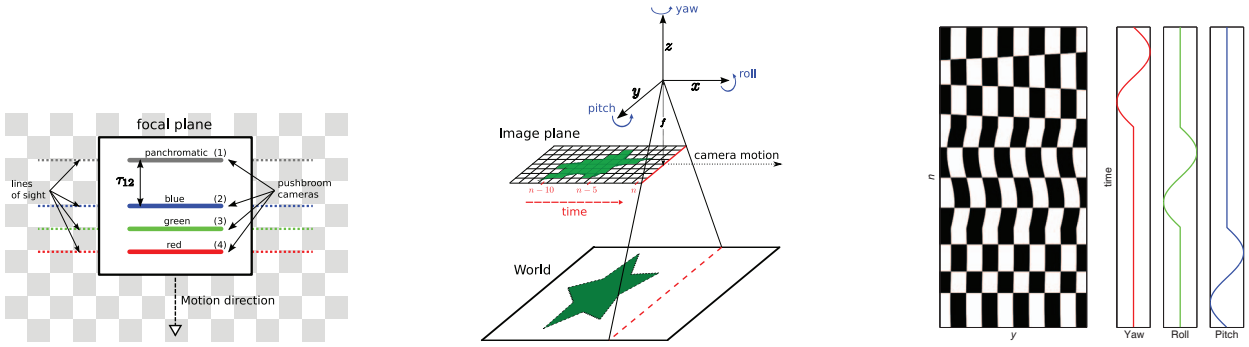
IMAGE-BASED SATELLITE ATTITUDE ESTIMATION

Régis PERRIER[*], *Elise ARNAUD*[†], *Peter STURM*[*], *Mathias ORTNER*[‡]

[*]INRIA Rhone Alpes, [†]Université Joseph Fourier, [‡] EADS Astrium

1. INTRODUCTION

In remote sensing applications for earth observation, pushbroom cameras are prefered to other type of cameras as they provide higher resolution images at a lower cost. For this purpose, spaceborne and airborne imagery combine multiple pushbroom cameras in a same focal plane to acquire multispectral information of the observed scene. Figure (1.a) shows a standard focal plane geometry of an observation satellite.
Concerning its principle, the pushbroom camera is a linear sensor that takes 1-D images at several time instants. The forward motion of the imaging plateform makes the sensor sweeping out a region of space; stitching together all 1-D images gives a complete 2-D image of the observed scene. This acquisition process is summarised in figure (1.b) where we define the three component of the attitude (the yaw, the roll and the pitch).



(a) (b) (c)

Fig. 1. (a) Standard focal plane geometry of an observation satellite with 4 pushbroom cameras: panchromatic, blue, green and red (respectively enumerated as 1,2,3 and 4). What is seen by the camera 2 at time n will be seen by the camera 1 at time $n + \tau_{12}$. (b) Pushbroom camera recording 1-D images over time denoted by n; the camera is moving straight along the x axis, y is defined as the camera axis and z the orthogonal axis to the image plane. We define the orientation of the camera with the yaw (rotation around the z axis), the roll (rotation around the x axis) and the pitch (rotation around the y axis). (c) Example of warps in a regular checkerboard when the pushbroom camera is tilting around its 3 rotation axis.

Stability of the focal plane is a crucial point during the pushbroom acquisition process to keep a constant attitude of the imaging plateform. Otherwise, small attitude variations can lead to noticeable geometric deformations of several pixels in the 2-D image and has to be corrected to properly represent the scene.
In airborne and spaceborne imagery, stability is an issue as the craft is exposed to dynamic disturbances caused by its engines as well as athmospheric and space turbulance. Inertial sensors can be combined with GPS information and Ground Control Points (GCP) to estimate camera orientation and thus rectify the images [1, 2]. However, such sensors are very costly for spaceborne application and besides, do not guarantee that attitude variations are fully estimated as their sampling frequency (4Hz) is lower than the sampling frequency of acquisition of the 1-D image acquisition ($2,500$Hz).
In this paper, we present an original method to estimate attitude variations (also called vibrations [3]) of the pushbroom camera which relies on image information only. To our best knowledge, only one paper describe a solution to estimate attitude variation directly from images. However this method uses local features to estimate warps; it does not take into account the linear

This work was funded by EADS Astrium (European Aeronautic Defence and Space Company).

geometry of the camera and cannot estimate directly an absolute attitude variation. In our case, we exploit the geometry of the focal plane and the stationnary property of the disturbance to register all acquired images in a same coordinate system. The estimation is embedded in a Bayesian setting which allows us to combine a prior model and an image data term to obtain the Maximum a Posteriori estimates (MAP) of the vibrations.

