Three-dimensional hyperspectral imaging technique

Jörgen Ahlberg^{a, b}, Ingmar G. Renhorn^{b, c}, Tomas R. Chevalier^d, Joakim Rydell^e, and David Bergström^e

^aComputer Vision Laboratory, Dept. of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden. www.cvl.isy.liu.se

^bGlana Sensors AB, Tegskiftesgatan 291, SE-583 34 Linköping, Sweden. www.glana.se

^cRenhorn IR Consultant AB, Odalgatan 8, SE-583 31 Linköping, Sweden

^dScienvisic AB, Diskettgatan 11A, SE-583 35 Linköping, Sweden. www.scienvic.se

^eFOI, Swedish Defence Research Agency, SE-164 90 Stockholm, Sweden. www.foi.se

ABSTRACT

Hyperspectral remote sensing based on unmanned airborne vehicles is a field increasing in importance. The combined functionality of simultaneous hyperspectral and geometric modeling is less developed. A configuration has been developed that enables the reconstruction of the hyperspectral three-dimensional (3D) environment. The hyperspectral camera is based on a linear variable filter and a high frame rate, high resolution camera enabling point-to-point matching and 3D reconstruction. This allows the information to be combined into a single and complete 3D hyperspectral model. In this paper, we describe the camera and illustrate capabilities and difficulties through real-world experiments.

Keywords: hyperspectral, remote sensing, 3d

1. BACKGROUND

Multispectral and hyperspectral imaging systems deliver valuable products to users in areas such as agriculture, mining, environmental monitoring, disaster assessment, and military reconnaissance. The rapid development toward low weight and affordable systems opens up a multitude of small scale applications using both ground-based and miniature unmanned aerial vehicle (UAV)-based systems. As the typical payload for mini-UAVs is 2 to 3 kg, a design goal for a hyperspectral imaging system must also be to make this technology adaptable to these smaller sensor platforms.

The optimal trade-off between spatial and spectral resolution depends on target range and target size as well as the spectral variability of the target and background materials. Solid materials mostly exhibit slow spectral variations and can, therefore, be more sparsely sampled. An example in which high spatial resolution is needed is the detection of intercropped plants, e.g., cannabis with maize, or in the detection of plastic litter in coastal regions.

Conventional hyperspectral sensors tend to be rather bulky and heavy. The wavelength separation mechanism often relies on dispersive or interferometric elements in combination with several sets of collimating and reimaging optics. New more compact systems are being developed to meet the requirements of low size, weight, power consumption, and cost, which includes both push-broom solutions and snapshot systems. In the design used in

Further author information: (Send correspondence to J.A.)

J.A.: E-mail: jorgen.ahlberg@liu.se, Telephone: +46 706 757 384

I.G.R.: E-mail: ingmar.renhorn@telia.com, Telephone: +46 706 91 31 96

D.B.: E-mail: david.bergstrom@foi.se

T.R.C.: E-mail: tomas.chevalier@scienvic.se

J.R.: E-mail: joakim.rydell@foi.se

this paper, high spatial resolution is achieved by mounting a linear variable bandpass filter (LVF) on top of a large focal plane array (FPA).

At the same time, the need for 3D mapping has increased, and it is therefore common to complement hyperspectral airborne mapping with a laser scanner. Combining data from the two sensors, a three-dimensional hyperspectral map is created. In recent years, methods for so-called passive 3D, which uses "standard" cameras for estimating 3D structure, have been introduced. However, such methods cannot be used with many existing hyperspectral cameras, as these observe one line at the ground, not an entire image. Thus they do not see a point on the ground from several angles, which is required for the estimate of the 3D structure.

For high spatial resolution hyperspectral imaging to be useful for remote sensing, accurate navigation is required. For push-broom systems, this typically results in a need of expensive inertial navigation systems. An alternative approach is to determine the egomotion using a camera attached to the platform, which with the proposed system is not needed. The pose of the sensor is incrementally determined from the changes in the images caused by the motion of the platform. In order to work well, the scene has to be well illuminated and there also has to be a large scene overlap between images in order to work well. Image based navigation is highly accurate with relative errors often less than 1% which makes it a viable alternative to global positioning systems and inertial navigation systems. This is especially important in GPS-denied environments.

