

DEM GENERATION FROM SPOT MULTIPLE VIEW IMAGES ,  
USING CORRESPONDENCE ANALYSIS.

H.R. Kostwinder, N.J. Mulder, M.M Radwan,  
I.T.C.

P.O.Box 6, 7500 AA Enschede  
the Netherlands

Commision III/4

ABSTRACT

Multiple viewpoint images from SPOT allow for stereo vision and automated digital elevation model (DEM) generation. In contrast with classical methods for "stereo correlation" which perform similarity detection at the signal processing level, we report on progress in the application of correspondence analysis at the image segment and object level. Correspondence is quantified on the basis of minimum distance in the featurespace of object and segment features (properties). The spatial search for corresponding objects is constrained by known maximum disparity changes (changes in elevation).

INTRODUCTION

Why DEM's from SPOT ?

- For the further operational use of SPOT and other RS data the problem of standard georeferencing must be solved. For resampling of the images to ortho images DEM's of the same resolution as the SPOT data are needed.
- A 10m or 20m DEM raster enhances (semi) automated image classification by providing an important element of scene understanding (2.5 dim. data). The derivatives of the DEM are often related to landuse and land suitability.

Why correspondence analysis ?

Correspondence analysis is the search (defined by a search domain) for corresponding objects based on spatial or spectral features characterising these objects [ref. 1].

In multiple view imagery the displacement (and scale difference) in the images is related to the elevation.

Inversion of the imaging model produces the elevations.

The most critical elements in correspondence analysis are the selection of features and the segmentation and object detection. As a measure of correspondence a weighted distance between a reference vector and a search vector in feature space is used. (Mahalanobis distance, or cityblock distance)

How does correspondence analysis relate to stereo correlation ?

In stereo correlation there is no element of feature extraction. A window of data defines a data vector. Often these data vectors are normalised on vector length. As a measure for similarity the inner product between a reference vector and the given vector is used. The inner product of two normalised vectors  $V_1, V_2$  is  $\cos(\text{angle}(V_1, V_2))$ . For the most interesting case of small angles  $(V_1, V_2)$  the sensitivity is minimal :

$$\cos(\delta) = 1 - 1/2*\delta^2$$

#### DATA PREPARATION

As multiple viewpoint imagery data are acquired from different satellite orbits, the heading in the orbits is also different and hence the scanlines are not running parallel. Thus resampling to epipolar lines is not possible. Resampling to horizontal image planes is possible if the sensor attitude is provided. For this purpose, SPOT-preprocessing level 1A data are preferred, which are corrected for radiometric errors only (sensor offset and -gain differences). These data have to be resampled to horizontal image planes using the physical imaging model and sensor attitude. Preprocessing level 1B data, which are used up to now, are radiometrically and geometrically corrected. However the resampling model is a third order warp,

which interferes with proper resampling to horizontal image planes. As the emphasis in this study is on the development of correspondence analysis, we accept the co-existence of x-parallax and y-parallax and defer the required corrections for non-horizontal image planes till after the generation of a (relative) DEM. The data are resampled using a first order distortion model and bilinear interpolation: the righthand data are resampled using corresponding groundcontrol points in lefthand- and righthand data. As far as possible these are selected at about equal elevation using a height contour map of the scene as reference, in order to minimize the influence of height displacements on the resampling process. Haze correction is performed to all spectral channels before resampling.

#### SPECTRAL FEATURE EXTRACTION

Visual inspection of anaglyphically displayed left/right image pairs of bands 1,2 and 3 indicates that the best depth perception is achieved using bands 2 (red, 0.61-0.68  $\mu\text{m}$ ) and the worst for bands 3 (near infrared, 0.79-0.89  $\mu\text{m}$ ). This is explained by the time difference of  $\pm 1.5$  month between left and right images, which gives rise to major changes in vegetation-areas. Vegetation is most prominent in the near infrared band.

For similarity detection on area objects, a weighted average of the three input bands per data set reduces this problem. A weighted average is the same as taking a weighted sum of sum-normed bands. In spectral data space this is equivalent to projecting all data vectors ( $B_1, B_2, B_3$ ) onto the normal vector of the vegetation plane. An estimation of this plane can be obtained by sampling vegetation vectors. The weights are determined by the components of this normal vector. The thus defined new spectral filter is insensitive for vegetation changes.

