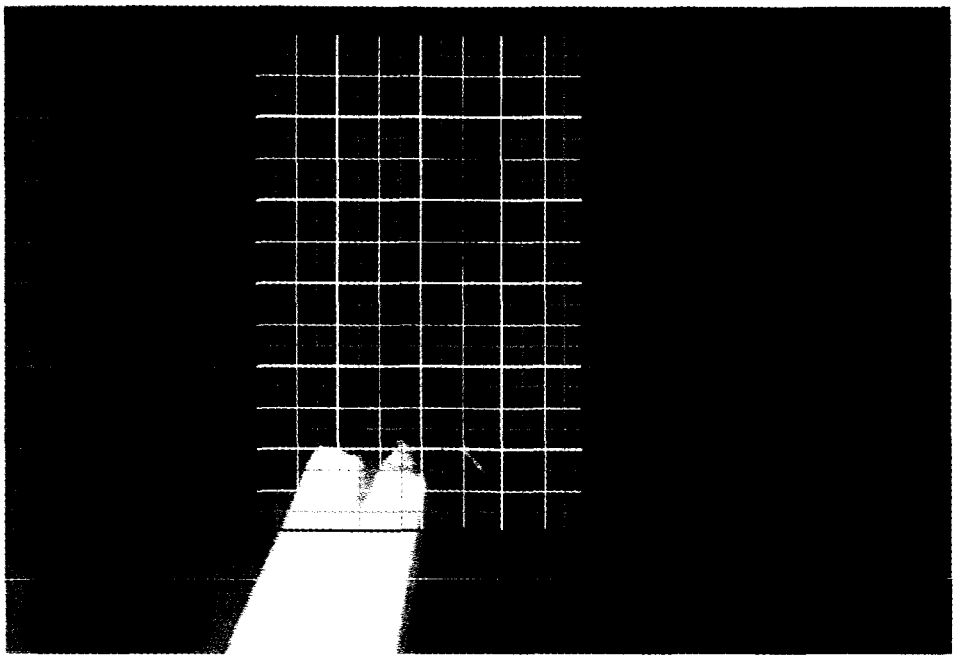

Figure 3. Sample image showing a key area outlined in red lines, 60-by-60 pixel blocks in thick white lines, 30-by-30 pixel blocks in thin white lines, and the 15-by-15 pixel blocks in thin gray lines.

Step 5:

Find the shift between the position of the key area in the key field and the corresponding area in the new video field. The first estimate of the image shift between the key field and a new video field is obtained by finding the corresponding area in the new video field that gives the best correlation with the key area. The horizontal shift is just the difference in the horizontal position between the position of the key area in its image and the position of the corresponding area in the new video field. The vertical shift is the difference in the vertical positions. Note that the corresponding area must be the same size and shape as the key area. We find the corresponding area that maximizes the correlation by first trying the one located at the previous position of the key area in its image and then trying area positioned several pixels side-to-side and up-and-down. We use the correlation coefficient calculated for each of these areas to continue our search until we find the area that gives the correlation coefficient closest to 1. For example, we might calculate correlation coefficients for the central area and corresponding areas 10 pixels to the right, 10 pixels to the left, 10 pixels up, and 10 pixels down. If the central area had the highest correlation coefficient, while the one on the left had a slightly higher coefficient than the one on the right, then we would try another area a little to the left of center to see if it had a higher correlation coefficient than the central one. We use the differences in the correlation coefficients to point the way to the area that maximizes the correlation coefficient. This step is the basic step

that is performed by many other image stabilization processes and it is used as an initial step in the VISAR process.

Step 6:

Find the shift between the positions of each of the smaller blocks of pixels in the key area and those in the new video field. The process described in Step 5 is repeated for each of the smaller blocks of pixels in the key area. The horizontal and vertical shifts determined in Step 5 for the entire key area are used as initial guesses for how the blocks of pixels are shifted. The actual shifts for each block are then determined in the same manner by looking for the position of the corresponding block of pixels in the other video field that maximizes the correlation coefficient. This step is repeated for each of the smaller sized pixel blocks using the shifts for the larger blocks as initial guesses. Thus, the 60-by-60 pixel blocks use the shifts for the full key area as initial guesses, the 30-by-30 pixel blocks use the shifts from their parent 60-by-60s, and the 15-by-15 pixel blocks use the shifts from their parent 30-by-30s. This step is also an innovation in the VISAR process. It is the key to our determination of how the new video field is rotated and magnified. The use of the nested blocks provides a quick and accurate determination of the horizontal and vertical shifts of each of the individual 15-by-15 pixel blocks.