In this paper, we demonstrate how the LVF-based sensor system can be used for simultaneous acquisition of 3D and hyperspectral data.

2. THE PROPOSED SENSOR AND METHOD

The proposed camera is an electro-optical sensor consisting of a sensor chip with a focal plane array (FPA) and read-out electronics, a linear variable filter (LVF), and optics. The sensor chip and optics can be standard components from existing cameras, such as high-end DSLRs or machine vision cameras. The LVF is attached on, or in the direct proximity to, the FPA. Alternatively the LVF could be manufactured as a layer of the FPA. In our prototype, we use a DSLR and modify it by removing the Bayer filter and adding the LVF. The sensor system is described in more in detail by Renhorn et al.¹

The LVF is a bandpass filter letting light pass only in a narrow wavelength band, centered at a wavelength λ_c . This center wavelength varies over the filter, so that the center wavelength is a function of the position (u, v) (pixel coordinates), see Figure 1, that is $\lambda_c = \lambda_c(u, v)$. The center wavelength varies continuously along one of the dimensions (here called u) of the filter, so that the center frequency is a function of u only, that is $\lambda_c = \lambda_c(u)$. Also, in the current implementation, $\lambda_c(u)$ is a linear function, and thus the filter is said to be a linear variable filter (LVF). However, for the purposes of our method, the filter is neither required to be linear or one-dimensional as long as it is known and continuous. Examples of views through the LVF are given in Figure 1b.

As the LVF is mounted on, or very close to, the FPA, the light registered by a sensor element at position (u, v) will only contain wavelengths close to $\lambda_c(u, v)$. When the camera and the observed surface are static, each point on the object will thus be observed in a specific wavelength. By rotating the camera, it can be used as a spectrometer, as described by Renhorn et al.¹ and Lundberg et al.²

When there is relative motion between the sensor and the surface, the projection by the optics of a certain point $\mathbf{p}$ on the object (that is, a point in the 3D world) will move across the FPA. That is, assume a point that is first observed by the sensor at one position and then by the sensor at another position. When the sensor at the first position observes the point $\mathbf{p}$, it will be projected on a sensor element at position (u_1, v_1) on the FPA and the light is filtered by the LVF at that same position. When the sensor at the second position observes the point $\mathbf{p}$, it will be projected on a sensor element at position (u_2, v_2) on the FPA and the light is filtered by the LVF at that position. When a point $\mathbf{p}$ is thus observed by different sensor elements, a spectral signature $\mathbf{S}(\mathbf{p})$ can be estimated for each such point $\mathbf{p}$. For an LVF, the vector $\mathbf{S}(\mathbf{p})$ has the same number of elements as the number of sensor elements along the u -direction on the FPA, so that a measurement on the i :th column on the FPA corresponds to the i :th value in the vector $\mathbf{S}(\mathbf{p})$. For a high resolution FPA, with thousands of sensor elements along each column, this requires large amounts of memory and processing power, as the vector $\mathbf{S}(\mathbf{p})$ would contain thousands of values for each tracked point $\mathbf{p}$. Note that this large amount of data is usually not needed

(a) The LVF mounted on the focal plane array (FPA).

(b) LVF examples. Left: View through the LVF of a white source using a camera with Bayer filter present but infrared cut-off filter removed. Middle: White screen in daylight. Variation in solar spectrum can be observed. Right: Green hedge in daylight. The green peak can be observed as well as the low reflectance in red and high reflectance in near infrared.

Figure 1: Linear variable filter (LVF).

– only the number of samples in corresponding to the sought spectral resolution – however, in the processing stage, spatial and spectral resolution are given by the same data, and cannot be downsampled initially.

3. ESTIMATION OF 3D STRUCTURE AND CAMERA PARAMETERS

In order to estimate 3D structure, we follow two different paths depending on the scenario and application. The principal difference is whether key features are *tracked* or *matched*.

