Terrestrial Imaging from Visible Light Mosaics and a Line-Scan Hyperspectral Sensor

Juan-Pablo Ramirez-Paredes *, David Lary †, Nicholas Gans ‡
The University of Texas at Dallas, Richardson, TX, 75080

Hyperspectral imagery contains many different frequency bands for each pixel, enabling advanced detection algorithms. The so called pushbroom sensor technology is the most widespread realization of hyperspectral camera. Such cameras capture a single row of pixels, but each pixel can have hundreds of associated frequency bands . Assembling a full image from these pixel rows typically requires having a precise knowledge of the movement of the camera in space while it scans the desired area. In this research we present a procedure to build hyperspectral images from line sensor data without camera position information. The approach relies on an accompanying conventional camera and exploits the homographies between images of planar scenes for image mosaic construction. The hyperspectral camera is geometrically calibrated by using a special target and a variation of the Direct Linear Transform algorithm. This enables mapping the hyperspectral lines to each of the images in the mosaic, and a hyperspectral image can be built using the same homographies computed for the mosaic. Ideally, this research can enable small UAVs to perform enhanced detection and navigation previously reserved for high altitude aircraft.

I. Introduction

While digital cameras are useful as airborne sensors for terrestrial surveying, the spectral information contained in their images is limited to a few broad bands, typically one for grayscale imaging or three in color imaging. Multispectral cameras partially overcome this limitation, but they also are restricted in the quantity of bands they make available per picture element. The hyperspectral camera is a development that provides potentially hundreds of different spectral bands per pixel. Since the 1970s, when field measurements were made supporting the analysis of Landsat-1 observations, imaging spectroscopy via hyperspectral cameras has steadily progressed.[1]

There have been many applications of hyperspectral sensors in imaging of land, ocean and atmospheric composition. For instance, in precision farming it is possible to model the response of plants to factors such as drought, disease or pollution by analyzing the reflectance and absorption of specific light wavelengths.[2] There have also been successful studies of mineral identification by processing hyperspectral data from instruments such as the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS).[3] Oil spills such as the *Deepwater Horizon* incident of 2010 can also be precisely delimited by spectroscopy.[4] Finally, there have been attempts at characterizing the human skin from hyperspectral data to get accurate identification of pedestrians, setting a precedent for surveillance or search and rescue applications.[5]

To enable imaging at hundreds of distinct frequency bands, the optics of hyperspectral cameras do not operate the same as a typical digital camera. Hyperspectral cameras are typically line-scanning and record the spectrum at all channels corresponding to a single image line in space. An example of a hyperspectral image is seen in Figure 1. Given the low vertical spatial resolution, a single hyperspectral image is often of limited use. Hyperspectral sensor technology has progressed towards solutions that allow 2D frame capture, where each pixel has an associated set of frequency responses. However, the pushbroom variety is reliable, has no moving parts and remains the lowest in cost and complexity.[6]

*Graduate Research Assistant, Department of Electrical Engineering, Richardson, TX, pablo.ramirez@utdallas.edu, Student Member AIAA
†Associate Professor, Department of Physics, Richardson, TX, david.lary@utdallas.edu
‡Associate Professor, Department of Electrical Engineering, Richardson, TX, ngans@utdallas.edu, Member AIAA

If the hyperspectral camera is moved at a precise speed orthogonal to the scan line and consistent with the camera frame rate, the scan lines can be compiled and processed into a single image. Ideally, the imaging line moves one pixel per sample time, and the collected data is a proper image. Typically, interpolation is performed to mitigate for imprecise motion. Therefore, this is often referred to as "pushbroom" imaging and "pushbroom" camera.[a]. This approach generally requires high altitude aircraft or Earth-orbiting instruments. Relevant examples with aircraft include Ref. 7, which imaged 32 spectral bands over farmland from an aircraft flying steadily at 1900m, and Ref. 8, which involved a camera imaging 21 spectral bands, moving precisely in a controlled lab environment. An example of an outdoors hyperspectral camera of the pushbroom variety not mounted on aircraft is presented in Ref. 9, in which a crane was used as a platform to mount the sensor. In all cases, the pushbroom camera must be moved smoothly and at constant speed. Such imaging on an UAV, which suffer from motor vibrations and external disturbances such as wind, are unsuitable for pushbroom imaging.

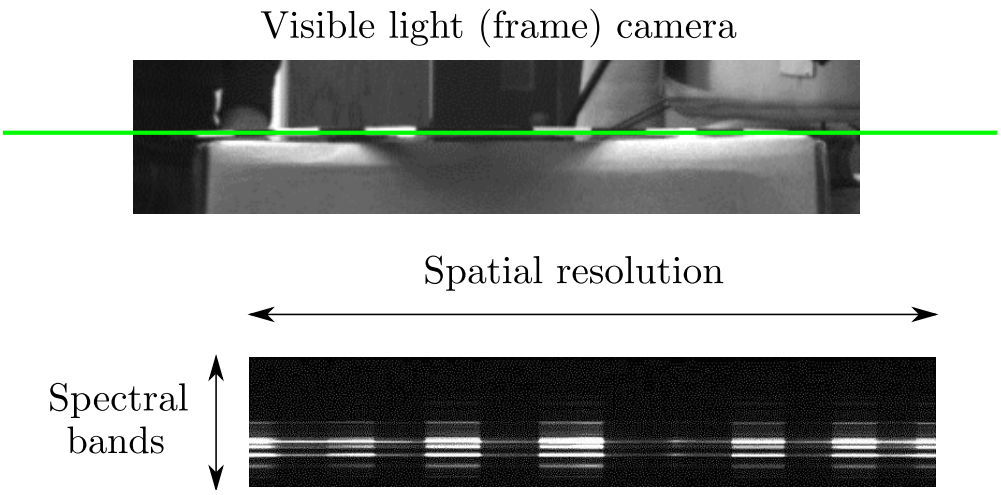

Figure 1. Comparison of a hyperspectral image line with a section of a visible light, conventional camera image under fluorescent illumination. The cameras are positioned side by side, and the green line corresponds to the line captured by the hyperspectral camera.

In this work, we propose a method to assemble images from collections of lines captured by a hyperspectral camera that is not moving smoothly. Furthermore, we do not require the addition of inertial sensors or external measurements of the camera pose. This is accomplished by using a conventional digital camera [b] rigidly attached to the hyperspectral camera as extra information to build the hyperspectral image.

With conventional cameras it is possible to stitch together different views of a scene by means of the homography.[10,11] This process is typically called image mosaicing. Computation of the homography requires point correspondences of a planar object and results in a simple linear relationship between point projections in both camera views. The homography contains information about the rotation and translation of one camera view relative to the other, and by exploiting this fact, our approach builds a hyperspectral image from a collection of hyperspectral lines.

The rest of this paper is organized as follows: Section II explains how the intrinsic parameters of the hyperspectral camera are computed. Section III deals with the construction of the hyperspectral and conventional camera rig, detailing its positioning restrictions and the method for aligning the cameras. Section IV explains how to map the pixels captured by the hyperspectral camera to their corresponding line in the conventional camera frame. Then Section V is devoted to the construction of an image mosaic for the conventional camera and applying that information to obtain a hyperspectral image. Our results are shown in Section VI, followed by a discussion in Section VII.

[a]While we present a method to generate hyperspectral images while not sweeping in a straight line, we will keep the common convention of calling such cameras "pushbroom" cameras.

