AD

AD-E404 419

Technical Report ARMET-TR-21028

FAST JITTER STABILIZATION OF STATIONARY HIGH-SPEED VIDEO

Ryan Decker
Steven Manole

December 2022

U.S. ARMY COMBAT CAPABILITIES DEVELOPMENT
COMMAND ARMAMENTS CENTER

Munitions Engineering and Technology Center

Picatinny Arsenal, New Jersey

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

1. REPORT DATE December 2022	2. REPORT TYPE Final	3. DATES COVERED START DATE August 2019	END DATE February 2021

4. TITLE AND SUBTITLE
Fast Jitter Stabilization of Stationary High-speed Video

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER 622624
5d. PROJECT NUMBER H1A00	5e. TASK NUMBER	5f. WORK UNIT NUMBER

6. AUTHOR(S)
Ryan Decker and Steven Manole

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army DEVCOM AC, METC Armaments Engineering Analysis and Manufacturing Directorate (FCDD-ACM-AA) Picatinny Arsenal, NJ 07806-5000	8. PERFORMING ORGANIZATION REPORT NUMBER N\A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army DEVCOM AC, ESIC Knowledge & Process Management Office (FCDD-ACE-K) Picatinny Arsenal, NJ 07806-5000	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S) Technical Report ARMET-TR-21028

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report discusses a fast algorithm for image stabilization and jitter reduction for stationary cameras subjected to moderate shaking or vibrational forces. The localized corner-tracking algorithm was developed to stabilize high resolution, mid-wave infrared video but can be used for cameras in other spectrums, including the visible spectrum. The stabilization transform developed can be used to align the camera's field of view to a reference image frame so that measurements throughout a sequence of video are consistent. The algorithm is shown to reduce the effects of camera shake to a level comparable to state-of-the-art methods such as speeded-up robust feature (SURF) registration, while processing in a small fraction of the time (at least 60x faster), making it useful for high-speed video applications.

15. SUBJECT TERMS
Computer vision Image processing Image stabilization Jitter reduction Real-time processing
Camera calibration Photogrammetry Speeded-up robust features SURF High-speed video

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	SAR	19

19a. NAME OF RESPONSIBLE PERSON Ryan Decker	19b. PHONE NUMBER (Include area code) (973) 724-7789

Standard Form 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

CONTENTS

FIGURES

ACKNOWLEDGMENTS

The authors wish to express their gratitude to the In-house Laboratory Independent Research (ILIR) panel at the U.S. Army Combat Capabilities Development Command (DEVCOM) Armaments Center (AC), Picatinny Arsenal, NJ, for their support in conducting this research.

INTRODUCTION

Digital video is used in a variety of applications to make measurements of the real world. One common application of videogrammetry involves the accurate tracking of an object in the field of view. In certain environments, vibrational shaking of the camera system may occur, causing the field of view to shift and adversely affect the measurements of a tracked object. Small-scale shaking, known as jitter, may be caused by a variety of sources, such as an internal fan, engine motion, nearby blast, or even the instability caused if the camera was held by a human camera operator. This report describes an approach to digitally correct for image jitter using a localized corner-tracking approach. The approach, while limited to moderate levels of shaking, has been shown to be as effective as global feature-based registration approaches, and it requires significantly less processing power, allowing for real-time processing of high-speed (120 Hz), high-resolution (1280 x 1024) video.

FEATURE-BASED REGISTRATION

One standard means of jitter reduction is through the use of global feature-based registration. Image features are regions of an image that can be mathematically described and matched between pairs of digital images. Features are usually located at corner points in an image, which are locations where the image gradients are high in both the horizontal and vertical directions, as opposed to edges that have a strong gradient in only one direction. The regions surrounding the corner locations are analyzed using image-processing techniques to develop what is known as a descriptor vector. The descriptor makes use of local gradient information, and once computed for one image, the descriptors from one image can be mathematically compared to numerous descriptors from another image to determine matches. Candidate pairs of matched feature locations are then passed through a nondeterministic random sample consensus (RANSAC) calculator to make a final determination of which pairs are valid matches and which pairs are spurious outliers (ref. 1). The final list of good matches is then analyzed, and a registration transform that maps points in one image to the other is developed.

Speeded-up robust features (SURF) is one of most established feature detection algorithms (ref. 2). For comparison reasons, the baseline image registration using SURF descriptors will be used in this report as the benchmark tool for verification of the new jitter-reduction method proposed. Numerous other approaches were also considered for this research, including features from accelerated segment test (FAST) and the Harris corner detector (refs. 3 and 4). Harris descriptors were found to demonstrate similar jitter reduction performance and processing time as SURF descriptors. The FAST descriptors required significantly less processing time but failed to provide the necessary precision because the corner locations are not computed with subpixel accuracy.

