Discussion of Image Geometry and Transformations

David L B Jupp
October 2018

1. Aligning image geometry using control points

Geometric operations on images take a number of forms. In most cases they seek to “align” one image with another and in others to modify the image to remove geometric “distortions”. These are not really very different operations in practice. To accomplish this the geometry is modelled using statistics of control points. Finally, the images can be transformed into new geometries and scales based on the match of points.

It is a good idea to distinguish “parameterised” transformations from “alignment” transformations. For example, a map projection is a parameterised transformation while an affine transformation that aligns an image with the (x,y) coordinates of a map using control points is an alignment transformation.

The situation can be illustrated by three cases:

Example 1

Take an example of a scanned image of an old map: It is a representation of a part of the earth. Since the earth surface is a roughly spherical the flat representation has used a map projection – in this case a conic projection with circular parallels and radial meridians:

The lines of constant latitude and longitude are not as they are on a sphere nor in the form of the (x,y) grid (Euclidean grid) which is best used for the filtering and other operations we will use for the theory of Linea Systems. To change the image of the map correctly to another specific projection means you need to know its scale and the parameters of its projection – as well as the way the scanned image relates to the map.

In general, a map is related to the Earth surface by a specific model such as:

$$x = f_x(\lambda, \varphi)$$
$$y = f_y(\lambda, \varphi)$$

The (x,y) values are in metres from a specific origin on the earth surface. The scale is the relationship of these metres in the projection to cm on the scanned map. But you may not know that scale. If you set up a map coordinate system for a position (i,j) (sample and line) in the image that agrees with the position of the pixel located on the map in its projection you can write (assuming the scanner is metric):

$$x'_i = x_0 + h_x i$$
$$y'_j = y_0 + h_y j$$

This assumes a simple shift of origin and scaling in x and y – so no rotations or non-linear distortions.

If a set of positions (i,j) are matched to (eg) crossings of the graticule in the map ($k = 1, n$) then it follows that:

$$x'_k = x_0 + h_x i_k$$
$$y'_k = y_0 + h_y j_k$$

and

$$x_k = f_x(\lambda_k, \varphi_k)$$
$$y_k = f_y(\lambda_k, \varphi_k)$$

The matching of image and map then resolves to finding map parameters and image alignment shifts and scale factors such that (for example):

$$RMS = \left(\sum_{k=1}^{n} \left(\left(x'_k - x_k \right)^2 + \left(y'_k - y_k \right)^2 \right) / 2n \right)^{1/2}$$

Is as small as possible.

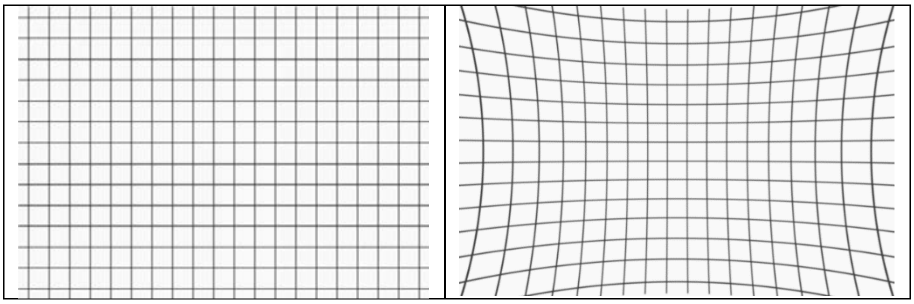
If this is done for the map of the example, the image may be re-projected (re-sampled) to a map in which the lines of latitude and longitude are now parallel:

Or it may be re-projected to match another map or be draped over a spherical earth in Google Earth etc.

Example 2

In Castleman there is an example of the re-projection of a hemispherical photograph to a "flat" standard test card photograph which we will discuss below. All lens and optical systems have some distortion from a metric image. Even electronic scanning systems can project with distortions as in the common "pincushion" distortion a TV set may have. It can be reduced using the controls on the set – or by the set itself automatically when a re-calibration is made.

For example, in a blog for people who use MTF Mapper software, regarding modelling such distortions, (http://mtfmapper.blogspot.com/2017/) the writer provided two model images – one with a metric set of line crossings and the other with the appearance of these lines in a pincushion distorted display.



The model used for the distortion has the form of an alignment model and a distortion model in the words of the writer:

"P(x, y, z) is projected onto the image plane at position p(x, y) as governed by the focal length f, such that:

$$p_x = (P_x - C_x) \times \frac{f}{(P_z - C_z)}$$
$$p_y = (P_y - C_y) \times \frac{f}{(P_z - C_z)}$$

where C(x, y, z) represents the centre of projection of the lens. We can express the point p(x, y) in polar coordinates as p(r, theta), where $r^2 = p_x^2 + p_y^2$; the angle theta is dropped, since we assume that the radial distortion is symmetrical around the optical axis.

Given this description of the pinhole part of the camera model, we can then model the observed radial position r_d as:

$$r_d = r_u \times F(r_u)$$

where the function $F(r_u)$ is some function that describes the distortion, and r_u is the undistorted radial position. Popular choices of $F(r_u)$ include:

- Polynomial model (simplified version of Brown's model), with $F(r_u) = 1 + k_1 \times r_u^4 + k_2 \times r_u^4$
- Division model (extended version of Fitzgibbon's model), with $F(r_u) = \dfrac{1}{1 + k_1 \times r_u^4 + k_2 \times r_u^4}$

Note that these models are really just simple approximations to the true radial distortion function of the lens; these simple models persist because they appear to be sufficiently good approximations for practical use.”

There are free parameters here to be used to model the distortion and fix it if needed. The blog does NOT match control points of actual images but rather models the distortion as above. In a future course with a Lab, examples like this can be discussed. In the meantime you are welcome to try and model such distortions.