[b]By "conventional digital camera" we mean a two-dimensional sensor with a rectangular grid as its image plane, where each element is sensitive to light in a broad band of frequencies hundreds of nanometers wide.

II. Hyperspectral camera calibration

In the pinhole camera model, a calibrated camera can measure the angles between individual light rays. For a two-camera array, calibration can be used to predict how an object imaged by one camera will be seen by the other, by means of the intrinsic and extrinsic parameters for each camera. Line sensors, such as hyperspectral cameras, can be calibrated by attaching them to a mechanical platform able to perform precise movements in a direction perpendicular to the scan line. If the line sensor scans the object at constant speed, stacking the image lines results in a full image frame. With this approach, the calibration procedure is similar to conventional camera techniques.[12]

When such a platform is unavailable, it is still possible to calibrate the camera in the horizontal direction. If the camera sensor technology is CCD or CMOS, the pixel size can be known with high certainty from manufacturer data. This fact, coupled with high quality optics, can be used to predict the angular resolution per pixel in the vertical direction from horizontal direction measurements.

Our novel approach to calibration is mathematically similar to an approach by Gupta and Hartley.[13] However, their calibration method requires a moving camera (or moving known target) with accurate measurements of velocity. Our method does not require any such motions, so it simplifies ex situ calibration. We consider the hyperspectral camera to be similar in operation to a conventional camera, with the restriction of having one pixel of vertical resolution. This approximation simplifies the camera model and provides precise results.

Finding the geometric calibration for a pushbroom camera is possible by using a special target. We consider a Cartesian frame $\mathcal{F}_t$ associated with the target. The target has a special pattern on the surface, containing a number of high contrast features we call *control points* with known coordinates in $\mathcal{F}_t$. It is important to avoid having every control point on a single line in Cartesian space. For such a configuration there is an ambiguity, and a certain image of the points can arise from an infinite number of camera poses. Therefore we construct the targets such that the control points are on two intersecting perpendicular lines, specifically a drawing triangle with an alternating pattern of black and white bars on its edges. The width of every bar is 30 mm. The target frame $\mathcal{F}_t$ has its X and Z axes as the intersecting lines containing the control points. The target and its hyperspectral image is seen in Figure 1, and a diagram of its shape is shown in Figure 2.

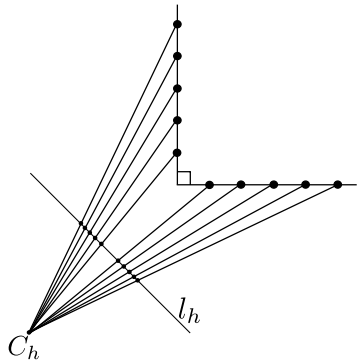

Figure 2. The calibration target used for the hyperspectral camera. Every point lies on the same plane, but not on the same line. The point pattern projection is unique, up to scale.

There is a Cartesian frame $\mathcal{F}_h$ attached to the hyperspectral camera body, with the origin at the camera focal point, the X axis parallel to the spatial pixels direction, the Y axis perpendicular to the spatial pixels direction, and the Z axis aligned with the line going from the focal point to the image line center and normal to the image line. The calibration target shape makes the control points visible to the hyperspectral camera only when the plane containing the points coincides with the plane formed by the hyperspectral camera's imaging line and its focal point.

Let us assume that no points are outside of the XZ plane of the hyperspectral camera frame $\mathcal{F}_h$, since they could not be imaged otherwise. For simplicity, and without loss of generality, assume the target frame $\mathcal{F}_t$ is the world/inertial frame $\mathcal{F}_w$. The homogeneous coordinates of a world point are $\mathbf{M} = [X\ 0\ Z\ 1]^T \in \mathbb{P}^3$,

and the homogeneous transformation from the world frame to the $\mathcal{F}_h$ camera frame is denoted by $T_w^h \in \mathcal{SE}(3)$. A projection model from the world to the hyperspectral camera is given by

$$\lambda \mathbf{u} = K \Pi_0 T_w^h \mathbf{M}, \tag{1}$$

where K is a calibration matrix

$$K = \begin{bmatrix} f & u_0 \\ 0 & 1 \end{bmatrix}$$

with f being the horizontal focal length of the camera and u_o its principal point in the horizontal direction. The projection of the point is $\lambda \mathbf{u} = [u \ 1]^T$, with λ being an unknown scale factor. The matrix Π_0 is a projection given by

$$\Pi_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Following the developments for conventional cameras, it is possible to adapt the Direct Linear Transform (DLT)[14] approach to the pushbroom camera case. Just as the original DLT inverts a map $P_2 : \mathbb{P}^3 \to \mathbb{P}^2$, the pushbroom case requires an inversion of a map $P_1 : \mathbb{P}^2 \to \mathbb{P}$. By exploiting the structure of the rigid body transformation between the imaging line and a set of points lying on the same plane, the problem can be formulated as a linear system of equations. Choosing the control points of the calibration target to be on the same plane as the hyperspectral camera line and focal point, it is possible to discard the y coordinate of the points by setting it to zero.

As the frames $\mathcal{F}_w$ and $\mathcal{F}_h$ must have the Y-axes aligned, the rotation in T_w^h must have the form

$$T_w^h = \begin{bmatrix} \cos\theta & 0 & \sin\theta & t_x \\ 0 & 1 & 0 & 0 \\ -\sin\theta & 0 & \cos\theta & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where θ is the angle between the X axes of $\mathcal{F}_h$ and $\mathcal{F}_w$.

Considering that $Y = 0$ for every point, let us multiply $\Pi_0 T_w^h$ and discard the Y dimension in our computations since it does not contribute to the projection. This gives rise to the simplified projection equation

$$\lambda u = \begin{bmatrix} f & u_0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta & t_x \\ -\sin\theta & \cos\theta & t_z \end{bmatrix} \begin{bmatrix} X \\ Z \\ 1 \end{bmatrix} = [R\,T]\mathbf{X}. \tag{2}$$

where $R \in \mathcal{SO}(2)$ and $T \in \mathbb{R}^2$. The intrinsic and extrinsic parameters can be combined to give a camera matrix that has six unknowns as

$$\lambda u = \begin{bmatrix} L_1 & L_2 & L_3 \\ L_4 & L_5 & L_6 \end{bmatrix} \begin{bmatrix} X \\ Z \\ 1 \end{bmatrix} = P_1 \mathbf{X} \tag{3}$$