Step 7:

Determine the change in magnification of the new video field. The magnification change between the key field and the new video field is determined by calculating how the individual 15-by-15 pixel blocks spread apart or move together. We calculate the differences in the horizontal shifts for pairs of pixel blocks on each row of blocks in the key area and the differences in the vertical shifts for pixel blocks on each column of blocks. For example, if the first block on a row moved to the left 10 pixels while the block 300 pixels away moved to the left 13 pixels then we would find that the magnification changed by 1% (a 3 pixel shift difference over a 300 pixel distance). The data mask is used to exclude poorly correlated or unwanted pixel blocks. A determination for the magnification change is found for each pair of unmasked pixel blocks (pixel blocks whose corresponding mask value is 1). The best estimate of the change in magnification is found using the statistics of these individual measurements. Measurements from pixel blocks that are widely separated are more sensitive to small variations and are given more weight by including multiple copies (repeated values) for the widely separated blocks. Some individual measurements will be quite different from the average and are excluded if this difference is too large. This use of the difference in the shifts of the pixel blocks to determine the change in magnification is an innovation of the VISAR process. The use of statistics and weighted averages on the multiple measurements is yet another innovation that provides more accurate determination of the magnification change.

Step 8:

Determine the angle of rotation of the new video field. The angle of rotation between the key field and the new video field is also determined by calculating how the individual 15-by-15 pixel blocks move with respect to each other. We calculate the differences in the *vertical* shifts for pairs of pixel blocks on each *row* of blocks in the key area and the differences in the *horizontal* shifts for pixel blocks on each *column*. For example, if the first pixel block on a row moved down 1 pixel while the block 300 pixels away moved up 2 pixels then we would find a rotation of 0.57° (a 3 pixel rise over a distance of 300 pixels gives an angle with a tangent of 3/300 or an angle of 0.57°). Here again the data mask, weighted averages, and statistics are used as they were for the determination of the change in magnification. This method of determining the angle of rotation is another innovation of the VISAR process.

Step 9:

Determine the horizontal and vertical shifts of the new video field. The horizontal and vertical shifts between the key field and the new video field are determined by finding the average horizontal and vertical shifts of those 15-by-15 pixel blocks that are unmasked. These individual shifts are first corrected for the change in magnification and the rotation of this video field relative to the key area as determined in Step 7 and Step 8. This correction is implemented by subtracting the horizontal and vertical shifts that each pixel block would have due to this previously determined magnification and rotation. The overall horizontal and vertical shift for the center of the key area is then obtained using the statistical average of these corrected individual shifts. This provides a more accurate measure of the horizontal and vertical shifts. This is another innovation of the VISAR process.

Step 10:

Repeat steps 1 and 4 through 9 for each of the remaining video fields in the sequence. For each video frame Step 1 is repeated to extract the video fields. Steps 4 through 9 are repeated using the results from the previous field for the initial guess at the position, orientation, and magnification of the new video field. The new video field is then pre-processed by shifting the image back, rotating it in the opposite direction, and de-magnifying it. This preprocessing produces an initial image that is already a much better match to the key field. The subsequent processing then provides corrections to the initial shift, rotation, and magnification. This is another innovation of the VISAR process.

This description indicates that many innovations are a part of the VISAR process. We have tested its capabilities using a series of test images. While the accuracy depends upon the size of the key area, we find that the registration is typically accurate to within 1/10th of a pixel, the magnification is accurate to a part in 1000, and the orientation is accurate to within $1/30^{th}$ of a degree. The results from dozens of real video cases support these values.

VISAR has already been used in dozens of criminal investigations as well as in engineering and medical imaging. The process is covered by two patents (#6,459,822 and #6,560,375) and the process is licensed for use as a forensic tool on Intergraph Corporation's Video Analyst workstations. While its current usage is as a forensic tool, its capabilities also make it an excellent tool for editing video in a wide range of applications...from engineering, medical, and military applications to video editing by both professionals and consumers.