This situation here is similar to example 1. In the case of the problem described by Castleman, how do you carry out such modifications? With a lens system, the answer is to photograph an image such as the one on the left and use the change in locations of lines or crossing points to model the distortion. It leaves a question of whether to photograph the test image (the grid of lines) with the given lens as well as a metric lens or just try and relate the photograph to the test image directly? In the example the distortion is modelled by photographing the test card and matching card and photograph. The alternative can be discussed.

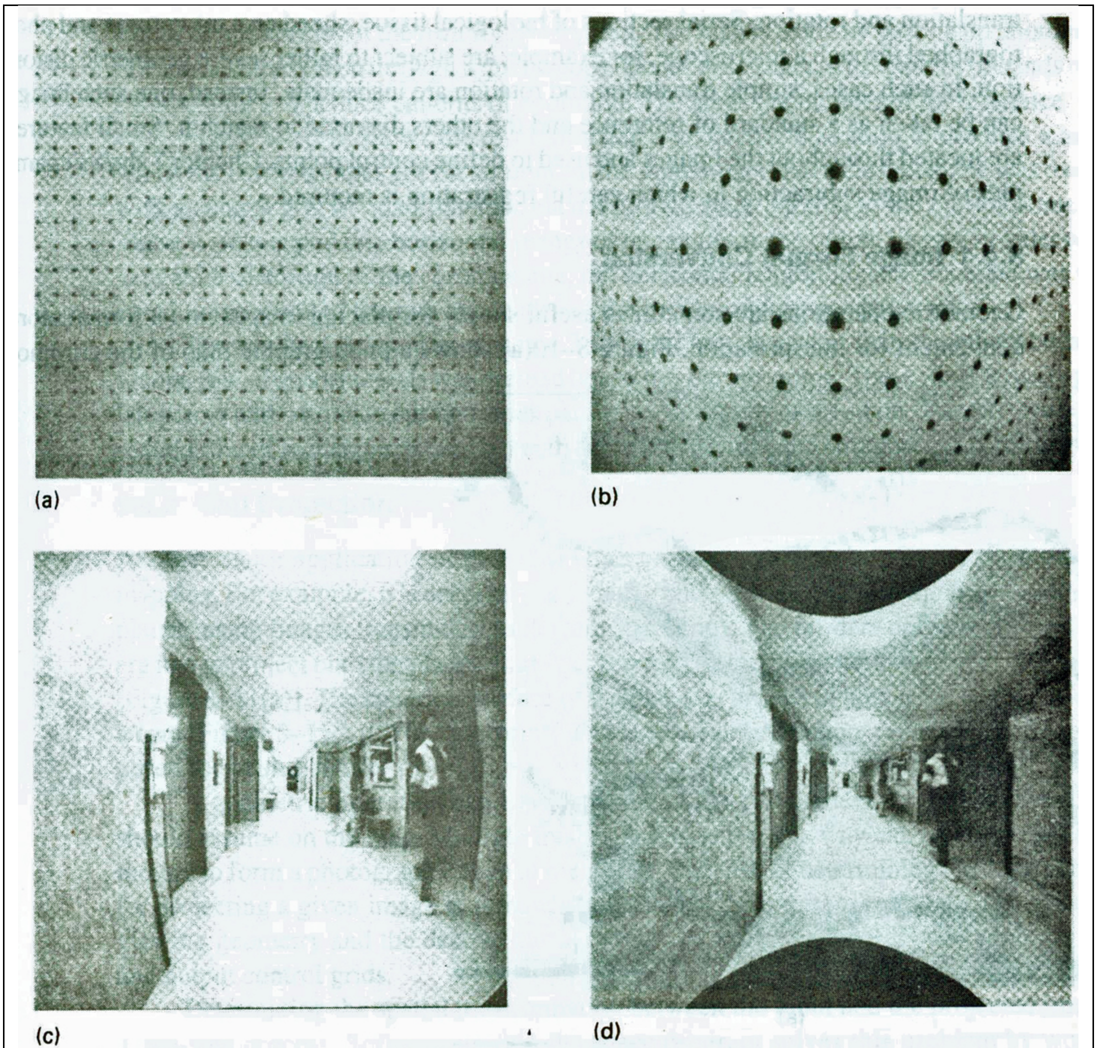
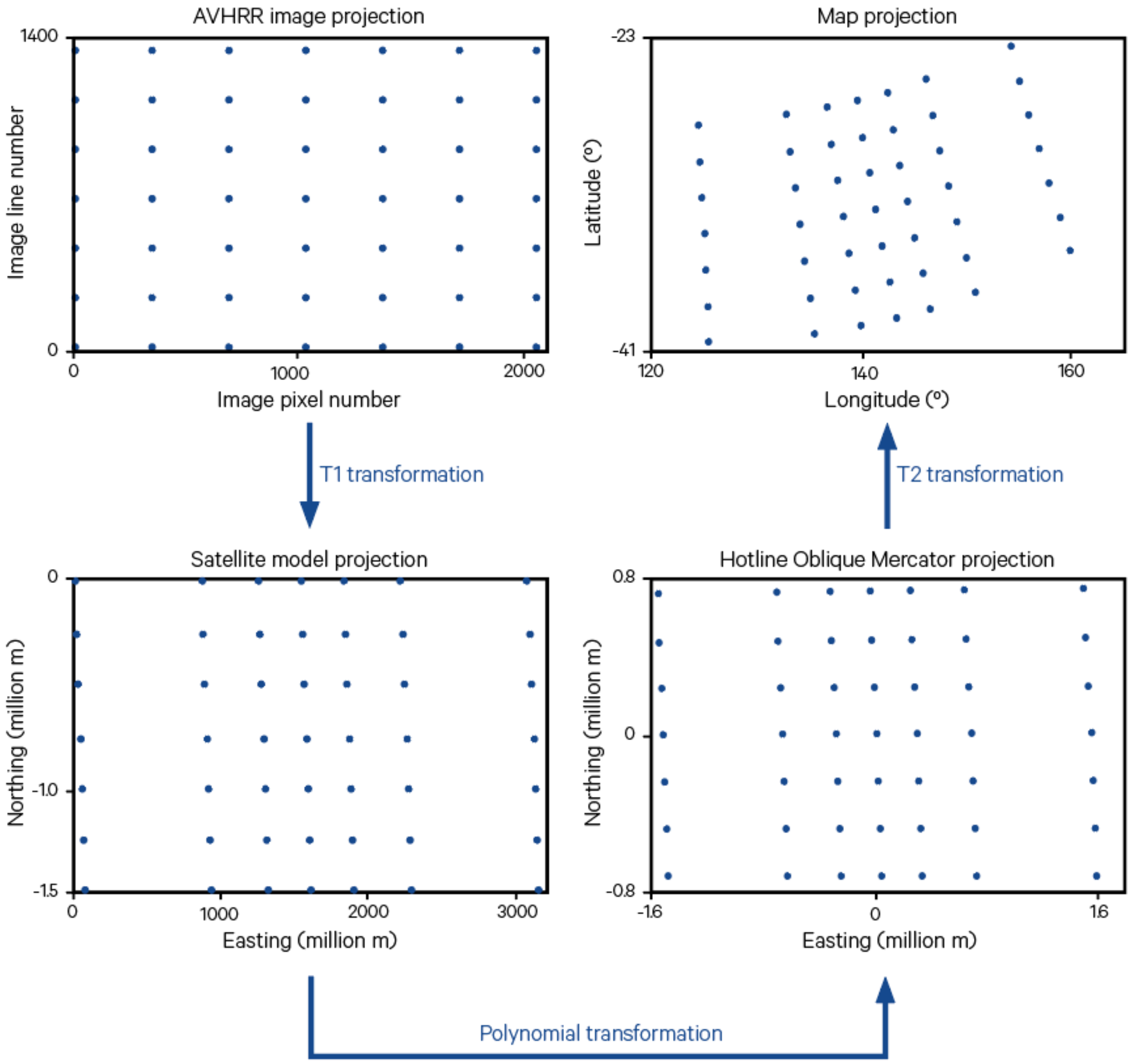