To cancel the scaling factor λ and rewrite the equations as a homogeneous linear system, we multiply both sides of (3) by $u^T Q$, where Q is a skew-symmetric matrix given by

$$Q = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

thus getting

$$u^T Q P_1 M = 0. \tag{4}$$

Arranging the terms the resulting equation gives

$$XL_1 + ZL_2 + L_3 - uXL_4 - uZL_5 - uL_6 = 0.$$

Stacking the equations for five or more point matches between the world and the projection line, we have the homogeneous system

$$\begin{bmatrix} X_1 & Z_1 & 1 & -u_1X_1 & -u_1Z_1 & -u_1 \\ X_2 & Z_2 & 1 & -u_2X_2 & -u_2Z_2 & -u_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_n & Z_n & 1 & -u_nX_n & -u_nZ_n & -u_n \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_6 \end{bmatrix} = 0 \tag{5}$$

$$SL_s = 0. \tag{6}$$

The matrix S can be at most of rank 5, so the system can be solved by finding the nullspace of the matrix S. From (2), we see that with proper scaling $L_4^2 + L_5^2 = 1$, so λ can be recovered up to a sign ambiguity, setting it to

$$\lambda = \pm \frac{1}{\sqrt{L_4^2 + L_5^2}}.$$

Once the matrix L has been properly scaled, the intrinsic and extrinsic parameters can be recovered as

$$u_0 = L_1L_4 + L_2L_5$$

$$f = \left[L_1^2 + L_2^2 - u_0^2\right]^{1/2}.$$

Any choice of sign for λ results in the same parameters f and u_0, but it leads to different sets $\{R, T\}$, giving rise to two solutions. Picking either of those two values for λ, the recovered poses are

$$R_1 = \begin{bmatrix} L_5 & -L_4 \\ L_4 & L_5 \end{bmatrix}$$

$$R_2 = -R_1$$

$$T_1 = \begin{bmatrix} \frac{L_3 - u_0 L_6}{f} \\ L_6 \end{bmatrix}$$

$$T_2 = -T_1.$$

Depending on the choice of control point coordinates in the target frame, it is possible to discard one of the poses by choosing the correct translation in the camera frame. The correct translation has the control points in front of the camera, i.e. positive t_z.

The results from a single picture of the calibration target are not precise enough most of the time. The method presented so far does not deal with radial distortion or image noise. Radial distortion is a lens induced phenomenon that causes a displacement in the projection of the world points from their ideal projections according to the pinhole model. In the vision literature, radial distortion is usually modeled as a polynomial of even degree where the only nonzero coefficients are those of even powers.[15] For the case of the pushbroom camera, radial distortion can be modeled as

$$u_d = u + K_1(u - u_0)^2 + K_2(u - u_0)^4 + \mathcal{H.O.T}. \tag{7}$$

The distorted point projection is u_d, while the projection predicted by the pinhole model is given by u and $K_1, K_2, \ldots$ are distortion coefficients. The number of coefficients to use depends on the degree of precision required. One coefficient provides acceptable results for our application.

To improve the parameter estimation, it is possible to follow up with a nonlinear optimization step. Taking several images of the calibration target and locating the control points for each, the measure to be minimized is the point reprojection error given by

$$J(\mathbf{u}, \mathbf{u}') = \sum_{i=1}^{n} \sum_{j=1}^{m} ||u_i^j - u_i^{j\prime}||^2. \tag{8}$$

The vector $\mathbf{u} \in \mathbb{R}^{nm}$ contains the projections of the calibration points as predicted by our model using the hyperspectral camera parameters, while $\mathbf{u}'$ are the actual projections of the calibration points. The point u_i^j represents the i-th point of the j-th calibration target image, and there are n such points and m images of the target. The Levenberg-Marquardt optimization is well suited for these problems, and the parameters obtained from the DLT part of the calibration can be used as initialization values.[16]

III. Camera array calibration

Our objective is to relate the hyperspectral camera line to some line on the conventional camera image. Let us assume that all of the world points captured by the cameras reside on a plane in space. Without knowledge of the plane depth relative to the cameras, it is not generally possible to predict which line of the conventional camera contains each of the world points that the line camera captures. This arises from epipolar geometry.[17]

Consider a pair of cameras facing the same direction, such as a stereo array. We will show that given the projection of a world point on the hyperspectral camera image plane, it is possible to predict the line on the conventional camera image plane onto which it is also projected. This line to line map generally depends on the depth of the world points in the camera frames. This dependency is eliminated only if the hyperspectral camera line is forced to coincide with an epipolar line.

In projective geometry, using the pinhole camera model, vectors in $\mathbb{R}^3$ can be used as homogeneous coordinates to represent both lines and points on the image plane. For points, the vector $\mathbf{p} = [u\ v\ 1]^T$ represents the coordinates of a point in the image plane, typically this denotes the intersection of the image plane and a ray of light from a 3D point through the camera focal point of the camera. For lines, their standard form $au + bv + c = 0$ can be represented as $l = [a\ b\ c]^T$. A given point $\mathbf{p}$ is part of a line l iff $\mathbf{p}^T l = 0$. Thus, the hyperspectral camera line can be expressed as $l_h = [0\ 1\ 0]^T$ in homogeneous coordinates in the hyperspectral camera frame, since only points $\mathbf{p} = [u\ 0\ 1]^T$ belong to it. By our hyperspectral camera model in (1), only points with 3D homogeneous coordinates $\mathbf{X} = [X\ 0\ Z\ 1]^T$ in the $\mathcal{F}_h$ camera frame can have a projection in the hyperspectral image. The projections of these points lie on l_h.

By making the epipole $\mathbf{e}_h$ to lie on l_h such that $\mathbf{e}^T l_h = 0$, l_h becomes an epipolar line. From epipolar geometry, it is known that each of the epipolar lines in one camera plane has a corresponding epipolar line in the other. To find the line l_c corresponding to l_h, consider any line k not crossing the epipole. The point were l_h and k intersect is $\mathbf{x} = [k]_\times l_h$, where $[k]_\times$ is the skew-symmetric matrix form used to compute the cross product between k and other vectors. Then $l = E\mathbf{x} = E[k]_\times l_h$, where E is the Essential matrix between both cameras.[18]

Since any nonzero vector in $\mathbb{R}^3$ can represent either a line or a point in projective geometry, a given point $\mathbf{p}$ can also be used as a representation of a line. Interestingly, the line equation expressed by $\mathbf{p}$ is such that $\mathbf{p}$ as a point does not belong to the line. This can be verified easily since $\mathbf{p}^T\mathbf{p} \neq 0$. We need a line k that cannot go through $\mathbf{e}_h$. Thus, the line with the same vector as the epipole $\mathbf{e}_h$ is a good choice for k. Then the epipolar line corresponding to l_h can be found as

$$l_c = E[\mathbf{e}_h]_\times l_h. \tag{9}$$

From the constraints placed on $\mathbf{e}_h$, the only translation between cameras that is allowed is $T = [X\ 0\ Z\ 1]^T$, in the hyperspectral camera frame. Rotation is unrestricted. To illustrate the different two-camera configurations consider Figure 3. It is only when the hyperspectral line becomes an epipolar line that the points projected onto the hyperspectral camera frame can be found on a corresponding epipolar line in the other camera regardless of the 3D world point depths.

With the conditions for a line to line mapping between cameras established, the next step is to formulate a way to align the cameras such that (9) holds. The calibration target from Section II can be used as an alignment aid. If every control point on the target is visible on the hyperspectral camera line, the conventional camera can be positioned to guarantee that all the control points lie on a single line in the conventional camera image. This is simpler to achieve if the conventional camera image line l_c is chosen to be approximately parallel to its horizontal pixel rows. The alignment procedure is depicted in Figure 1, although only a small section of the conventional camera image is shown.

Note that between the two camera frames, $R \neq I$ and $T = [t_x\ 0\ t_z]^T$ in the $\mathcal{F}_h$ frame. However, their relative poses, and by extension E, are not required to be explicitly computed. To justify this, note that E will map any imaged point on l_h to l_c, but due to the lack of depth information, it does not provide a point-to-point map between camera image planes. This issue and its solution are discussed further in Section IV.

