14th Congress of the International Society of Photogrammetry, Hamburg 1980.Invited paper Comm.III Wg.2.

# A VIEW ON DIGITAL IMAGE PROCESSING

N.J. Mulder - International Institute for Aerial Survey and Earth Sciences
P.O. Box 6, Enschede, the Netherlands

## ABSTRACT

The paper consists of two major parts. In the first part a top-down structured view is given in which image processing is presented as part of a decision making procedure. The importance of a wide view on decision making and data preparation is stressed and some examples are given of different application fields with the same basic decision theory. The role of Remote Sensing as a unifying concept is discussed.

The second part consists mainly of examples of applying the concept of mapping to the intensity, multispectral, spatial and temporal domains. For human decision making, knowledge of visual perception is important in mapping colour and pattern features into the "brain domain".

It is shown that image processing for automated decision making and human decision making is essentially the same. The human interpretor should have more knowledge of elements of decision theory.

## KEYWORDS

Decision Theory, Image Processing, Pattern Recognition Colour coding, Feature extraction. Image enhancement, Remote Sensing.

## INTRODUCTION

When asked to review a certain field, one is tempted to do just that. One would have to catch up with a year's unread literature, sort things and present who's been writing what.

A disadvantage of review type of papers is that they tend to direct one's attention backward. Another problem specially found in the field of image processing applied to remote sensing data is the huge confusion in terminology, most of the review would have to be spent on definition of terms. Many wheels are reinvented but get different names. (Remote sensing is a typical intellectual escape field).

As an alternative to a review paper, I will present a "view" paper. My aim is to present digital image processing as a mental tool kit which is to be used for problem solving. In problem solving and invention it is necessary to be able to work with different levels of abstraction of our "problem world". It is equally important to work with abstractions of digital image processors. Present day computers are mostly unsuitable for image processing. We should not let our thinking be limited by Von Neumann machines.

My view will include more fields than are usually understood to belong to DIGITAL IMAGE PROCESSING. In the section with examples of current standard processes I will stay within the more conventional boundaries and even omit some of the more interesting examples like the relation between Digital Image Processing, Photogrammetry and Cartography.

## DIGITAL IMAGE PROCESSING - WHAT IS IT?

Firstly we recognise that it is a complex combination of methods and techniques applied to somthing which is very hard to describe. Many views are possible and to capture the complete picture, many angles of view must be studied from a level of sufficient high abstraction.

Abstraction is essential for understanding, it enables us to see the common elements in apparently unrelated fields. In order to communicate our abstractions we need a common language, code, symbolism, formalism. Two way's of abstract thinking can be discerned: the formal logic way, which produces strings of causes and effects, in the notation of an algebra of logic entities and the pictorial, geometrical way of thinking, the mental manipulation of 1D, 2D and 3 Dimensional objects.

I prefer to think in the second way, to move and transform "things" in multidimensional spaces, because it allows me to think multidimensional, I can handle complex objects as a whole. When something has to be proved formally or when a computer algorith has to be defined, I go down to logic formalism and produce essentially 1 dimensional strings of causes and effects (with associated branchpoints of course).

Using this approach, it will be possible to explain all idea's in Digital Image Processing by the use of pictures (2-D-projections). The corresponding formalism is the formalism of mapping, vector algebra on scalar- and vectorfields and some elements of decision theory, pattern recognition, perception theory and even some elements of physics.

"Digital Images", or rather digital representations of images. A common characteristic of all "Digital Images" is that they are generated by spatial sampling of spatial continues radiation fields. The measured Intensity on Energy per sample area is digitised to integer values.
The visual representation of one spatial sample area is called a pixel (picture element).

In the monospectral case, our abstract view shows a scalar field. In the multispectral case we "see" a vector field. With repetitive image cover a time component is introduced which can either be treated as a parameter or as an extra dimension.

Take notice that through abstraction our view has become tremendously wide. We can connect now the processing of all sorts of scalar and vector fields with image processing. Just to give some examples: digital terrain models → hill shading, the differential equations of electrical-, temperature- and density fields, the theory of membrane vibration, optimum routing of transport, 2-D transforms and filter theory, stereo terrain models, photogrammetry, cartography, graphics (geo)databases, etc.

Digital Processing of images encompasses a wider field even than Digital image processing. As already indicated in its abstract form it includes the processing of scalar and vector fields, however, people like to make a distinction between processing which results in "normal" images with many greyscale or colour levels and between classified image data with few colours or symbols thematic mapping etc.

From an abstract point of view there is only a small difference between image processing, pattern recognition clustering and automatic classification. The difference being: mapping from many → many states (image processing) and mapping many → few states (clustering, automatic classification).

The question from an educational point of view is: can we place image processing and classification in a common abstract framework and indicate possibly a hierarchy of concepts?

## A HIERARCHY OF CONCEPTS

If we want to evaluate idea's from literature or if a we want to give courses at application or academic level we need a clear view on how different abstract idea's are interrelated.

DECISION MAKING

This is the most general concept. The final aim of our activities is to be able to make decisions based on all available data and knowledge.
DECISION MAKING → the ART of DECISION MAKING
→ the THEORY of DECISION MAKING

The ART of decision making is usually a component of management and photointerpretation courses.
The THEORY of decision making is explicitly found in Operations Research and Game Theory and is implicitly found in Pattern Recognition, Artificial Intelligence and Applied Statistics.

It is to be hoped that a closer relation between art and theory will grow in future.

IMAGE PROCESSING

The theory and methods of MAPPING scalar- and vectorfield data into representation domains which can be discontinues:
AUTOMATIC CLASSIFICATION, CLUSTERING, REGION FINDING or continues:
IMAGE ENHANCEMENT.

IMAGE SEGMENTATION

A parallel concept/process is segmentation of image data fields into regions which have something in common. This involves already some human or algorithmic decision.

IMAGE ENHANCEMENT

This is a very general term, which includes techniques applicable to preparation for human- as well as machine decision making. I would call the common concept in both fields:
PROBABILITY CODING, LIKELYHOOD CODING
"class" probability can be coded in colour, in symbols or purely numeric in a probability array. A practical example of likelyhood coding is our method of Multispectral Correlation Colour coding.

FEATURE EXTRACTION, FEATURE ENHANCEMENT

If we assume some knowledge with the user of the system, he will be able to map the raw input data into data which is much more specific and useful in the decision making or likelyhood estimation phase of the project, then would be possible with the raw data.
Feature extraction usually results in:
DATA REDUCTION and SIGNAL TO NOISE RATIO improvement.
Features can be SPECTRAL features or SPATIAL features or a combination of both such as TEXTURAL features.
Characteristic changes of signals with time give TEMPORAL features.