Figure 8–9 Geometric rectification of an image taken with a fish-eye lens: (a) test target, (b) fisheye image; (c) original, (d) rectified hallway image (Courtesy Shishir Shah, The University of Texas at Austin, from [20])

In the Castleman example, the image (c) is the hemispherical (or fish-eye) photograph of a room. The person with the camera also took a photograph (b) of a square grid of dots (a). The method for correcting the image is to match points of (a) with points of (b), to set up an alignment model of the two frames (a) and (b) and fit a model distortion to obtain the camera and lens parameters. The distortion model used could be obtained from the paper referenced by Castleman or other models investigated in Hughes et al. (2008). The models used by the MTF Mapper software are also relevant here.

Example 3

As a third example. In Harrison et al. (2018) is an example of how image transformations can be used to make the alignment model as simple as possible. At the top are a set of (x,y) coordinates in what is taken to be an AVHRR image. On the top right is the distribution of the points in a geographic projection. It is possible to set up a simple scan and scale model for the satellite data (including distortions of view geometry). It is also possible to represent the target image in a Hotine Oblique Mercator (or Space Oblique Mercator) projection where the geometries of the points are well matched.

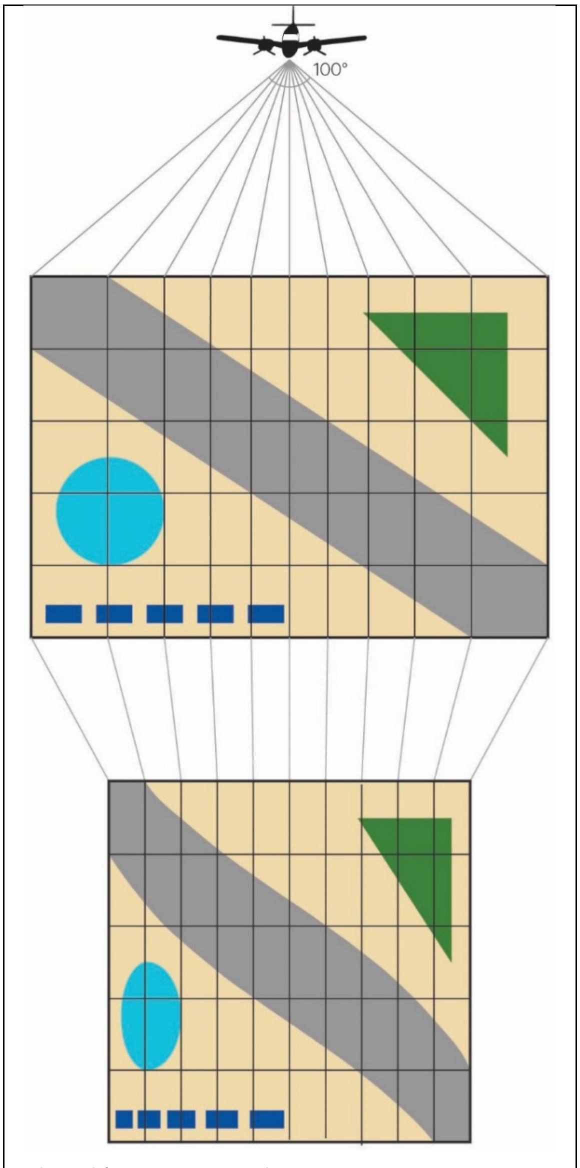


In this case the control points will be to find points in the image and map or orthophoto (or Google Earth) in both source image and a target image, set up alignment and parameter models and fit one to the other using control points. Finally, one or more images are transformed into a different aligned geometry.