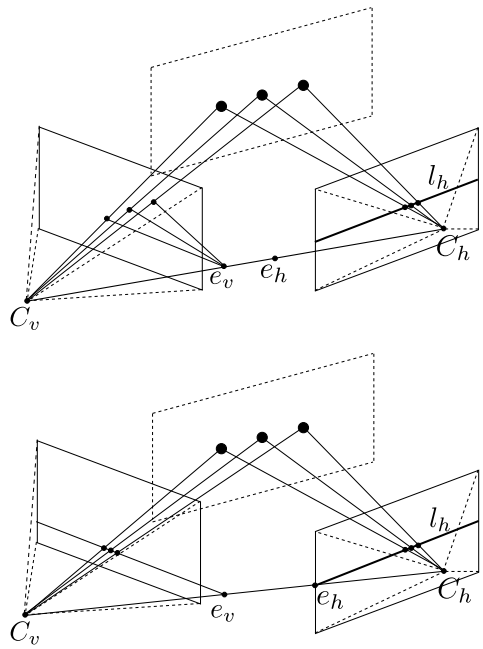

Figure 3. Effect of the world point depth over the projection of the hyperspectral camera line onto the conventional camera, with different alignments between cameras. Top: The world points lie on the same plane and are projected on a line on the conventional camera, but the position of this line varies with the point depth. Bottom: The hyperspectral line is an epipolar line. The points project over the same line in the conventional camera regardless of their depth.

IV. Hyperspectral line to image mapping

With the two cameras aligned as in Section III, the points projected on their image planes are related via a homography when capturing world points that lie on a plane $\mathbf{\Pi}$. Let us consider the cameras to be calibrated, so that the coordinates of the image plane points for both are expressed in the same units. For a given world point, its projections on the hyperspectral and the conventional camera are denoted by x_h and x_c, respectively. Then $x_h = H_c^h x_c$, where H_c^h is the homography between both cameras for $\mathbf{\Pi}$. There also exists a homography between any two images of points in $\mathbf{\Pi}$ captured by a single camera from two different positions. If the points captured by the conventional camera from two poses are x_c and x_c', respectively, they are related by $x_c' = Hx_c$. Consider the points captured by the hyperspectral camera as well. The points x_h, x_h' captured by the hyperspectral camera while x_c and x_c' were captured are related by

$$x_h' = H_c^h H [H_c^h]^{-1} x_h.$$

Thus, it is clear that the effect of H_c^h is significant, and depends on the pose difference of the cameras as well as on the distance and normal vector to $\mathbf{\Pi}$. Figure 4 illustrates this situation.

To compensate for this effect, we propose an optimization task that finds the mapping between cameras for each different scene. On Section III it was established that the search space for corresponding points is a single line on each image plane. Each line can be treated as a one-dimensional signal. Using this information we find the scaling and displacement of the hyperspectral signal for each image from the camera pair prior to constructing the hyperspectral image.

The spectral sensitivities of the cameras are usually supplied by the manufacturers. Although this information may not be precise enough to be considered for radiometric calibration, it can be used as an approximation. In our case this is aggravated by the fact that we were not provided with any information on the prism or diffraction grating of the hyperspectral camera, so we are assuming that this component does not modify the imaging sensor sensitivity substantially. The conventional camera captures grayscale intensity images, so every pixel performs an integration of the light intensity over the pass band of the sensor. In the case of the hyperspectral camera, the response of the intensity image from the conventional camera can be approximated at each pixel by adding the responses to every frequency band. This allows

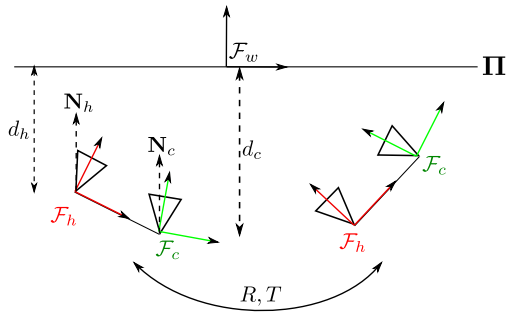

Figure 4. The plane $\mathbf{\Pi}$ and the cameras as projected onto the XZ plane of $\mathcal{F}_h$. The homography for two views of $\mathbf{\Pi}$ is different for each camera. Only the conventional camera can be used to determine its corresponding homography for two views via the four-point algorithm.

a comparison between intensity values from both cameras, even though they may differ by some unknown scalar factor and offset.

From (9) we know the line to line mapping between the hyperspectral and the conventional camera. However, the data in the hyperspectral image is a line segment in l_h, which corresponds to a line segment in l. The exact mapping between line segments is not known if the depth and orientation of the plane $\mathbf{\Pi}$ w.r.t. $\mathcal{F}_h$ and $\mathcal{F}_c$ is unknown. Finding a mapping between line segments involves finding a scale factor n and displacement u, which scale and shift the line segment along l.

Each hyperspectral image captured is a matrix $\mathbf{G}_k$, where each element $\mathbf{g}_{i,j}$ corresponds to the i-th frequency band and the j-th spatial component. There are f frequency bands, and there are $k = 1, \ldots, N$ such matrices for a N image sequence. Consider the matrix $\mathbf{V}$, where each row $\mathbf{v_k}$ is given by

$$\mathbf{v_k} = \sum_{i=1}^{f} \mathbf{g}_{i,k}. \tag{10}$$

Similarly, consider a matrix $\mathbf{W}$ where the i-th row is the line of pixels of the conventional camera corresponding to the hyperspectral line. This matrix has m columns, each a different sample point over the line. To match each of those lines to the corresponding hyperspectral line, it is necessary to minimize the following cost function

$$J(\mathbf{u}, \mathbf{n}, \mathbf{V}, \mathbf{W}) = \sum_{k=1}^{N} \frac{\sum_{i=1}^{m} v_{k,i} \cdot w_{k,\lfloor (i-u_k)\frac{m}{n_k} \rfloor}}{||\mathbf{v_k}|| \sqrt{\sum_{i=1}^{m} \left[w_{k,\lfloor (i-u_k)\frac{m}{n_k} \rfloor} \right]^2}}. \tag{11}$$

The vectors $\mathbf{u}$ and $\mathbf{n}$ are the displacements and the scale factors for each line correspondence, respectively. The elements of the scale factors vector are denoted as n_k. This cost function is based on the cosine similarity between signals.

This optimization task is nonlinear in nature and we have obtained our best results by applying the Nelder-Mead algorithm to the data.[19] For initialization it is possible to estimate an initial displacement by using normalized correlation between the scaled hyperspectral lines and their conventional camera equivalents. The initial scaling factor to use is the ratio between the conventional camera focal length and that of the hyperspectral camera.

V. Homography-based image registration

When the scenes to be captured contain objects far away from the camera, it can be valid to approximate that every world point lies on a single plane. This assumption is of special interest while performing aerial or satellite photography. The mapping of pixel between different images of such scenes can be characterized as a homography. It is possible to create image mosaics by combining many images and the homographies between pairs.

The homography matrix between two camera views can be recovered by the four-point algorithm from point correspondences,[20].[21] Ground-based control point markers can be used in the case of aerial photography

if manual correspondences are established. Coupling the four-point algorithm with RANSAC can be used to get a good homography matrix approximation in the presence of noise and mismatches from many candidate point correspondences.[22] This is useful if the correspondences are obtained in an automated manner with algorithms such as SIFT or SURF. In our approach we make use of SURF.[23] Its performance is comparable to SIFT while being computationally less complex, and there are open implementations available.[24] Matching of features can be done either exhaustively or through an approximation algorithm, depending on the computing time available. We use the exhaustive approach with thresholding of the feature vector matches, since the computation is performed offline.