For any stabilization algorithm to work effectively (including the one described next), most of the objects in the image field of view must remain still. This means that if most of the image consisted of swaying trees, then the video cannot be digitally stabilized accurately. Likewise, the scene must be feature rich, meaning mostly sky, ocean, or constant backdrop would be insufficient for effective stabilization.

LOCALIZED CORNER-TRACKING APPROACH

The SURF image registration algorithm is robust to size, rotation, and image intensity, which means it can be used to find matches between images, even those taken with different cameras from different locations, within reasonable limits. However, the processes of determining suitable feature locations, calculating descriptors, and robustly comparing pairs of candidate matches require significant computational expense. In cases where the scene changes relatively little (in terms of lighting intensity, camera position, and camera orientation), much of the global matching and processing is unnecessary. For the analysis of high-speed (faster than 60 Hz), high-resolution (greater than 1 megapixel) video where the camera is held mostly stationary, a simpler approach is possible, similar to that proposed by Raut et al. (ref. 5).

The proposed corner-tracking algorithm for jitter reduction of stationary high-speed, high-resolution camera video involves first initializing locations in the image where robust feature registration is possible. To do this, a reference image is chosen, and conventional featured-based registration with robust SURF or Harris feature matching is used on a subsequent (or previous) image. This identifies regions of the image where robust matches will be made. The details of conventional feature-based registration can be found in reference 4 and are not included here. This step can be conducted before real-time tracking or other processing is conducted.

The robustly matched feature regions of the reference image are then processed using the Harris corner detection algorithm. A tracking window size is determined based on the expected amount of camera movement and is a critical parameter of the algorithm. Using a large window allows the jitter reduction algorithm to compensate for larger amplitudes of camera shaking but significantly degrades computation efficiency, as will be discussed in the performance section of this report. Also, a tracking window that is too large may increase the likelihood that an erroneous nearby corner is found instead of the intended corner within each search window. In addition, corners that are too close to the reference image border must be ignored, so the tracking window size determines how far a corner must be from the image border to be correctly tracked. It was found that a handheld camera exhibits field-of-view movement of many pixels, whereas a stationary camera on a tripod exposed to the shock of a nearby blast only moved around 3 to 4 pixels (0.03 deg). Therefore, it was determined that a movement window size of $k = 7$ was sufficient for the videos captured near the gun blast. Since the overall window size follows the relationship of $2k + 1$ by $2k + 1$, the overall corner-tracking window used was 15 rows by 15 columns of pixels. Although this amount of movement may seem small, failure to account for movements of this magnitude would totally invalidate tracking measurements.

To track the movement of the corners during a period of camera shake, the original image coordinates of the robustly identified corners are processed to find the new subpixel Harris corner location. This algorithm uses the horizontal (I_x) and vertical (I_y) image gradients of the search window (I) to quantify how well each location in the search window constitutes a corner. To do this, the image gradients I_x and I_y are computed by two-dimensional (2D) filtering (correlation or convolution) with a Sobel kernel (ref. 6) as shown in equation 1:

$$I_x = I \otimes \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \qquad I_y = I \otimes \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{1}$$

Next, the Hessian matrix (H) is computed, according to equation 2, over some small summation window (w) within the corner-tracking search window. By experiment, a 5 x 5 region surrounding each point (u, v) was found to be effective and quick for computation of the H matrix. Because of this summation window size, corners cannot be found within 2 pixels of the image border, which must be considered when choosing the size of the corner-tracking search window.

$$H(u,v) = \begin{bmatrix} \sum_{(u,v)\in w} {I_x}^2 & \sum_{(u,v)\in w} I_x I_y \\ \sum_{(u,v)\in w} I_x I_y & \sum_{(u,v)\in w} {I_y}^2 \end{bmatrix} \tag{2}$$

Eigenvalue decomposition of the H matrix (computed at each point in the image) results in two eigenvectors (λ_1, λ_2) that represent the directions of the largest single-directional change and the direction where change is greatest in both directions for each pixel location in the image. The eigenvalue that corresponds to the second eigenvector (λ_2) is thus a metric of how well that location experiences changes in two perpendicular directions, also known as a corner. Furthermore, the relative ratio between the two eigenvalues can be investigated to emphasize points that are corners, and not just strong edges using either of the following relationships:

$$f_{HarrisMetric} = \frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2} = \frac{determinant(H)}{trace(H)} = \frac{H_{11}H_{22}-H_{12}H_{21}}{H_{11}+H_{22}} \tag{3}$$

Points in the image where $f_{\text{HarrisMetric}}$ from equation 3 is largest represent the strongest corners in the image.