Spectral sum normalisation per data set:

$$B'_i = B_i / (B_1 + B_2 + B_3) \quad , i=1,2,3$$

For similarity detection on area objects the shadowing of areas

must be removed. Size, shape and position of shadow segments are not similar in left and right data set since the sun angles are different (azimuth, elevation). Similarity detection will result in false matches on shadow segments. Dividing each band by the sum of bands, i.e. the total number of detected photons, illumination (and shadowing) changes over the scene are removed. The three spectral reflectances remain, which are dependant on groundcover material only. For humans it is much more difficult to perceive depth in anaglyphically displayed left/right image pairs of sum normalized bands. Human stereo vision extracts depth information from disparity and from shadows. These informations are then combined with (prior) knowledge resulting in depth impression. The sum of bands of multispectral SPOT (XS) data is radiometrically similar to panchromatic SPOT(PAN) data (0.5-0.89 um for XS, 0.51-0.73 um for PAN). In order to take optimum advantage of the 10 meter resolution for DEM generation, both XS and PAN data must be combined for feature extraction (spectral resolution of XS, spatial resolution of PAN).

A problem will arise for scenes with one kind of groundcover (e.g. desert sand dunes), because the spectral reflectances are almost constant for those scenes. Segmentation of sum normalized bands is bound to fail. The intensity (sum of bands) is the only information source in this case. As the study area shows a variety of groundcover (vegetation, sand, rock), we extracted sum normalized band 2 spectral feature for similarity detection on areas (segments). We consider this selection as an approach to the weighted sum of sum normalized input bands.

The Laplacian effect in the sum normed image of band 2 is explained by the sensor displacement per spectral band in the CCD detector array ("pushbroom"). The sum of bands has a wider point spread function than each individual band. As the eye takes the logarithm, a ratio becomes a difference :

$$\log((B_2 / (B_1 + B_2 + B_3))) = \log(B_2) - \log(B_1 + B_2 + B_3)$$

This is equivalent with a -Laplacian operator in horizontal direction. A drawback of this effect is the enhancement of the minor vertical striping. It can cause severe problems for

similarity detection.

For similarity detection on line shaped objects we selected band 2 (non normalized). This band is little sensitive to vegetation changes and shows lines with good contrast, compared to the other bands. Sum norming is not preferable as it distorts or removes (parts of) linear objects in horizontal direction due to the Laplacian effect.

#### SPATIAL FEATURE EXTRACTION AND SIMILARITY DETECTION ON AREAS

For similarity detection on areas research on segmentation filters is still going on. We developed the following specifications for the segmentation filter :

- 1) independant of global radiometric thresholds
- 2) (radiometric) topology preserving
- 3) segmentation by merging only [ref. 1]

So the merging has to be performed in a topological consistent way. The topology is defined by the zero-crossings of the Laplacian. Taking this image as a decision plane one can change the edges into more steplike edges using the following algorithm. The edges are representing segment boundaries, at which positions we want to generate elevations. On a 3x3 neighbourhood at position (i,j):

```
IF Laplacian>0 THEN new(i,j) := RANK[2]
IF Laplacian=0 THEN new(i,j) := RANK[5]
IF Laplacian<0 THEN new(i,j) := RANK[8]
```

After a certain number of iterations the output becomes stable, containing piece-wise constant segments (due to the step edges) and segments with constant radiometric slopes only. A number of variations to this filter are under study now. The main problem is the geometric accuracy of segment boundary positions in left and right image, which are important as they determine the disparity and thus the elevation (results will be presented on a ISPRS postersession, 1988.)

Subsequent segment merging of the pre-segmented image is controlled by a symbol image, describing the radiometric topology. This symbol image is derived from the zero-crossings of the Laplacian. Each segment boundary element has a label describing whether it is a node- or arc element (graph representation in raster), e.g. L=line-element, P=single point, X=crossing, etcetera. Local inconsistencies in the topology can be removed: apply local operators to the pre-segmented image using the symbol image as conditionplane (label relaxation).  
Example :

```
IF symbol(i,j)=P THEN new(i,j) := MEDIAN
```

The number of segments (and their average size) can be controlled by the size of the Laplacian kernel.