To construct a mosaic from several images of a planar scene, it is possible to find the homographies for every image pair. Then the images can be transformed to fit a reference image chosen among them. The problem with this approach is that errors accumulate and damage the results. To enhance the resulting mosaic we use the approach of Ref. 25. Consider a collection $\mathbf{I}$ of $k = 1 \ldots N$ images I_k of a planar scene, with enough overlap and common feature points between pairs to successfully recover the homographies. The feature points homogeneous coordinates for each I_k are a $3 \times M$ matrix p_k, where M is the number of features. The homographies are then the set $\mathbf{H}$, with each element H_i^j being the homography between images I_i and I_j that satisfies $p_j = H_i^j p_i$. For this set, $i = 1, \ldots, N$, $j = 1, \ldots, N$, and $i \neq j$. These relationships can be exploited in a nonlinear optimization scheme to enhance the quality of the mosaic. The function to minimize is the following

$$J(\mathbf{H}, p_1, \ldots, p_N) = \sum_{i=1}^{N} \sum_{j=1}^{N} ||p_j - H_i^j p_i||^2. \tag{12}$$

The resulting homographies can be converted to their Euclidean equivalents by applying the calibration matrix for the camera, as $H_e = K^{-1} H K$.

At this stage there is enough information to construct the hyperspectral image. Using the line-to-line correspondence between cameras, as well as the scaling factors $s_k = m/n_k$ and the displacements $\mathbf{u}$, the positions of the hyperspectral camera points projected onto the conventional camera can be recovered. Using these new point coordinates for the hyperspectral data and applying the homographies from the previous step, a single hyperspectral image with several frequency band layers is recovered. To achieve this, each of the projected point positions is recalculated to fit a place in the final hyperspectral image. This results in scattered data consisting of triplets of the form $\{x, y, F_{1\ldots f}\}$, where $F_{1\ldots f}$ is the set of all the intensity values for each of the frequency bands from the hyperspectral sensor. There are nonuniform interpolation techniques that can then be applied to the data, such as inverse distance weighting, nearest neighbors, kriging, etc.[26,27] The choice of interpolation algorithm is up to the specific application for the images. Since this work had no particular conditions to meet with regard to image applications, the algorithm chosen was linear interpolation.

VI. Results

To illustrate our hyperspectral imaging approach, we present four outdoors experiments. We employed a Sony XCG-V60E monochrome conventional camera and a custom-made hyperspectral camera manufactured by Surface Optics, based on the AVT Prosilica GC 655. Both cameras have 640 pixels of horizontal resolution. The conventional camera has 480 pixels of vertical resolution, while the hyperspectral vertical information contains the frequency response in 120 different bands. Our camera array is mounted on an aluminum rail, shown in Figure 5. No radiometric calibration was performed between cameras, i.e. the exact relative spectral sensitivities of both cameras were unknown. For each experiment, we show five images. The first row shows two images with the image mosaic from the conventional camera on the left, along with an overlay of the intensity image derived from the hyperspectral data on the right side. This intensity image is composed of the sum of every hyperspectral channel. The second row contains three images, each cropped and zoomed to show detail of the hyperspectral data with a small surrounding portion of conventional camera imagery. The second row consists of the intensity image from the hyperspectral camera on the left, a pseudocolor image from three spectral channels in the middle and the Normalized Difference Vegetation Index (NDVI)[28] on the right. To build the pseudocolor image we took channels 80, 90 and 100 to represent the colors red, green and blue respectively. This choice was based on the intensity peaks observed for such channels from an LCD monitor displaying a white screen.

Figure 5. Experimental camera array. The hyperspectral camera has a cylindrical shape and it is shown in the middle, with the smaller conventional camera to its left.

The NDVI combines the reflectance of visible and near-infrared light wavelengths to estimate the vegetation content of a region. Healthy vegetation absorbs most of the visible light that reaches it, while reflecting most of the infrared radiation. Unhealthy or dormant vegetation reflects more visible light with respect to near-infrared. For each imaged point, the NDVI is computed as

$$NDVI = \frac{NIR - VIS}{NIR + VIS}$$

where NIR stands for near-infrared intensity and VIS is the visible light intensity. The range of values for the NDVI is $[-1, 1]$, where 1 represents healthy, dense vegetation. Values near 0 may represent concrete, dirt or otherwise vegetation depleted regions, and values near -1 typically correspond to deep water or clear sky. There is no single choice of visible and near-infrared bands to calculate NDVI, so we empirically selected spectral bands to represent them. For the following NDVI images, we used channel 10 to represent the near-infrared intensity response, while the visible light portion was represented by channel 80.

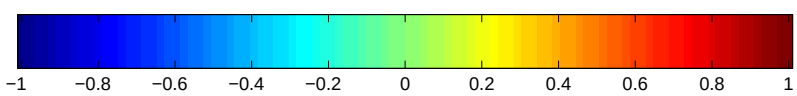

Figure 6. Color map used for the NDVI measurements.

To illustrate the NDVI of the scenes, we used the colormap shown in Figure 6, where blue tones correspond to areas with no vegetation. Healthy plants begin to appear as red tones start, with dark red being the most lively. The images that make up these scenes were acquired with the camera array being hand-held. The scenes are nonplanar, so the assembled mosaics and hyperspectral images show some effects of parallax, such as ghosting. Nevertheless, the approach works well and demonstrates it effectiveness for general imaging. High-altitude imagery, such as from a UAV, should mitigate these effects. The first scene, presented in Figure 7, is a natural setting with a few man-made objects such as trailers and containers with a coat of paint. In this scene, in the image mosaic from the frame camera, the trailers are not clearly distinguished from the trees. The pseudocolor image shows them in their actual tones, as seen in the middle. The NDVI image displays the trees in intense red tones, while the trailers appear near the zero value from Figure 6 and the sky appears closer to -1. The conventional camera images show an overexposed sky, but the hyperspectral images preserve the cloud structure for some channel combinations.

The scene at Figure 8 shows more man-made objects. A yellow steel pillar is visible on the left side of the image mosaic, with a concrete slab on the lower right quarter. Some details lost in the conventional camera images can be seen in hyperspectral mosaics, such as lines painted on the concrete (for its use as a basketball court). The pseudocolor scene matches the yellow tone of the pillar and allows the observer to