As illustrated in figure 1, the top of the nostril represents the strongest corner location, as the gradients in both directions are large at that point. However, the Harris corner metric is only computed on a per pixel basis. To achieve subpixel accuracy of the best-fit corner location, the 3 x 3 region surrounding the pixel with the highest $f_{\text{HarrisMetric}}$ for each corner region is fitted with a two-dimensional (2D) second order polynomial to find the peak value. The location of the maximum value of the polynomial is then chosen as the subpixel estimate of the corner location. An analytical solution in reference 7 provides an efficient subpixel estimate for the peak corner location in each search window. The graphic in figure 2 illustrates surface and 2D image plots of corner detection with subpixel accuracy in a 15 x 15 local tracking image window.

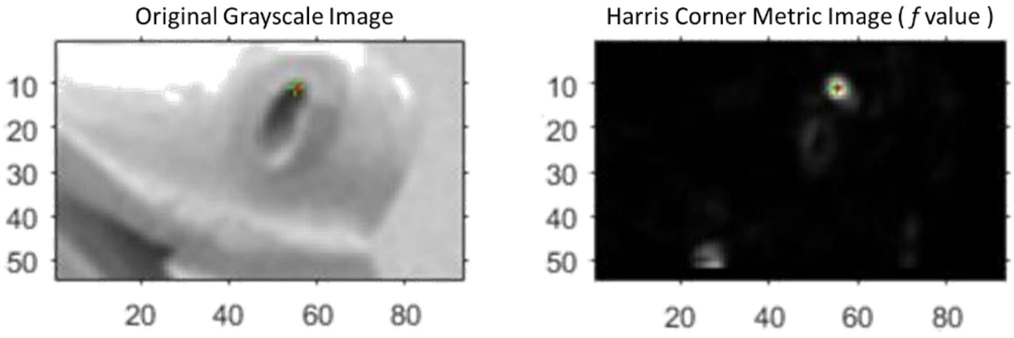

Figure 1
Conversion to Harris corner metric and peak finding from an original grayscale image

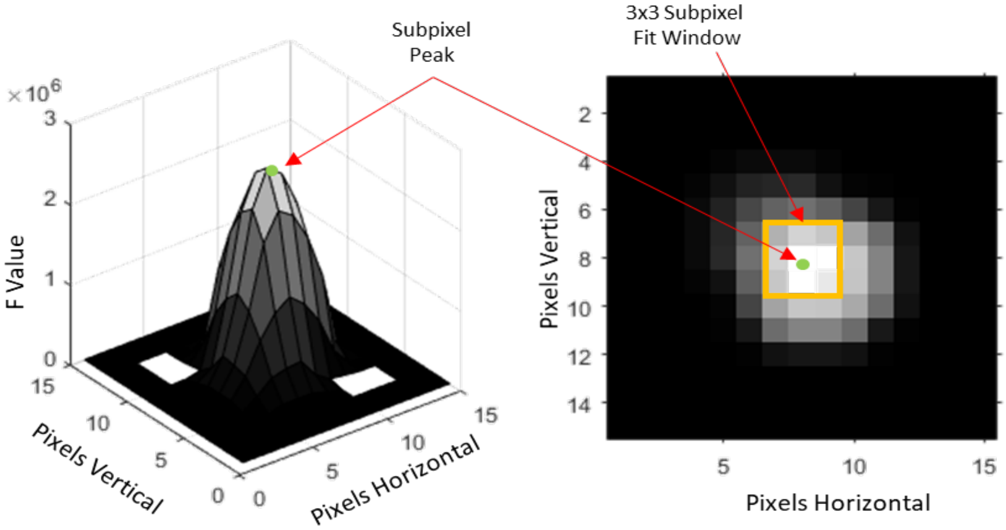

Figure 2
Plots of the Harris corner metric *f* as a surface plot (left) and 2D image (right)

The subpixel corner-locating process is repeated for each of the originally matched robust corner positions for each new frame of the video. Instead of performing the entire feature-matching and registration procedure globally, the new localized subpixel peaks are calculated only in the search windows for each corner being tracked, significantly reducing computation time.