Example 4

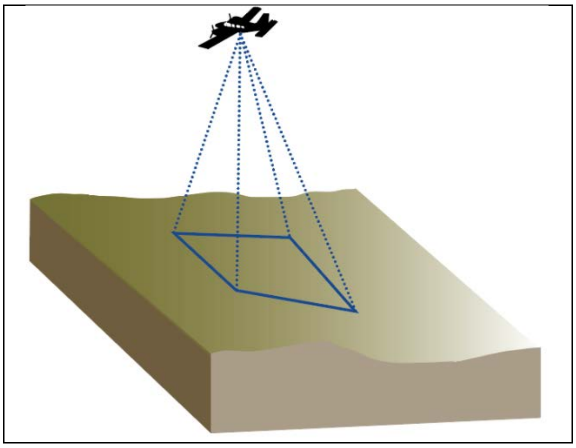
Example 4 is the geometry of images from an airborne scanner. Sometimes these can be very badly distorted. However, if information is collected at the same time from an inertial navigation system (INS) and a GPS, the image can be resampled to a much less distorted form before using ground control and a DEM to obtain ortho-corrected images.

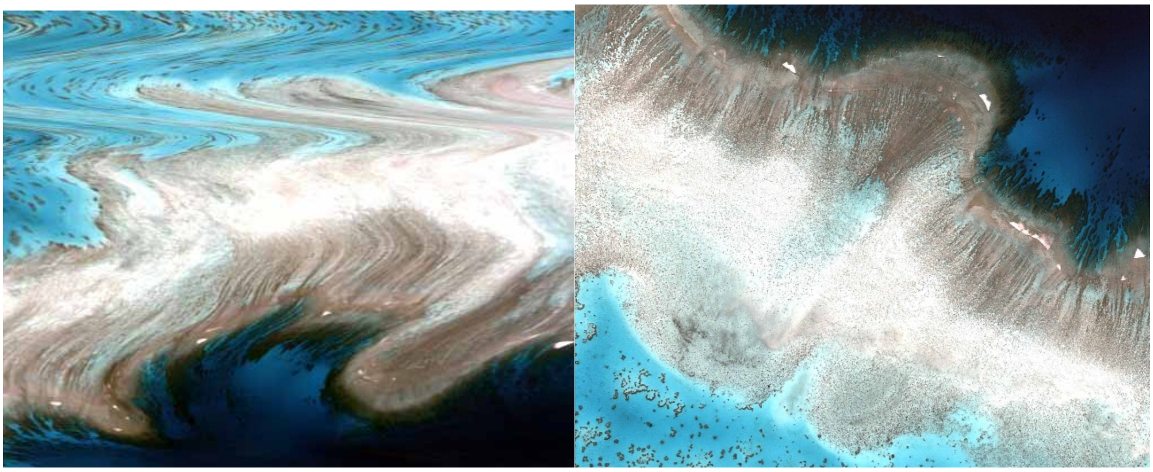
Even if the airborne platform is stable, the imagery will be distorted by the scan geometry:



Adapted from: Harrison and Jupp (1992) Figure 18

The “S-bend” effect is clear in this example. However, if the aircraft rolls and skew with the wind or difficult transects the resulting images can be much worse. For example the following CASI data is from a reef of Australia’s Great Barrier Reef:



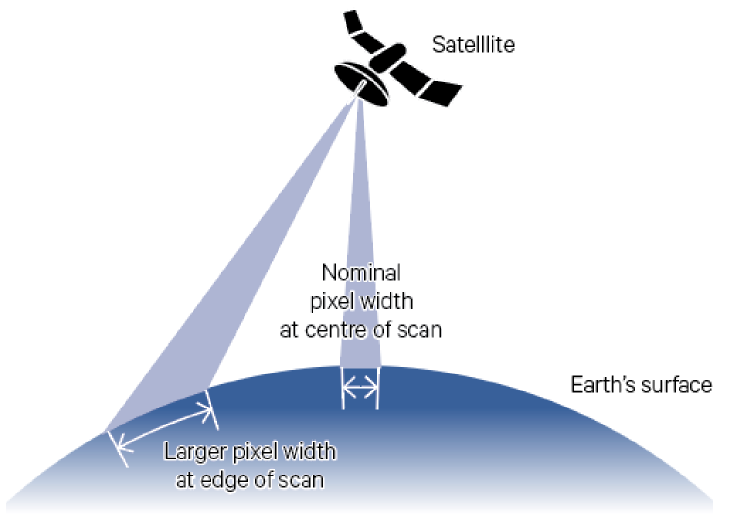


Background image: CASI airborne imagery of Heron Island, Queensland, acquired in June 2002 (see Excursus 2.1). A portion of the original image is shown on the left and the rectified image, resampled to 1 m grid, is shown on the right.

Source: © Nearmap

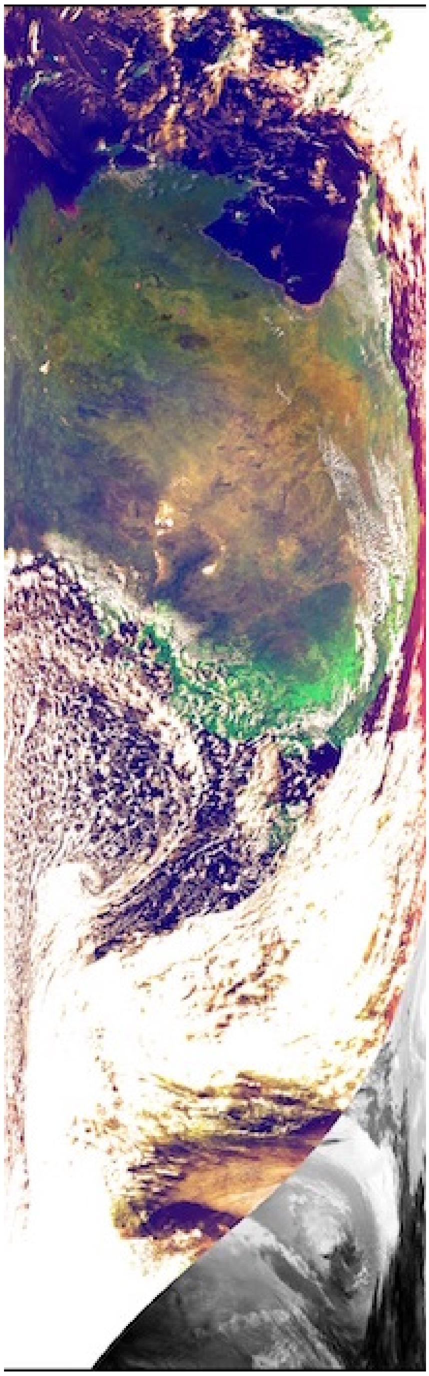
In this case, only data on the pitch, roll and yaw of the aircraft and from the GPS plus knowledge of the scanner geometry are used to correct the data. The resampling puts North at the top as well. The final product can be rectified using GCPs or registered with another map much more easily than directly from the image on the left.