In the general case, when for example flying with a drone around a building of which the 3D structure is going to be reconstructed, methods based on extraction and matching of image features, such as SIFT³ or SURF,⁴ are used. Using these image matches, the intrinsic camera parameters (focal length, aspect ratio, etc.) are estimated once for all images in the acquired data set, and the extrinsic camera parameters (position, rotation) are estimated once for each image. We use commercial software from the company Spotscale AB* to compute a 3D model of the building as well as the camera parameters. The drawback with the general method is that whereas it is robust and allows us to find many observations of a certain point on the studied object (such as a building), the precision is not always high enough to give us the wanted certainty in the following spectral reconstruction. The output from the 3D reconstruction is a wireframe model with $\mathbf{M}$ vertices (3D points) and associated surface normals $\{\mathbf{p}_i, \mathbf{n}_i\}_1^{\mathbf{M}}$ and $\mathbf{N}$ camera parameter sets $\{\mathbf{P}_j\}_1^{\mathbf{N}}$ (one set of camera parameters for each image $\mathbf{I}_j$). For details, see, for example, the textbook by Szeliski.⁵

In a more constrained case, such as flying over mostly flat ground or inspecting objects on a conveyor belt, we estimate the 3D structure as a depth map. From the motion of each observed point as it moves through the image sequence, its depth (that is, how much it extrudes from the ground/belt plane) can be computed. A requirement is of course that those points can be tracked with high precision through the sequence, which is more difficult when the observed wavelength varies with the image position.

In the following, we will focus on 3D structure estimation using matching, that is, the general case.

*www.spotscale.com

4. ESTIMATION OF SPECTRAL SIGNATURES OF 3D POINTS

In the general case mentioned above, that is, when the camera parameters are estimated for each image, we need a mapping from cameras to spectra. The procedure is as follows:

Assume that we want to reconstruct the spectral vector of a point $\mathbf{p}_i$ on the 3D model given by the 3D reconstruction software mentioned above. Using the common terminology from the computer vision community, each estimated set of camera parameters is denoted as one camera, defined by the camera matrix $\mathbf{P}_i$ (see below). We search the set of $\mathbf{N}$ estimated cameras and find the ones where that point is within the field of view and where the surface normal $\mathbf{n}_i$ is pointing towards the camera. For each such camera, we can compute the image coordinates of the point, and retrieve the pixel value from the corresponding image. We remind that the image coordinate also maps to the wavelength at which the point was observed.

That is, the j :th camera, used to acquire the image $\mathbf{I}_j(u, v)$, is given by the camera matrix

$$\mathbf{P}_j = \mathbf{K} [\mathbf{R}_j | \mathbf{t}_j], \quad (1)$$

where $\mathbf{K}$ contains the intrinsic parameters and $\mathbf{R}_j$ and $\mathbf{t}_j$ the extrinsic parameters. The intrinsic parameters are focal length, optical center, etc., and needs only to be estimated once per physical camera. The extrinsic parameters consist of a rotation matrix and a translation vector, together defining the camera pose. For more details, see a textbook on computer vision, such as Hartley and Zimmerman⁶ or Szeliski.⁵

The image coordinates (u_{ij}, v_{ij}) of a 3D world point $\mathbf{p}_i = (x_i, y_i, z_i)$ are given by[†]

$$\begin{bmatrix} X_{ij} \\ Y_{ij} \\ Z_{ij} \end{bmatrix} = \mathbf{P}_j \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} \quad (2)$$

$$\begin{bmatrix} u_{ij} \\ v_{ij} \end{bmatrix} = \frac{1}{Z_{ij}} \begin{bmatrix} X_{ij} \\ Y_{ij} \end{bmatrix}. \quad (3)$$

Thus, for a 3D point $\mathbf{p}_i$ we can find the subset of $\mathbf{C}_i$ cameras where $\mathbf{p}_i$ is projected onto the image $\mathbf{I}_j$ as

$$\mathbf{C}_i = \{ \mathbf{P}_j : 0 < u_{ij} < U, 0 < v_{ij} < V, \frac{\mathbf{t}_j - \mathbf{p}_i}{\|\mathbf{t}_j - \mathbf{p}_i\|} \cdot \mathbf{n}_i > a \} \quad (4)$$

and get a set of $\mathbf{N}'$ spectral value-wavelength pairs $\{ \mathbf{I}_j(u_{ij}, v_{ij}), \lambda_c(u_{ij}) \}$. From this set we can interpolate the wanted spectral vector. The threshold a ensures that the surface normal is pointing towards, not away from, the camera ($a = 0$) and could also be used to ensure a suitable observation angle ($0 < a < 1$).

