Video Image Registration Evaluation for a Layered Sensing Environment

Olga Mendoza-Schrock
Sensors Directorate
Air Force Research Lab
WPAFB, OH 45433

James A. Patrick
Sensors Directorate
Air Force Research Lab
WPAFB, OH 45433

Erik P. Blasch
Sensors Directorate
Air Force Research Lab
WPAFB, OH 45433

Abstract—**In this paper, several methods to register and stabilize a motion imagery video sequence under the layered sensing concept are evaluated. Utilizing the layered sensing paradigm, an area is surveyed by a multitude of sensors at many different altitudes and operating across many modalities. Utilizing a combination of sensors provides better insight into a situation than could ever be achieved with a single sensor. A fundamental requirement in layered sensing is to first register, stabilize, and normalize the data from each of the individual sensors. This paper extends our previous work [1] to include experimental analysis. The paper contribution provides an evaluation of four registration algorithms now including the (1) *Lucas-Kanade* (LK) algorithm, (2) the Ohio State University (OSU)[1] *correlation-based* method, (3) *robust data alignment* (RDA), and (4) *Scale Invariant Feature Transform* (SIFT). Results demonstrate that registration accuracy and robustness were achieved with the LK and correlation-based methods over the others for image-to-image registration, restricted adaptive tuning, and stabilization over warped images; while the SIFT outperformed the others for partial image overlap.**

Keywords: *Image registration, image registration evaluation, layered-sensing*

I. INTRODUCTION

Image registration methods (IRM) are fundamental for automatic target detection, recognition, and identification [2]; image and data fusion [3]; and video processing, compression, and storage [4]. IRM aligns one or more images taken of the same scene. These images are usually obtained from different sensors, on different platforms, from different viewpoints. 'Image Registration' is a term that captures many products including: image-to-image registration [5][6], cross-modal image-to-image registration [7][8], video registration [10], orthorectification [9], image geo-registration [9][10], and image mosaicking [11].

As the name suggests, image-to-image registration [5], involves aligning two images. Video registration is similar, except the images are typically from the same platform and the same sensor, but from a slowly changing viewpoint [6]. Orthorectification is transforming a set of images so that they have the same viewpoint; it differs from image-to-image registration because the images are first transferred to a similar viewpoint, most often the nadir viewpoint, and then transformed to one another, versus transferring one image to another image [9]. Image geo-registration is the most valuable product involving correlating image pixels with world coordinates to appropriately map objects of interest. Finally, image mosaicking involves stitching two or more images together to create one large image [5].

All of these processes are difficult problems due to changes in the operating conditions (i.e. sensor, targets, and environments) of the images [12]. However, although image registration is a difficult process, it is a requirement for technologies such as data fusion, especially as applied to Geographic Information Systems (GIS) data, change detection, target tracking, recognition, wide area surveillance, and situational awareness which are captured in the Layered Sensing (LS) concept shown in Figure 1.
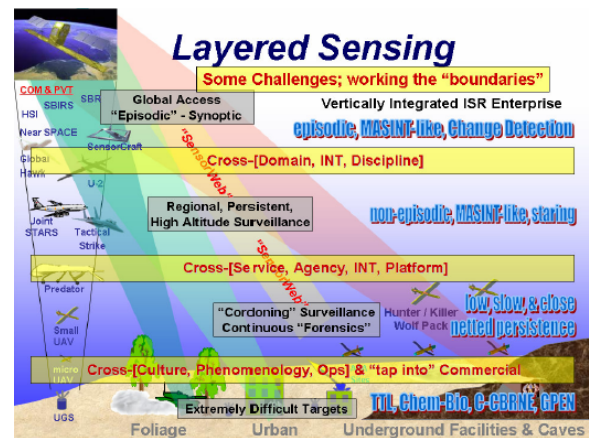

Figure 1.   Layered Sensing Paradigm [13].

Multi-layered Sensing or simply, Layered Sensing (LS), is defined in [13]. The goal of multi-layered sensing is to fuse a combination of sensors to provide timely, trusted, and relevant information. The combination of sensors is multi-dimensional with sensors flying on different platforms, at different altitudes, and capturing data of different phenomenology. The goal of LS is:

> *"Layered Sensing provides military and homeland security decision makers at all levels with timely, actionable, trusted, and relevant information necessary for situational awareness to ensure their decisions achieve the desired military/humanitarian effects. Layered Sensing is characterized by the appropriate sensor or combination of sensors/platforms, infrastructure and exploitation capabilities to generate that situation awareness and directly support delivery of "tailored effects" [13].*

In order to fully exploit Layered Sensing, there is a need to improve enabling technologies such as sensor resource management (SRM), image registration, tracking, and target identification. Image registration is crucial to the Layered sensing construct as it enables simultaneous tracking and target identification.