Constructing scanner and platform models for satellite data is also very important. Some notes on the Landsat satellite model are provided in case people have an interest but the benefits of using the satellite model of the MODIS satellite to first present the data in a suitable geometry and then using GCPs and/or correlation patches with or without ortho-rectification are clear:



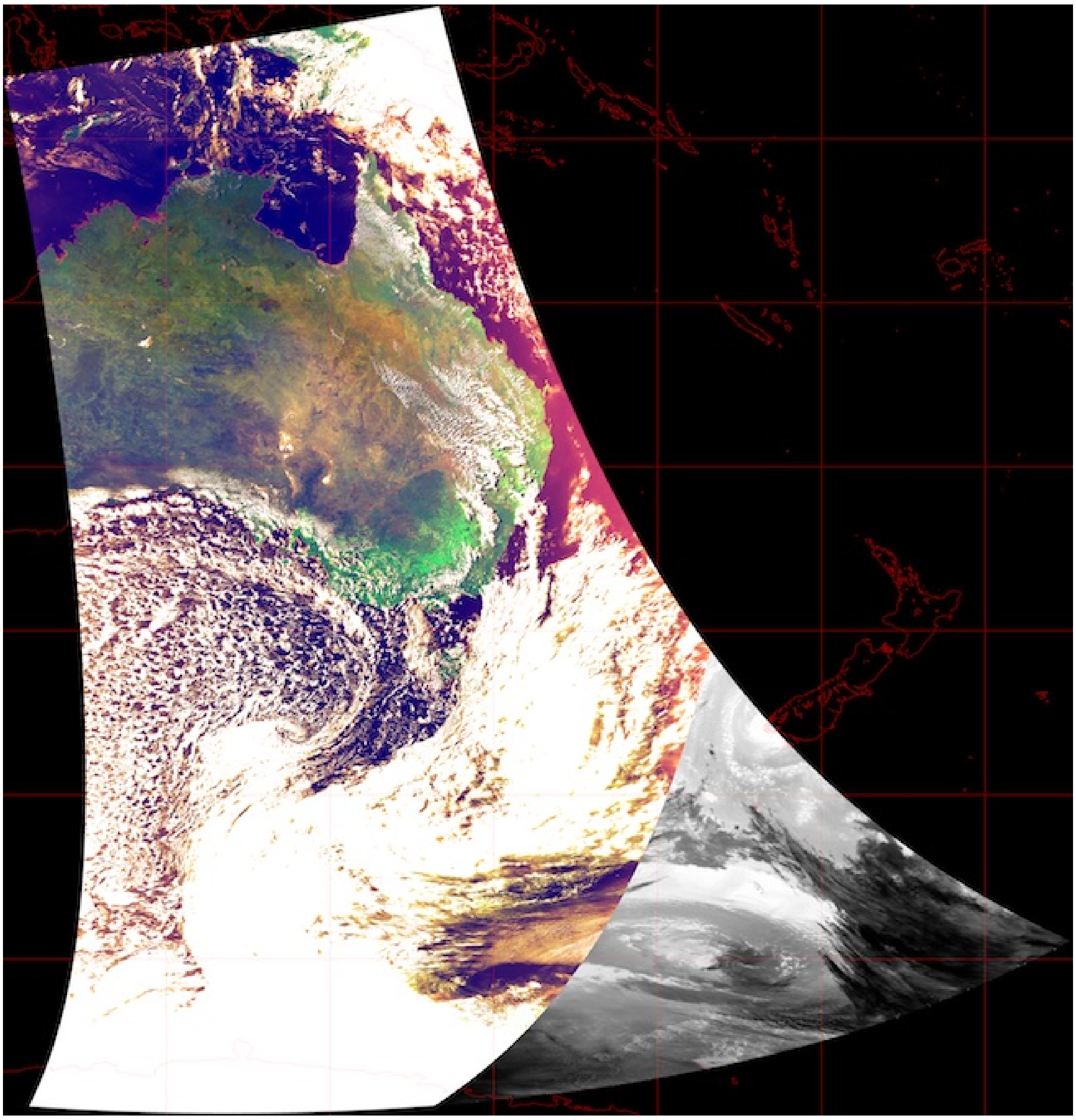
Source: Harrison and Jupp (1992) Figure 17

a. Original image swath (after flipping left to right)



The satellite is in ascending mode and the wide angle scanning induces panoramic distortion. It is only in the “satellite projection”.

b. Rectified Image swath resampled to 0.01° grid (~1 km pixel size) and aligned with geographic projection.



Source: Edward King, CSIRO

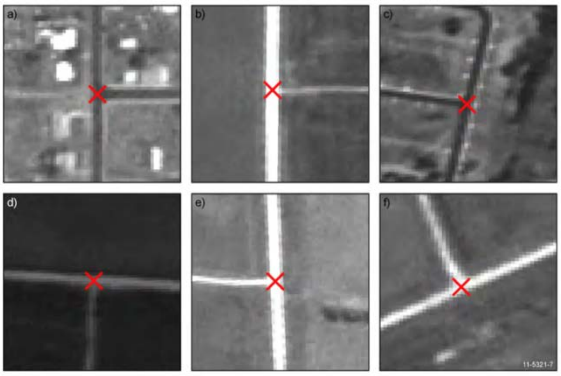
Now it is a geographic projection and more easily co-registered with the map and with other data sources.

