GEODETIC ALIGNMENT OF AERIAL VIDEO FRAMES

Y. SHEIKH, S. KHAN, M. SHAH
Computer Vision Laboratory
Computer Science Department
University of Central Florida
Orlando, FL 32816
{yaser, khan, shah}@cs.ucf.edu

AND

R. CANNATA
Harris Corporation
GCSD, P.O. Box 37
Melbourne, FL 32902
rcannata@harris.com

1. Introduction

With the sophistication of artificial vision systems, the need to endanger human lives for many hazardous activities is increasingly proving to be avoidable. From aerial reconnaissance missions to space exploration, many projects stand to benefit, in particular, from the sophistication in techniques to precisely find world positions of objects present in video data. Unfortunately, mechanical automation of such a task is complicated by the narrow fields of view of video data and the inaccuracy of mechanical information available describing the position of the camera in the world. Instead, computer vision techniques can be used to successfully align any given video frame with pre-calibrated reference imagery. After alignment, a video frame inherits pixel-wise calibration and as a consequence objects in the frame are *exactly* placed in the world. This ability to accurately position objects like buildings, roads, landing sites and spatial landmarks in general, facilitates precise automation of actions that previously required human intervention. The core challenge then is to develop techniques to autonomously align video sequences to pre-calibrated reference imagery.

Significant research effort has been expended towards frame-to-frame registration (the spatial alignment of successive frames of a video sequence) and it is now largely acknowledged to be a solved problem. However, frame-to-frame registration techniques are not easily generalized to frame-to-*reference* registration (alignment of video frames with pre-calibrated reference imagery), since the reference image may be captured from a different viewpoint, through a different modality or at a different time altogether. As a result, the mapping between corresponding pixels in the frame and reference is often highly complex and unmodelled. In particular, the large duration of time that may elapse between capturing of the reference and the video frame can produce distortions from extreme change in illumination to the total absence of certain visual features in either one of the two images. Furthermore, inconsistencies of textured areas like forests or plateaus may be introduced due to seasonal changes, due to changes in illumination or simply because of intrinsic differences in cameras. Clouds, blurring, and occlusion by vehicle parts may exacerbate these problems even further. As these problems are not encountered in frame-to-frame registration problems, related registration techniques do not take them into consideration. Furthermore, from a conceptual point of view, however accurate frame-to-frame registration may be, it can only provide positional information of a given object relative to the camera. In order to accurately recover the absolute position of an object in the world (in the form of geo-coordinates or any fixed world coordinates), some accurate standard of reference is required.

However, despite these limitations, the *framework* for frame-to-frame registration is useful in approaching frame-to-reference registration as well. Image registration, in general, can be defined as a search for the ideal spatial transformation between two images. If $I_1(\vec{x})$ and $I_2(\vec{x})$ are the two image arrays, their relationship is defined as

$$I_1(\vec{x}) \equiv I_2(f(\vec{x})), \tag{1}$$

where $f(\vec{x})$ is the set of allowable transforms for each image $I_2(\vec{x})$. Within the taxonomy of [14], parametric alignment is achieved by a search over the transformation parameters, $\vec{p}$, that would maximize some global measure of 'fit' or similarity, between $I_1(\vec{x})$ and $I_2(W(\vec{x};\vec{p}))$, where $W(\vec{x};\vec{p})$ is the set of allowable *parametric* transformations.

The situation addressed in this chapter is the geo-registration of an incoming video frame with precisely calibrated reference imagery. The video frame is captured by a camera mounted on an aircraft and is referred to as the Aerial Video Frame, $I_{video}(\vec{x})$. As shown in Figure 1, the angle at which an aerial photograph is taken is often used to classify the photograph into one of three types: Nadir, High Oblique and Low Oblique. Photographs are classified as Nadir when the camera axis points directly downwards, as High

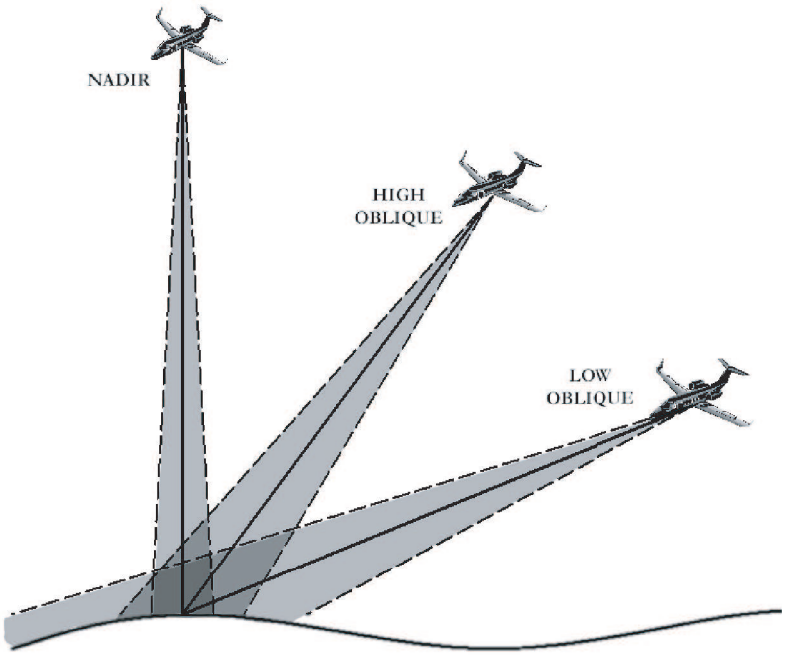

Figure 1. Aerial Photograph Classifications. Depending on the angle of the optical axis the aerial photograph can be classified into one of the following three categories: Nadir, High Oblique and Low Oblique.

Oblique when the camera axis makes a large angle with the ground and Low Oblique when the horizon is visible in the photograph. These angular differences are computed using the position and attitude of the camera relative to a point in the real world which is detailed in the telemetry (meta-data) accompanying each video frame. Telemetry is an automatic measurement of data that defines the position of the camera in terms of nine parameters: vehicle latitude, vehicle longitude, vehicle height, vehicle roll, vehicle pitch, vehicle heading, camera elevation, camera scan angle and camera focal length. This telemetry information can be used in conjunction with a sensor model to place the video frame relative to the Reference Imagery in a world coordinate (or vice versa). The Reference Imagery is a high-resolution orthographic image, usually with a Ground Sampling Distance of $\sim$1 (meaning a pixel corresponds to 1 m^2 on ground). This Reference Imagery is geodetically aligned, and has an associated Digital Elevation Map (DEM), so that each pixel of the Reference Imagery has a precise longitude, latitude, and height associated with it. Figure 2 pictorially explains the nature of the Reference Imagery available along with its associated DEM. The Reference Imagery, which covers a substantial area, is cropped on the

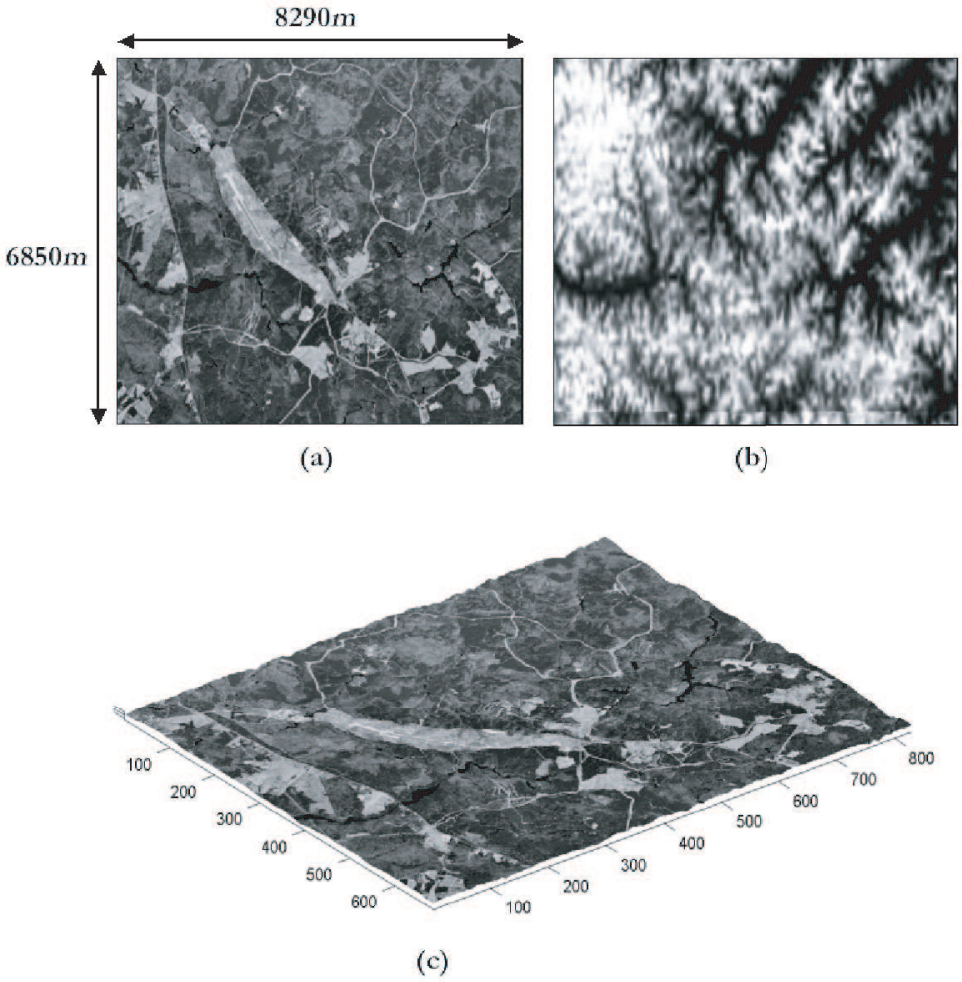

Figure 2. The Reference Imagery and its associated DEM. The Reference Imagery is geodetically aligned, i.e. each pixel has a longitude, longitude, and elevation associated with it. (a) The Reference Image is a high-resolution intensity array (6856 x 8292) with each pixel corresponding to $1m^2$ in the real world. (b) The Digital Elevation Map (DEM) is of lower resolution compared to the Reference Imagery but can be interpolated to provide an elevation for each reference image pixel. Array elements of higher elevation are shown to have brighter intensity values. (c) The 3 dimensional display of the texturized elevation map. The axes are 10^2 m i.e. 1 pixel corresponds to $1m^2$ in the real world.

basis of the telemetry data to a smaller area corresponding to $I_{video}(\vec{x})$ (see Figure 4). This Cropped Reference Image is subsequently referred to as $I_{ref}(\vec{x})$.