In each new frame, the recalculated corner positions are subtracted from the original subpixel locations to determine the movement of the corners. This movement is not always consistent throughout the entire image due to local variations in intensity and the shaking motion itself. In addition, some of the newly computed subpixel peaks may be erroneous, so to account for this, a simple mean-shift algorithm (ref. 8) is implemented to find the 50% of corners that exhibit similar movement. If the simple mean is used instead of a robust mean-shift measurement, it has been demonstrated that outliers can significantly affect the results as shown in figure 3. Notice that the solid lines (simple mean) experience rapid jumps during some frames, while the points (mean shift) demonstrate consistent motion. Once identified, the movement of the inlier corner locations can be used to develop the coordinate transform back to the reference image.

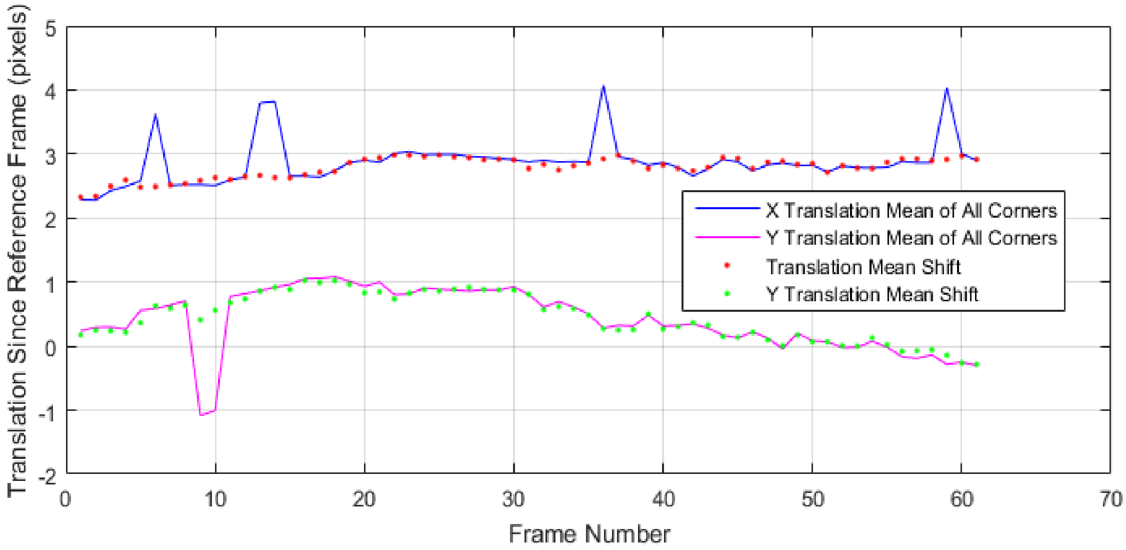

Figure 3
Results illustrating occasional gross errors if robust mean calculation is not used

It should be noted that many consider the shaking motion to have a strong rotational component, but it has been found through experimentation that applying a robust translation only back to the original reference image is sufficient to correct for small-scale jitter effects.

The algorithm works by compensating for jitter motion in both visible spectrum high-resolution video and mid-wave infrared high-resolution video. The plots in figures 4 and 5 illustrate the outlier rejection process using the mean-shift algorithm to identify inlier corner movements, which are used to calculate the transform of the new image from the reference image.

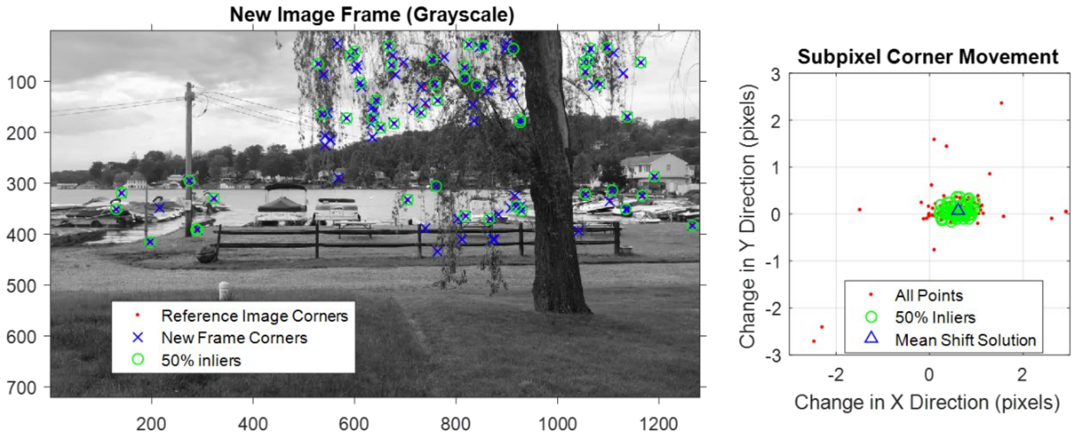

Figure 4
Tracked corner movement and inlier selection for visible-spectrum camera with vibrational shaking