RADIOMETRIC- and GEOMETRIC CORRECTIONS

In order to reduce this "noise" component in the data it has to be corrected. Sensor variation and atmospheric influences necessitate RADIOMETRIC corrections. Platform altitude variations and sensor nonlinear scan require GEOMETRIC corrections, which link digital image processing with PHOTOGRAMMETRY.

One keyword missing in the list of concepts is the word REMOTE SENSING. The reason for this is that I personally feel REMOTE SENSING is only a general concept on the sensor side. Its main role in DIGITAL IMAGE PROCESSING is to provide us with most of our data (Landsat mainly at present). The question is then: isnt'it important to know the source of the data and the details of the sensor system?
The answer of this question is: the provider of the data should worry about correcting the data which he supplies. The user should not have to worry about low-level image processing often called PRE-PROCESSING. (However he should understand image processing well enough to tell TELESPATIO why Nearest Neighbour Geometric Correction is visually not acceptable on vertical features, although the Root Mean Square thinkers proclaim it a very reasonable method).

The only remaining connection between REMOTE SENSING and IMAGE PROCESSING is in the domain of PHYSICS.
A working knowledge of Physics will guide the user in his selection of meaning full mapping transforms from all possible (spectral) mappings. I will illustrate this in section with examples.

CONCLUSIONS ON CONCEPTS

Privately thinking about the present situation in the Processing of Remote Sensing data, I feel that we should restructure our approach to research and education in emphasising the combination of DECISION MAKING and IMAGE PROCESSING. The link with REMOTE SENSING is rather circumstantial. The links with GAME THEORY, OPERATIONS RESEARCH, PATTERN RECOGNITION, THEORY OF CELLULAR AUTOMATA, ARTIFICIAL INTELLIGENCE, HUMAN PERCEPTION are more important, just to mention a few.

EXAMPLES OF THE APPLICATION OF THE CONCEPT OF MAPPING TO CURRENT "STANDARD" IMAGE PROCESSING ALGORITHMS

First we need a (conventional) definition of our data world then we will define four domains in and into which mapping operations occur.

Single (spatial band) images are treated as scalar fields. They allow mapping of scalar → scalar in the intensity domain and mapping of a neighbourhood of pixels into a new pixel (at a new location) in spatial domain. Multispectral images introduce one extra dimension. The ordered scalar values per individual band form together a multispectral vector. All sorts of mapping can be applied to such vector fields.

Time is the fourth variable and is related to e.g. temporal changes in landuse, vegetation, temperature, seastate etc.

Treating Multispectral data as vector fields is allowed most of the time but one must be carefull not to put apples and horses in the same vector.
Distinction must be made between measurements which depend mainly on material properties (reflection) and those that depend on a combination of material properties and a state-variables like temperature (→ emission).

DEFINITION OF OUR DATA WORLD

As already indicated we will be mainly concerned with Digital Image Processing as applied to Remote Sensing data. We should not restrict application of Digital Image Processing to Remote Sensing data only, such as Multispectral Scanner (MSS) data, Thermal scanner (THS) data, Digital Side Looking Airborne Radar (dSLAR) data, but Graphics (Cartographic data), digital terrain models and other geo-data bases belong also to the "problem-world" of digital image processing.

Why include graphics in image processing?
It is difficult to deny that in graphics we also work with "images" the only special thing about graphics is data representation. In the old times when computer storage was expensive, images were first compressed into line images, which were further compressed into line-strings (with attributes). Nowadays with the availability of colour raster-scan graphic systems with high resolution (e.g. 4096 x 4096) the distinction between computer graphics and image processing disappears. Computer graphics is rediscovering most existing image processing algorithms. A general view must include all sorts of data which can be represented on a grid, even if intermediate storage is in string format.

## PROBLEM DEFINITION

In general we want to map available digital data into a presentation which is optimal for a certain group of users.

Digital input has often many possible states, which are not all relevant to the users definition of information. Mapping will mostly be from many states to fewer states. Two different aims can be discerned ·

a1 to present all available data in such a way to the human eye-brain combination that he can use his unique capabilities to classify the image (photo interpretation)

a2 to present the data in a computer classified form, with main emphasis on the statistics of the resulting data.

We will concentrate our examples on processing the data for human decision making, but it must be understood that the same processing concepts are used in preparing the data for automatic classification.

## FOUR DOMAINS

If we agree that digital image processing is just a matter of mapping input into some output presentation we only have to define the domains and the mapping rules.
First the domains :

1 Intensity domain, usually a range of integer numbers which have a one-to-one relation with Remotely Sensed intensities in some part of the E.m. spectrum. Most data is available in the byte range (∅..255), scalar $B_i$ (except for digital SLAR (∅..64.∅∅∅))

2 Multispectral domain:
With one elementary sample area on the ground, intensity values for many spectral bands may be related. We may order MS-data into MS vectors, with the ordered intensity value's per MS band as elements $\bar{B} = (B_1, B_2, \dots B_N)$

3 Spatial domain
Each sample (loosely called PIXEL=picture element) has a position vector, related to the centre of the pixel, associated with it, $\bar{x}$, $\bar{y}$ etc., $\bar{B}(\bar{x})$. Often we will have to map a neighbourhood $N\bar{x}$ of $\bar{x}$ into a new pixel at position $\bar{y}$ : $\bar{B}(N\bar{x}) \rightarrow \bar{C}(\bar{y})$

4 Time domain
As satellite data is inherently repetitive, time must be included in image processing. Time is treated as an extra dimension or as a parameter

$$\bar{B}(\bar{x},t) : \bar{B}(\bar{x},t_1) - \bar{B}(\bar{x},t_2) \rightarrow \bar{C}(\bar{x},\Delta t_{12})$$

## MAPPING IN AND FROM INTENSITY DOMAIN

Typically we only consider one MultiSpectral (MS)Band at a time say $B_i(\bar{x},t)$ and $\bar{x},t$ or the other $B_i$-Bands do not influence the mapping of e.g.

$$B_i(\bar{x},t) \rightarrow C_i(\bar{x},t)$$

A special case is the mapping of a simple Band B into colour (colour vector $\bar{C}=(C_{BLUE}, C_{GREEN}, C_{RED})$

$$B(\bar{x},t) \rightarrow \bar{C}(\bar{x},t)$$

The nature of possible mappings is scalar → scalar and scalar → vector. Let us first look at ways to map scalar x → scalar y. We are used to functions like $y = 1/x$, $y = x^2$, $y = \sqrt{x}$, $y = \log x$ etc.; in fact we map $x \rightarrow y(x)$. The mapping is usually defined as $y = f(x)$ but can equally well be defined by a Look-Up-Table(LUT) which stores for each possible x the corresponding y. In digital image processing the use of LUT's is possible because x is most often defined as an integer in the byte range ($0 \leqslant x < 256$). Result y can be rounded or scaled and rounded of to the nearest integer. Depending on the number of consecutive mappings, y will be stored in byte range or double byte range ($0 \leqslant y < 2^{16}$).