There are several traditional registration techniques widely used. Fundamentally there are *intensity-based* approaches and *feature-based* approaches [5][6]. Over the years, developments have gone from signal registration [14], to tie-point matching such as mutual information [2][15][8], to combining Fourier and Multiple Classifier (e.g. MUSIC) techniques to aid registration [16]. Different techniques are compared for confidence, accuracy, timeliness, throughput, cost, and robustness. Image registration evaluation requires metrics [17], comparisons to image sets [18], and performance estimation and bounds [19]. Critical to IRM evaluation is a set of data for joint comparisons.

A.   Columbus Large Image Format (CLIF) Data

To research Layered Sensing, two multi-layered data collections are available for public release, namely the *Columbus Large Image Format* (CLIF) 2006 and 2007 data collections. These collections are available on the Sensor Data Management System (SDMS) website at https://www.sdms.afrl.af.mil/datasets/clif2006/. The CLIF 2007 data collection includes three 'layers' of data from multi-layered assets--a high-flying asset, surrogate UAVs, and building mounted sensors. These sensors imaged the same area at the same time—The Ohio State University (OSU) Campus. The high-flying asset provides electro-optical (EO) data from the Large Area Image Recorder (LAIR) sensor—a Wide-Area Surveillance Platform (WASP). The LAIR data is unique as it has a 3.5-4 mile diameter footprint with one-meter resolution, i.e. it can image several city blocks at a time. Furthermore, the asset is flying circular paths so it is able to maintain persistent surveillance over an area. The surrogate UAV layer, which includes EO and mid-wave infrared (MWIR) data are closer-in sensors flying at 2500-6000 feet, and have sub-foot resolution. The final layer is EO and IR data from building mounted sensors which are at 70 feet above ground and have sub-foot resolution. The CLIF 2006 data collection contains similar data but contains two 'layers'—the LAIR layer and the building mounted layer. Figure 2. shows examples of all three layers.
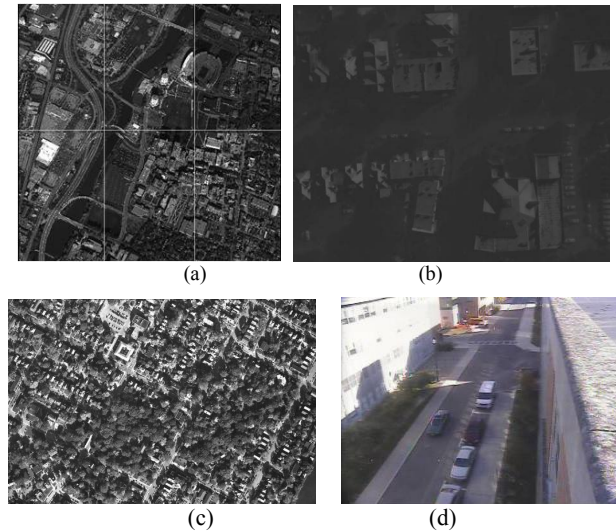

Figure 2.   Example data from the CLIF data collections (a) EO data from the LAIR sensor (b) EO data from the surrogate UAV (c) EO data from the building mounted sensor (d) IR data from the surrogate UAV. Notice that (a) – (c) are viewing the same area and that the vechilcle shown in (c) can also be seen in (a).

The CLIF datasets are very rich and provide a good data set to test algorithms with applications to layered sensing. There are six image registration product needs for the data set:

(1) Very accurate *image-to-image registration*, i.e. frame-to-frame registration, to at least within a pixel accuracy but sub-pixel accuracy is preferred [5].

(2) *Orthorectification* of the data to the nadir (bird's eye) view [9].

(3) *Geo-registration* of the data to world coordinates. This entails aligning the frames to either a digital elevation model (DEM) or another GIS product containing world coordinates of the imaged area.[9]

(4) *Image mosaics* of the data, specifically of the LAIR data. The LAIR is a system composed of six cameras with overlapping areas. The result is a 'frame set' or rather six overlapping frames taken at the same time. The mosaic would create one large image from the six overlapping frames and thus the 'frame set' would just become a single frame [11].

(5) *Geo-registration* of the EO data to 3D Geographical Information System (GIS) model, also referred to as Fusion Level 0 [20].

(6) The ultimate product is *video stabilization* of each of the videos from each of the layers in the data collection which would entail a combination of the previous products [21].

This paper provides an evaluation of the first product, accurate image-to-image registration for different registration

algorithms. The next section provides a general overview of the registration algorithms used in the comparison.

Figure 3. Example data from the CLIF LAIR sensor. The six different frames from one frame set are laid side-by-side to illustrate the need for an image mosaic.