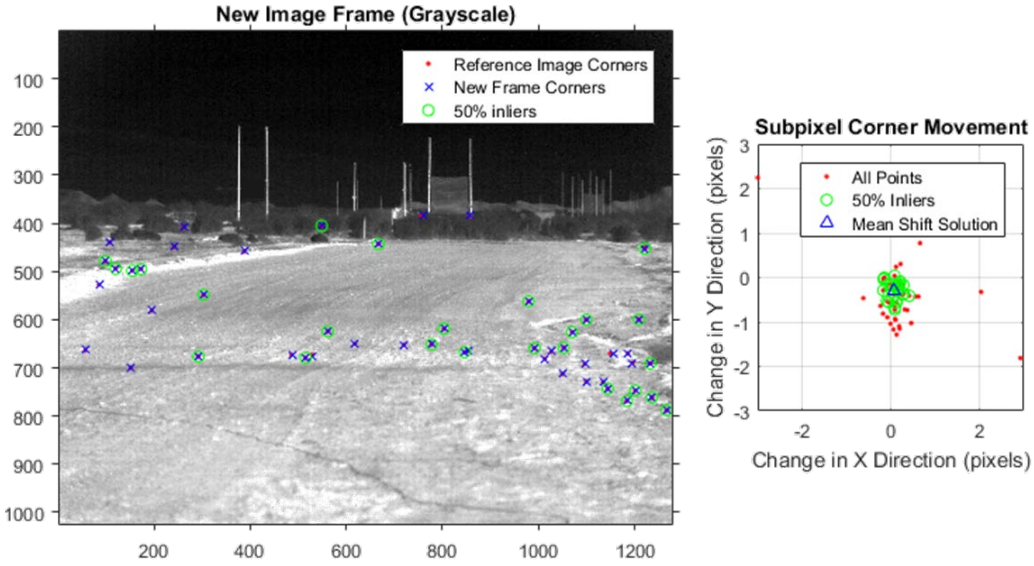

Figure 5
Tracked corner movement and inlier selection for infrared camera near artillery muzzle blast

The algorithm as presented so far is fast and accurate but limited to small translations of the field of view. If larger jitter translations are present (due to larger magnitudes of shake or drift of the field of view), it is possible to augment the algorithm by periodically repeating the initialization process with global feature-based registration between the original reference image and the current image. It is proposed that the amount of total translation since the latest reference image be monitored, and if the cumulative translation from the original reference image is approaching the limit of the current corner-tracking window allowance, then initialization should be repeated. When repeated, the reference image can be replaced by a newer image, or the original reference image can be maintained, but the starting position for corner tracking (center of the tracking windows) should be offset to the new corner-tracking positions.

JITTER REDUCTION PERFORMANCE IN THE MATLAB ENVIRONMENT

A MATLAB (ref. 9) implementation of the corner-tracking algorithm was tested using two different camera systems. One was a visible-spectrum camera (shown in fig. 4) subjected to handheld fluctuating motion (manually stabilized by a human operator). The other was a mid-wave infrared camera (shown in fig. 5) mounted on a tripod but subjected to the supersonic muzzle blast of a 120-mm artillery cannon roughly 70 m from the camera system. In both cases, the shaking motion was compensated for by using two methods: (1) conventional SURF feature-based registration and (2) the new localized corner-tracking method for jitter correction. The translational movement relative to the reference image using SURF and the corner-tracking method show excellent agreement as shown in figures 6 and 7. Although it appears that in a few cases the corner-tracking method slightly overpredicts (within 0.10 pixels) oscillations in scene movement, this could very well be an underprediction on the part of the SURF approach. The difference is so small that even the scintillation around tracked objects in the scene is on the same order of magnitude, making it impossible to determine which method is doing a better job of estimating the true jitter motion. Jitter-compensated videos generated using both approaches appear to the eye to be perfectly compensating for the jitter motion.