Note that for estimating the spectral vector of a point, we thus need to search the set of cameras and then access a given subset of images in order to extract one (or a few, if we do some interpolation) pixel values for each. This is a very time consuming operation, and we thus prefer to do part of it off-line; we traverse all the $\mathbf{M}$ points and for each point we search the $\mathbf{N}$ cameras and store a list of cameras that observe that particular point as well as the image coordinates. Then, when a spectral vector for a 3D point is asked for, a chosen subset of images can be accessed and the spectral vector created at the wanted spectral resolution.

Given a set of value-wavelength pairs, the spectral vectors should be interpolated. Note that these pairs can be very irregularly sampled depending on the flying pattern during the acquisition, and the interpolation should be done with some care. In this paper, we use simple interpolation methods (splines or linear) in order to visualize results, but this will be further elaborated in the future.

5. EXPERIMENTS

We have performed three data acquisition experiments with a prototype camera. The first by mounting the camera on a car and acquiring images from a building when passing by; the second by mounting the camera on a small electrical drone and flying over a field with reference objects; and the third flying around an office building as shown in Figure 4a.

[†]In practice, the lens distortion is also parameterized and estimated, but omitted here for simplicity.

Figure 2: The first 3D reconstruction test from a drive-by.

5.1 Experiment 1: Car passing building

The purpose of the first experiment was to try out the 3D reconstruction using image matching. The result was a somewhat crude 3D model shown in Figure 2. The main purpose of the experiment was to see if the software and algorithms designed for common RGB imagery would work well enough for this kind of imagery, and to develop the software for the recovery of spectral signatures of the 3D model points (as described above). Several difficulties were revealed, in hindsight somewhat obvious that they would appear, such as robust point matching on homogeneous surfaces, handling of sharp corners (building corners), and partial occlusion. The latter was a significant problem in this particular scene where trees and parked cars sometimes block the line of sight between the camera and some points on the studied building facade, resulting in spectral peculiarities.

The conclusion from the experiment was that the 3D reconstruction works for this kind of data, and that the next step was to try airborne image acquisition.

5.2 Experiment 2: Drone over field

The purpose of the experiment was to arrange a quite simple scene to test and demonstrate the 3D and hyperspectral reconstruction capability when flying. 3D reconstruction using matching was used, which is troublesome in such an environment due to highly irregular objects (bushes, trees) and ambiguous spatial structure. Example results are shown in Figure 3. As is apparent from the figure, the 3D reconstruction works well, however, the projections of the 3D points on the image planes (that is, the u_{ij}, v_{ij} coordinates) are not estimated with enough accuracy. By manually inspecting the images from the cameras observing a selected point, such as one of the marked points in Figure 3b and plotting the image coordinates, it becomes visible that the point "moves around" a few decimeters. To compensate for that, the estimated spectra in Figure 3c are the median spectra of a few neighboring 3D points, which in practice make the useful resolution of the sensor smaller. As we will show in the third experiment, we can mitigate this by low-level image processing. Another option would be to implement the 3D reconstruction scheme using tracking, as mentioned above.

5.3 Experiment 3: Drone circling building

The third experiment was also performed using a drone. The drone flew several circuits around an office building, at different altitudes, covering all walls with overlapping images. Commercially available software from Spotscale AB was used to create the 3D model, see Figure 4. A reader with very keen eyesight will notice a small flower pot just to the left of the entrance to the building. For the more normal reader, a part of one picture from the airborne hyperspectral camera is shown in Figure 5b and a picture on the same flower pot taken from close distance with an ordinary consumer camera is shown in Figure 5a. In a similar way to above, we can find the images that contain the flower pot, and track one of the small flowers through that set of images (the flowers are high-contrast easily tracked objects). The extracted spectrum of one of the flowers is shown in Figure 5c. Thus, with additional high-accuracy tracking, in this case manually initialized, we can obtain a good spectral estimation.