Summary of Examples

To summarise, the geometry of image pixel location may be transformed in various ways and the parameters of the transformations as well as the alignments of the target image and other images may be modelled using control points. The modelling is a statistical process. When it is done, the image geometry may be modified so that the image aligns with other images, with maps or with other data sources. The process of map modification is often called image resampling.

As illustrated above, if care is taken, the model fitting and alignment can be accomplished with minimal complexity when image distortions are reduced. Specific models and model forms that take account of scanner models, map projections and camera models should all be used as much as possible. If there is no model of the projections or distortions, a complex polynomial may be fitted but it can result in an unstable fit to images.

As described above, at some stage images and maps are aligned and registered using control points. A control point is often at the precise location of a feature on the ground. The feature may be manmade – such as a corner reflector for SAR Radar control or an existing opportunity such as a road intersection. Such locations are surveyed using GPS and surveying.



Source: Lewis *et al.* (2011) Figure 9

Sometimes the control points may simply be special points of the image or map. The examples above had these kinds of points – crossings of the graticule of latitudes and longitudes. Again, a control point may simply be the same location or feature in two or more images. One use is to register images to a "base image" which may be a map. But general points that simply identify the "same" point in two or more images have a special use as "tie points". They can be used to tie maps and images in a mosaic or time series with great precision. They try and ensure that the same place or feature is always in the same position in the images after resampling. This is obviously important in both image mosaics and in time series.

The practical aspects of the task of fitting the parameters of models or registering images using the control points involves:

(i) General models - polynomials
(ii) Statistical fitting of general image models
(iii) Resampling images

2. General models for the alignment of images

This is a presentation of arbitrary models from Book 2B:

In this section we will not consider the perspective and division models of the lens distortion examples. It is just the polynomial part of the arbitrary image distortion model that is considered. The general transformation is considered to be of the form:

$$x' = f_x(x,y)$$
$$y' = f_y(x,y)$$

The smooth polynomial models are often sufficient, however, more advanced models including the perspective models and rational transformations. These can be left for later. The polynomial models considered are affine, bilinear (or ruled or first order polynomial), quadratic (or second order polynomial); or cubic (or third order polynomial).

Affine

$$x' = a_1 + b_1 x + c_1 y$$
$$y' = a_2 + b_2 x + c_2 y$$

Bilinear

$$x' = a_1 + b_1 x + c_1 y + d_1 xy$$
$$y' = a_2 + b_2 x + c_2 y + d_2 xy$$

Quadratic

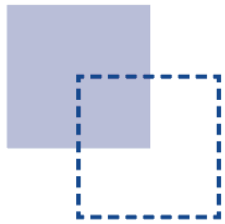
$$x' = a_1 + b_1 x + c_1 y + d_1 x^2 + e_1 xy + f_1 y^2$$
$$y' = a_2 + b_2 x + c_2 y + d_2 x^2 + e_2 xy + f_2 y^2$$

Cubic

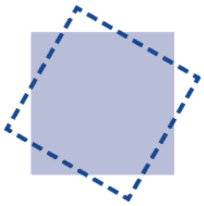
$$x' = a_1 + b_1 x + c_1 y + d_1 x^2 + e_1 xy + f_1 y^2 + p_1 x^3 + q_1 x^2 y + r_1 xy^2 + s_1 y^3$$
$$y' = a_2 + b_2 x + c_2 y + d_2 x^2 + e_2 xy + f_2 y^2 + p_2 x^3 + q_2 x^2 y + r_2 xy^2 + s_2 y^3$$

a. An affine model changes origin (shift), orientation (rotation), scale (enlargement or reduction) and one-directional linear skew (in X or Y direction).

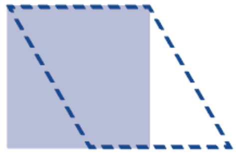
Shift of origin



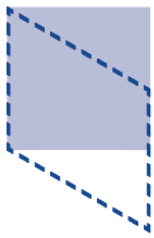
Rotation of axes



Skew in X direction

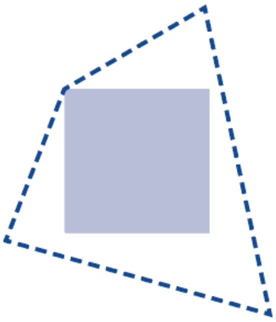


Skew in Y direction

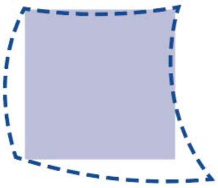


Scale change

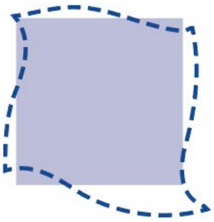
b. Bilinear model: changes origin, orientation, scale and two-directional, linear skew



c. Quadratic model: changes origin, orientation, scale and two-directional, non-linear skew with one point of inflection



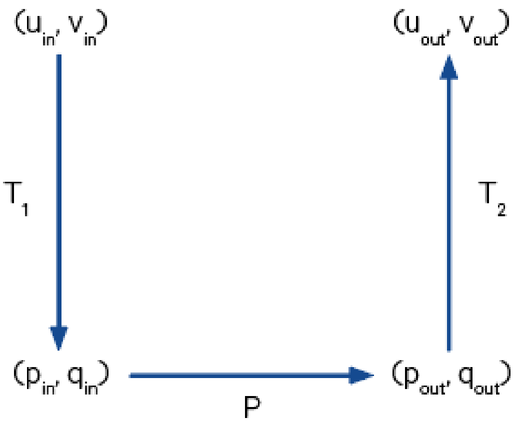
d. Cubic model: changes origin, orientation, scale and two-directional, non-linear skew with two points of inflection



Source: Harrison and Jupp (1992) Figure 24b, 26, 27 and 28

Figure 3.1 Generalisation of three-stage registration model

P is a polynomial transformation from (p_{in}, q_{in}) to (p_{out}, q_{out})
T_1 is a nominal model which converts from (u_{in}, v_{in}) to (p_{in}, q_{in})
T_2 is a nominal model which converts from (p_{out}, q_{out}) to (u_{out}, v_{out})



Source: Harrison and Jupp (1992) Figure 29

The general models can be combined with specific or nominal models to create the alignment capacity described above. The critical objective is always to ensure that the order of an arbitrary transformation is as small as possible. This is done through the selection of the T1 and T2 transformations above.