Example 1: $y = 2^x$

The corresponding LUT: y(x) has the following structure for x = ∅ to 15

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| x = | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | . | .... 15 |
| y = | 1 | 2 | 4 | 8 | 16 | 32 | 64 | . | . | ........ |

Using LUT's is a very efficient way of mapping. Instead of computing $2^x$ for each of the say $10^6$ pixels in a typical image file, x is used as the address for getting data y(x) stored in array y.

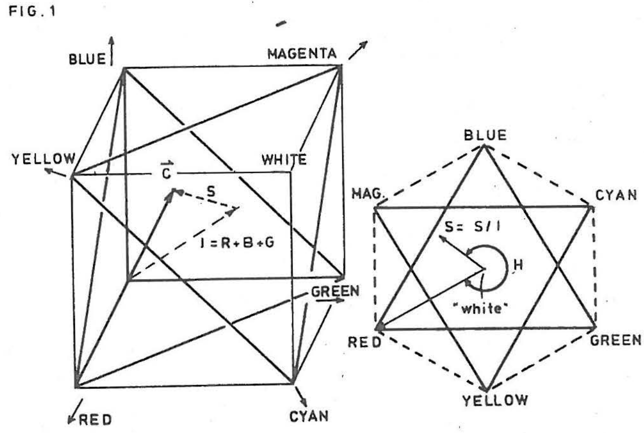
Example 2 : $y_j = a_j + b_j x_j$, radiometric correction.
A linear radiometric correction can be performed in two ways. Practicality of each method depends on the number of different sensors per spectral Band.
In case of Landsat, each of the 6 sensors in a swath can have its own $y_j(x_j)$ table, which needs not only contain a linear correction but might as well include antilog decompression.
In case of a CCD array we have too many sensors, we would need too many LUT's. It is more efficient to store the $a_j$ and $b_j$'s in a table and perform the multiplication and addition with a special fast(integer) processor (e.g. 200 ns/correction).

Understanding mapping one band into colour assumes some knowledge of colour theory.

## COLOUR THEORY

In the practice of colour t.v. screens and colour film writers like the Optronics C-4300, we only have to define the colour intensities or densities in three colours. Usually technical systems work with a RGB set (Red, Green, Blue). Instead of a 7(colours of the rainbow)dimensional 7-D space, we only have to worry about a 3-D RGB colour space, as shown in Fig. 1. By using LUT's for each colour or by using hybrid electronic antilog devices we can linearize the relation between digital R, G, B values and perceived intensities. Remember that most biological sensors have a logarithmic response → theory of perception.

FIG. 1




RBG _ Colour cube HSI _ Coördinates

Fig.1. 3-D colour space,with base vectors RBG Red,Green, Blue on (0,255).In the colour cube 2 colour triangles can be constructed,the RGB triangle and the CMY triangle. Transformation to polar co-ordinates gives HSI vectors (Hue,Saturation,Intensity),I = R+G+B, S =I-normalised radial distance from the white point, H = angular pointer to the spectral colour.

A physiologically meaningful transform supported by colour t.v. practice is from RGB coordinates to HSI (Hue, Saturation, Intensity, 2 angles + 1 radius) coordinates. I = R + G + B and defines the diagonal plane in which the colour vector $\bar{C}$ = (R,G,B) is located, S: Saturation is the radial distance from the main diagonal to $\bar{C}$, if S = ∅ the colour is grey/white, if S = max then we have a maximum saturated colour; S is normalized by division by I, S/I→S is an angular measure. Hue is also an angular measure, it indicates the sort of colour (points to part of the rainbow).

Back to mapping intensity into colour.
Example 3: thermal false colour, given digitized thermal values from Ø to 7, assign a colour to each thermal level (density slicing). The following LUT will perform the trick. As an exercise trace the path of mapping the 1-D line value Ø to 7 on the 3-D colour cube.
Bi CLUT:

| Bi | Ø | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| C Red | Ø | Ø | Ø | Ø | 255 | 255 | 255 | 255 |
| C Green | Ø | Ø | 255 | 255 | 255 | Ø | Ø | 255 |
| C Blue | Ø | 255 | 255 | Ø | Ø | Ø | 255 | 255 |
| Colour | Bk | Bl | Cy | Gr | Yel | Red | Mg | Wh |

Given a basic understanding of mapping and the use of LUT's it should be easy to understand the following mapping procedures in Intensity domain:

Radiometric correction, range compression-decompression, log-antilog, gamma correction, linear stretch, linear compression, inverse, square, square root, scaling, density slicing with or without colour coding, film sensibility corrections, mapping for linear perception, intensity to density mapping.

A more or less special case is histogram equalisation/ entropy maximisation mapping. The aim of this mapping is to generate an output Y(X) with a flat frequency distribution Py, maximising

$$H = \sum_{y} Py\ {}^{2}\log Py$$

which is a (poor) measure for total information content of a picture. It maximizes surprise when looking only at one pixel at a time, forgetting the neighbour pixels.

Examples of: histogram equalisation.
An image of e.g. $10^7$ points has a histogram on Band 1 = Bi as shown in Fig. 2. All data is in the range (0,63). The eye can only discern about 16 grey steps. Map (0,63) into (0,15) using the cumulative histogram of Fig. 2.

FIG. 2

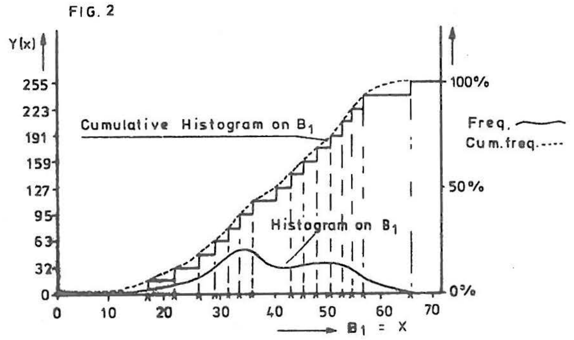



Fig. 2. Histogram equalisation X→Y(X) using a cumulative histogram on X. A division on Y(X) in 16 equal intervals is mapped through $\Sigma f(X)$ into unequal intervals on X. A LUT is built up with constant Y(X) within the thus found intervals.