(a) One of the acquired images of the field.

(b) 3D reconstruction of the field scene seen from two angles. The colors are mapped from the z -coordinate, just to make the 3D-structure visible. Note the three markers (blue, red, green).

(c) Estimated spectra of the example points marked in Figure 3b.

Figure 3: Acquired and computed data from the Experiment 1: Drone over field.

(a) Acquiring image data by flying in circuits around an office building.

(b) Two views of the resulting 3D model created by Spotscale AB, in these pictures shown with texture from an "ordinary" RGB camera.

Figure 4: Acquiring data in Experiment 2: Drone circling an office building.

In order to evaluate the spectral estimation, another experiment was performed. Assuming a uniform surface, covering a large part of the field of view, we can immediately recover the its spectrum by simply reading one line of the image data. Thus, one single image covering the wall of the building we are observing can be used to recover the spectral signature of the bricks, as illustrated in Figure 6. Some care need to be taken to exclude measurements between the bricks; this has been done manually in this experiment.

Then, a part of the brick wall has been studied, see Figure 7, and the spectra of three points (vertices of the 3D model) have been recovered using the methods described above. As shown in Figure 7b, the recovered spectra fit the reference spectrum well.

(a) Close-up picture of the flower pot using a hand-held consumer camera.

(b) The flower pot seen by the airborne hyperspectral camera. The arrow points at one of the flowers chosen for example spectrum extraction.

(c) Estimated spectrum of the flower marked in Figure 5b.

Figure 5: Acquired and computed data from Experiment 3: Office building.

Figure 6: One picture from the drone gives the possibility to recover the spectral signature of the bricks in the wall over a large part of the wavelength range. Using two images, the spectrum for the entire range can be recovered and a polynomial fitted to use as a reference.

(a) The studied part of the brick wall.

(b) Estimated spectra (dots) of three points on the brick wall compared to the reference spectrum (line).

Figure 7: Acquired and computed data from Experiment 3: Office building.

6. CONCLUSION

We have demonstrated the use of a hyperspectral sensor able to recover 3D structure of the observed objects. We have performed real-world experiments to test and verify the capability to recover the 3D structure using the proposed sensor (experiment 1), to estimate the hyperspectral signature of points on the 3D model given by the 3D reconstruction (experiment 2), and to estimate the hyperspectral signature of arbitrary points on the model's surface (experiment 3). We have shown that all these things can be achieved in practical situations, but that there is a significant amount of work before the process is completely automated and with the sought-for accuracy.

ACKNOWLEDGMENTS

This work has partly been funded by the Linköping University Innovation Office and by the Swedish Armed Forces Research & Technology Programme. We also thankfully acknowledge Spotscale AB for their assistance with drone operation and 3D model reconstruction.

REFERENCES

- [1] Renhorn, I. G. E., Bergström, D., Hedborg, J., Letalick, D., and Möller, S., "High spatial resolution hyperspectral camera based on a linear variable filter," *Optical Engineering* **55**(11), 114105 (2016).
- [2] Lundberg, M., Rattfält, S., Gustafsson, D., Axelsson, M., Petersson, H., and Bergström, D., "Blood trace detection using a hyperspectral sensor based on a linear variable filter," in [*Swedish Symposium on Image Analysis*], (March 2017).
- [3] Lowe, D. G., "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision* **60**, 91–110 (Nov. 2004).
- [4] Bay, H., Ess, A., Tuytelaars, T., and Gool, L. V., "Surf: Speeded up robust features," *Computer Vision and Image Understanding* **110**(3), 346–359 (2008).
- [5] Szeliski, R., [*Computer Vision: Algorithms and Applications*], Springer, New York (2010).
- [6] Hartley, R. I. and Zisserman, A., [*Multiple View Geometry in Computer Vision*], Cambridge University Press, ISBN: 0521540518, second ed. (2004).