2. ATTITUDE ESTIMATION

In this paper, we denote the set of 4 images (red, blue, green and panchromatic) as $\mathbf{I} = \{I_1, I_2, I_3, I_4\}$. Each image is defined by its pixel coordinates (y, n) in the pixel set $\mathcal{S}$, with $n \in [0, N-1]$ being the discretised time and N the total number of time steps (which is also equivalent to the number of lines in the acquired images). We call $\theta(n) \in \Theta$ the unknown attitude of the satellite at time n; it is a (3×1) vector whose components are respectively the yaw $\theta_y(n)$, the roll $\theta_r(n)$ and the pitch $\theta_p(n)$. We call $\boldsymbol{\theta}$ the $(3N \times 1)$ vector that gathers all attitudes for all time instants.

2.1. MAP formulation

Given the observation images $\mathbf{I}$, we seek to maximise the *a posteriori* probability of the vibrations $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p(\boldsymbol{\theta}|\mathbf{I}) \propto \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \quad \underbrace{p(\mathbf{I}|\boldsymbol{\theta})}_{\text{Image data term}} \quad \underbrace{p(\boldsymbol{\theta})}_{\text{Prior term}} .$$

The first term of equation (1) is usually refered as the data term whereas the second term represents a prior on the vibrations. This last term can also be understood as a penality function which constrains the solution to follow specific properties (depending on the shape of this penality function).

2.2. Image data term

We first assume that all images are radiometrically calibrated; we discuss the way we handle multi-modalities in the results section. Also, we start by comparing two images and will derive a global formulation thereafter.
Let σ_i^2 being the variance of a zero mean Gaussian acquisition noise i.i.d. over all pixels in images. Given I_1 and I_2 spaced by τ_{12} in the focal plane, we expect to have the following relationship:

$$\Big[I_1\big(W(y, n; \theta(n))\big) - I_2\big(W(y, n+\tau_{12}; \theta(n+\tau_{12}))\big)\Big] \sim \mathcal{N}(0, \sigma_i^2) \tag{1}$$

where $W : \mathcal{S} \times \Theta \to \mathcal{S}$ is a warp function that maps pixel coordinates to a new position depending on the attitude of the pushbroom camera. Fixing I_1 as the reference image, equation (1) is equivalent to the following one:

$$\Big[I_1(y, n) - I_2\big(W(y, n+\tau_{12}; \theta(n+\tau_{12}) - \theta(n))\big)\Big] \sim \mathcal{N}(0, \sigma_i^2). \tag{2}$$

Let $*$ denotes the convolution operator. From previous equation, one can notice that:

$$\theta(n+\tau_{12}) - \theta(n)) = (\theta * k_{12})(n), \text{ with } k_{12} = [-1 \quad \underbrace{0 \quad \dots \quad 0}_{\tau_{12}-1 \text{ zeros}} \quad 1]. \tag{3}$$

k_{12} is a kernel built upon the time gap between the acquisition of I_1 and I_2. This basically shows that trying to estimate $\boldsymbol{\theta}$ from any pair of images $\{I_i, I_j\}$ acquired by the focal plane leads to the deconvolution of $\boldsymbol{\theta}$ by a kernel k_{ij}.
In a global formulation, let $I_i^{\tau_{ij}}(y, n) = I_i(y, n+\tau_{ij})$ be the shifted image in time with factor τ_{ij}, $\mathbf{y} = [y, n]$ the vector of pixel coordinates, and K_{ij} the matrix convolution operator with kernel k_{ij}, then the log-likelihood arises immediately:

$$\log(p(\mathbf{I}|\boldsymbol{\theta})) \propto \sum_{i,j;i\neq j} \sum_{\mathbf{y}\in\mathcal{S}} \Big(I_i(\mathbf{y}) - I_j^{\tau_{ij}}\big(W(\mathbf{y}; K_{ij}\boldsymbol{\theta})\big)\Big)^2 + \text{cst.} \tag{4}$$

This last equation is the expression of a pixel-based registration method, also known as the Lucas-Kanade method [4, 5]. The outer sum takes into account all possible pairs of images while the inner one is a summation over all pixels in the images. The additional constant term refers to the normalizing term of the Gaussian p.d.f. which is independent of $\boldsymbol{\theta}$, and thus of no importance for the minimization procedure. Let us stress that this expression relates the *absolute* vibrations to all image data, thus ensuring the global coherence of the estimation with respect to all image pairs. This is opposed to [3] where *relative* vibrations between each pair of images are first estimated, and then filtered with a Wiener deconvolution operator to recover the *absolute* vibrations; this leads to a suboptimal estimate [6].