Once a satisfying segmentation process has been developed, similarity detection can be performed on segments. The best match between a reference segment in the lefthand image and a number of candidates in the righthand image is then defined by the minimum distance in feature space. The dimension of this feature space is determined by the number of segment properties, such as area, centre of gravity, perimeter, and other shape descriptors. The distance measure (Euclidian, Mahanalobis) which has to be applied is dependant on the nature of the segment-property data (ordinal/nominal), and the (an)isotropy of feature space [ref. 1]. The search space (in the other image) is limited by the maximum possible disparity (known from prior knowledge) and is restricted to one direction in case of epipolar geometry (shifts over a number of columns). The RAG (region adjacency graph) of left and right image is the tool to find the neighbouring segments.

After segment matching, their boundaries must be matched. The idea is to calculate the minimum energy needed in replacement of chaincode-elements of chaincoded boundaries of the corresponding segments (equivalent to local stretching of the raster). This will result in disparity values per scanline, in case of epipolar geometry. The chaincodes can be computed from the topology symbol image.

## SPATIAL FEATURE EXTRACTION ON LINEAR OBJECTS

For line-extraction we used the line-detector from the standard Image Processing Software of the ContextVision GOP system called LINE [ref.5]. By comparing the convolution results of a subimage with a set of line- and edge-detecting 15\*15 kernels, it evaluates the hypothesis of a linear feature passing through (the centre) of the subimage. The result consists of a measure of certainty for the validity of the hypothesis, as well as an indication for the orientation of the line (if there is one).

Thresholding the certainty measure defines the linear shaped objects, which are stored in a bitplane :

```
IF certainty(i,j)>5% THEN object(i,j) := 1
                               ELSE object(i,j) := 0
```

The (1-dimensional) feature space for search of minimum distance in similarity detection is defined by the "object" property (local) orientation. As this is a local property, based on a 15\*15 neighbourhood, it is not a property of entire objects such as line-length.

## SIMILARITY DETECTION ON LINEAR OBJECTS

The absolute difference of orientations(left,right) at object positions defines the (Euclidian) distance measure in (ordinal) feature space. A search along "epipolar" scanlines in the righthand object image for each reference object in the lefthand image gives all candidate matches (search area for stereo matching). The best match is then selected from the minimal distance in feature space. The search area is limited by the maximum possible disparity, determined by the maximum elevation difference in the scene. The pseudo algorithm :

```
FOR line:=1 TO all lines DO
FOR col:=1 TO all columns DO
  minimumdistance=max;
```

```

IF objectleft(line,col)=1 THEN
FOR shift:=minimumshift TO maximumshift DO
  IF objectright(line,col+shift)=1 THEN
    distance:=|orientleft(line,col)-orientright(line,col+shift)|
    IF distance<minimumdistance THEN minimumdistance:=distance;
                                shiftmin:=shift;
  NEXT shift;
disparity(line,col):=shiftmin;
NEXT col
NEXT line

```

The output image ("disparity") contains the shift-values (disparities) at object positions for the best match. They determine the elevation. Results of this application of similarity detection will be shown in the poster session.

#### REFERENCES

- [1] Mulder, N.J., Radwan, M.M., Correspondence analysis applied to image matching and map updating, ISPRS 1988, Kyoto, Japan.
- [2] N.J.Mulder, H.Middelkoop, J.Miltenburg:"Progress in knowledge engineering for image interpretation and classification", ISPRS 1988, Kyoto, Japan.
- [3] Gorte, B.G.H., Mulder, N.J., Topographic feature extraction based on context dependant filters, ISPRS 1988, Kyoto, Japan.
- [4] SPOT image, The SPOT CCT format.
- [5] Hedlund, M., Granlund, G.H., Knutsson, H., 1982, A Consistency Operation for Line and Curve Enhancement, Proc. IEEE PRIP-82, Las Vegas, June 1982.