There are several challenges specific to Aerial Video Geo-Registration

that can be identified individually. First, it should be noted that the two imageries are in different projection views: $I_{video}(\vec{x})$ is an image of perspective projection, whereas $I_{ref}(\vec{x})$ is an image of orthographic projection. While the telemetry information can be used with a sensor model to bring both images into a single projection view, telemetry noise present at high altitudes can cause geo-positioning errors of up to $100m$. Second, because of the large duration of time that elapses between the capturing of the two images, data distortions like severe lighting and atmospheric variations and object changes in the form of forest growths or new construction cause a high number of disjoint features (features present in one image but not in the other). Third, it should also be noted that remotely sensed terrain imagery, in particular, has the property of being highly self-correlated both as image data and elevation data. This includes first order correlations (locally similar luminance or elevation values in buildings), second order correlations (edge continuations in roads, forest edges, and ridges), as well as higher order correlations (homogeneous textures in forests and homogenous elevations in plateaus). Therefore, a central challenge in achieving precise geo-registration is the reliable handling of the outliers caused by the data distortions and ambiguities that have been described in this paragraph.

The objective of this work is to recover a meaningful adjustment of the sensor parameters based on the spatial registration of Aerial Video Frames with Reference Imagery. Furthermore, pixel-wise assignment of precise three-dimensional locations can be computed for an incoming video frame as it would be aligned with the geodetically calibrated Reference Imagery. The overlaying of a registered frame on the reference environment is illustrated in Figure 3. Figure 4 shows the difference between the Aerial Video Frame and the Reference Image, and successful geo-registration between the two. It can be observed that not all buildings present in the Aerial Video Frame are present in the Reference Image and vice versa. Thus such 'geo-registration' can be effectively used for updating aerial maps, accurate targeting and providing accurate geo-locations for objects of interest. Geo-registration can be used for creating geo-mosaics [30] and annotation of video data as well [25].

The remainder of this chapter is structured as follows. Section 2 reviews and categorizes related work. Section 3 discusses procedures employed in bringing both images into a common viewing space. Section 4 describes the processes involved in the geo-registration of images, which is followed by a discussion of the results and conclusion in Sections 5 and 6, respectively.

Figure 3. Draping an Aerial Video Frame. A video frame is registered in the real world with the reference image. As the geodetic position of each pixel in the reference image is accurately known, the corresponding pixels in the aerial image inherit this information.

2. Related Work

In the past, substantial research has been directed towards determining the geo-location of objects from an aerial view. Several systems such as Terrain Contour Matching (TERCOM) [2], SITAN, Inertial Navigation/Guidance Systems (INS/IGS), Global Positioning Systems (GPS) and most recently Digital Scene-Matching and Area Correlation (DSMAC) have already been deployed in applications requiring geo-location. While each of these systems has had some degree of success, several shortcomings and deficiencies have become increasingly apparent. By understanding the limitations of these systems, we can acquire a better appreciation for the need of effective image-based systems.

As the name suggests, TERCOM (Terrain Contour Matching) fixes the position of airborne vehicles by matching elevation contours detected by radar, with stored digital contour data. TERCOM operates on the premise that the elevation contours of a given terrain area uniquely distinguish it from any other. This premise is evidently violated in plateaus and wherever terrain relief is below sensor detection capability, as well as areas containing ridge-like relief (the aperture effect). Moreover, systems like TERCOM are not a 'passive' means of geo-location as they require the emission of interceptable electromagnetic emission. One passive alternative is INS, which is a gyroscopic-based technology that has the ability to accurately mea-

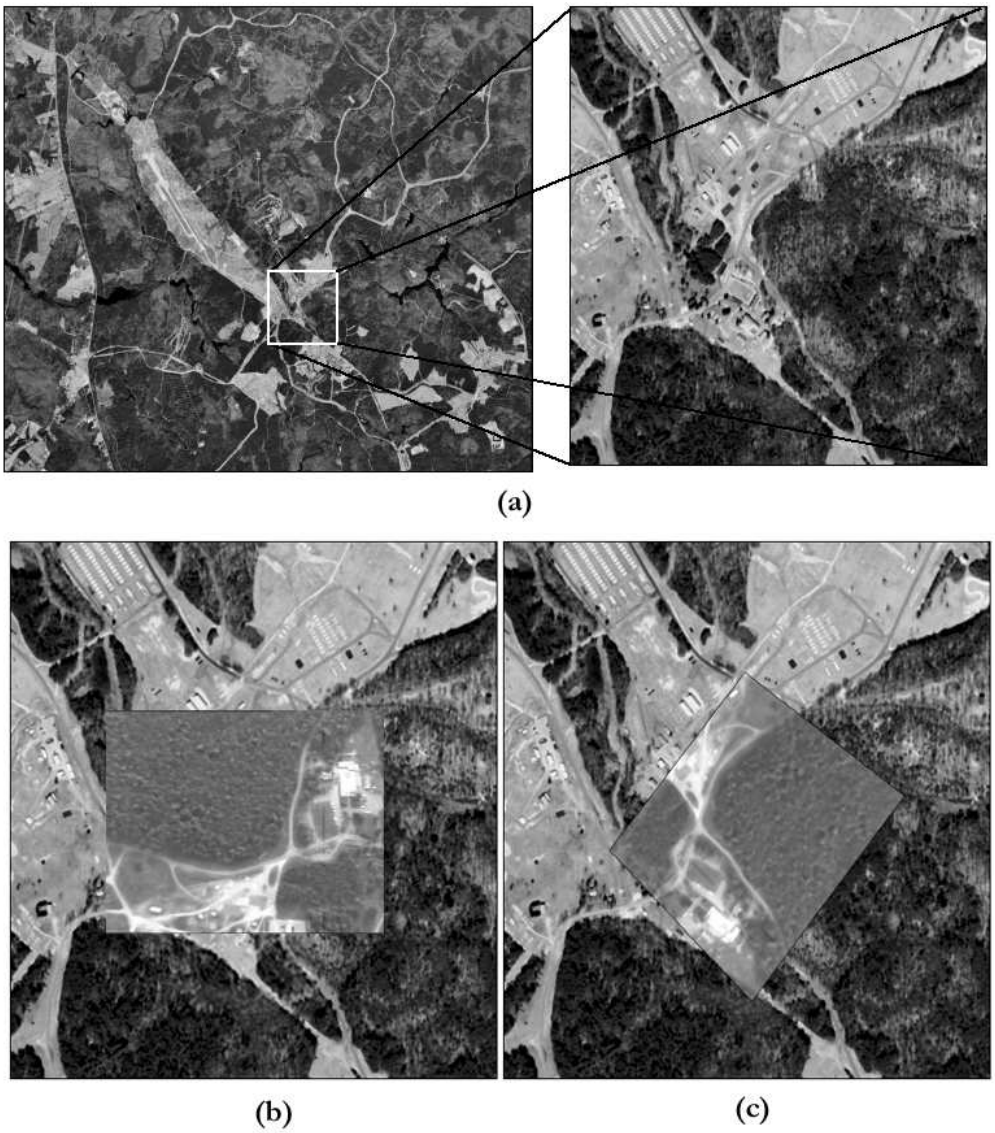

Figure 4. Registration of the Aerial Video Frame with the Cropped Reference Image. (a) Based on the telemetry data, that specifies the corresponding area of the Reference Imagery the camera is capturing, the Reference Image is cropped. (b) The aerial video frame before and (c) after geo-registration with the Cropped Reference Image. It should be noted that the Reference Image is an Orthographic Image while the Aerial Video Frame is a Perspective Image.

sure telemetry information of an airborne vehicle. However, despite the fact that this system produces only slight errors, these errors are cumulative and furthermore the most accurate systems are usually too expensive for widespread use. Global Positioning Systems are a reliable and cheaper alternative that use the concept of trilateration to estimate the position of a GPS receiver, however GPS systems are susceptible to interference and may be blocked altogether.

Although the aforementioned technologies were able to revolutionize many military and domestic functions, the problems highlighted here motivated the advent of image-matching techniques to exactly recover the position of an airborne vehicle in the real world. Image-based geo-location has two properties in particular that motivate its use: First, it is a passive positioning approach, i.e. it does not rely on electromagnetic emission that may be distorted or blocked. Second, it allows geo-positioning per frame, so the small errors that may be incurred are not cumulative. One example of a deployed image-based technology is DSMAC. It converts the scenes picked up by the missile's camera to simplistic binary images, and cannot rotate or scale images. Due to this crudeness of the matching model this attempt has had limited field success. It was quickly realized that even small mechanical vibrations that are commonplace on such vehicles could cause significant error in high altitude cameras, and therefore solving the resulting problem of image alignment was a difficult one. As a result, many assorted approaches of aligning images to recover geo-location information, or 'geo-registration' as it is often called, have subsequently been proposed, some strictly using computer vision concepts, and others implicitly incorporating earlier contour matching and inertial navigation concepts into their geo-registration algorithms.

Two types of approaches can be distinguished: Elevation-Based Correspondence and Image-Based Correspondence. Elevation-Based approaches have the general drawback that they rely on the accuracy of recovered elevation from two frames, a task found to be notoriously difficult. Furthermore the contour-based approach in [13] is unlikely to find correct matches in areas of self-correlated elevation like plateaus and ridges when correspondence is difficult to establish. On the other hand, the research literature of image-based correspondence is quite vast; [15] is a general survey of some of these registration techniques. However, conventional techniques are liable to fail because of the inherent differences between the two imageries. 'Direct methods' of alignment typically minimize a parametric error function specified in terms of some image measurable quality such as brightness constancy (corresponding pixels will have equal intensity values as in [14], [17]). These methods are liable to fail since many corresponding pixels are often dissimilar. In such a case, there is little statistical correlation

between the imageries *globally*. Alignment by maximization of Mutual Information [20] is another frequently used registration approach, and while it provides high levels of robustness it also allows many false positives when matching over a search area of the nature encountered in Geo-Registration. Furthermore, formulating an efficient search strategy is difficult. On the other hand, specific to geo-registration, several intensity based approaches to geo-registration intensity have been proposed. We will investigate previous work on geo-registration subsequently, followed by a description of our work.