2.3. Prior term

Variations in the attitude of the satellite are mainly due to its engines. One basic and general assumption is the stationary property of the vibratory signal, meaning that it keeps the values of its harmonic frequencies constant over time. Smoothness prior is usually well adapted to vibrations as we are not expecting to have strong non-stationary variations in the signal.
The Tikhonov prior encourages global smoothness in the solution by penalizing its Euclidean norm:

$$\log(p(\boldsymbol{\theta})) \propto \|\boldsymbol{\theta}\|^2 + \text{cst} \tag{5}$$

The constant term accounts for the normalizing factor of the Gaussian p.d.f.. Due to its linear form with $\boldsymbol{\theta}$, equation (5) is well suited to solve ill-posed least-squares problems; the solution is closed-form and thus easy to compute.

2.4. Global equation

Deduced from eq.(4) and eq.(5), we need to minimise over $\boldsymbol{\theta}$ the following equation:

$$\log(p(\mathbf{I}|\boldsymbol{\theta})p(\boldsymbol{\theta})) = \sum_{i,j;i\neq j} \|I_i(\mathbf{y}) - I_j^{\tau_{ij}}\big(W(\mathbf{y}; K_{ij}\boldsymbol{\theta})\big)\|^2 + \lambda\|\boldsymbol{\theta}\|^2. \tag{6}$$

where λ is a trade-off scalar between the data and the prior term. Due to the non linear form of the image data term, this expression has to be minimised iteratively. In pratice, we use a Gauss-Newton algorithm to estimate $\boldsymbol{\theta}$ for its good performance in registration methods [5].

One should notice that using a gradient descent algorithm does not require the knowledge of an analytical form of the warp W as long as derivatives of the warp $\frac{\partial W}{\partial y}$ and $\frac{\partial W}{\partial n}$ are available. In our experiments, we use numerical derivatives of W computed using known calibration pattern on earth. This has several advantages as Digital Elevation Model (DEM) and optical distortions of the lens can be directly added to those numerical derivatives of W.

3. RESULTS AND DISCUSSION

We present experimental results on two satellite datasets to show performance of our algorithm. They both have been simulated by ASTRIUM, meaning that the ground truth on vibrations is available. The simulation process creates real-life condition data acquisition and can be considered as difficult as could be real data.
Each dataset is composed of 4 multispectral images: green, red, blue and panchromatic. All images are of size (2564×900) pixels; the pushbroom sensor size being 900 pixels. For all experiments we use a Matlab implementation on a Core2 duo at 3GHz with 3.8GiB, and obtain a computational time of approximatively 200 seconds.
We choose to handle multi-modalities using a linear model to correct radiometric differences between images. This method showed slightly better results than filtering methods; this can be linked to the fact that panchromatic and RGB modalities are quite well correlated by default.

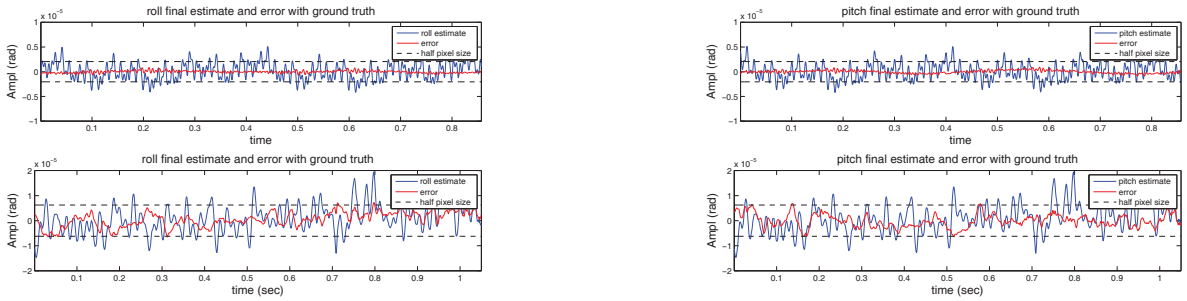


Fig. 2. The top and bottom figures show respectively results for the first and the second dataset. The curves are estimates of attitude variations in radian for the roll (left) and pitch (right) compared to ground truth and depending on the time of acquisition; the error is the substraction between the estimation and the real attitude.