II. REGISTRATION ALGORITHMS

There are several traditional registration techniques widely used. Fundamentally there are *intensity-based* approaches and *feature-based* approaches.[5][6]

The most common technique is to *match image intensities* based on cross-correlation.[14] Cross-correlation (CC) suffers from variations in operating conditions. For example, images taken from EO videos still have changes in lighting conditions due to the difference in the time of day when they are collected. These changes in lighting challenge correspondence accuracy. Typical operating conditions include: differing viewpoint, shadows, weather conditions, sensor type, terrain type, image quality, lighting variation, reference image differences, and quality of metadata. However, CC typically works well for frame-to-frame image registration for video stabilization. Less constrained scenarios require more advanced registration techniques that are invariant to the operating conditions.

Another popular method is *feature-based* registration. The approach extracts salient features from the images and aligns the images by aligning the features. Features, such as points, edges, regions, and most commonly corners, are more robust to changes in lighting and viewpoint than just image intensity correlation.

There are some steps that are common to either technique that a registration algorithm will need to solve: (1) feature extraction (either pixel-based or feature-based 'features'), (2) establishing correspondence between features, (3) calculating a transformation, (4) align images by resampling, and (5) measuring accuracy.

In this paper, we examine four different registration techniques. Three techniques were developed under the Revolutionary Automatic Target Recognition and Sensor Research (RASER) grants and the fourth, the *Lucas-Kanade* (LK) algorithm, was adapted from [10-12] and tailored to the CLIF data set for registration comparison and evaluation. The *correlation-based* approach is intensity-based, the *robust data alignment* (RDA) approach is feature and intensity-based, the *Scale Invariant Feature Transform* (SIFT) is feature-based,

and the LK is intensity-based. The first three algorithms were described in detail in [1] but for convenience this section gives a quick overview of the algorithms with emphasis on the Lucas-Kanade algorithm.

A. The OSU Correlation-Based Appraach

The Ohio State University (OSU) Correlation-Based approach was explored by Professor James W. Davis and Mark Keck in cooperation with the authors [1]. The original algorithm was written in MATLAB and the authors adapted it and rewrote it in C++. The goal is to register and stabilize a series of images taken from a video sequence. In order to do this efficiently, a *Gaussian image pyramid* technique is implemented where the images are first convolved with a Gaussian kernel and then downsampled, for example by a factor of 2 or 4, and the registration process is done at the coarsest resolution. Starting at the coarsest resolution helps to increase the computational efficiency of the algorithm.

Using the images with the coarsest resolution, uniformly distributed 'feature points' are extracted from the target image by dividing the image into a grid and extracting features located at the grid intersections. These features are not traditional features such as corners or edges but rather small, square areas of pixels where the area is referred to as the *Template Area*. The images are assumed to be texturally rich and thus the areas are assumed to be quality areas. The reference image is gridded in the same fashion (using the same grid size) and features are similarly extracted. A projective transformation is then calculated between the pixel area in the target image and the pixel area in the reference image. The target image is then warped by the projective transformation to create a warped image. The *similarity measure* of normalized cross-correlation is then calculated between the warped image and the original reference image.

The goal is to find the *projective transformation* that gives the best value for normalized cross-correlation. Since the images were taken at different times from a moving camera, the corresponding points are assumed to be displaced by some small distance (at the coarsest resolution), thus the image intensity values at corresponding points are not equal. To more accurately estimate the true locations in the reference image, a small region around the coarser correspondences is searched at the finer resolution (via normalized cross-correlation of small image patches). This area is referred to as the *Search Area*. The best matching points continue to propagate down to the finer levels of the image pyramid and the process is repeated, but within a restricted search space. The use of the pyramid constrains the search to a relatively small area, rather than using an extremely large search window at the original resolution. After processing the full-resolution images, corresponding point sets are established. Given these two sets of corresponding points, a projective transformation matrix between the images is estimated. Additionally, the *RANSAC* (RANdom SAmple Consensus) algorithm is used to eliminate outliers in the correspondence set and provide the near optimal inlier set.

B. The Robust Data Alignment approach

The Robust Data Alignment (RDA) approach is a novel approach that transforms the image registration problem into an optimization problem. This section gives an overview of the algorithm but for a detailed explanation of the approach see [22].

The RDA first extracts features from the target and references images. The approach is not tied to a specific feature extractor but the implementation used by the authors, courtesy of Dr. Ümit Özgöner and Sangil Jwa of OSU, uses the Kanade-Lucas-Tomasi (KLT) algorithm that essentially extracts corners from the image. Using the KLT features from the target image a distribution function of those features is created and likewise a probability distribution function (PDF) is created for the reference image. Then the joint PDF for the template and reference image is calculated. The calculation of the two marginal PDFs and the joint PDF is similar to the steps taken for mutual information technique. As in mutual information, the big picture idea is when the images are aligned; their PDFs will be most similar. Hence the problem is to define how to best measure the similarity between the PDFs.