2.1. ELEVATION BASED GEO-REGISTRATION

Elevation based algorithms attempt to achieve alignment by matching the DEM with an elevation map recovered from video data. Rodrequez and Aggarwal in [13] perform pixel-wise stereo analysis of successive frames to yield a recovered elevation map or REM, as the initial Data Rectification step. Next, to bring the REM and DEM into a common representation both are converted into 'cliff maps', which are the contours of zero crossings of the elevation map after convolution with a Laplacian of Gaussian Filter. Along these cliff contours (expressed in terms of chain code), local extrema in curvature are detected to define critical points. To achieve correspondence, each critical point in the REM is then compared to each critical point in the DEM. A hypothesis/verification scheme is used, where a match is hypothesized if the mean squared error between the REM and DEM critical point neighbourhood is small. From each hypothesis instance, a transformation between REM and DEM contours can be recovered. After transforming the REM cliff map by this transformation, alignment verification is performed by finding the fraction of transformed REM critical points that lie near DEM critical points of similar orientation. While this algorithm is highly efficient and lends itself easily to real-time implementation, it runs into similar problems as TERCOM i.e. it is likely to fail in plateaus, ridges and depends highly on the accurate reconstruction of the REM. Recovering elevation from stereo is a challenging task and no relevant solution was proposed. In [31], Sim and Park propose another geo-registration algorithm that reconstructs a REM from stereo analysis of successive video frames. Normalized Cross Correlation based point-matching is used to recover the elevation values. Both elevation maps are rectified into a relative elevation map with respect to a pre-defined maximal feature point. To establish correspondence, a set of sample feature points are selected along a fixed row with equal intervals and a search area of 5x5 pixels is defined between the relative REM and DEM. For each possible match, an evaluation of cumulative difference between the relative REM at each feature point,

and the associated relative DEM at the search instance is computed. The translation that minimizes this cumulative difference is then chosen to be the correspondence between the REM and DEM. In another approach proposed by the same group ([18]) a relative position estimation algorithm is applied between two successive video frames, and their transformation is recovered using point-matching in stereo. As the error may accumulate while calculating relative position between one frame and the last, an absolute position estimation algorithm is proposed using image based registration in unison with elevation based registration. The image based alignment uses Hausdorff Distance Matching between edges detected in the images. The elevation based approach estimates the absolute position, by calculating the variance of displacements. These algorithms, while having been shown to be highly efficient, restrict degrees of alignment to only two (translation along x and y), and furthermore do not address the conventional issues associated with elevation recovery from stereo.

2.2. INTENSITY BASED GEO-REGISTRATION

Intensity-based approaches to geo-registration use intensity properties of both imageries to achieve alignment. Work has been done developing image-based techniques towards registration of two sets of reference imageries [16], as well as the registration of two successive video images ([14], [17]). However, it was found that for frame-to-reference registration a different set of issues needed to be tackled. As the video data and the reference imagery are usually in different projection views the initial view rectification module is usually required. In [27], Cannata *et al* use the telemetry information to bring a video frame into an orthographic projection view, by associating each pixel with an elevation value from the DEM. As the telemetry information is noisy the association of elevation is erroneous as well. However, for aerial imagery that is taken from aircrafts of nadir orientation the rate of change in elevation may be assumed low enough for the elevation error to be small. By ortho-rectifying the aerial video frame, the process of alignment is simplified to a strict 2D registration problem. Correspondence is achieved by taking 32×32 pixel patches uniformly over the aerial image and correlating them with a larger search patch in the Reference Image, using Normalized Cross Correlation. As the correlation surface is expected to have a significant number of outliers, four of the strongest peaks in each correlation surface are selected and consistency measured to find the best subset of peaks that can be expressed by a four parameter affine transform. Finally, the sensor parameters are updated using a conjugate gradient method, or by a Kalman Filter to stress temporal continuity.

An alternate approach is presented by Kumar *et al* in [22] and by Wildes

et al in [29] following up on that work , where instead of ortho-rectifying the Aerial Video Frame, a perspective projection of the associated area of the Reference Image is performed. This approach avoids the errors involved in associating elevations with each aerial video pixel on the basis of the telemetry information and therefore does not make any assumptions about the rate of change of the elevation information. In [22], two further data rectification steps are performed. Video frame-to-frame alignment is used to create a mosaic providing greater context for alignment than a single image. For data rectification, a Laplacian filter at multiple scales is then applied to both the video mosaic and reference image. To achieve correspondence, two stages of alignment are used: coarse followed by fine alignment. For coarse alignment salient (feature) points are defined as the locations where the response in both scale and space is maximum. Normalized correlation is used as a match measure between salient points and the associated reference patch. One feature point is picked as a reference, and the correlation surfaces for each feature point are then translated to be centered at the reference feature point. In effect, all the correlation surfaces are superimposed, and for each location on the resulting superimposed surface, the top k values (where k is a constant dependant on number of feature points) are multiplied together to establish a consensus surface. The highest resulting point on the correlation surface is then taken to be the true displacement. To achieve fine alignment, a 'direct' method of alignment is employed, minimizing the SSD of user selected areas in the video and reference (filtered) image. The plane-parallax model is employed, expressing the transformation between images in terms of 11 parameters, and optimization is achieved iteratively using the Levenberg-Marquardt technique.

In the subsequent work, [29], the filter is modified to use the Laplacian of Gaussian filter as well as it's Hilbert Transform, in four directions to yield four oriented energy images for each aerial video frame, and for each perspectively projected reference image. Instead of considering video mosaics for alignment, the authors use a mosaic of 3 'key-frames' from the data stream, each with at least 50 percent overlap. For correspondence, once again a local-global alignment process is used. For local alignment, individual frames are aligned using a three-stage Gaussian pyramid. Tiles centered around feature points from the aerial video frame are correlated with associated patches from the projected reference image. From the correlation surface the dominant peak is expressed by its covariance structure. As outliers are common, RANSAC is applied for each frame on the covariance structures to detect matches consistent to the alignment model. Global alignment is then performed using both the frame to frame correspondence as well as the frame-to-reference correspondence, in three stages of progressive alignment models. A purely translational model is used at the coarsest

level, an affine model is then used at the intermediate level, and finally a 2D projective model is used for alignment. To estimate these parameters an error function relating the Euclidean distances of the frame-to-frame and frame-to-reference correspondences is minimized using the Levenberg Marquardt Optimization.

The major limitation of the intensity based approaches are the assumptions that are made. In [27] such an assumption is made implicity through the choice of an orthographic system model, since the error of ortho-rectification increases with the magnitude of terrain relief. While such an error is avoided by use of perspective projection in [29], strong assumptions of scene planarity are made during correspondence, first with a translational local matching, followed by the progressive pyramid proposed. Though these assumption may hold in many cases, and they may simplify computation significantly, they are liable to introduce error when scene relief increases. Furthermore, since generic transformation models are being used, transformations that are not physically realizable (like single dimensional shears or scalings) are included within the set of allowable transformations that is searched.

2.3. OUR WORK

In this chapter, we outline a method to recover geodetic alignment for a video sequence, while plausibly adjusting the sensor telemetry parameters. An intensity-based approach was favored over an elevation-based one because recovering elevation from a video sequence has proven to be unreliable, particularly when the scene is as highly self-correlated as aerial video often is. A salient aspect of earlier work in intensity-based approaches was the generation of local correlation surfaces by translating a template. Instead of imposing such a strict translational constraint on motion so early on in the alignment estimation process, we propose an algorithm that computes local similarity measures, and utilizes them to directly estimate global similarity. Since we do not generate similarity surfaces, our method can recover larger rotation, shear and scaling and does not degenerate when higher order parametric models of motion are used or when scene relief is high. We correct the correlation coefficient to allow coefficient addition by the use of Fisher's Z-transform and detail a modification of the error function to inherently allow optimal registration in the presence of outliers caused by disjoint features or the dissimilarity in sensors. Finally, since the estimation procedure is performed by adjustment of the telemetry parameters, an update of telemetry information is output, along with pixel-wise calibration of the aerial video image. The general workflow is diagrammatically expressed in Figure 5.

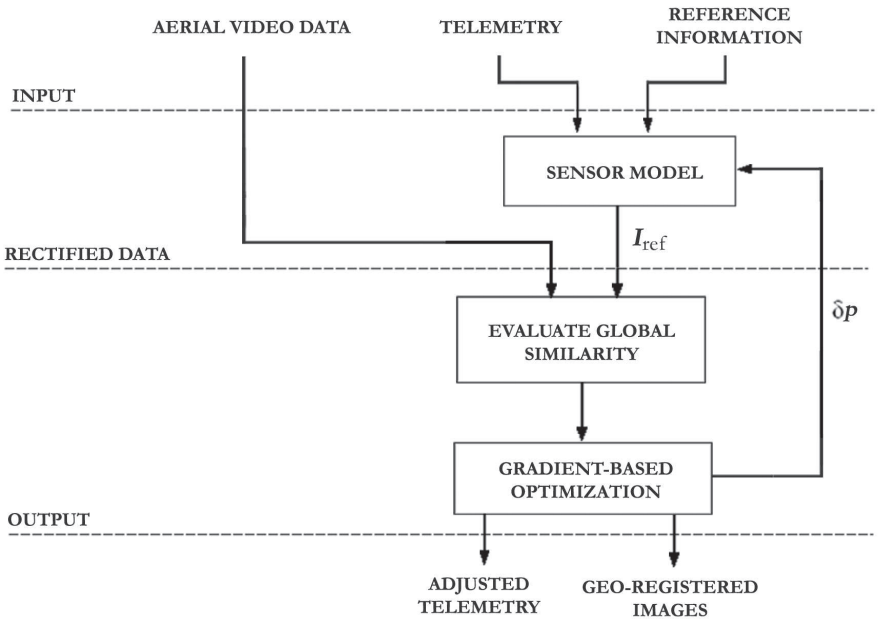

Figure 5. The Geo-registration Work-Flow. This diagram gives a general overview of the inputs, processing and output of the proposed algorithm. δp is the iterative sensor parameter update predicted by the optimization algorithm.