3. Statistical fitting of general image models

The general models involve T1 and T2 transformations to minimize the order needed for the general P transformation. For each case the objective is to obtain as effective an estimate of P as possible. This can sometimes/often/normally involve also allowing parameters of T1 and T2 to vary. When such complex modelling is attempted then statistics can help you to obtain a good result.

This simple introduction to methods and needs is not the place to go into this phase in great detail. However, for interested people, some brief discussions written as part of the microBRIAN image processing system support have been provided. But first I will indicate why these ideas became researched and later used. It is “Somporn’s Story”.

In the late 1980’s and early 1990’s, CSIRO was involved with a commercial image processing system called “microBRIAN”. One system was at Melbourne University. Somporn was a Thai PhD student who was looking at using Landsat to map changes of land cover in northern Thailand. Her difference image made no sense so she asked us what may be happening. Most images of Thailand have cloud so the areas she had available to pick control points were different in the two images and they were not many. She used some higher order polynomial models to fit the points (separately in each image) to make the RMS error small. She then took the difference of the two images. The change areas were very strange and (because she knew the area) definitely made no sense. What was the problem?

If the control points are located in a sub-area of an image, the sub-area will be well controlled but other areas will not be as well controlled. The higher order the polynomial the less stable will be the rectification model in areas where there are few control points. The images had different areas of control and were fitted and resampled separately. As a result the images were very poorly co-registered and the difference was difference between different places in the images. All this despite the RMS errors being very small. So what could be done?

The result was new software. One aspect was to get a better measure of the RMS error in terms of the ability of the model to "predict" rather than just "fit". This is called a predictive error model which has an associated statistic called the "GCV". The second was to get an idea of how badly the model would fit in areas outside the control point cluster. This will help by showing places where some extra control is ***really*** needed. In Somporn's case she found extra points where they were needed most. Thirdly, because the most critical thing was for the images at different times to match and because it was hard to find extra "true" GCPs where a ground map value could be allocated, the images were fitted at the same time by using "tie points". Tie points were features that could be found in both images. Ensuring they were the same in both images after resampling was very important. This use of tie points is the same as in a mosaic of images which overlap. It was accomplished using a program called "MOSMOD". Finally, when it is so hard to find good control and tie points it was found that some may be bad points. Bad control or tie points can do a lot of damage. So we needed some tool to help establish when all the points were "good" or which were "bad". A bad control point (we found) can sometimes have a ***very small*** apparent error! In this case it is doing most damage to the model.

Given the importance of well distributed control points and the need to keep models as stable and low order as possible, the methods were developed for the microBRIAN system and the documents written are attached. These are:

1. Predictive error and predictive variance

The document prede_v.pdf has title "Predictive Error and Prediction Variance". These two predictive error measures are not quite the same.

Predictive Error is a measure of fit to data that is not the same as RMS error we came across earlier:

$$RMS = \left(\sum_{k=1}^{n} \left\| \underset{\sim}{d}_k - \underset{\sim}{m}_k \right\|^2 / 2n \right)^{1/2}$$

In this equation $\underset{\sim}{d}_k$ the k'th "data" control point (a 2-vector) and $\underset{\sim}{m}_k$ is the k'th "model" control point. A "best" model could be the one that minimizes RMS. But people noticed that sometimes the residual error at a point is "too small". This often happens for bad data values. So another statistic was developed in which the error at a point is the error you get when you fit the model to all of the other points and predict the value at the given point. The sums of squares of these n estimates is called the "Predictive Error" or PRESS. It is larger – especially at bad points. Since high order polynomials are "unstable" it also gets larger when fitting higher order models.

The new error is used to (1) locate bad data points and (2) measure when the order of the model is getting too high for a stable prediction. Both of these were needed to help Somporn as well as many others. Fortunately, the complex fitting of n models is not needed as the equations can be calculated mathematically. This is done in the document. The document also describes a simpler (but related) statistic called GCV that has been used when the distribution of the points is not an issue. It is used in ANUDEM which you will know of.

Predictive Variance is a different but related idea. If you have all of your GCPs and fitted the model and are happy there is still a question of how well the model is fitting at other places in the image? Predictive variance measures possible variations that occur at every location of the image if the values at the GCPs vary with a specified variance. So it “predicts” the instability. It is found that for higher order models the predictive variance rises quickly away from control. In this case it can tell you where you need extra control or tell you a lower order model may be safer.

The situation is summarised in Harrison et al. (2018) as follows (with some changes):