For the RDA, they were able to construct a cost criterion based on information theory for this application. Thus, given the two PDFs the idea is to find an affine transformation that minimizes the following cost function,

$$J(\zeta) := -\frac{1}{N_y} \sum_{i=1}^{N_y} \log \left\{ \frac{1}{N_x} \sum_{j=1}^{N_x} G(\hat{x}_i - x_j) \right\} \approx D(X \| \hat{X}) + H(X), \quad (1)$$

where $\hat{x}_i = \zeta(y_i), \hat{X} = \zeta(Y)$, is a transformed feature point by an affine transformation ζ, and G corresponds to the Parzen-window as described in [22]. An impressive result of the RDA is that finding the transformation that minimizes the cost criterion, is equivalent to minimizing relative entropy; and that it has been show to be more efficient than the mutual information approach [22].

C. The OSU SIFT Approach

The OSU SIFT Approach is an image registration technique that relies on SIFT features. For a detailed explanation of the approach see [1]. The OSU SIFT approach is based on the *Scale Invariant Feature Transform* (SIFT) algorithm, originally developed by David Lowe at the University of British Columbia in conjunction with an implementation of the RANSAC algorithm and a Linear Least Squares fit [4]. The original application of the OSU SIFT Algorithm was to geo-register the CLIF LAIR data. For this paper, the OSU SIFT approach is used for frame-to-frame registration to register the UAV data and evaluate the performance.

The SIFT algorithm is unusual because the detector and descriptor are more closely related than in most algorithms. The SIFT descriptor relies on an orientation assignment given to the detector, so the detector and descriptor are not easily separable.

SIFT looks for local extrema in the scale-space which consists of several Difference of Gaussian (DoG) images. Extrema that are along edges or in low contrast regions are eliminated because they are considered to be unstable. The algorithm then convolves the area immediately around the keypoint with a Gaussian kernel and uses the result to calculate the local image gradients. The image gradients are then used to assign an orientation to the keypoint.

The descriptor is then computed as a 4×4 array of histograms. Each histogram contains 8 bins according to the original implementation of SIFT. The key to the descriptor is that it is calculated relative to the orientation of the keypoint by first resampling the area around the keypoint so that the orientation of the keypoint will be in one of the cardinal directions. By computing the descriptor relative to the orientation of the keypoint, rotational invariance is achieved. The descriptor is computed by convolving the area around the keypoint with a Gaussian kernel with sigma equal to 1.5 times that of the keypoint sigma. The image gradients at each pixel are calculated and the magnitude is added to the correct bin in the histogram, based on the direction of the image gradient.

The OSU SIFT approach then uses RANSAC to find the optimal affine transformation between the image pairs based on the matching SIFT descriptors. The best transformation found by RANSAC is then used to select the points that are closest to each other after the transformation is applied. These points are then used to generate the final transformation between the two images by using a linear least squares fit.

D. Lucas-Kanade Algorithm

The Lucas-Kanade image alignment approach is unique compared to the other algorithms explored in this paper because it doesn't align images by selecting features and solving the correspondence problem [10-12]. Instead LK uses a gradient descent optimization technique, Gauss-Newton, to align images by making use of image intensity differences along with intensity gradient information [8-10]. The approach starts with an initial estimate of the global projective transformation and uses the intensity gradients to improve the transformation estimate in an iterative fashion. The Gauss-Newton optimization relies on local linear approximations and is easily influenced by the initial transformation estimate and local extrema so convergence is not guaranteed. In order to provide more robustness a coarse-to-fine, the Gaussian pyramid, strategy was implemented.

Registration involves aligning an input image $I(x)$ with a template image $T(x)$ using a global warp $W(x; p)$ where x represents pixel coordinates and p represents the global parameterized transformation [27]. For our implementation we chose our transformation/warp, $W(x; p)$, to be a projective homography as seen in Equation (2). Other various parameterized transformations, including affine, are discussed in [27].

$$W(x; p) = \frac{1}{p_7 x + p_8 y + 1} \begin{bmatrix} p_1 x + p_2 y + p_3 \\ p_4 x + p_5 y + p_6 \end{bmatrix} \quad (2)$$

The Lucas-Kanade algorithm tries to minimize the squared differences between the warped image and the template image at each iteration.

$$\sum_x \left[I\big(W(x; p + \Delta p)\big) - T(x) \right]^2 \qquad (3)$$

From [27], the algorithm outline used in our method consists of:

(1) Warp $I(x)$ with $W(x; p)$ to get $I\big(W(x; p)\big)$

(2) Calculate the error image $T(x) - I\big(W(x; p)\big)$

(3) Warp $\nabla I(x)$, gradient of $I(x)$, with $W(x; p)$

(4) Evaluate the Jacobian $\dfrac{\partial W}{\partial p}$

(5) Compute $\nabla I(x)\dfrac{\partial W}{\partial p}$

(6) Compute Hessian using Equation (4)