The correct histogram equalisation table Y(X) is found by first mapping 16 "intensities" with an interval of 16 through the cumulative histogram graph Fig. 2 into irregular intervals on X.
In applying X→Y(X) X-intervals will be small where Frequency (X) Fx→Px is high. When Px is low a larger interval on X is needed before 1/16 of the total number of data points has accumulated.
The output frequency distribution Fy will consist of spikes at 0, 15, 31 etc. with empty space in between. The spikes will have approximately the same height (100% / 16).

When using Histogram mapping, care must be taken to make sure the output histogram is constant in perception space.

In ease no histogram is available one can assume a histogram e.g. Gaussion, log-Gaussion, Poisson etc. and use the corresponding cumulative distributions to map the desired output distribution into the input intervals of the Y(X) LUT.

In general man can always define a better intensity mapping for visual inspection than the crude maximum Entropy rule can provide. We use manual interval-on-X setting or define the mapping functions interactively with a graphic screen and X-Y table, trackbell or such.

A rest-group of intensity mapping to be mentioned are:
level slicing, bit slicing, sawtooth mapping, thresholding.

Level slicing can be regarded as a very primitive way of classification (we map many→few states). The most extreme case is mapping from many intensity levels into 2 levels. Corresponding classes are: object and background. A binary valued image is produced.

In Thermal Infrared image processing it could be desirable to map radiation intensities into equivalent black body temperatures by use of Plank's law. In practice the I.R.sensor is calibrated with a black body source of known temperature and a Tbl (I) LUT can be contructed to map Intensity I→Tblack.

MULTISPECTRAL DOMAIN MAPPING

In multispectral domain each pixel is now associated with an ordered number of spectral bands $Bi(\bar{x},t)$. With four MSS bands a 4-D space can be defined with 4-D vectors $\bar{B} = (B4,B5,B6,B7)^T$. Vectorfield $\bar{B}(\bar{x},t)$ can be transformed (mapped) in many different ways. This section will be limited to pixelwise tranformations $\bar{B}(\bar{x},t) \rightarrow \bar{C}(\bar{x},t)$. Therefore we can omit parameters $\bar{x}$ and t in our discussion on MS-Domain mapping; $\bar{B} \rightarrow \bar{C}$.

Spectral Bands are the result of mapping the continues Energy-wavelength through a set of filters into a series of discrete values, one value for each filter. Figure 3 illustrates what happens. Each filter is defined by its transmission $Ti(\lambda)$ over a limited wavelength interval. Given a spectrum $B(\lambda)$ the sum of the energy transmitted power through the filter is:

$$Bi = \int_{-\infty}^{\infty} E\lambda . Ti\lambda . d\lambda$$

For a sampled spectrum over e.g. the wavelength interval from 0.4 µm to 1.0 µm with d = 60 nm, E and Ti become vectors $\bar{E}(1)$ and $\bar{T}i(1)$ with each 600/60 = 10 elements. Rewriting our formula for Bi we get:

$$Bi = \sum_{l=1}^{10} El.Til$$

This is a vector "dot" or "in" product, Bi is a scalar, it can be interpreted as the Spectral Correlation of unknown $\bar{E}$, with spectral mask $\bar{T}i$.
Traditionally filters $\bar{T}i$ have been designed on vague technical grounds, without a direct link with image processing. My proposal has been to design a set of filters such that $Ti(\lambda)$ coincides with the $E(\lambda)$ of globally occurring spectral classes like Water, Vegetation and Bare Soil. In this way class-probability coding will already occur at sensor level.

FIG 3

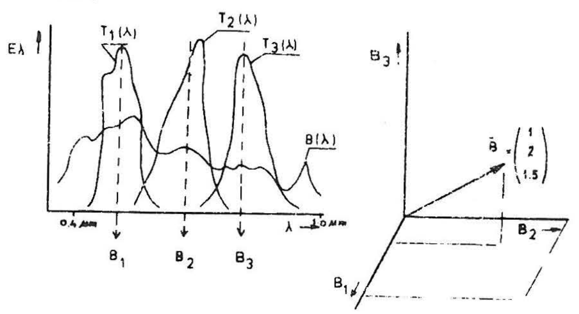



Fig. 3. Filters T1,T2,T3 map spectral signature $B(\lambda)$ into 3-D vector $\bar{B} = (B1,B2,B3)^T$. Spectral filters can be regarded as stored spectra which are correlated with input spectra $B(\lambda)$.

The probabilities for other classes than Water, Soil and Vegetation can be derived from linear combinations of Pw, Ps and Pv. Class probabilities can be mapped into colour space or through a classifier into class labels which in turn can be mapped through a colour LUT into a colour coded classification map.

## LINEAR TRANSFORMS

Shift, rotation, linear projection are linear transforms in feature space (or measurement space). Shift of axis or shift of origin is implied in additive haze correction. The haze contribution in MS-vector $\bar{B}$ is also a vectorial quantity $\bar{H}$ which is added to signal $\bar{S}$,

$$\bar{B} = \bar{S}+\bar{H}$$

to get $\bar{S}$ we apply

$$\bar{S} = \bar{B}-\bar{H}$$

This essentially means a shift of coordinate system in measurement space which is specially important in central projection mapping and mapping into angular coordinate systems.

The Principal Components (PC) Transform is a rotation of the measurement space axis on axis of maximum variance of a covariance matrix on the MS domain of a sample set of MS pixels. As shown in Fig. 4 the new PX-axis coincide with the main axis of an elliptical cluster.

FIG 4

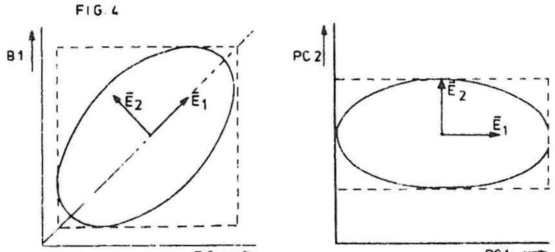



Fig. 4. Mapping a cluster from vector base $\bar{B}1,\bar{B}2$ on vector base $\bar{E}1,\bar{E}2$: Rotation. $\bar{E}1$ and $\bar{E}2$ are the eigenvectors of a covariance matrix defined on the cluster in B1,B2. After rotation the data is mapped to Principal Components (PC) axis.