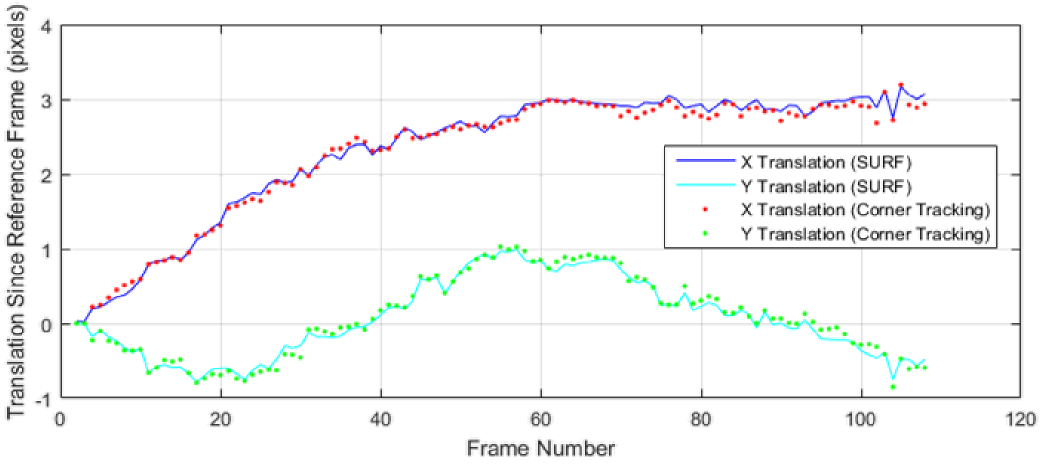

Figure 6
Performance comparison of SURF and localized corner-tracking approaches to jitter reduction on handheld, visible-spectrum video

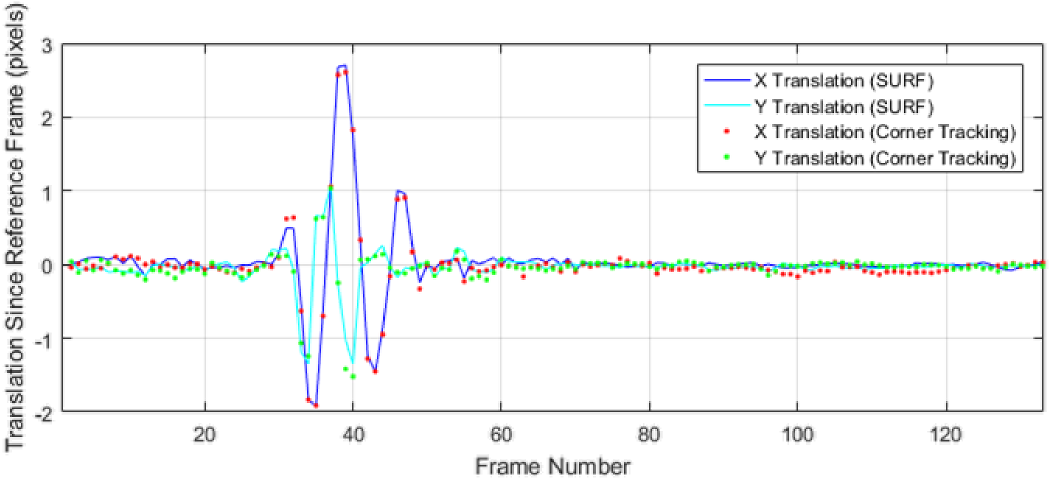

Figure 7
Performance comparison of SURF and corner-tracking approaches to jitter reduction on tripod-mounted infrared video camera subjected to nearby blast disturbance

IMPROVING THE PERFORMANCE AND SPEED OF THE ALGORITHM

As mentioned earlier, a window size metric, k, of 7 (meaning a full window size of 15 x 15 pixels) was chosen for the analyses in the previous section. This window size allows corners to move 5 pixels from the reference frame (because the two rows/columns of border pixels are lost in the calculation of the Harris corner metric). Even with severe blast disturbance and vibrational shaking, this was more than sufficient, as in most cases the corners moved less than a few pixels.

Changing the size of the corner-tracking window significantly affects the computation time as shown in figure 8. In addition, limiting the number of tracked features also has an effect on computation time as shown in figure 9. If computation time is critical and the expected amount of shake can be quantified, it is recommended to use the smallest window size suitable for the application. A noticeable drop in the jitter compensation performance was noticed when the number of tracked features dropped below 30.