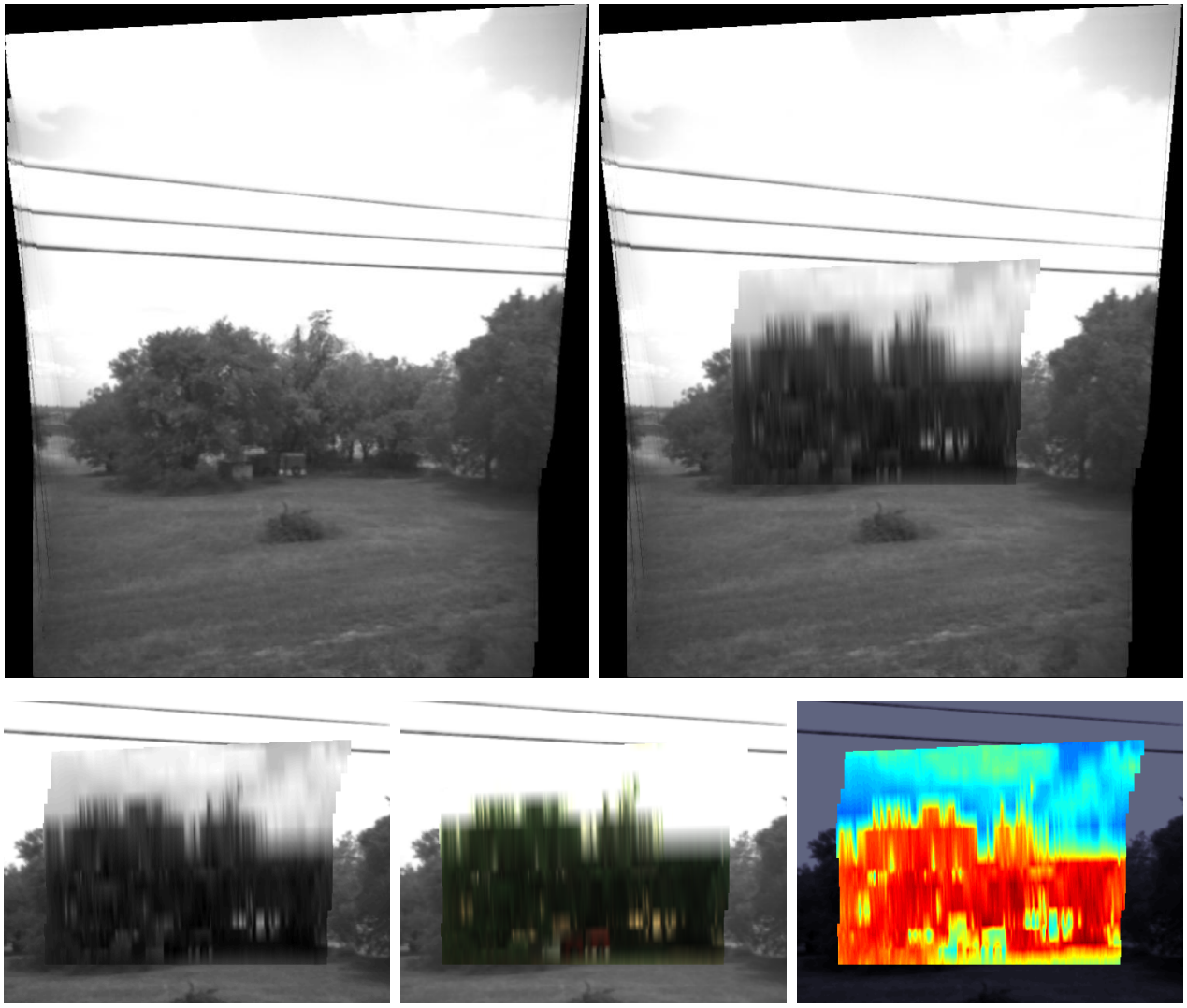

Figure 7. Hyperspectral imaging of a rural/urban landscape. Top left: the visible light mosaic as captured by the Sony XCG-V60E camera. Top right: sum of every spectral channel from the HS image, overlaid on top of the visible camera mosaic. Bottom row, left to right: summation over every channel of the data cube; pseudocolor from red, green and blue channels; and Normalized Difference Vegetation Index.

visualize healthy and dry grass patches. The NDVI image makes those distinctions clearer, and the tree line is highlighted in red tones.

For the third scene, shown in Figure 9, we selected a site that contained prominent foreground objects, complicating the mosaicing task. A storage shed stands in the middle of a grass field, and a chain-link fence closer to the cameras holds a few plants. One interesting feature of this scene is the small bush in front of the shed, which is not clearly visible on the image mosaic. It remains hidden in the shadows cast by the building even in the pseudocolor image. The NDVI image shows its location clearly.

One last scene showing the results of smooth camera motions given in Figure 10. Unlike the previous examples, this image sequence was captured with the camera pair mounted on top of a tripod with a swiveling head. The tripod head tilt was moved in small increments, resulting in a sharp mosaic image. The same yellow steel pillar and concrete slab from Figure 8 appear in this scene, captured from a different point of view. Again, the pillar appears dark blue on the NDVI image, while the grass is separated into healthy and dry patches. Unlike previous scenes, the sky on the NDVI image is not dark blue but remains close to zero, indicating an overcast sky.

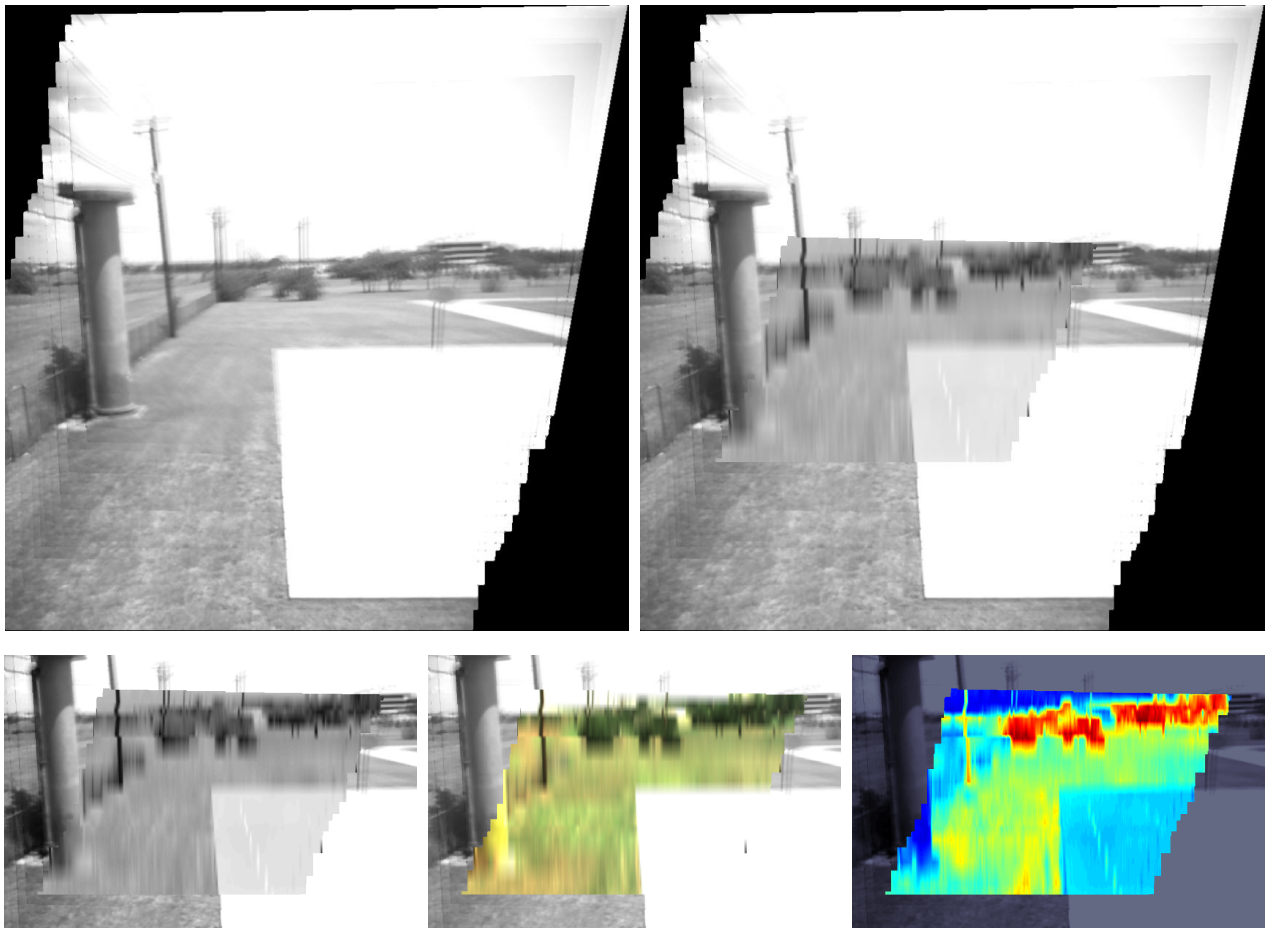

Figure 8. Basketball court scene, depicting concrete, painted metal and grass.

