Photogrammetric Record, 14(82): 595–611 (October 1993)

FEATURE EXTRACTION IN DIGITAL PHOTOGRAMMETRY

By W. FÖRSTNER
Universität Bonn

(Paper read at a Technical Meeting of the Society on 17th March, 1992)

Abstract

Feature extraction is the first and crucial step of all image analysis procedures. Together with their mutual relations, features form perceptual structures on which hypothesis generation and reconstruction processes are built. The article gives a framework for feature extraction within an interpretation system, discusses a possible object and image model useful for building up perceptual structures and gives examples for the use of various features for standard photogrammetric tasks.

INTRODUCTION

DIGITAL PHOTOGRAMMETRY is the field of research and application dealing with the extraction of geometric and thematic information from digital or digitised images. Without being obvious, it became part of the broad area of computer vision and image understanding. On one hand this requires a new definition of what is typically "photogrammetric": the geometric aspects of image analysis, the still photographic data capture and/or the photogrammetric fields of application, especially topographic mapping? On the other hand, the enrichment of the tool box of image analysis opens the door to full or at least partial automation of image interpretation. This requires a redefinition of the basic concepts in our field; the physical aspect of the imaging process can no longer be reduced to the perspective geometry including the necessary calibration procedures, but it has to be modelled from the very beginning thus closing the research gap between photogrammetry and remote sensing. At the other extreme, an explicit modelling of the objects to be extracted with image analysis tools is necessary, linking photogrammetry with photo-interpretation, with data modelling in geoinformatics and with knowledge representation in artificial intelligence.

This article discusses the impact of automatic feature extraction using techniques from pattern recognition and image understanding for use in present and future digital photogrammetric systems. It tries to illustrate the new changes with respect to concepts and procedures, using examples from recent photogrammetric research and development. The topic of feature extraction seems to be best suited for this purpose as it is at the heart of all automatic and semi-automatic systems for image analysis and hence also for digital photogrammetry.

In the following account we will first place the task of feature extraction into the larger framework of an image interpretation system. We thereby discuss the many facets of feature extraction and show their role as links between the image(s) and the objects to be extracted. Then we discuss a possible framework for feature extraction and illustrate its use for the extraction of points, lines and regions including their relations. The use of various types of symbolic image descriptions obtained in this way will then be shown for standard photogrammetric tasks, namely aerial triangulation, digital elevation model generation, control point location and object reconstruction.