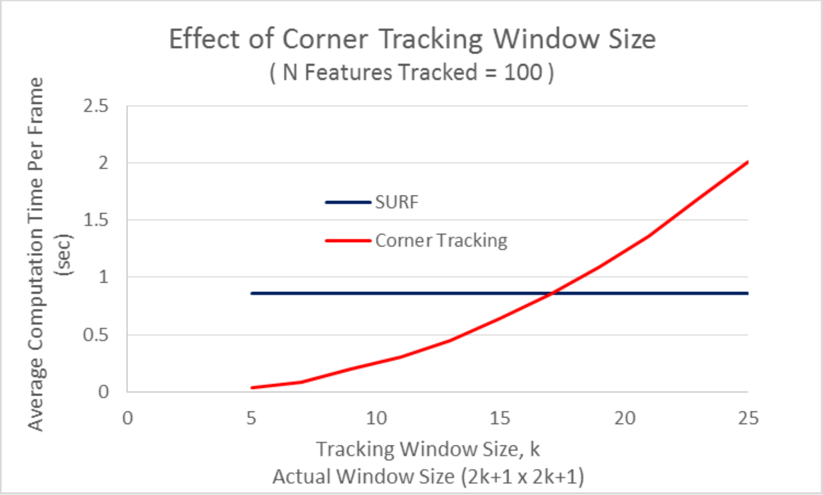

Figure 8
Sensitivity to window size on computation time

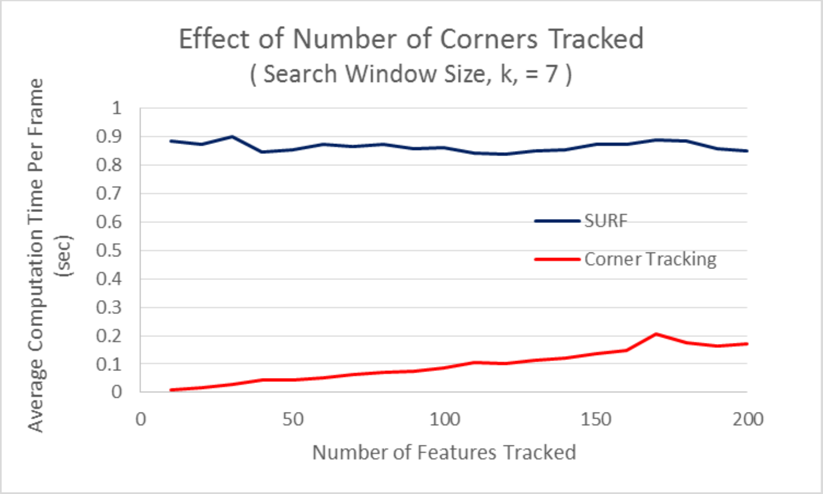

Figure 9
Sensitivity to number of tracked corners on computation time

There are numerous ways to further improve the speed of the corner-tracking method for jitter reduction. Parallel processing or graphics processing unit (GPU) computing are simple to implement as the corners can all be processed independently. This is unlike SURF, which while some elements can be parallelized, such as descriptor calculation or independent RANSAC iterations, the majority of the code is simpler to implement on a single processor. It should also be mentioned that once converted and compiled as C code, the relative improvement in computation time for the local corner-tracking method would be far greater than that of SURF, as many of the SURF registration functions built into MATLAB are already precompiled when run in the MATLAB environment.

The sensitivities to the number of tracked features and window size will be reevaluated in a follow-on effort once converted to run in a C-code implementation. Given that when run in the MATLAB environment the improvement is roughly 60 times faster than SURF (for k = 7 window size), the authors expect a similar implementation to have a greater improvement when run in C code. It should also be noted that when run in the MATLAB environment on a single 2.4-GHz processor, the corner-finding jitter correction runs at roughly 30 Hz. When implemented in C code, the authors expect the algorithm to run as much as 100 times faster (3,000 Hz may be possible).

A final advantage of the corner-tracking method is the ability to avoid regions of the image where dead pixels are located. Most commercially available image sensors exhibit dead pixels, which incorrectly sense the number of photons captured relative to the correctly functioning pixel sensors. For the purpose of generating eye-friendly images, dead pixels are usually digitally masked. To do this, a 3 x 3 median filter replaces the intensity value collected by individual dead pixels (or sometimes all the pixels) by taking the median value of all the adjacent pixels. For quantitative videogrammetry, however, median filters can be problematic, as they introduce error and can significantly interfere with any algorithms that make measurements at the subpixel level. If a dead-pixel map is available, corners in regions with dead pixels can be ignored to increase the validity of the corner location measurements.

Additional augmentation of the method can be used to periodically reinitialize the small tracking windows to allow for larger drifts or amplitudes of camera shaking.