The box needed for packing the cluster in B-space is bigger than the box needed in PC-space. A PC transformation orders the data on variance and has inherent "data compression" properties or better "data packing" properties. Each PC component corresponds to a PC picture. Starting with the PC1 picture with biggest variance each successive PC will have a lower variance.

By visual inspection we may often conclude that e.g. in the 4-D case of Landsat, PC3 and PC4 contain no information which is not already included and well expressed in PC1 and PC2. Variance is however never equivalent with information. It only relates to possible information storage capacities. This sort of findings lead to the concepts of intrinsic dimension and feature extraction. Our general finding is that the reflectance spectra of natural materials can completely be described with 2 or 3 reference spectra viz. the intrinsic dimension or degree of freedom in reflection spectra is 2 or 3.
The PC-transform can be used as a default MS mapping for ordering the data in feature space and dimensionality reduction, which is essential for understanding the nature of data and the position of clusters. A PC-transform is completely determined by the choice of the sample set on which the covariance matrix will be computed.

α-Transforms are a succession of rotations in FS (Feature Space) in which at each time the axis and the amount of rotation about that axis has to be specified. It can be used in an interactive way and ensures the orthogonality of the FS axis.

Linear Projection, MS Correlation, MS Filters are different names for the same mathematical trick. Take a filter or correlation vector $\bar{F}i$ for each vector $\bar{B}$ in the file:

$$fi = \bar{B}.\bar{F}i$$

fi can be interpreted as a projection of $\bar{B}$ on $\bar{F}i$ or a MS correlation of $\bar{B}$ with a characteristic vector $\bar{F}i$ or as a digital MS filter. In Multispectral Correlation Colour Coding Fig. 5 we first correlate each $\bar{B}$ with stored signatures $\bar{W},\bar{V},\bar{S}$ and then map the result through a colour triangularisation transform. The combined mapping is a linear transform:

$$\bar{C} = \begin{bmatrix}R\\B\\G\end{bmatrix} = \begin{bmatrix}S4,S5,S6,S7\\W4,W5,W6,W7\\V4,V5,V6,V7\end{bmatrix}\begin{bmatrix}2&-1&-1\\-1&2&-1\\-1&-1&2\end{bmatrix} 1/2 \begin{bmatrix}B4\\B5\\B6\\B7\end{bmatrix}$$

from 4-D $\bar{B}$ vector to 3-D $\bar{C}$ vector in colour space. The three 4-D axis on which all data is projected are not orthogonal but fit well to the problem.

FIG. 5

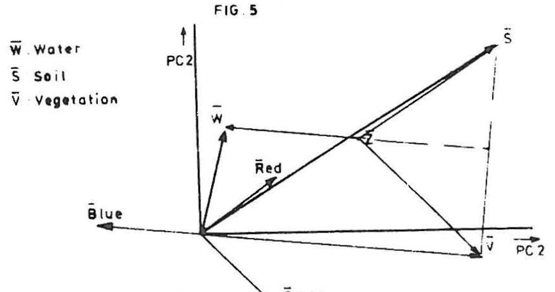



Fig. 5. MS Correlation Colour Coding, shown in a PC1,PC2 subspace. $\bar{B}$ is first correlated with $\bar{S},\bar{V}$ and $\bar{W}$ "stored spectra", next the data is transformed to colour axis using $\bar{Z} = 1/3\ (\bar{W}+\bar{V}+\bar{S})$. $\bar{R}$ed $= \bar{S}-\bar{Z}$ $\bar{B}$lue $= \bar{W}-\bar{Z}$ $\bar{G}$reen $= \bar{V}-\bar{Z}$.

So far no physics was involved in our selection of mapping procedures. However, some knowledge of outdoor physics is useful. Sunangle variation, shade and shadows are usually not features of the classes or materials we want to discriminate between. A transformation which eliminates illumination variations is useful.

Taking the ratios of 2 bands in case of 4 bands MSS is proof of feature space blindness, 12 different ratios are possible and the results have to be interpreted as tangents of angles on 2-D subspaces of a 4-D cube.

A much better solution is to normalise the data on

$\Sigma B = \sum_{i=1}^{N} Bi$ : total reflected energy

$B'i = Bi/\Sigma B$ : $\sum_{i=1}^{N} B'i = 1$

The corresponding feature space mapping is shown in Fig. 6. All data points are projected through the origin O into a diagonal "plane" $\bar{P}.\bar{B}' = 1$. "Plane" can only be used in the 3-D case, in general in an N-D space the sum norm transform will map the data into a N-1-Dim. subspace. We retain $\Sigma B$ as an intensity image which is useful for topological features, it gives a strong relief impression in mountainous terrain.

Before this sort of mapping through the origin or central projection mappings the right origin should be used by performing the haze correction:

$$\bar{B}' = (\bar{B}-\bar{H})/\sum_{i=1}^{N}(Bi-Hi)$$

otherwise the angles implied in $\bar{B}'$ are influenced by an artifact and are not representative for the data.

FIG 6

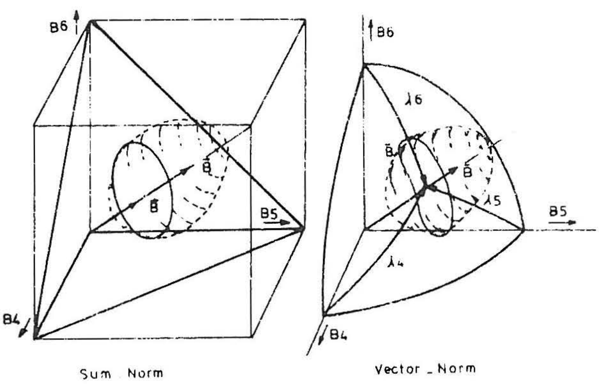



Fig. 6. Sum-Norm (left) and Vector-Norm (right). Norming the data projects a cluster of data from 3-D into a 2-D subspace, which is a "plane" in case of the Sum-Norm and a "spherical surface" in the case of Vector-Norm. Angular separation of data vectors $\bar{B}$ does not change. Vector length $|\bar{B}|$ only is affected.

Mark the close correspondence between the HSI colour transform and the sum-norm transform. The only problem remains dimension reduction. We use a PC-transform or a MS-correlation transform to map e.g. 10-band sum-normed MS data into a 2-D Hue and Saturation plane (diagonal plane in colour cube). The sum can be mapped onto I or treated as a separate black & white relief overlay on the colour data. Further rotations and stretches in 3-D colour space can be used to enhance the image presentation.