3. Rectifying the Projection View

Since the aerial video data is a sequence of perspective images and the Reference Imagery is high-resolution orthographic image, the transformation between any arbitrary video frame and the reference image can be quite large. As a result, robustly recovering these transformations is a difficult and unsolved task, even when the images have high visual similarity. Fortunately, each aerial video frame is accompanied by telemetry (meta) data detailing the position and orientation of the sensor (camera). By using the telemetry and elevation data to generate a sensor model, the two imageries can be projected into a common projection view. Ideally, if the telemetry data were noiseless, there would be no need for further correspondence, but due to mechanical vibrations and turbulence, image rectification using the telemetry provides only coarse alignment. While the estimate provided by the telemetry information is sufficiently close to make the problem tractable, the visual differences between the images are still acute enough to make the precise adjustment a challenging task. In this section we compare two approaches to bringing the images into a common projection view: ortho-rectification and perspective projection, followed by details involved

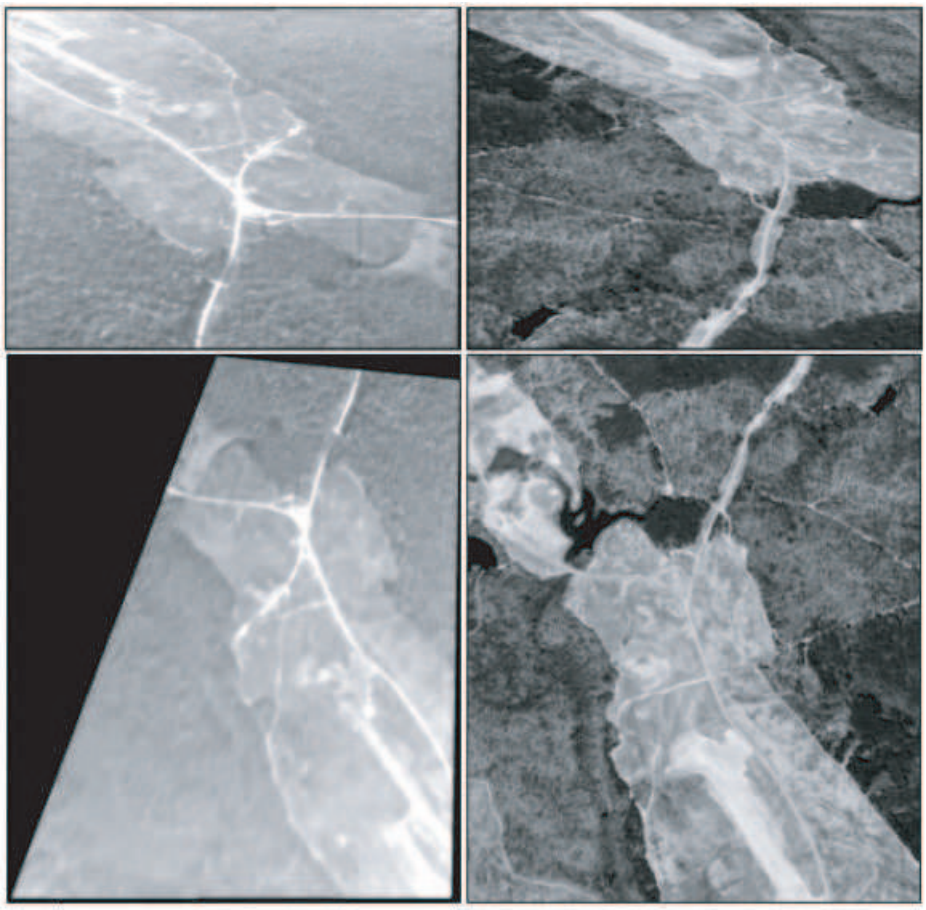

Figure 6. Projection Views. Top Left: Original Aerial Video Image in Perspective View. Bottom Left: An orthographic view of the Aerial Video Image. Bottom Right: Cropped Reference Image in Orthographic View. Top Right: Perspective View of Reference Image.

in using the telemetry information to bring both image into a common view projection (shown in Figure 6). Such projection constitutes the first module of our geo-registration algorithm.

3.1. ORTHO-RECTIFICATION VS PERSPECTIVE PROJECTION

Image projections (transformations of the 3D world onto a 2D projection plane) are either perspective or parallel. The distinction between these two projections is the position of the center of projection relative to the image plane. In perspective projection, the center of projection lies close to the image plane and therefore the lines of projection (all of which converge at

the center of projection) may meet the image plane at different angles (i.e. they are not parallel). The visual manifestation of this phenomenon is perspective fore-shortening, where objects at larger distances appear smaller than similar objects up close. In parallel projection views, like orthographic projection, a line at infinity is defined, and the center of projection is said to lie on that plane. The projection lines are therefore parallel, since they all 'intersect' at the plane at infinity. Figure 7 (a) and 7 (b) illustrate the differences between both projections. Since the aerial video data is received from a camera mounted on an aircraft, it is a perspective image. On the other hand, the Reference Image is a high-resolution photograph taken from high altitude cameras, available as an orthographic image (in parallel view). To analyze both images in a common projection view, two options emerge: (1) An ortho-rectification of the Aerial Video Frame as in [27], or (2) a perspective projection of the Reference Image as performed in [13], [22], [29]. These two alternatives have already been shown in Figure 6 and the spatial relationship between them is illustrated in Figure 7 (c). Since the sensor and elevation model of the scene are available it is possible to perform such rectification of projection. Ideally, if the telemetry and elevation information are accurate, both these projections should be equivalent, but noise in the telemetry brings about certain differences in each approach.

The general importance of Orthographic Projection is that this projection preserves both distances and angles, and there is no distortion of shape or distance in any two-dimensional transformation. For purposes of rectification, the utility of orthographic projection lies in its independence from depth values of pixels. If both the Aerial Video Frame and the Reference Image are projected as orthographic images, transformations can be restricted to two-dimensions, and these transformations are easier to estimate robustly. It is also important to note that by definition any two-dimensional transformation of an orthographic or accurately ortho-rectified image should not reveal hidden surfaces, nor occlude currently exposed surfaces. The main drawback of working exclusively in the orthographic view, however, is that the process of ortho-rectifying the Aerial Video Frame requires the elevation values corresponding to each image pixel. Since the telemetry is noisy, a projection error results when each pixel is traced to the Digital Elevation Map (to recover its elevation), and thus the projected ortho-rectified image will not strictly be an accurate orthographic representation of the Aerial Video Frame. This error can often be assumed to be negligible if the camera is nadir or the environment is one of low elevation rate of change. However, for environments of moderately high rates of change, and more so for low flying aircrafts, such an assumption is often violated.

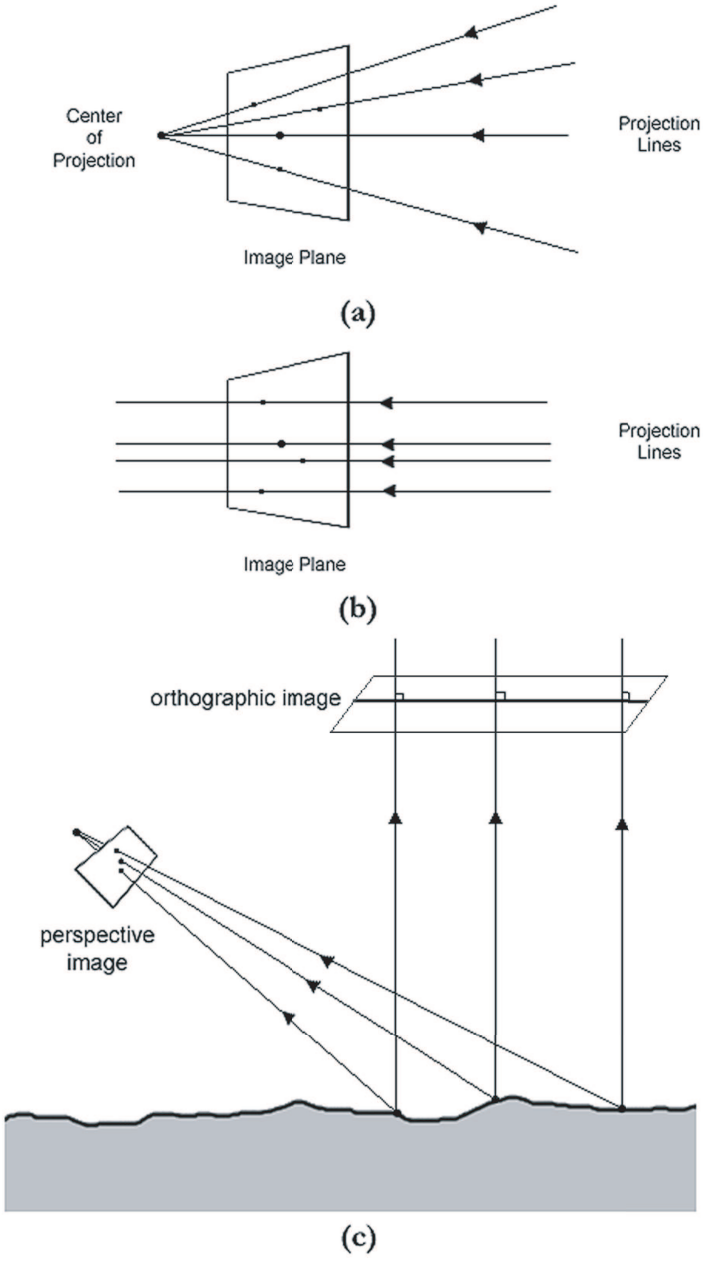

Figure 7. Image Projections. (a) In perspective projection the projection lines converge at the center of projection close to the image plane. (b) In Parallel projection the projection lines remain parallel and converge at infinity. (c) The relationship between the perspective projection lines and the parallel or orthographic projection lines. To re-create an orthographic image from a perspective view, the corresponding elevations of the perspective image pixels in the world are required.

For the general case, inclusive of oblique cameras and environments of high rate of change in elevation, this inaccuracy can be avoided altogether if the Cropped Reference Image is instead perspectively projected. Since both the Digital Elevation Map and the Reference Image are geodetically co-registered, the Reference Imagery can be used to texturize the DEM (Figure 2), effectively assigning each reference pixel an accurate elevation value (as far as the accuracy of the DEM allows). Thus while viewing the Reference Image from the perspective projection view, the elevation value for each pixel is known and the view as it should appear from the camera according to the telemetry information can be generated accurately. The drawback involved with working in the perspective projection view mainly pertain to an increase in complexity, since accurately aligning large displacements requires the estimation of three-dimensional transformations, and accuracy can be lost by making assumptions of scene planarity. However, this nominal increase in complexity is outweighed by the errors avoided in many potential situations. Therefore, in the interest of maintaining a *general* framework for geo-registration, we employ a perspective projection of the Reference Image. It should be noted that if the perspective projection model is precisely followed any transformation may expose hidden surfaces, or alternately occlude exposed ones.