(7) Compute Δp using Equation (5) and update p

(8) Repeat until convergence, $|\Delta p| \leq \varepsilon$

$$H = \sum_x \left[\nabla I \frac{\partial W}{\partial p} \right]^T \left[\nabla I \frac{\partial W}{\partial p} \right] \qquad (4)$$

$$\Delta p = H^{-1} \sum_x \left[\nabla I \frac{\partial W}{\partial p} \right]^T \left[T(x) - I\big(W(x; p)\big) \right] \qquad (5)$$

where

$$\frac{\partial W}{\partial p} = \begin{bmatrix} \dfrac{x}{\lambda} & \dfrac{y}{\lambda} & \dfrac{1}{\lambda} & 0 & 0 & 0 & -\dfrac{\alpha x}{\lambda^2} & -\dfrac{\alpha y}{\lambda^2} \\[2mm] 0 & 0 & 0 & \dfrac{x}{\lambda} & \dfrac{y}{\lambda} & \dfrac{1}{\lambda} & -\dfrac{\beta x}{\lambda^2} & -\dfrac{\beta y}{\lambda^2} \end{bmatrix} \qquad (6)$$

and

$$\alpha = p_1 x + p_2 y + p_3 \qquad (7)$$

$$\beta = p_4 x + p_5 y + p_6 \qquad (8)$$

$$\lambda = p_7 x + p_8 y + 1 \qquad (9)$$

Steps (1) and (3) required warping the image and its gradient image. In order to perform the resampling, our implementation uses bilinear interpolation which involves linear weighting based on the 4 nearest pixel intensity values.

III. EXPERIMENTS AND RESULTS

The evaluation consisted of five experiments as explained in the author's previous publication [1]. For convenience those experiments are quickly outlined in the next section. Furthermore, the results of those experiments are presented.

A. Experiment 1(Image-to-Image Registration)

For the first experiment, each registration algorithm was used for *frame-to-frame registration* (F-F) of the EO unmanned aerial vehicle (UAV) data. For the F-F experiment, a transformation was found between Images 1-2 and then Images 2-3 and so on. The reference image was then warped to the reference image using four different types of interpolation routines--nearest neighbor, bilinear, bicubic spline, and Lanczos. In [1], Mendoza-Schrock et al provides an explanation of these routines. The root mean square error (RMSE) was then calculated for each of the pairs. First, a sample dataset was used to find the optimum parameters for the algorithm. The training set consisted of three sub-datasets chosen to be representative of the entire dataset. Each sub-dataset contained unique structure, for example the first was over a body of water, the second was over an urban environment, and the third was over an area with high levels of vegetation. There were a total of 296 frames, of size 4004x2672 pixels, used for the experiments where Frame 1 is entitled "EO_Run01_s2_301_15_00_31.993190_1" in the original CLIF 2007 EO UAV data collection, and Frame 2 is entitled "EO_Run01_s2_301_15_00_31.993190_2" and so on. The results of this experiment are shown in Table 1. The bicubic spline interpolation routine consistently gave slightly better results (on average one tenth of a pixel) so the results using the bicubic spline routine are reported in Table I.

TABLE I.

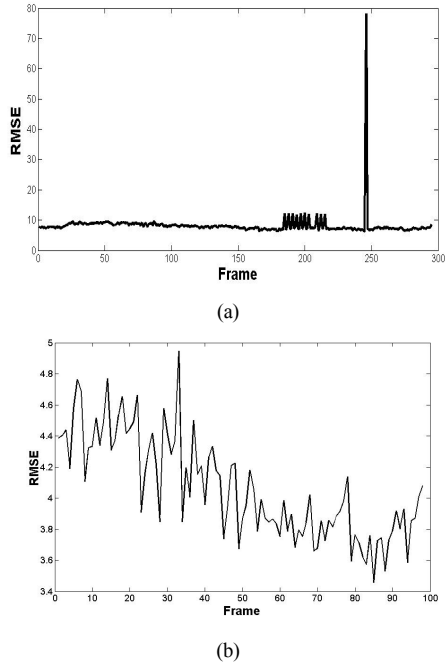| RMSE Results for Experiment 1			
Algorithm	*Median RMSE*	*Mean RMSE*	*Standard Deviation*
Correlation-Based Approach	7.6862	8.2159	4.9661
RDA Approach	9.1542	9.8452	5.2024
SIFT Approach	9.2506	9.4069	2.6130
LK Approach	7.4788	7.9908	4.2682

(a)

(b)

Figure 4. Evaluation of the registration accuracy and robustness of the LK approach. (a) Results for frame-to-frame UAV EO data. (b) Results for LAIR EO data using tuning parameters of UAV EO data.