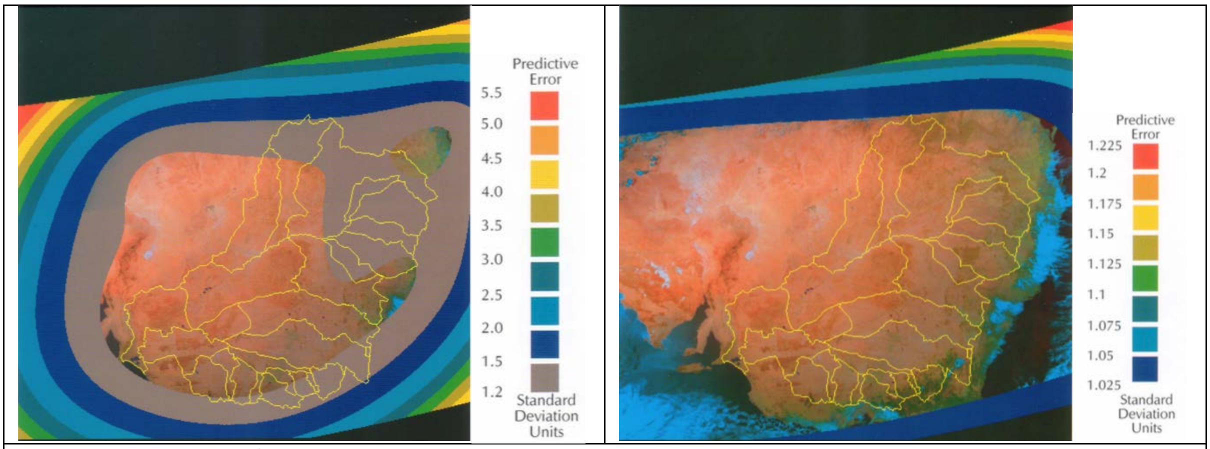
“In this case the 'best' model depends on the number of model parameters the sample datasets can support and what the model will be used for. Three types of model selection criteria are relevant to image rectification:

The “optimum” model is the polynomial with minimum predictive error value (PRESS). The predictive error of a point indicates the importance of each data point to the fitted model. If the sample points are well distributed over the whole image, this model should be used for image resampling and point location.

The “maximum” model is the highest order model that can be fitted to the data. In many image processing systems, this is can be up to a cubic polynomial (provided there are at least 11 control points). The maximum model is the most sensitive to the range and spread of the sample data values used to fit it. The maximum model is therefore useful for checking the extent of control provided by the sample points over the image. “

An example of how GCPs were found to be inadequate because of the behaviour of the maximum model and the result of working to obtain more control can also be found in Harrison et al. (2018) and is:

Predictive error channel on AVHRR colour composite, acquired on 25 December 1987. Catchment boundaries are shown as yellow. Note that colour scales for predictive error are different in each image.
a. High predictive errors result when the cubic transformation model is derived from 32 visually identified GCPs that are poorly distributed, especially in northern catchments.
b. Lower predictive errors over the Murray Darling Basin indicate good control for a cubic transformation model derived from 150 GCPs. These points were systematically computed by the receiving station and well distributed across the image.



Source: McVicar and Mashford (1993) Plates 2 and 3

2. Mosmod

Document Mos_bas.pdf.

The above example was from a report that created a time series of images. Not only did the individual images need to be fitted well and be stable and have good control over the whole area but also the time series needed to be tied strongly together so that features that could be recognised in two or more images were in the same place in the time series.

The method of tying images in mosaics or time series is described mathematically in detail in the document. However, it is useful to quote its discussion and examples:

“Mosmod is based on the idea that when you mosaic a set of images it is a good idea to take into account all of the associations between the images to derive image to image models. The situation is like photogrammetric “bundle adjustment” and offers advantages over more common image to base map transformations that do not take into account overlapping areas or other matches between images.

In the formulation that follows we have used the terminology of “Page’s” which are usually images or a map or the section of the world that is being mapped. In the Mosmod approach, there is one Page nominated as the “Base Page” or “Reference Page”. The Base Page can be looked at as the collating base for the Mosaic or simply a base coordinate system underlying the modelling.

Assume that the image pages use geometric models defined by a polynomial, spline or other linear function. This would be the case with most satellite data or airborne data if the Pages were frames from which the major geometric distortions had been taken.

In this case, the Base Page can be the map and the assignments of points to the Base Page are “Ground Control Points” or GCPs. In this formulation, the issue is to fit the GCPs in each frame or Page as well as to match the overlapping Pages to the same degree of accuracy in the coordinate frame of the map.

If the goodness of fit criterion is least squares then the solution is given below. The advantage here is that, just like “Bundle Adjustment” the GCPs that you have between a Page and the map can pass control into adjacent Pages through the tie points. The success of this can be tested by analysing the consequent least-squares matrices as described below.

Conversely, the Base Page can be an individual image and the Map included among the Pages. Then a transformation is obtained from the Map to the image Page. Such an “inverse” mapping is important when it is necessary to find a pixel corresponding to a Map coordinate – such as in image resampling. With Mosmod, it is also the case that all of the Pages tie together with equal accuracy in the coordinates of the image Page.

This formulation also covers the case of time series data where the mosaic is a set of frames at different times that are closely registered. In many systems, each image is registered to a base page separately or to its nearest time image(s). However, by forming Tie points (by automatic correlation methods if possible) it is possible using the complete formulation to simultaneously register all the images pairwise to each other. This is a great advantage for time series products where ties between Pages are generally more crucial than accurate location in a map.

It is especially useful when some of the image Pages in the time series have features that can sometimes be located and not at others – such as fallow fields or when cloud or pointed images exclude features in specific Pages. Such missing matches can be well compensated by the additional ties.”

3. Siever

Document name: Siever_frag.pdf

The document about Siever is more complex but included for interested people. It relates to how you generally analyse a set of control and tie points to locate bad points or systematic variations in the collections. When a large number of automatic GCPs are collected by (eg) correlation patches this becomes quite important.

All of these methods are described in Harrison et al. (2018) and it is left to there for interested people to find more information.

What about Somporn? Using these software items she was able to co-register a number of Landsat images into a time series even with clouds in each and obtain useful and sensible change images – as well as a PhD. But she was not the only user and the use of Mosmod for time series proved a very significant benefit of using microBRIAN.

4. Resampling Images

Will rely on the Wang notes and overheads for now. To fully understand resampling it is best to have already been through the sampling sections of Castleman and understand the frequency domain methods of analysing the effects.

5. Reference

Hughes et al. (2008) Review of Geometric Distortion Compensation in Fish-Eye Cameras. ISSC 2008, Galway, June 18-19. (Conference proceedings). (10.1.1.140.4542.pdf)