CONCLUSIONS

The localized corner-tracking approach described in this report provides an accurate means to correct for small-scale jitter or camera shaking. This type of correction is critical for cases where high-precision measurements must be made from digital images or video. The method was successfully implemented for cameras that record in either the visible spectrum or infrared mid-wave spectrums and would be expected to work for other types of cameras not tested. Because of the ease of computation, the approach works well for real-time applications using high-speed and high-resolution video. It was demonstrated that the method can run at 30 Hz on a single processor in the MATLAB environment. Once implemented on a faster platform, it is believed that the algorithm could run even faster by several orders of magnitude.

REFERENCES

1. Fischler, M. and Bolles, R., “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography,” Communications of the ACM, vol. 24, no. 6, pp. 381-395,1981.

2. Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L., “SURF: Speeded Up Robust Features,” Computer Vision – ECCV 2006, ECCV 2006, Lecture Notes in Computer Science, vol. 3951, Springer, Berlin, Heidelberg, 2006.

3. Rosten, E. and Drummond, T., “Fusing points and lines for high performance tracking,” Tenth IEEE International Conference on Computer Vision (ICCV’05), vol. 1, pp. 1508–1515, 2005.

4. Harris, C. and Stephens, M., “A Combined Corner and Edge Detector,” Proceedings of the 4th Alvey Vision Conference, pp. 147–151, 1988.

5. Raut, S., Shimasaki, K., Singh, S., Takeshi, T., and Ishii, I., “Real-time high-resolution video stabilization using high-frame-rate jitter sensing,” Robomech Journal, 6, 16, 2019.

6. Sobel, I. and Feldman, G., “An Isotropic 3x3 Image Gradient Operator,” <https://www.researchgate.net/publication/281104656_An_Isotropic_3x3_Image_Gradient_Operator>, 2015.

7. O'Haver, T., “peakfit.m,” Version 9.0, MATLAB Central File Exchange, <https://www.mathworks.com/matlabcentral/fileexchange/23611-peakfit-m>, January 2018, Retrieved 30 November 2022.

8. Cheng, Y., “Mean Shift, Mode Seeking, and Clustering,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(8), pp. 790–799, 1995.

9. “Computer Vision Toolbox,” MATLAB Version 9.4 (r2018a), MathWorks, 2018.

DISTRIBUTION LIST

U.S. Army DEVCOM AC
ATTN: FCDD-ACE-K
FCDD-ACM-AA, R. Decker
S. Manole
FCDD-ACM-LC, A. Hassan
FCDD-ACW-SB, J. Bittner
Picatinny Arsenal, NJ 07806-5000

Defense Technical Information Center (DTIC)
ATTN: Accessions Division
8725 John J. Kingman Road, Ste 0944
Fort Belvoir, VA 22060-6218

GIDEP Operations Center
P.O. Box 8000
Corona, CA 91718-8000
gidep@gidep.org

REVIEW AND APPROVAL OF CCDC AC REPORTS

THIS IS A:

☒ **TECHNICAL REPORT** ☐ **SPECIAL REPORT**

☐ **MEMORANDUM REPORT** ☐ **ARMAMENT GRADUATE SCHOOL REPORT**

FUNDING SOURCE Mostly S&T Funded, some ILIR (for writing)
[e.g., TEX3; 6.1 (ILIR, FTAS); 6.2; 6.3; PM funded EMD; PM funded Production/ESIP; Other (please identify)]

FAST JITTER STABILIZATION OF STATIONARY HIGH SPEED VIDEO
Title

Advanced Lethality Kinetic Energy
Project

Ryan Decker
Author/Project Engineer

2021-03-XX
Report number/Date received (to be completed by LCSD)

X7789
Extension

B94
Building

FCDD-ACM-AA
Author's/Project Engineers Office (Division, Laboratory, Symbol)

PART 1. **Must be signed before the report can be edited.**

a. The draft copy of this report has been reviewed for technical accuracy and is approved for editing.

b. Use Distribution Statement A_X__, B____, C____, D____, E____, or F____ for the reason checked on the continuation of this form. Reason: ____________________

1. If Statement A is selected, the report will be released to the National Technical Information Service (NTIS) for sale to the general public. Only unclassified reports whose distribution is not limited or controlled in any way are released to NTIS.

2. If Statement B, C, D, E, or F is selected, the report will be released to the Defense Technical Information Center (DTIC) which will limit distribution according to the conditions indicated in the statement.

c. The distribution list for this report has been reviewed for accuracy and completeness.

12/8/2022

X Chris Stout

Division Chief (Date)

PART 2. To be signed either when draft report is submitted or after review of reproduction copy.

This report is approved for publication.

12/8/2022

X Chris Stout

Division Chief (Date)

LCSD 49 supersedes SMCAR Form 49, 20 Dec 06