Vector-Norm or Direction-Cosine is a related technique to sum-norm it is mathematically more elegant but computationally more involved. It is also not in the same direct manner related with the HSI colour coordinate system as the sum-norm is. Definition: vectorlength

$$V = |\bar{B}| = \sqrt{\sum_{i=1}^{N} Bi^2}$$

$$\bar{D} = \bar{B}/|\bar{B}| = \bar{B}/V, \; Di = Bi/|B| = \cos \lambda i$$

All data points $\bar{B}$ are projected through the origin on the surface of a sphere. A N-D cluster will be projected into a (N-1)-Dim. cluster on the sphere.
In Fig. 6 the angle between Bi and $\bar{B}$ is $\lambda i$. $\bar{B}/|\bar{B}|$ has also the meaning of a cosine $\lambda i$ is usually called a direction cosine, $\Sigma i \cos^2 \lambda i = 1$. $\lambda 4, \lambda 5$ and $\bar{B}$ are the polar coordinate equivalent components of B4, B5 and B6 in Fig. 6.

## Probability Colour Coding

The linear transforms used in the previous paragraphs can also be viewed as mapping the $\bar{B}$ vector data into a set of distances from straight lines.
Figure 7 shows how projection and in-products can be treated as giving distance measures to a complementary set of lines. We use a PC1,PC2 subspace as a convenient 2-D space for examples. The method applies of course to other feature spaces as well.

FIG 7

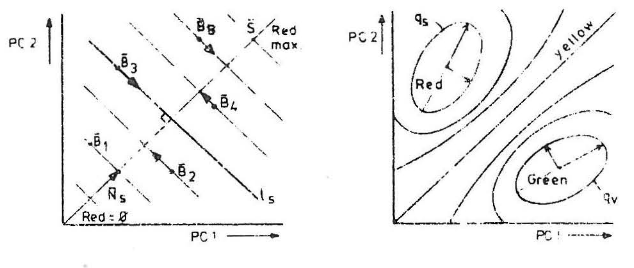



Fig. 7. Substitution in linear and quadratic functions for probability encoding. Projection of $\bar{B}i$ on $\bar{n}_s$ is equivalent to calculating the distance from $\bar{B}i$ to $l_s$ except for an addive constant. "Distance" to a line can be used as a probability measure. Ellipses can be mapped into Gaussian probability estimates.

Figure 7 (left) mapped $\bar{B}i$ i=1 to 5 into a red intensity which increases towards $\bar{S}$. In our example

$$l_s : \bar{n}_s.\bar{B}-2.8 = \emptyset \quad , \quad \bar{n}_s = 1/2 \quad (1,1)^T$$

$$Red = \bar{n}_s.\bar{B}-2.8 \text{ if } Red > \emptyset \text{ else } Red = \emptyset$$

The reader can easily verify that the measure for Red is equal to projecting $\bar{B}$ on $\bar{n}_s$ and subtracting 2.8. $\bar{S}$ will be the most red pixel in the set. Imagine a complete colour coded PC1,PC2 feature space.

Figure 7 (right) shows an example of the use of quadratic distance functions for probability colour coding. Elliptical functions Qs and Qv are interactively defined e.g. on a graphic screen with a scattergram of the sample set on PC1 and PC2. Based on the clusters in the scattergram approximate ellipses are drawn or axis indicated for those classes of interest. A point $\bar{P}$ in feature space will have distance e.g. ds to ellipse Qs, dv to Qv, etc. Distances ds, dv etc. are mapped into e.g. Gaussian probabilities Ps and Pv. Each P is mapped into a colour intensity which is maximum at the centre of the ellipses and decreases outward. In our example Fig. 7 e.g. Ps→Red and Pv→Green, which leads to a quadratic set of equi-colour lines, with yellow on the (1,1) diagonal. The above principle is applicable to more than 2 classes, with any colour for the centres of the classes.

## 2-Dimensional LUT's

Procedures like probability colour coding and classification involve for each pixel in the file, quadratic equations, division, subtraction and exponentiation. If the problem can be reduced to a 2-D feature space, it is much more efficient to apply the procedure first to a 2-D space and store the result in a 2-D LUT and map the data file through the LUT. In general

$$\bar{P} = (PC1,PC2)^T \xrightarrow{\Gamma} \bar{C} = (R,G,B)^T \; LUT \; (PC1,PC2)$$

With a continuous range and variation of $\bar{C}$ in 2-D space we are still in image processing. If our 2-D feature space is segmented according to some decision rule into few regions with discontinues colour assignment we are in the domain of automatic classification.
I would put the boundary between image processing and multispectral classification at about 16 discrete colours or classes.

## SPATIAL DOMAIN MAPPING

In spatial domain mapping the position pointer $\bar{x}$ in $\bar{B}(\bar{x},t)$ is important and the relation of a pixel to its neighbours becomes important. We need a definition of Neighbourhood. On rectangular sample rasters N-4 or N-8 neighbourhoods are used, hexagonal rasters have nice symmetric N-6.

| N-4 | N-8 | N-6 |
|---|---|---|
| N1 | N1 N2 N3 | N1 N2 |
| N2 x N3 | N4 x N5 | N3 x N4 |
| N4 | N6 N7 N8 | N5 N7 |

Time t in $\bar{B}(\bar{x},t)$ will not play a role as a feature and will be omitted in the notation of this chapter.

Two major distinctions can be made in Spatial Domain mapping; Global, Integral Transforms and Local transforms.

Global, Integral transforms are e.g. Digital Fourier Transform (or Integral), Hadamard Transform, Cosine transform, Karhunen Loève or PC transform etc.

The main reason for the use of global transforms are: availability, electrotechnical brainwashing and laziness. Many people are conditioned to think in terms of high, low and band pass filters and try to design beautiful filters in frequency domain which give horrible results in spatial domain.

Example 1 : a linear Fourier transform on a line of 512 pixels maps a 512-D vector through a 512x512 complex valued matrix (Fourier Kernel) into a 512-D complex vector in frequency domain. This sort of mapping is only useful if it leads to feature extraction or data reduction. However, image features are usually local features, global transforms only mix things up. The only use for Fourier transforms in feature extraction is in applying it to images with sine or cosine spatial variations in intensity.

A general yard-stick for the usefulness of global spatial mappings is their possibility for feature extraction. Eigenfunctions of the spatial process should be optimum in this sense but spatial features are hardly ever global and usually local.