3.2. PERSPECTIVE PROJECTION OF THE REFERENCE IMAGE

The first step in perspective projection is setting up the reference environment. The elevation data is triangulated to form a mesh-surface, and subsequently texturized with the Reference Imagery. In this way, each reference pixel is exactly calibrated with a latitude, longitude and elevation. This accurate co-registration of the reference image and the elevation map is the basis for perspective projection (which will be elaborated presently). Using information from the telemetry, the point of intersection between the camera projection axis and the reference surface is defined as the Reference Origin. A world coordinate system is defined around this Reference Origin as

$$\vec{X}_{world} = [X_{world}, Y_{world}, Z_{world}]. \tag{2}$$

Next, a sensor model is defined. The sensor (camera) is mounted on an aircraft, and Figure 9(a) shows the camera's 3-D coordinate system. This camera coordinate system is defined, relative to the camera's Center of Projection, as

$$\vec{X}_{camera} = [X_{camera}, Y_{camera}, Z_{camera}]. \tag{3}$$

Telemetry information is then used to recover the position and orientation of the aircraft, with respect to the world coordinate system. This

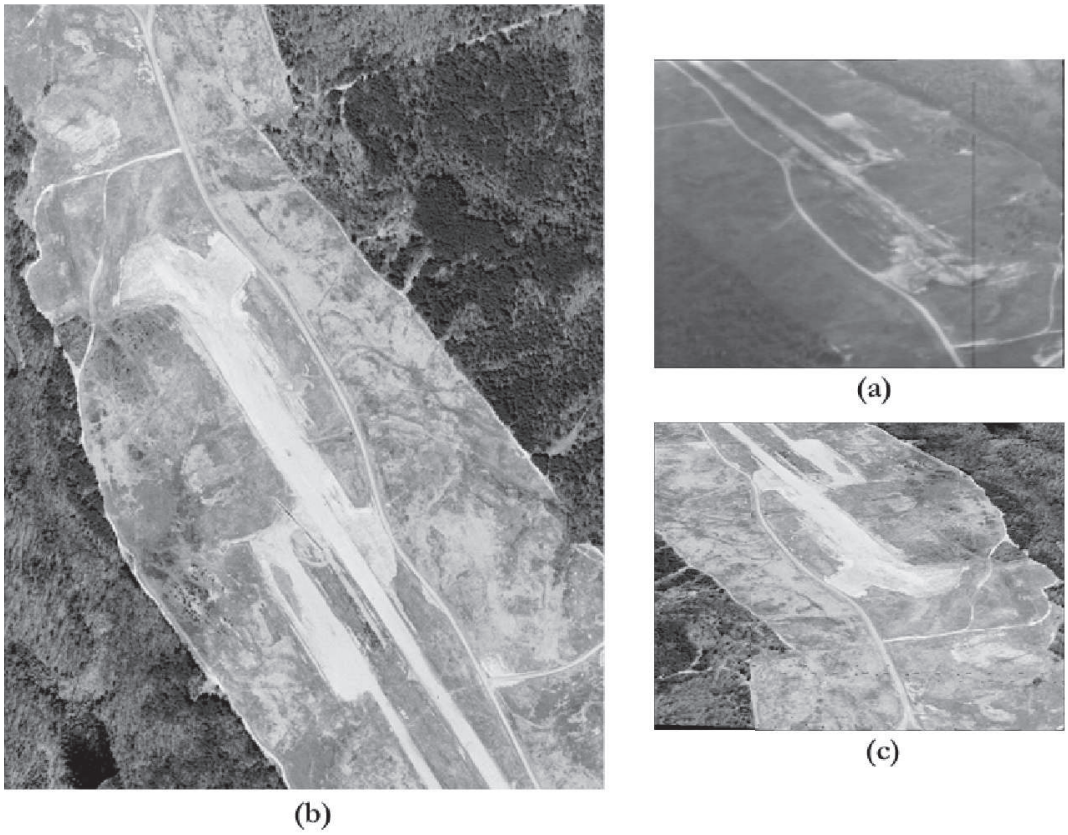

Figure 8. Perspective Projection of the Reference Image. (a) The Aerial Video Frame displays what the camera *actually* captured during the mission . (b) Orthographic Footprint of the Aerial Video Frame on the Reference Imagery (c) The Perspective projection of Reference Imagery displays what the camera *should* have captured according to the telemetry.

relationship is expressed as

$$\vec{X}_{camera} = \Pi_t \vec{X}_{world}, \tag{4}$$

where the coordinate transformation matrix Π_t is

$$\Pi_t = \begin{bmatrix} \cos\omega & 0 & -\sin\omega & 0 \\ 0 & 1 & 0 & 0 \\ \sin\omega & 0 & \cos\omega & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos\tau & -\sin\tau & 0 & 0 \\ \sin\tau & \cos\tau & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\cdot \begin{bmatrix} \cos\phi & 0 & -\sin\phi & 0 \\ 0 & 1 & 0 & 0 \\ \sin\phi & 0 & \cos\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\beta & \sin\beta & 0 \\ 0 & -\sin\beta & \cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
$$\cdot \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 & 0 \\ \sin\alpha & \cos\alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & \Delta T_x \\ 0 & 1 & 0 & \Delta T_y \\ 0 & 0 & 1 & \Delta T_z \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{5}$$

or more concisely,

$$\vec{X}_{camera} = G_y G_z R_y R_x R_z T \vec{X}_{world}, \tag{6}$$

where G_y is a rotation matrix in terms of the camera elevation angle ω, G_z is a rotation matrix in terms of the camera scan angle τ, R_y is a rotation matrix in terms of the vehicle pitch angle ϕ, R_x is a rotation matrix in terms of the vehicle roll angle β, R_z is a rotation matrix in terms of the vehicle heading angle α, T is the translation matrix derived from the vehicle latitude, longitude and height. The details of converting vehicle longitude and latitude to meter distances from the given reference point can be found using many cartographic texts and for the scope of this paper, it is assumed that the vehicle displacements ΔT_x, ΔT_y and ΔT_z are either available or have been computed. Figure 9(b) shows the relationship between the camera and world coordinate systems. Once the camera image plane has been placed, it is possible to establish correspondence between Aerial Image Pixels and elevation data (DEM) by use of a simple ray tracer. It is reiterated here that since the telemetry data is noisy the correspondence yielded by the ray tracer is erroneous as well.

Therefore, instead of ortho-rectifying the Aerial Image using erroneous elevation correspondence, we perspectively project the Reference Image using the exact elevation correspondence (since it is co-registered with the DEM). To achieve this perspective projection of reference coordinates, we define the homogeneous reference coordinates as

$$\vec{X}^{ref}_{world} = [X_{world}, Y_{world}, Z_{elev}, 1], \tag{7}$$

where z_{elev} is taken from the co-registered DEM. Furthermore, the homogeneous perspective coordinates are defined as

$$\vec{X}^{ref}_{perspective} = [X_{perspective}, Y_{perspective}, Z_{perspective}, 1]. \tag{8}$$

Finally to project the reference image the following camera matrix is used

$$\vec{X}^{ref}_{perspective} = \Pi_c \vec{X}^{ref}_{world}, \tag{9}$$

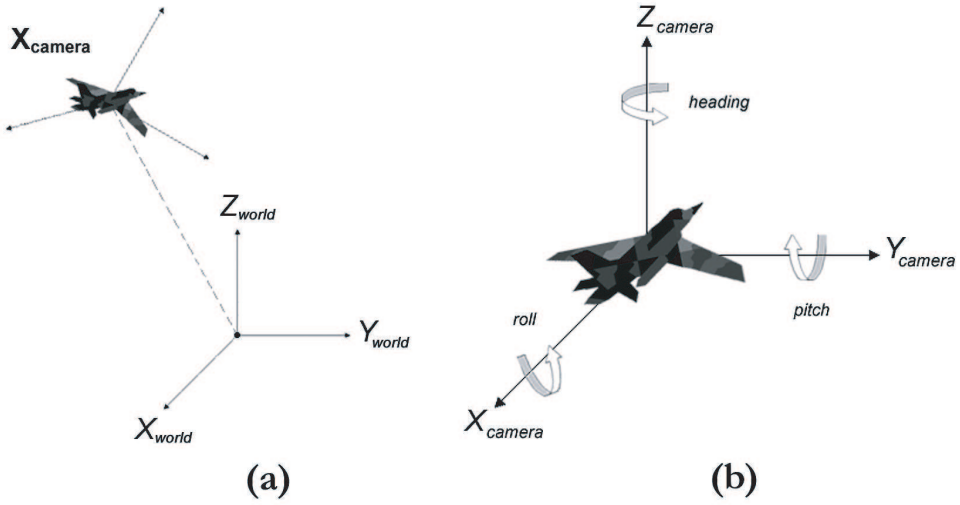

Figure 9. Coordinate Systems. (a) The aircraft coordinate system is shown here. This is the coordinate defined relative to an origin point in the world. For clarity, the additional coordinate system of the camera relative to the aircraft has been omitted. (b) The aircraft coordinate system shown relative to the origin in the real world. Based on the parameters defined by the telemetry, the aircraft coordinate system is placed in within the real world.

where Π_c is the Camera Matrix. It is calculated as

$$\Pi_c = P\Pi_t, \tag{10}$$

where P is the perspective matrix (defined by the focal length) and Π_t is as defined in (6). The perspectively projected image can thus be generated by matching each projected pixel to its corresponding reference pixel using (9). The spatial relation between the projected image and the reference data is shown in Figure 10. At this point it is instructive to note that had the telemetry information been precise, computing the geodetic coordinates of each aerial image pixel would have been a trivial exercise of matching Reference Values with Pixel Coordinates. However, since the telemetry information is noisy, the elevation and positional values assigned to each pixel may be misaligned by up to a hundred pixel elements. It is this displacement that is compensated for using spatial registration.