B. Experiment 2 (Robustness over tuning)

For the second experiment, each registration approach was used for frame-to-frame registration of the EO LAIR data, again using all four different types of interpolation routines. However, the algorithms were *not re-tuned* for the new dataset. The goal of this experiment was to see how well the algorithm performed without being re-tuned for this specific dataset. This gives an indication of the algorithm's robustness. There were a total of 100 frames, of size 2672x4008 pixels, used for this experiment where Frame 1 is entitled "Camera 3 Frame 319" in the original CLIF 2006 EO LAIR data collection. The results are shown in Table 1I. Again the bicubic spline interpolation routine consistently gave slightly better results (on average one tenth of a pixel) so the results using the bicubic spline routine are reported in Table II.

TABLE II.

RMSE Results for Experiment 2			
Algorithm	*Median RMSE*	*Mean RMSE*	*Standard Deviation*
Correlation-Based Approach	3.8114	3.8461	0.3123
RDA Approach	9.6577	11.0055	5.4599
SIFT Approach	9.8566	10.4649	3.6886
LK Approach	3.8114	3.8461	0.3123

C. Experiment 3 (Robustness over Stabilization)

Next, the UAV EO data was *stabilized* using each algorithm. Pairwise transformations were concatenated

throughout the sequence to register each frame back to the first frame of the sequence, i.e. instead of producing the results for frame-to-frame registration; the transformations were multiplied to produce a stable video sequence. Thus, to get the transformation for Frame 3, the transformation from Frames 1-2, and the transformation from Frames 2-3 were multiplied, then to get the transformation for Frame 4, the transformation from Frames 1-2, the transformation from Frames 2-3, and the transformation from Frames 3-4 were multiplied, and so on. Since the flight paths were linear, dog bone patterns, the video was restarted every time half the image was black, i.e. 50% of the resulting frame was black pixels. This resulted in multiple sequences of stabilized video, where the number of sequences varied by algorithm. This provides a micro-analysis into the data, and shows the specific part of the data that the algorithm had trouble with. The results for the Lucas Kanade approach are shown in Table III. To eliminate unnecessary repetition, the results for the other three approaches are found in [1].

TABLE III.

Results for Experiment 3					
Seq.	*Frames*	*Number of Frames in Sequence*	*Median Intensity RMS Error*	*Mean Intensity RMS Error*	*Standard Deviation*
1	001-032	031	7.1346	8.6345	2.4024
2	033-062	029	8.4412	8.4293	0.3245
3	063-094	031	8.3374	8.4110	0.7531
4	095-124	029	7.3415	7.6948	0.8706
5	125-152	027	7.3109	7.4660	0.5682
6	153-180	025	6.4080	6.6068	0.7795
7	181-213	032	7.5853	8.6602	2.4238
8	214-240	026	6.9169	7.3652	1.2392
9	241-246	005	6.0108	6.2413	0.5859
10	247-276	029	7.9719	9.2153	2.5755
11	277-296	019	6.9050	7.2071	1.0973
Overall Mean		295	7.3058	7.8120	1.2382

D. Experiment 4 (Robustness over warping)

Experiment 4 also provides an indication of the algorithm robustness. The images were warped using a transformation consisting of the product of the previous transformations as described in Experiment 3. The warped image at each stage was then compared to the first image and the RMSE was calculated, and the result is shown in a graph. The RMSE is expected to increase over time as errors will accumulate through the propagation. Sources of error will be from the registration algorithm, moving vehicles, and the 'parallax problem'. The 'parallax problem' is a byproduct of trying to perform 2D registration of a 3D object. Typically one focuses on the ground plane, and the taller an object is, the more out-of-plane movement there will be between images and the more error in trying to find a 2D transformation to describe the change in viewpoint. The results for the Lucas Kanade approach are shown in Figure 5. This shows the propagation of error over time for the first 30 frames compared to the first frame. Here flatter trends are more desirable, and the curve appears to follow a logarithmic trend with an R^2 value of

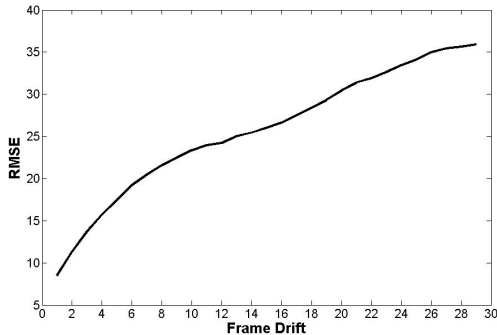0.9735. To eliminate unnecessary repetition, the results for the other three approaches are found in [1].

Figure 5.   Results of Experiment 4 for the LK approach.

E.   *Experiment 5 (Robustness over partial overlap)*

Experiment 5 is another indication of the robustness of the algorithm, specifically in regards to partial overlap. Specifically, a transformation was found between the first frame and the *n*th frame, where *n* ran from 2 to 295. Eventually the partial overlap, or the translation and rotation change will be so large, that the algorithm will not be able to find a suitable transformation. The more robust an algorithm is, the farther it will progress before it breaks down.

TABLE IV.