VII. Conclusion

Based on an auxiliary conventional camera, we have presented a method to acquire images with a linescan hyperspectral sensor. Our approach maps each hyperspectral line to a line in the conventional camera image plane. We showed that this can only be achieved for arbitrary world point depths when the hyperspectral line is an epipolar line too. Two variables that need to be solved for even after this line to line mapping has been found are the scale of each hyperspectral line segment, and its displacement with respect to the conventional camera line segment over which it is mapped. Nonlinear optimization through direct search methods yields the best results when dealing with such a task.

The final assembly of a hyperspectral image involves finding the homographies between conventional camera image pairs. This is used to create an image mosaic and at the same times the hyperspectral image. The use of conventional mosaicing techniques means that our approach is subject to some of their same restrictions, such as parallax errors and sensitivity to Z-axis rotation. Flyover sequences with nadir-pointing cameras would yield better results for our homography-based approach, and they constitute the next testing ground.

Acknowledgments

This work was financially supported by the Mexican National Council for Science and Technology (CONACyT) through scholarship grant number 215161.

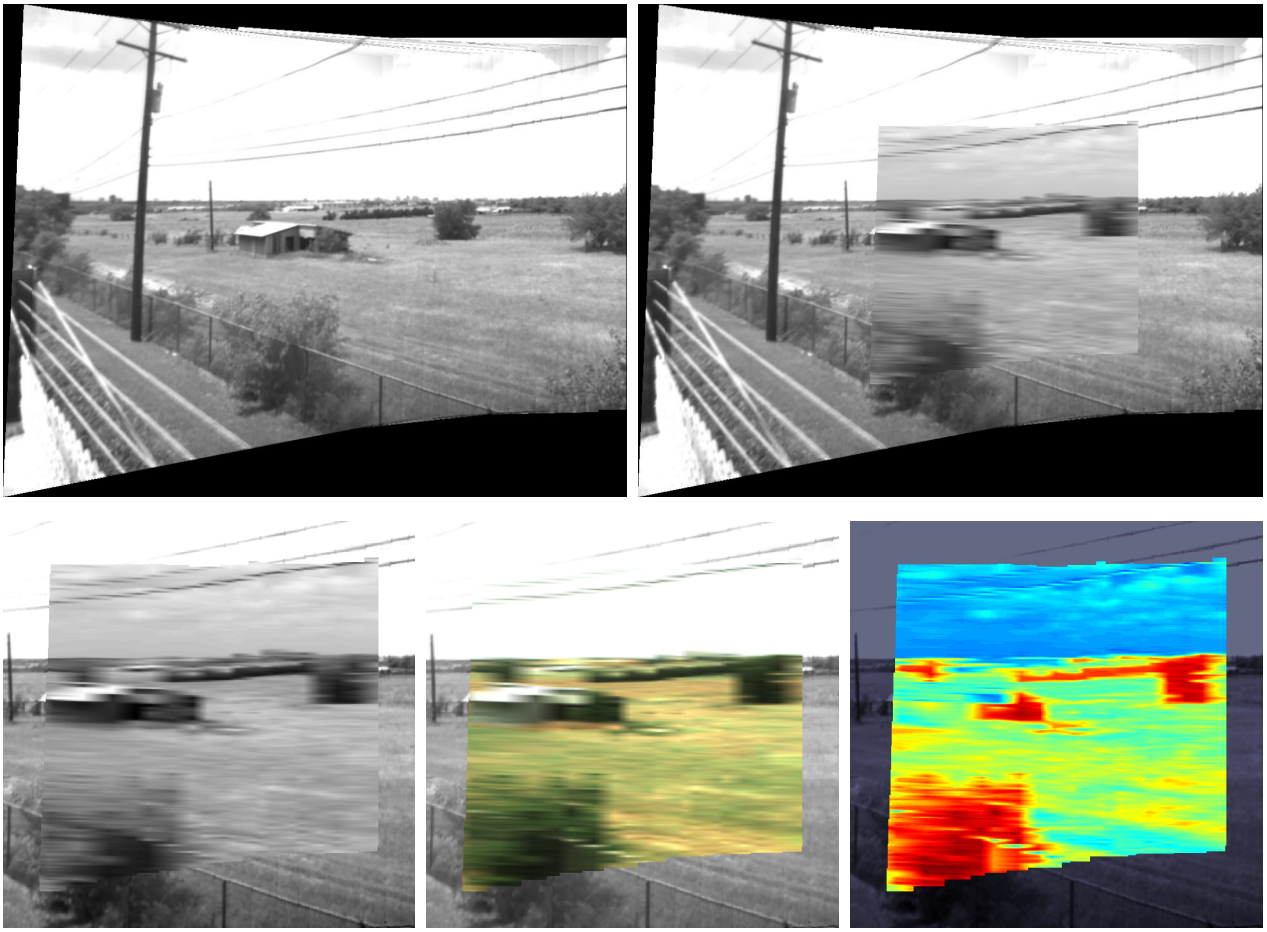

Figure 9. An scene depicting objects at different depths, showing the effects of parallax error from the mosaicing procedure.

References

[1]A. Goetz, "Three decades of hyperspectral remote sensing of the Earth: A personal view," *Remote Sensing of Environment*, vol. 113, no. 1, pp. S5–S16, 2009.

[2]K. Smith, M. Steven, and J. Colls, "Use of hyperspectral derivative ratios in the red-edge region to identify plant stress responses to gas leaks," *Remote Sensing of Environment*, vol. 92, no. 2, pp. 207 – 217, 2004.

[3]F. Kruse, J. Boardman, and J. Huntington, "Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1388 – 1400, june 2003.

[4]P. Sidike, J. Khan, M. Alam, and S. Bhuiyan, "Spectral unmixing of hyperspectral data for oil spill detection," in *SPIE Optical Engineering+ Applications*. International Society for Optics and Photonics, 2012, pp. 84 981B–84 981B.

[5]J. Herweg, J. Kerekes, and M. Eismann, "Hyperspectral imaging phenomenology for the detection and tracking of pedestrians," in *2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, july 2012, pp. 5482 –5485.

[6]F. Vagni, "Survey of hyperspectral and multispectral imaging technologies," *NATO Research and Technology*, no. TR-SET-065-P3, 2007.

[7]A. Asmat, E. Milton, and P. Atkinson, "Empirical correction of multiple flightline hyperspectral aerial image mosaics," *Remote Sensing of Environment*, vol. 115, no. 10, pp. 2664 – 2673, 2011.