Fast Fourier Transforms are often used for filtering. In some cases it may be faster than the equivalent digital convolution. Because F.F.T. is a matrix multiplication with many coefficients, calculations must be done with high (double or triple) precision floating point processors. On the other hand digital convolution can be done with faster integer processors with expected improvement in processing time through the use of parallel processing. F.F.T. has of course the same conceptual disadvantages as the traditional Fourier Transform.

Local Transforms

The information of a pixel and its local neighbours is transformed into a new pixel value maybe mapped in a different location:

$$\bar{B}(N\bar{x}) \rightarrow \bar{D}(\bar{x})$$

A simple example of this sort of mapping is
Example 2 : resampling and geometric corrections.
Given two Landsat MSS scanlines with sample distance in-scan = 57m, scanline 1 is No. 6 of the previous swath and scanline 2 is scanline 1 of the next swath. In between swaths a 24m shift occurs due to Earth-rotation. We want the data skew-corrected and resampled to an 80m square grid. (The Landsat sample grid of 57m by 79m is ridiculus in view of the point spread function of about 100m in-scan. The only plausible explanation for the 57/79 ratio seems to me compliance with IMB lineprinter output (1960-ies product).

FIG. 8

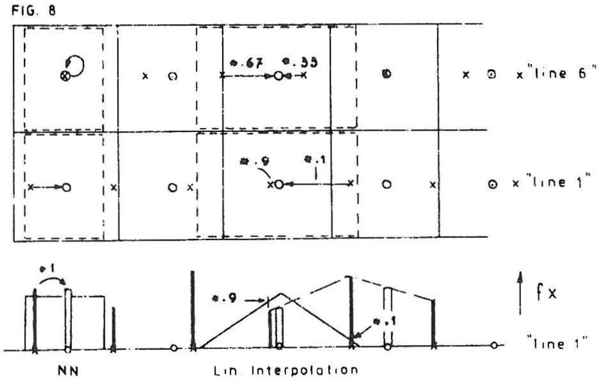



Fig. 8. Resampling of Landsat raster "x" with skew into new square raster "o". Nearest Neighbour maps the old "x" contained in a 57x79 box centered on "o" into "o". Linear interpolation maps "x" values in a box of 114x158m² via a triangular weighting function into "o".

As shown in Fig. 8 there are two components in the mapping: address calculation and intensity mapping. $\bar{X} \rightarrow \bar{Y}$ and $B(N\bar{X}) \rightarrow B(\bar{Y})$.
Mapping $\bar{X} \rightarrow \bar{Y}$ is a simple photogrammetric problem. For $B(N\bar{X}) \rightarrow B(\bar{Y})$ we illustrate NN-Nearest Neighbour and L. INT. Linear interpolation mapping. In NN-mapping a 57m x79 box, centered on $\bar{Y}$ = "o" has a weight = 1 inside and weight = 0 outside the box. For each new "o" the box function is multiplied with the values at "x". The dimensions of the box ensure that only one "x" value will be in each box. The one "x" value is mapped into "o" position. In Linear Interpolation mapping the box is twice as high and wide but the weighting function is triangular with value 1 in the centre and Ø at the borders of the box. The size ensures that always two "x" points are mapped into "o". The closest "x" has the highest weight.

Without saying so we have used the concept of digital convolution. Although two images can be convolved we usually will convolve an image with convolution-operator: a box Neighbourhood with a weight for each point of the Neighbourhood. Spatial Correlation is equivalent with convolution except for cases of asymmetric operators which should be mirror reversed in case of convolution. Convolution is also related to operator-Algebra which is applied in the field of systems analyses and specially in systems which are described by differential equations or difference equations in digital computations.

Difference-Operators which are of much use in digital image processing are:

$D^{-1}$ : $\begin{matrix} 1 & 1 \\ 1 & 1 \end{matrix}$ running average, smoothing operator

$D^{0}$ : 1 identity, original image

$D^{1}$ : $-1 \ \ 1$ d/dx ; $\begin{matrix} -1 \\ 1 \end{matrix}$ d/dy } grad→edge detector

$D^{2}$ : $\begin{matrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{matrix}$ $\frac{d^2}{dx^2} + \frac{d^2}{dy^2}$ 2-D Laplace operator texture enhancement

Most useful image enhancement "filters" can be constructed as a linear combination of $D^x$ operators or repeated self and cross convolution of $D^x$ operators.

A minimum orthogonal set of spatial correlation "vectors" in N-3 are:

$$h\emptyset = \begin{matrix} 1 & 1 \\ 1 & 1 \end{matrix},\ h1 = \begin{matrix} 1 & 1 \\ -1 & -1 \end{matrix},\ h2 = \begin{matrix} 1 & -1 \\ -1 & 1 \end{matrix},\ h3 = \begin{matrix} 1 & -1 \\ 1 & -1 \end{matrix}$$

Each pixel and its 3 Neighbours can be treated as one new data vector. Mapping hØ-3 in the same way gives

$$H = \begin{matrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{matrix}$$ the local 4-D Hadard Transform

4-D data vectors are mapped in 4-D Hadamard domain feature space. This feature space can be used for further mapping. Local features can be coupled through image syntax or by iteration in a pyramid fashion using 1/4 of the pixels in each next iteration on the previous hØ map.

The idea of Operator Algebra can be extended to include not only in-products within the operator Neighbourhood but also include logical functions and state transition LUT's.

Example 3 : The Game of Life.
This game is played on a binary valued (0,1) image $I(\bar{X},t)$ with N = 8. The number of "1" neighbours in N-8 is first determined by convolution with:

$$\begin{matrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{matrix} * I(\bar{X},t) \rightarrow N(\bar{X},t)$$

Logical mapping:

if $I(\bar{X},t) = \emptyset$ and $N(\bar{X},t) = 3$ then $I(\bar{X},t+1) = 1$
else $I(\bar{X},t+1) = 0$

if $I(\bar{X},t) = 1$ and $N(\bar{X},t) = 2$ or 3 then $I(\bar{X},t+1) = 1$
else $I(\bar{X},t+1) = 0$

Equivalent State LUT : I(t+1)LUT(It,Nt)

| It = 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Nt = | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |

Using the concept of a state transition LUT it is very easy and efficient to program all sorts of local image mappings e.g. boundary finding, sceletonising region finding etc.