4. Alignment

In the introduction various sources of visual differences between the Aerial Video Data and the Reference Imagery were pointed out. As a result of these differences, brightness constancy constraints are regularly violated

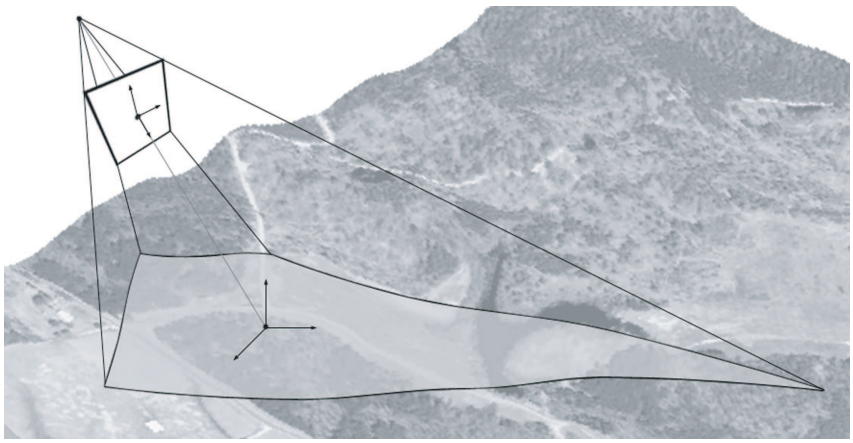

Figure 10. The relationship between the camera plane and the projected area on the reference imagery. Placing a bounding box on the extremities of the footprint creates the cropped reference image.

and therefore conventional direct alignment techniques that globally minimize the brightness constancy constraint usually meet failure. While there may not be strong global correlation between two corresponding images, within small patches that contain corresponding image features, statistical correlation has been shown to be significantly higher ([23]). In this section, we present a direct alignment technique that globally maximizes an average measure of local similarity by searching over the parameters of the telemetry. We begin with a discussion of sensor model update strategy, followed by the construction of the error function using Normalized Cross Correlation and the outlier rejection mechanism. Finally, we discuss the error minimization strategy.

4.1. SENSOR MODEL UPDATE

In section 3.2 we described the sensor model in terms of eight transformation parameters and the camera focal length. For each Aerial Video Frame, I_{video} these parameters are approximately detailed within the associated telemetry data. To compensate for the misalignment between I_{video} and I_{ref} (projected using the telemetry), we adjust these parameters directly to maximize alignment. Since the telemetry detail includes three vehicle translational parameters, three vehicle rotation parameters, and two camera rotation parameters, each with their individual co-variance values, we use these co-variance values to perform a weighted optimization. The transformed coordinate location $\vec{x}_{n+1}$ is defined as

$$\vec{x}_{n+1} = \Pi(\vec{x_n}; \vec{a}, \vec{\varphi}), \tag{11}$$

where Π is the transformation, defined with respect to the pixel location $\vec{x_t}$ at iteration t, the sensor parameters $\vec{a}$, and the co-variance values $\vec{\varphi}$.

Optimizing over the parameters of $\vec{a}$ has two advantages over the use of generic parametric transformation (e.g. affine, projective). First, along with alignment, the telemetry information is simultaneously refined as well. Thus the system can be used not only as a means to calibrate what is being viewed but also to passively determine where the scene is being viewed from. Second, searching over the telemetry parameters inherently excludes physically unrealizable transformations. The set of allowable transformations within a generic transformation matrix includes transformations like single dimensional shears and scalings that cannot be realistically achieved by the sensor setup. Because of these advantages, the sensor parameters are updated to optimize a measure of alignment, defined as the optimization function, between the two images.

4.2. OPTIMIZATION FUNCTION

A pixel's intensity, while actually a measurement of the brightness at a certain receptor on the CCD-array, is often treated as a pixel's *identity* and is used to measure similarity. The implicit assumption of brightness constancy approaches is that pixels can be identified on the basis of their brightness. In practice, of course, this is not always the case as illustrated lucidly by the so-called 'aperture-problem' described by Stumpf (e.g. linear ambiguity due to linear features). Although not a complete solution to the aperture problem (which is an inherent ambiguity), the identity of a pixel is often more uniquely expressed by the set of pixels centered at a given pixel coordinate rather than just the one pixel itself. It directly follows that similarity can be measured more reliably if the pixel identities are more unique. One simple demonstration can be made with counting rules. If local similarity is being measured between two images, a single pixel at a coordinate can have one of a maximum of 256 'identities'. By expanding the coordinate representation to a 3×3 patch instead, not only do the number of possible 'identities' increase drastically, important structural information is captured as well. Hence, in order to have a stronger local measure, we compute similarity between two patches at each location rather than simply comparing two pixels. This point can also be made in terms of solving a system of equations. Since, two dimensional flow estimation equations are underconstrained for a single pixel, the Lucas and Kanade optical flow estimation technique [4] assumes that neighboring pixels in a small window have the same flow vectors. The system is then solved as an overconstrained system. While Lucas and Kanade estimate local motion by looking over a pixel neighborhood, we estimate local similarity over a pixel neighborhood

and compute global motion by maximizing the sum of local similarity.

The objective function we use is a measure of global alignment between the current Aerial Video Frame and Reference Image. For any state of the sensor parameters this measure of global fit is defined *locally*, since stronger correlation is likely to exist locally. However, unlike previous approaches, we do not locally convolve correlation templates to recover correlation surfaces. The similarity measure of choice is Normalized Cross Correlation, since it is invariant to local contrast changes and closely approximates the statistical correlation between two patches. Between $I_{video}(\vec{x})$ and $I_{ref}(\Pi(\vec{x_n}; \vec{a}, \vec{\varphi}))$ the measure of similarity is defined as

$$F(\vec{x}) = \sum_i r(\vec{x}_i; \vec{a}, \vec{\varphi})), \tag{12}$$

where r is a correlation coefficient between two patches centered at each pixel location. In order to ensure that F is a quantity to be minimized, we define r as $1 - \|\rho\|$ (the Normalized Cross Correlation Coefficient).

However, the Normalized Cross Correlation coefficient is not a linear function of the relational magnitude between the images [1], and as a result, correlation coefficients cannot simply be averaged. As a statistic, the r has a sampling distribution (if n sample pairs from two signals were correlated over and over again the resulting r's would form a sampling distribution). This distribution has a negative skew (negative bias). A transformation called Fisher's Z-transformation converts r to a value that is normally distributed and is defined as

$$z_i = \frac{1}{2}[\ln(1 + \|r\|) - \ln(1 - \|r\|)]. \tag{13}$$

As a result of having a normally distributed sampling distribution, there is an equal probability of detecting different correlation magnitudes and hence they can be meaningfully added.

4.2.1. *Incorporating an Outlier-Rejection Mechanism*

Disjoint image features, local motion and photometric ambiguity may all contribute to causing outliers. We propose a methodology to minimize the effect of outliers on the global average of local similarity, by making an observation about pixel identities. Changes in pixels of high dissimilarity are given less importance than pixels with higher similarity. We observe that the larger the dissimilarity between two pixels, the more likely they are to represent different artifacts (disjoint features, local motion etc). In order to ensure that similarity variations in areas of low similarity have less of an effect on the global similarity measure than variations in areas of high similarity we use a sigmoid response function. Since we use gradient

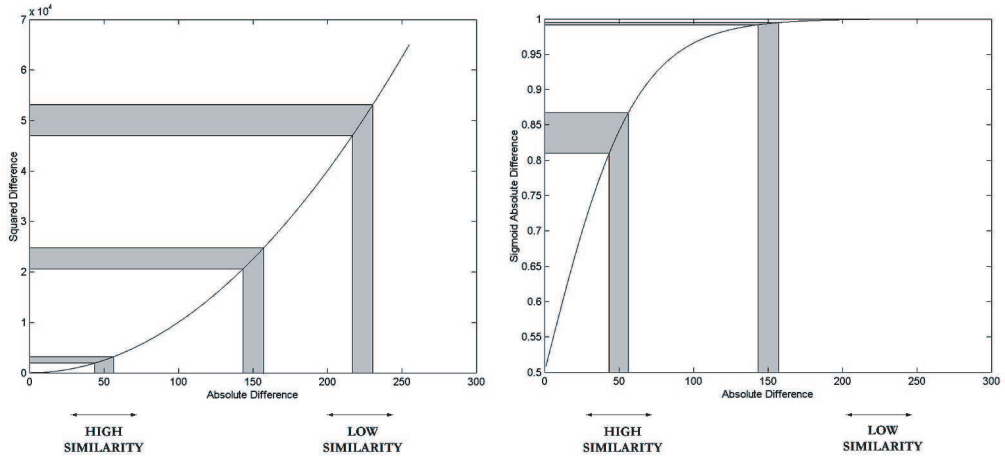

Figure 11. Rectifying the Error Function. The error response of Sum of Squared Difference (left) encourages variations in pixels of low similarity. After rectifying with a sigmoid response (right), pixels with high similarity have a larger effect on the error function.

information during optimization, the gradient behavior of our similarity measure is of prime importance. We therefore modify the similarity measure to ensure that minor changes in areas of large difference are ignored. The sigmoid correlation function is

$$\eta(\vec{x_i}, \vec{a}) = \frac{1}{1 + (\frac{1-\|r_i\|}{1+\|r_i\|})^{\frac{b}{2}}}, \tag{14}$$

where b is a constant that represents sensitivity to noise. Substituting the value of z from equation (12) gives the final similarity measurement. At every pixel (x_i, y_j), a similarity score, $\eta(\vec{x_i}; \vec{a})$ is calculated between two patches of size $w_x \times w_y$ centered at (x_i, y_j). Since the similarity score is additive, the global similarity measure F is redefined summing for all (i,j) as

$$F(a) = \sum_i \eta(\vec{x_i}; \vec{a}, \vec{\varphi})). \tag{15}$$

4.2.2. *Optimization Strategy*

Optimizations are defined in terms of an objective (error) function, system parameters that are to be adjusted to optimize the objective function and an exit condition to terminate the search. In order to find the optimal alignment between the Projected Reference Image and the Aerial Video Frame

we directly adjust the parameters of the sensor model defined in (9). By adjusting these parameters we try to maximize a measure of 'goodness' of alignment between the two images under inspection, the objective function, defined in (15). The search for the optimal state of the sensor parameters is performed using Finite-Difference Quasi-Newton Optimization. This algorithm iteratively builds up curvature information to formulate a quadratic model problem. Gradient information is required, which is provided using finite differences. This method involves perturbing each of the sensor parameters, and thus calculating the rate of change of the objective function. The algorithm was implemented in a hierarchical fashion over a pyramid, since this provides an escape from local extrema and also performs analysis at multiple frequencies. Typically, three major iterations are performed at each level of a five level pyramid. Several options for exit conditions may be used, like number of iterations, error thresholds, but the most often employed exit condition is fired the change in error falls below a threshold.

The steps of the algorithm may be summarized as follows:

1. For each coordinate position (i, j) calculate the local similarity $\eta(\vec{x}_i; \vec{a})$ between the two 5×5 block around $I_{ref}(\Pi(\vec{x_t}; \vec{a}, \vec{\varphi}))$ and $I_{video}(\vec{x_i})$ using normalized cross correlation. Sum $\eta(\vec{x}_i; \vec{a})$ for all (i, j) to evaluate the global measure of similarity.
2. Calculate $\delta\vec{a}$, the update for $\vec{a}$, using Quasi-Newton Maximization of objective function.
3. Update $\vec{a'} = \delta\vec{a} \cdot \vec{a}$.
4. Return to step one and repeat until exit condition is fulfilled.