[8]Y. Schechner and S. Nayar, "Generalized mosaicing: wide field of view multispectral imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 10, pp. 1334 – 1348, oct 2002.

[9]S. T. Monteiro, Y. Minekawa, Y. Kosugi, T. Akazawa, and K. Oda, "Prediction of sweetness and amino acid content in soybean crops from hyperspectral imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, no. 1, pp. 2–12, 2007.

[10]R. Szeliski, "Video mosaics for virtual environments," *Computer Graphics and Applications, IEEE*, vol. 16, no. 2, pp. 22–30, 1996.

[11]F. Caballero, L. Merino, J. Ferruz, and A. Ollero, "Unmanned aerial vehicle localization based on monocular vision and online mosaicking," *Journal of Intelligent and Robotic Systems*, vol. 55, no. 4-5, pp. 323–343, 2009.

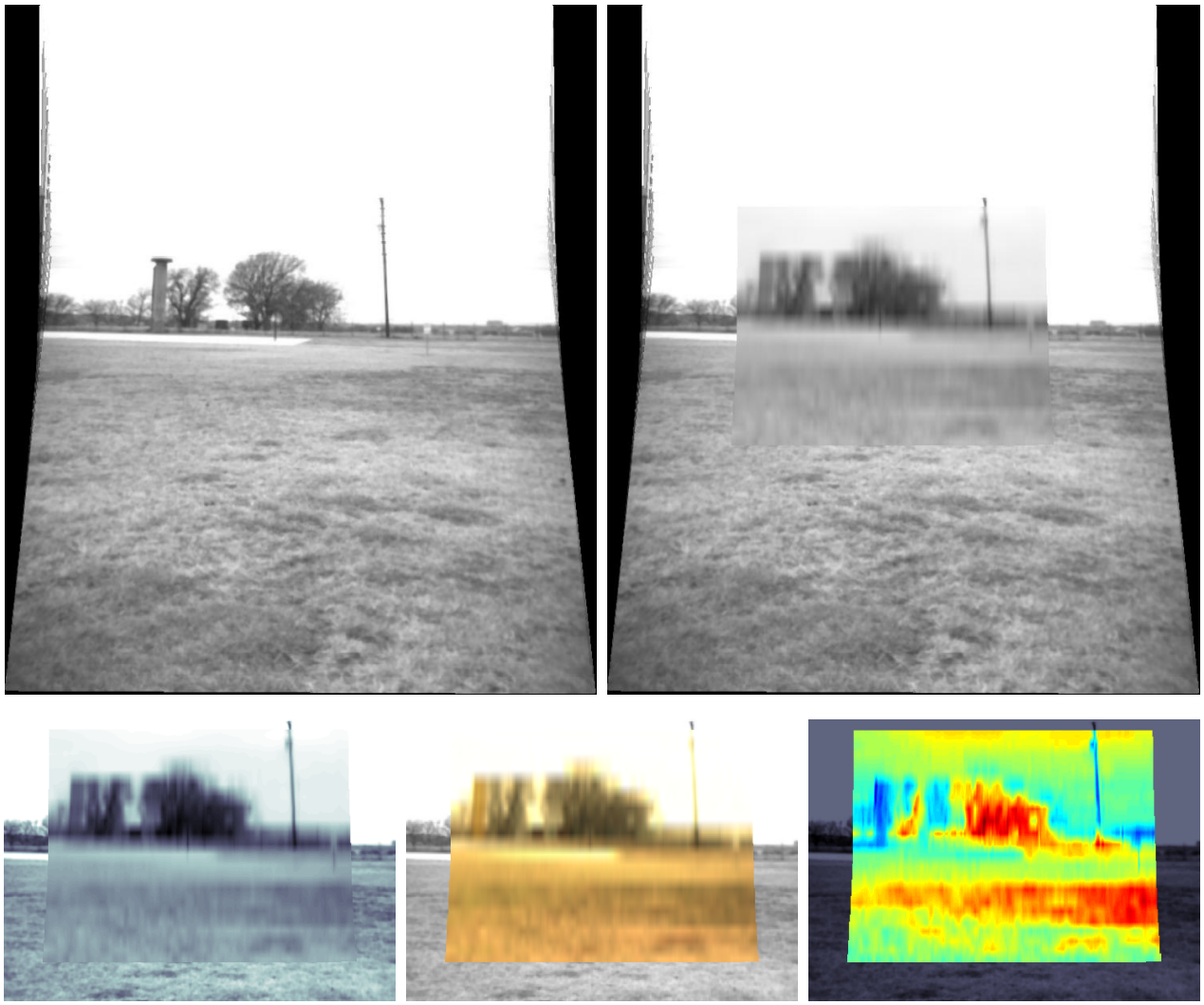

Figure 10. Scene captured by smooth rotational motion of the cameras.

[12]J. Draréni, S. Roy, and P. Sturm, "Plane-based calibration for linear cameras," *International Journal of Computer Vision*, vol. 91, no. 2, pp. 146–156, 2011.

[13]R. Gupta and R. I. Hartley, "Linear pushbroom cameras," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 9, pp. 963–975, 1997.

[14]Y. Abdel-Aziz, "Direct linear transformation from comparator coordinates in close-range photogrammetry," in *ASP Symposium on Close-Range Photogrammetry in Illinois*, 1971.

[15]J. Salvi, X. Armangué, and J. Batlle, "A comparative review of camera calibrating methods with accuracy evaluation," *Pattern recognition*, vol. 35, no. 7, pp. 1617–1635, 2002.

[16]Z. Zhang, "A flexible new technique for camera calibration," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 11, pp. 1330–1334, 2000.

[17]R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK., 2003.

[18]T. S. Huang and O. D. Faugeras, "Some properties of the E matrix in two-view motion estimation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 12, pp. 1310–1312, 1989.

[19]J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.

[20]O. D. Faugeras and F. Lustman, "Motion and structure from motion in a piecewise planar environment," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 2, no. 03, pp. 485–508, 1988.

[21]Y. Ma, S. Soatto, J. Košecká, and S. Sastry, *An Invitation to 3D Vision*. Springer-Verlag, New York, 2004.

[22]P. H. Torr and A. Zisserman, *Vision Algorithms: Theory and Practice*, ser. Lecture Notes in Computer Science. Springer-Verlag, September 1999, no. 1883, ch. Feature based methods for structure and motion estimation, pp. 278–294.

[23]H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[24]J. Bauer, N. Sunderhauf, and P. Protzel, "Comparing several implementations of two recently published feature detectors," in *Proc. of the International Conference on Intelligent and Autonomous Systems*, vol. 6, no. pt 1, 2007.

[25]H.-Y. Shum and R. Szeliski, "Systems and experiment paper: Construction of panoramic image mosaics with global and local alignment," *International Journal of Computer Vision*, vol. 36, no. 2, pp. 101–130, 2000.

[26]D. Ruprecht, R. Nagel, and H. Müller, "Spatial free-form deformation with scattered data interpolation methods," *Computers & Graphics*, vol. 19, no. 1, pp. 63–71, 1995.

[27]M. A. Oliver and R. Webster, "Kriging: a method of interpolation for geographical information systems," *International Journal of Geographical Information System*, vol. 4, no. 3, pp. 313–332, 1990.

[28]C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote sensing of Environment*, vol. 8, no. 2, pp. 127–150, 1979.