The first dataset is a typical case where our algorithm gives its best performance. 5 iterations are needed to obtain the final MAP estimates of attitude vibrations. Figure 2 shows the error on the vibrations estimation which is negligible with a standard

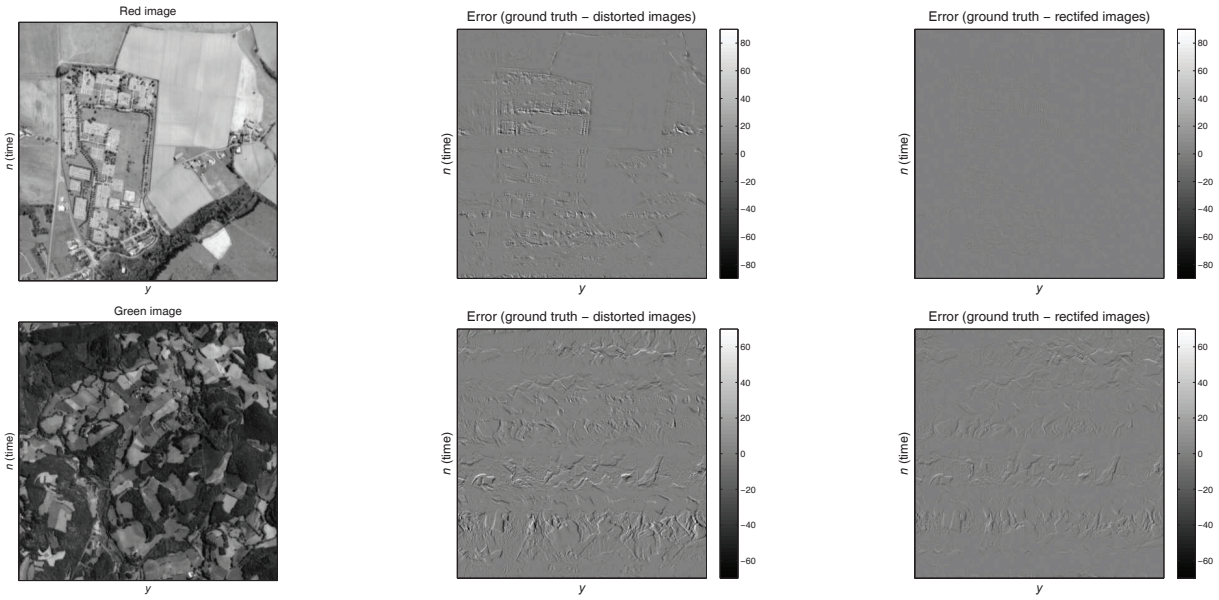


Fig. 3. The top and bottom figures show respectively extracted patches of (400×400) for the first and second dataset. The error images are the differences of intensity between the real image and the distorted image (the one which is acquired) in the middle, and the differences between the real image and the rectified image on the right.

deviation below $\frac{9}{100}$ in pixel unit.
The second dataset is a challenging case as the vibration signal has a low frequency component. The algorithm converged in 5 iterations. In this case, the error between the ground truth and the estimate reveals low frequency variations which were not restored during the registration process (the standard deviation is below $\frac{21}{100}$ in pixel unit). Due to the convolution kernels k_{ij}, most of the low frequency information is lost and the deconvolution process becomes highly ill posed. Thus, the Tikhonov prior helps regularizing the estimate in those frequencies.
These experimental results demonstrate the performance and limits of our algorithm, the latter appearing mainly in the low frequency estimation of attitude variation. Also, the method presents very good results in high frequency estimation, which can not be obtained by the use of inertial sensors alone, as they only provide low frequency information. Our next step will be to fuse measurements from images and inertial sensors so as to get the best estimate of attitude variations for all frequencies. The Bayesian setting we used easily allows us to combine prior as well as observations by other sensors in a coherent manner.

4. REFERENCES

[1] Daniela Poli, "General model for airborne and spaceborne linear array sensors," in *Int. Archives of Photogrammetry and Remote Sensing*, 2002, vol. 34.

[2] Michael Cramer, Dirk Stallmann, and Norbert Haala, "Direct georeferencing using GPS/inertial exterior orientations for photogrammetric applications," in *Int. Archives of Photogrammetry and Remote Sensing, Vol. 33*, 2000, pp. 198–205.

[3] F. de Lussy, D. Greslou, and L. Gross Colzy, "Process line for geometrical image correction of disruptive microvibrations," in *Int. Society for Photogrammetry and Remote Sensing*, 2008, pp. 27–35.

[4] Bruce D. Lucas and Takeo Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81)*, 1981, pp. 674–679.

[5] Simon Baker and Iain Matthews, "Lucas-Kanade 20 years on: A unifying framework," *International Journal of Computer Vision*, vol. 56, no. 1, pp. 221 – 255, 2004.

[6] S. Farsiu, M. Elad, and P. Milanfar, "Constrained, globally optimal, multi-frame motion estimation," in *Proc. of the 2005 IEEE Workshop on Statistical Signal Processing*, 2005, pp. 1396 – 1401.