5. Results

To demonstrate the algorithm described in this chapter, experimental results are presented in this section. Despite the substantial illumination change to the extent of contrast reversal, examination of the results shows a precise pixel-wise alignment. Figure 13, 14, 15, and 16 show the initial Video Frame and Reference Imagery before and after registration. Visual inspection reveals significant misalignment after perspective projection of the reference image using the telemetry and sensor model. Attempts at minimizing this misalignment using brightness consistency constraints fails, but with the proposed algorithm proposed in this chapter, accurate alignment is achieved.

On the first clip, a pre-registration average error of 47.68 meters with a standard deviation of 12.47 and a post-registration average error of 4.34 meters and standard deviation of 3.19 per frame was recorded. On the second clip, a pre-registration error of 51.43 meters with a standard deviation of 14.66 and a post-registration average error of 3.46 and a standard deviation

of 2.91 was recorded. As ground truth was not available to assess the error automatically, manual measurement was performed per frame. The results on the two 30 key-frame clip is shown in Figure 12. The frames in the clip contained adequate visual context to allow single frame registration.

The portion of the image set on which the algorithm presented did not perform accurately, were of three types. The first type was images without any features at all, like images of textured areas of trees. Since there was little information providing constraints for alignment, it was difficult to judge a successful alignment. The second problem faced was the linear aperture problem, and thus only a single dimensional constraint could be retrieved from them. The most convincing method of addressing both these issues is using some form of bundle adjustment, as was used in [29]. These methods were not used in this work since only video key-frames with little or no overlap were available. The last problem faced was that of occlusion by vehicle parts like tires and wings. This was addressed by calculating the fixed positions of the vehicle parts with respect to the camera in terms of the camera parameters (camera elevation angle, camera scan angle, and camera focal length). The portion of the image is then ignored or if it happened to cover too much of the image space then the image is ignored.

6. Conclusion

The objective of this chapter was to present an algorithm that robustly aligns an Aerial Video Image to an Area Reference Image while realistically updating the sensor model parameters. As input the algorithm receives Aerial Video Data, noisy telemetry information, the DEM and its associated area reference image. The major problems tackled here were rectifying the images to bring them into a common projection view, geodetic assignment for aerial video pixels, and plausible sensor model parameter adjustment. Various forms of distortions were tackled, adjusting for illumination, compensating for texture variation, handling clouds and occlusion by vehicle parts. The first step in the algorithm was the perspective projection of the Reference Image using telemetry, elevation information, and the sensor model to bring both images into a common projection view. Alignment was then performed directly using normalized cross-correlation without the use of a translating template. Instead local correlation values were summed to calculate an estimate of global similarity, a measure then minimized using Quasi-Newton Minimization by Finite-Differences. Instead of relying on planar transformation models, we perform per iteration rendering to compute updates of the *original* telemetry parameters. To compensate for the significant number of outliers, an intuitive outlier rejection mechanism was used to reject outlying information directly. It is

Figure 12. Misalignment errors. (a) Clip One: The pre-alignment and post-alignment errors for 30 frames. A pre-registration average error of 47.68 meters with a standard deviation of 12.47 and a post-registration average error of 4.34 meters and standard deviation of 3.19 per frame was recorded. (b) Clip Two: The pre-alignment and post-alignment errors for 30 frames. A pre-registration error of 51.43 meters with a standard deviation of 14.66 and a post-registration average error of 3.46 and a standard deviation of 2.91 was recorded.

to be expected that the sensor data will improve with the forward march of technology, bringing with it the possibilities of more sophisticated models for the geo-registration problem. Any improvement in the accuracy of elevation data in particular would allow more confident use of three-dimensional information and matching. Future directions of the work include solving the initial alignment robustly in the perspective viewing space using more realistic rendering, and performing registration without continuous telemetry information.

References

1. J.K. Wani, *"Probability and Statistical Inference"*, Appleton-Century-Crofts, New York, 1971.
2. J. P. Golden, *"Terrain Contour Matching (TERCOM): A cruise missile guidance aid"*, Proc. Image Processing Missile Guidance, vol. 238, pp. 10-18, 1980.
3. B. Horn, B. Schunk, *"Determining Optical Flow"*, Artificial Intelligence, vol. 17, pp. 185-203, 1981.

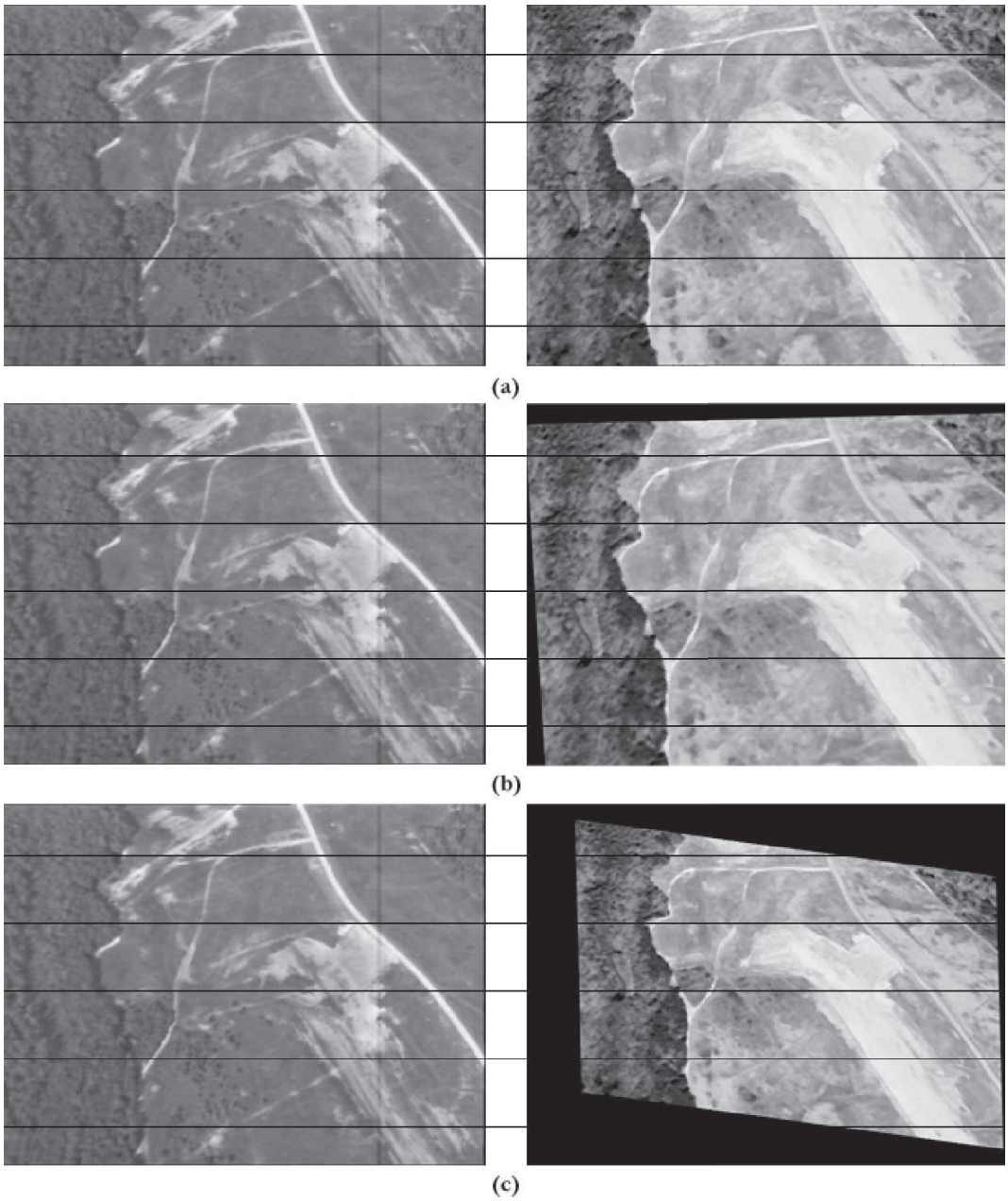

Figure 13. Geo-registration despite visual dissimilarity. Aerial Video Frame and Projected Reference Images shown before (a) and after (b) registration. (c) Affine Frame-to-Frame alignment algorithm fails.