Local Histogram mapping : $N\bar{X}$, N = 3x3 or N = 5x5 is mapped into a histogram for each $\bar{X}$, the rank order in the histogram is used as a criterion for the new value of the central pixel, e.g.

$$B(N\bar{X}) \rightarrow Hist(N\bar{X}) \rightarrow C(\bar{X}) = \text{modulus}(Hist(N\bar{X})).$$

Can also be used interactively e.g. for improving local consistency.

Local Probability Relaxation

In the final stages of an image processing procedure we should have colour coded class probabilities. So far we only have considered probabilities derived from Multispectral features only. One should include second order statistics also. Some probability vectors are locally compatible (mixed pixels) some are not. One solution to increase local consistency is to change the probability vector of each pixel a small amount in each pass. The direction of the change is guided by a probability consistency matrix. The process is repeated until no improvement is achieved anymore. In that case the probability vectors can be mapped into colour domain.

Deconvolution, image enhancement.

In many sensors smearing of the image occurs for reasons of optical limitations, electronic bandwidth or platform motion. The response of the system to e.g. a small light source on the ground (delta function) will not be one non zero value but a group of values like:

$$\text{Delta } 24 \longrightarrow \begin{matrix} 1 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 1 \end{matrix} \quad \text{PRF}$$

The aim of Deconvolution is to find an operator which transforms the Point Response Function (PRF) as close as possible back to the original single pixel value 24 (delta function). It is easy to prove that the deconvolution operator must have some negative Neighbourhood values. Always a compromise has to be found between sharpness (delta function) and "ringing" ripples around the central value.
As resampling with NN or interpolation involves smoothing we combine the operations:
geometrical correction, resampling, interpolation and deconvolution into one generalised "convolution" operator.
Deconvolution operators and comparable operators like $(2D^{\emptyset}-D^{2})$ produce enhanced, more brilliant and sharp images. The eye-brain system is not much bothered by the traditionally predicted noise enhancement.

TEMPORAL DOMAIN MAPPING

The classical error in the processing of MT Multitemporal delta is to map $\bar{B}(\bar{X},t)$ and $\bar{B}(\bar{X},t+1)$ into the same feature space and apply the local scientific subroutine library.
A simple meaningful way of treating MT data is e.g.

$$\bar{B}(\bar{X},t) - \bar{B}(\bar{X},t+1) \rightarrow \bar{C}(\bar{X},t+1)$$

this is the process of change detection which is important for monitoring processes which are supposed to be constant or show a predictable change.
An improvement on this schema is prediction-correction mapping:

$$\bar{B}(\bar{X},t) \rightarrow \bar{P}r(\bar{X},t+1).$$
$$\bar{P}r(\bar{X},t+1) - \bar{B}(\bar{X},t+1) \rightarrow \bar{C}(\bar{X},t+1) \text{ etc.}$$

This process can be repeated e.g. through the season.
In case a prediction $\bar{P}r$ cannot be made with high enough accuracy $\bar{P}r$ takes the form of a set of hypothesis $\bar{H}y(\bar{X},t)$, each hypothesis leads to a prediction. Predictions are then compared with measurements and the hypotheses are updated with extra information. This is the process of converging evidence or sequential decision making. The process should be interactive.

$$\bar{B}(\bar{X},t) \rightarrow \bar{H}y(\bar{X},t) \rightarrow \bar{P}hy(\bar{X},t+1)$$
$$\bar{P}hy(\bar{X},t+1) - \bar{B}(\bar{X},t+1) \rightarrow \bar{C}(\bar{X},t+1)$$
$$\bar{C}(\bar{X},t+1) \rightarrow \bar{H}y(\bar{X},t+1) \rightarrow \bar{P}hy(\bar{X},t+2) \ldots\ldots \text{ etc.}$$

All the above processing takes place in probability domain which can be visualised in colour.

In feature space we have to think in terms of clusters moving in a yearly cycle for vegetation classes and jittering a bit for "constant" classes because of imperfect radiometric and geometric corrections.

Accurate relative geometric corrections using a sophisticated convolution operator or cubic convolution is a first requirement for operational use of satellite R.S. data.

SUMMARY

Starting with a general concept of image processing as part of a decision making procedure. I have given examples on how the mathematical tool of mapping is used to convert raw data into colour coded class- or state-probabilities. Spatial neighbourhood mapping can be used to improve local consistency of class probability and help in spatial image segmentation as pre-classification. The most interesting problems occur where we include the dynamics of processes on the earth's surface as a movement of vectors in a feature space. Using the concept of probability vectors the use of predictor-corrector methods is indicated which may include hypothesis building and testing in a converging evidence method.

In my view, more emphasis should be placed on the decision preparation aspects of digital image processing and less on traditional map making. In teaching and understanding, the concept of feature spaces is very important.
More emphasis should be placed on monitoring and forecasting, with integration of other image data such as meteorological data, existing topo- and other maps, statistical surveys, etc.

## REFERENCES

The list of references is limited to the category recommended reading and papers which refer to some concepts mentioned in this paper but published elsewhere.
For reading on current topics:

IEEE Transactions on PATTERN ANALYSIS AND MACHINE INTELLIGENCE.

Excellent illustrations of image processing and a reasonable processing phylosophy can be found in

J.P.L. PUBLICATIONS and TECHNICAL REPORTS.

A tutorial article on the meaning of the Principal Component Transform and the use of Probability coding is:

N.J. Mulder - "HOW TO ESCAPE FROM 4-DIMENSIONAL SPACE, or STATISTICS VERSUS KNOWLEDGE". ITC-Journal 1975-3, pp. 352-359.

An application of our phylosophy to low technology environments can be found in:

N.J. Mulder & N.H.W. Donker - "POOR MAN'S IMAGE PROCESSING - A STIMULUS TO THINKING" Proc. Int. Sym. Image Proc. Graz. Oct. 1977, pp. 155-162.

A proposal on on-board spectral correlation with colour coding and spatial correlation with feature extration and line coding can be found in a paper for the I.A.F. (Internation Aeronautical Federation) Münich, Sept.1979.

N.J. Mulder - "FEATURE SPECIFIC SPATIAL- AND SPECTRAL SENSORS - ARTIFICIAL RETINAS IN SPACE" I.A.F. paper, Sept. 1979, Münich.

The "Turing" computer of digital image processing can be found in:

N.J. Mulder - "THREE ELEMENTARY TOOLS IN THE PROCESSING OF REMOTE SENSING DATA - A UNIFYING APPROACH" 3RD B.I.S. Computers and Space Technology Conference, Appleton lab., Slough, Nov. 1979. To be published in B.I.S. Journal.

END