RMSE Results for Experiment 5			
Algorithm	*Ending Frame*	*Starting Frame*	*Percent Overlap*
Correlation-Based Approach	8	1	85
RDA Approach	3	1	95
SIFT Approach	58	1	16
LK Approach	6	1	89

IV.   DISCUSSION AND FUTURE WORK

In the previous section, five experiments were explained and the results for the experiments using four different registration algorithms were reported (CC, RDA, SIFT, and LK). The LK approach outperformed the other three registration algorithms in the first three experiments. These results are not surprising since the KL approach is very similar to the OSU Correlation based approach and the OSU correlation-based approach outperformed the other algorithms in similar experiments [1]. The LK was specifically set-up to optimize the RMSE and to perform well on single modality data so it should naturally outperform the other algorithms. For Experiment 2, where the LAIR data was registered using the tuning parameters from the UAV data, the LK approach was expected to outperform all the other algorithms. As expected it outperformed the RDA and the SIFT but tied with the LK. For Experiment 5, the SIFT algorithm still outperformed all the other algorithms. This is not surprising as

the SIFT features are very salient and persist over a long sequence of frames affording video registration.

For frame-to-frame registration of the EO UAV data, all algorithms produced RMSEs in the range of 8-9 m. All four algorithms had trouble with the Frames 185–215. Frame 200 is shown in Fig. 1(c). These frames were predominately imaging vegetation in residential areas. For all algorithms, there was a unique spike in the intensity RMSE from Frame 245-246. What occurred here is that there were a couple of lost frames in the original data. This is a real-world event that occurs often, and illustrates a need for robust registration algorithms that can recover from the loss of frames.

Future work will focus on *registration evaluation metrics* in a sensitivity study (e.g. registration robustness versus operating conditions). RMSE provides a reasonable measure with the assumption that frame rates are high enough and scene changes (parallax and moving vehicles) are minimal. However, instead of using RMSE perhaps median or mean of the difference of the absolute values would be a better metric since it would not be as easily affected by the outlier frames such as Frame 245, where a loss of frames occurred. Additionally, a ground-truth effort will be performed so non-intensity-based metrics can be employed. Further, the algorithms will be used to produce georegistered results where data is fused into a GIS. The image-GIS fusion will solve the problem of running into the 'black' pixel issue with video stabilization as described in Experiment 3. Also, bundle adjustment will be utilized to reduce error propagation.

Preprocessing work such as image orthorectification also is needed so that all images are aligned in regards to viewpoint. Improvements will include more sophisticated multi-dimensional optimization techniques with each algorithm so that optimum parameters can be more accurately identified (and evaluated) instead of optimizing for one parameter at a time, independent of the other parameters.

Additionally recall that, the OSU Correlation-Based approach solved for a *projective transformation*, while the others solved for an *affine transformation*. Also, the OSU Correlation-Based approach has to some degree a built-in capability to self-tune to find the optimal parameters, i.e. after setting the initial conditions it self-tunes for number of features, window size, and RANSAC parameter. Self-tuning a feature implemented by the authors. In order to be able to compare apples-to-apples, all algorithms should solve for similar transformations and all algorithms should be adjusted to have self-tuning capabilities. Finally, other RASER Grant algorithms will be brought into the evaluation, as well as other state-of-the-art-algorithms found in the literature.

V.   SUMMARY

There are several traditional registration techniques based on feature and intensity based methods. This paper summarizes four different algorithms. The first approach is intensity-based—The OSU *Correlation-Based* approach. The second approach is feature and intensity-based—the *Robust Data Alignment* (RDA) approach. The third approach is a

feature-based approach—the OSU *Scale Invariant Feature Transform* (SIFT) approach. The fourth approach—the *Lucas-Kanade* (LK) approach—is intensity-based. All approaches were utilized over a common data set for registration evaluation for frame-to-frame registration in a series of experiments using the CLIF data collections—specifically the EO UAV data and the EO LAIR data.

In future work, multi-modality data will be brought into the experiments to demonstrate the strength of the RDA approach as it is specifically designed to handle multi-modality data. Both of the CLIF data collections have overlapping IR and EO data, making these collections well suited for further experiments.

REFERENCES

[1] O. Mendoza-Schrock, J. Patrick, and M. Garing, "Exploring image Applications registration algorithms for layered sensing," Evolutionary and Bio-Inspired Computation: Theory and Applications III. Edited by Teresa H. O'Donnell, Misty Blowers, and Kevin L. Priddy. *Proceedings of the SPIE*, May 2009.

[2] E. P. Blasch and M. Bryant "Information Assessment of SAR Data For ATR," *Proceedings of IEEE NAECON*. Dayton, OH, 1998.

[3] S. Gupta, K. Ramesh and E. Blasch, "Mutual Information Metric Evaluation for PET/MRI Image Fusion," *IEEE NAECON*, 2008.