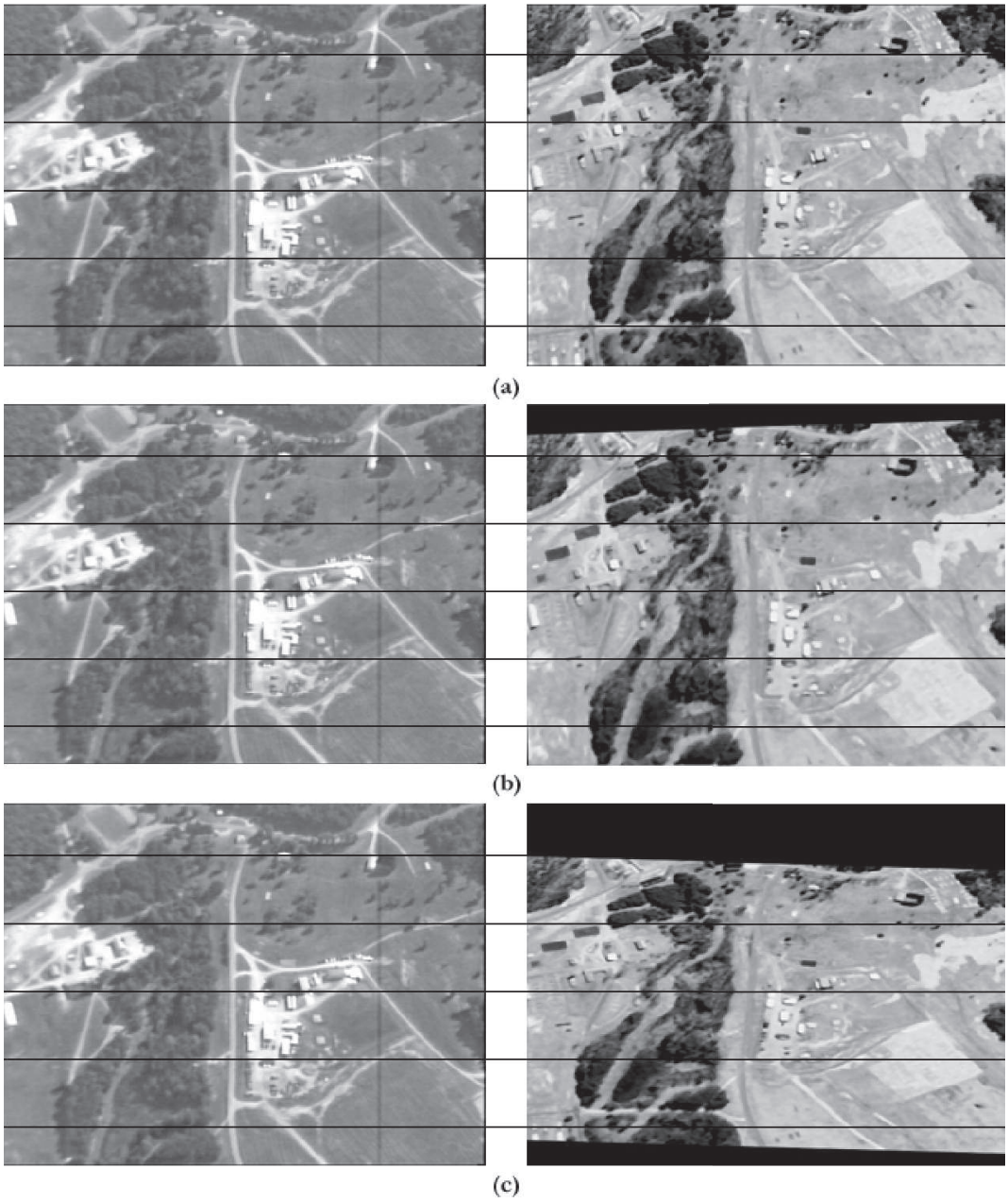

Figure 14. Geo-registration despite disjoint features. Aerial Video Frame and Projected Reference Images shown before (a) and after (b) registration. (c) Affine Frame-to-Frame alignment algorithm fails.

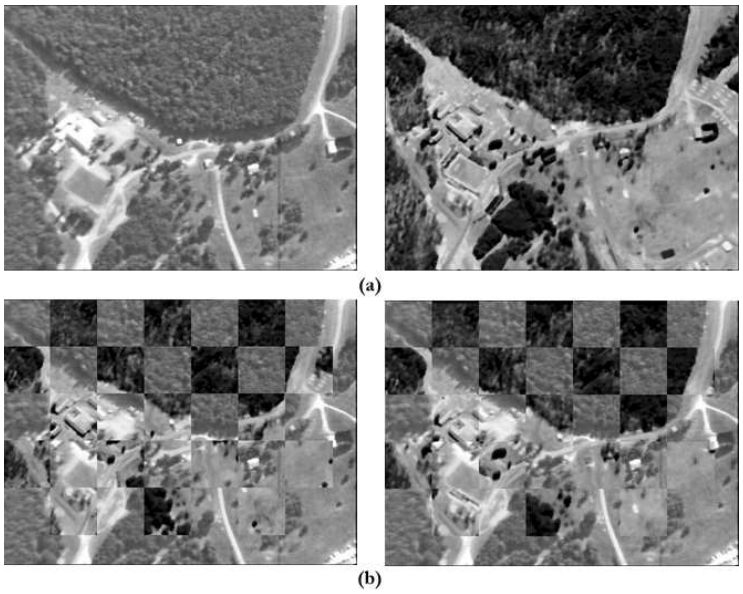

Figure 15. Checker Board Comparison. (a) Original Images (b) Before (right) and after (left) registration.

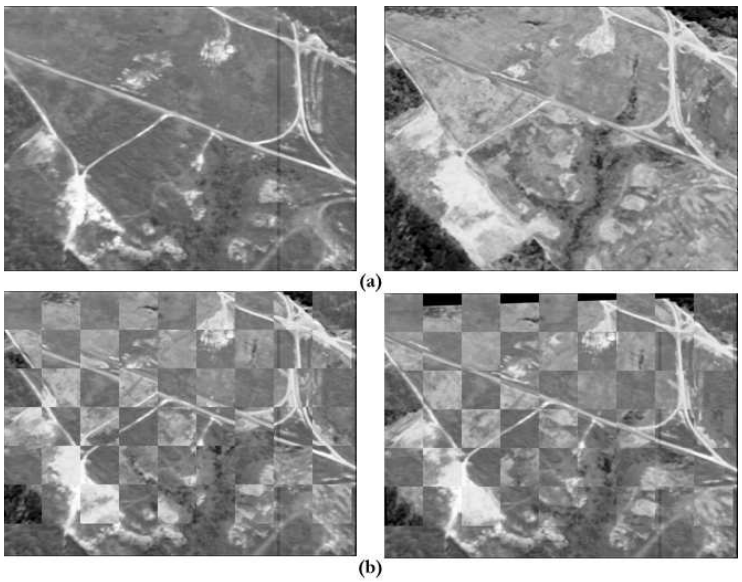

Figure 16. Checker Board Comparison. (a) Original Images (b) Before (right) and after (left) registration.

4. B. Lucas and T. Kanade. *"An Iterative image registration technique with an application to stereo vision"*, Proceedings of the 7th International Joint Conference on Artificial Intelligence, pp. 674-679, 1981.
5. C. Baird and M. Abrarnson, *"A comparison of several digital map-aided navigation techniques"*, Proc. IEEE Position Location and Navigation Symposium, pp. 294-300, 1984.
6. P. Curran, *"Principles of Remote Sensing"*, Longman Group Limited, 1985.
7. S. J. Merhav, Y. Bresler, *"On-line Vehicle Motion Estimation from Visual Terrain Information Part I: Recursive Image Registration"*, IEEE Trans. Aerospace and Electronic System, 22(5), pp. 583-587, 1986.
8. Y. Bresler, S. J. Merhav, *"On-line Vehicle Motion Estimation from Visual Terrain Information Part II: Ground Velocity and Position Estimation"*, IEEE Trans. Aerospace and Electronic System, 22(5), pp. 588-603, 1986.
9. D.H. Field, *"Relations between the statistics of natural images and the response properties of cortical cells"*, JOS A, vol 4, pp. 2379-2394, 1987.
10. B. Kamgar-Parsi, J. Jones, A. Rosenfeld, *"Registration of multiple overlapping range images: scenes without distinctive features"*, Computer Vision and Pattern Recognition, pp. 282-290, 1989.
11. P. Anandan, *"A computational framework and an algorithm for the measurement of visual motion "*, International Journal of Computer Vision, vol.2, pp. 283-310, 1989.
12. J. Foley, A. van Dam, S. Feiner, J. Highes, *"Computer Graphics, Principles and Practices"*, Addison-Wesley, 1990.
13. J. Rodriquez, J. Aggarwal, *"Matching Aerial Images to 3D terrain maps"*, IEEE PAMI, 12(12), pp. 1138-1149, 1990.
14. J.Bergen, P. Anandan, K. Hanna, R. Hingorani, *"Hierarchical model-based motion estimation"*, Proc. European Conference on Computer Vision, pp. 237-252, 1992.
15. L. Brown, *"A Survey of Image Registration Techniques"*, ACM Computing Surveys, 24(4), pp. 325-376, 1992.
16. Q. Zheng., R. Chellappa. *"A computational vision approach to image registration"*, IEEE Transactions on Image Processing, 2(3), pp. 311 -326, 1993.
17. R. Szeliski, *"Image mosaicing for tele-reality applications"*, IEEE Workshop on Applications of Computer Vision, pp. 44-53, 1994.
18. D.-G. Sim, S.-Y. Jeong, R.-H. Park, R.-C. Kim, S. Lee, I. Kim, *"Navigation parameter estimation from sequential aerial images"*. Proc. International Conference on Image Processing, vol.2, pp. 629-632, 1996.
19. S. Manna nd R.W. Picard, *"Video orbits of the projective group a simple approach to featureless estimation of parameters"*, IEEE Transactions on Image Processing, 6(9) , pp. 1281 -1295, 1997.
20. P. Viola and W. M. Wells, *"Alignment by maximization of mutual information."*, International Journal of Computer Vision, 24(2) pp. 134-154, 1997.
21. R. Szeliski, H. Shum, *"Creating Full View Panoramic Image Mosaics and Environment Maps"*, Computer Graphics Proceedings, SIGGRAPH, pp. 252-258, 1997.
22. R. Kumar, H. Sawhney, J. Asmuth, A. Pope, S. Hsu, *"Registration of video to georeferenced imagery"*, Fourteenth International Conference on Pattern Recognition, vol. 2. pp.1393-1400, 1998.
23. M. Irani, P. Anandan, *"Robust Multi-Sensor Image Alignment"*, International Conference on Computer Vision, 1998.
24. J. Nocedal, S. Wright, *"Numerical Optimization"*, Springer-Verlag, 1999.
25. K. Hanna, H. Sawhney, R. Kumar, Y. Guo, S. Samarasekara, *"Annotation of video by alignment to reference imagery"*, IEEE International Conference on Multimedia Computing and Systems, vol.1, pp. 38 - 43, 1999.
26. V. Govindu and C. Shekar, *"Alignment Using Distributions of Local Geometric Properties"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(10), pp. 1031-1043, 1999.

27. R. Cannata, M. Shah, S. Blask, J. Van Workum, *"Autonomous Video Registration Using Sensor Model Parameter Adjustments"*, Applied Imagery Pattern Recognition Workshop, 2000.
28. J. Le Moigne, N. Netanyahu, J. Masek, D. Mount, S. Goward, M. Honzak, *"Geo-registration of Landsat Data by robust matching of wavelet features"*, Proc. Geoscience and Remote Sensing Symposium, IGARSS, vol.4, pp. 1610-1612, 2000.
29. R. Wildes, D. Hirvonen, S. Hsu, R. Kumar, W. Lehman, B. Matei, W.-Y. Zhao *"Video Registration: Algorithm and quantitative evaluation"*, Proc. International Conference on Computer Vision, Vol. 2, pp. 343 -350, 2001.
30. S. Hsu, *"Geocoded terrestrial mosaics using pose sensors and video registration"*, Computer Vision and Pattern Recognition, 2001. vol. 1, pp. 834 -841, 2001.
31. D. Sim and R. Park, *"Localization based on the gradient information for DEM Matching"*, Proc. Transactions on Image Processing, 11(1), pp. 52-55, 2002.
32. D-G. Sim, R-H Park, R-C. Kim, S. U. Lee, I-C. Kim, *"Integrated Position Estimation Using Aerial Image Sequences"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(1), pp. 1-18, 2002.