[4] J. Patrick, R. Brant, and E. Blasch, "Hyperspectral Imagery Exploitation and Fusion Evaluation over Compression," *ISIF Fusion08*, 2008.

[5] A. Goshtasby, *2-D and 3-D Image Registration: for Medical, Remote Sensing, and Industrial*, Wiley-Interscience, 2005.

[6] L. Brown. "A survey of image registration techniques", *ACM Computing Surveys*, 24(4), pp. 276-325, 1992.

[7] B. Fei, *et. al*, "Regisrtation and Fusion fo SPECT, High-Resolution MRI, and Interventional MRI for Thermal Ablation", *IEEE Trans. On Nuclear Science*, Vol. 51, No. 1., 2004.

[8] H-M. Chen and P. K. Varshney, "Mutual Information-based CT-MR Brain Image Registration Using Generalized Partial Volume Joint Histogram Estimation," *IEEE Trans. On Medical Imaging*, Vol. 22, No. 9, 2003.

[9] E. M, Mikhail and J. S. Bethel, *Introduction to Modern Photogrammetry*, John Wiley & Sons, Inc., 2001.

[10] K. Sankaranarayanan and J. Davis, "A Fast Linear Registration Framework for Multi-Camera GIS Coordination," *IEEE Conf. Adv, Video and Signal-based Surveillance*, 2008.

[11] T. Wang, Z. Zhu, and E. Blasch, "Bio-Inspired Adaptive Hyperspectral Imaging for real-Time Target Tracking " *IEEE Sensors Journal*, 2009.

[12] B. Kahler, E. Blasch, and L. Goodwon, "Operating Condition Modeling for ATR Fusion Assessment SPIE Defense and Security 2007, Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications Conference, Orlando, FL, April 2007.

[13] M. Bryant, P. Johnson, B. M. Kent, M. Nowak, and S. Rogers, "Layered Sensing: Its Definition, Attributes, and Guiding Principles for AFRL Strategic Technology Development", http://www.wpafb.af.mil/shared/media/document/AFD-080820-005.pdf , 2008.

[14] A. Goshtasby, S. H. Gage, and J. F. Bartholic, "A Two-Stage Cross Correlation Approach to Template Matching," *IEEE Pattern Analysis and Machine Intelligence*, Vol. 6, No.3, 1984.

[15] J. S. De Bonet and A. Chao, "Structure-driven SAR image Registration," *Proc. of SPIE*, 1998.

[16] M. Xu and P. K. Varshney, "A subspace Method for Fourier-Based Image Registration," *IEEE Geoscience and Remote Sesning Letters*, Vol. 6, No. 3, 2009.

[17] O. L. Mendoza, G. Arnold and P. Stiller, "Further Exploration of the Object-Image Metric with Image Regisrtation in Mind," *Proc. Of SPIE*, Vol. 6974, 2008.

[18] D. A. LeMaster, "A comparison of Template Matching Regisrtation Methods for Polarimetric Imagery," *IEEE Aerospace Conf.*, 2008.

[19] M. Xu, H. Chen, and P. K. Varshney, "Ziv-Zakai Bounds on Image Registration," *IEEE Trans. On Signal Processing*, Vol. 57, No. 5, 2009.

[20] E. Blasch, I. Kadar, K. Hintz, J. Salerno, C. Chong, J. Salerno, and S. Das, "Resource Management Coordination with Level 2/3 Fusion," *IEEE AES Magazine ,* Mar. 2008.

[21] J. S. Jin, Z. Zhu, and G. Xu, "Digital Video Sequence Stabilization Based on 2.5D Motion Estimation and Inertial Motion Filtering," *Real-Time Imaging,* Vol. 7, No. 4, Academic Press, 2001.

[22] S. Jwa, T. Zhijun, and Ö. Ümit, "Robust Data Alignment Based on Information Theory and Its Applications in Road Following Situation," *IEEE International Conference on Intelligent Transportation Systems*, Toronto, Canada, 2006.

[23] K. Redmill, J. Martin, and Ö. Ümit, "Aerial image registration incorporating GPS/IMU data", Evolutionary and Bio-Inspired Computation: Theory and Applications III. Edited by Teresa H. O'Donnell, Misty Blowers, and Kevin L. Priddy. *Proceedings of the SPIE*, May 2009.

[24] D. Lowe, "Object recognition from local scale-invariant features." In *Int. Conf. on Computer Vision*, Corfu, Greece, pp. 1150-1157, 1999.

[25] B.D. Lucas, *Generalized Image Matching by the Method of Differences*, Ph.D. diss., Carnegie Mellon University, July 1984.

[26] B.D. Lucas, and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision (IJCAI)," *Proceedings of the 7th International Joint Conference on Artificial Intelligence* (IJCAI '81), pp. 674-679, April 1981.

[27] S. Baker, and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework: Part 1," Robotics Institute, Carnegie Mellon University, tech. report CMU-RI-TR-02-16, July 2002.