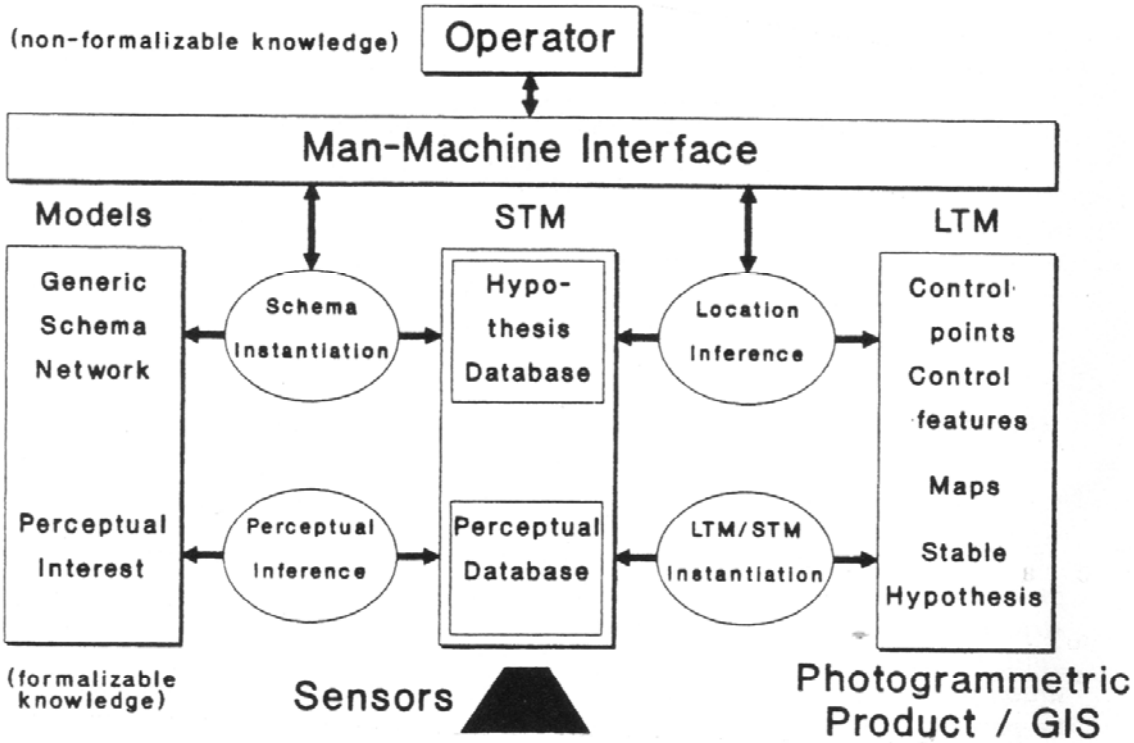


FIG. 1. A possible set-up for an interactive interpretation system (adapted from Lawton *et al.*, 1987). The *Oxford English Dictionary* defines INSTANTIATION as "the action or fact of instantiating; representation by an instance." The term appears to have originated among philosophers in the 1940s and 1950s. Lawton *et al.* used the term in a computer science context of giving a specific value to a variable.

CONCEPTS FOR AN INTERPRETATION SYSTEM

The connexion of computer and image content allows the inclusion of automatic image interpretation modules in a photogrammetric system. Mensuration and orientation procedures then appear as submodules in a larger system which, more or less autonomously, selects these procedures among others for identifying objects. As identifying objects, say points, like the corner of a building, is an interpretation task which has, up to now, been performed by the human operator, interpretation appears to occur prior to accurate mensuration. The measurement appears to be simple when we consider the complexity of the corresponding algorithms for interpretation and the amount of knowledge required.

Up to now, the process of image interpretation has not been well understood in detail. However some components which seem to play a central role are known and, perhaps better still, are agreed upon. Fig. 1 shows a possible set-up of an interpretation module, adapted from a system for autonomous vehicle guidance (Lawton *et al.*, 1987). It explicitly refers to the image *features*, here called *perceptual structures*, stressing the necessity of rich image descriptions for image interpretation.

The concept distinguishes between modules which contain possibly changing information or knowledge (boxes) and modules which operate on these information or knowledge sources (ellipses). There are three groups of information sources.

Models

Interpretation of images (or other data) consists of matching models of the objects to be extracted with the data obtained by some sensor. Therefore the *object models* have to contain a rich enough structure in order to be able to represent all possibly visible parts. They therefore have to contain relations between parts of the objects (roof *is part of* house), classification (house *is a* building) as well as relations between objects (house *is near* road). The *appearance* of objects needs to be stored as well, as it is the link to the measurements to be taken. Obviously no detailed description of the objects is possible, only some generic knowledge may be encoded in the models leaving both

structural details (garage *right or left or behind* house) as well as actual values of parameters unmodelled (length = 10 m).

As the perspective projection preserves certain relations between different objects (neighbour relation), these may be used in the interpretation process and therefore have to be stored, thus to be modelled explicitly. The neighbour relation and the knowledge about likely occlusions are the basis for an *invocation network* (cf. Bruce *et al.*, 1989). The inferred existence of one object (e.g. sky) leads to the set up of an hypothesis of another object (e.g. telephone wire) to be searched for in the image. As this search is usually limited to a certain area, such as at the boundary of the known object, the *perceptual interest* of the system may be guided by this part of the scene model.

Long Term Memory

The goal of the interpretation is to find a coherent description of the scene. This may have to be stored for later analysis, as in a geographic information system (GIS). *Prior knowledge* about the scene, such as maps or control points again possibly available in a GIS, may be used to trigger the interpretation or just to find the orientation of the camera. One may interpret this type of knowledge as belonging to the long term memory of the system. It is gradually updated by inserting a *stable hypothesis* derived from the image data. Obviously these hypotheses have to have the structure given by the object models and thus are instances of these models.

Short Term Memory

The link between the models and the long term memory is established by data derived from the images. This is achieved by image descriptions on a level where they can be linked to the appearance models of the objects to be extracted. These image descriptions consist of more or less rich *perceptual structures* or *features* which may be derived using no or only minimal specific knowledge about the scene. The structures then may be used to set up hypotheses about the image content (such as using spectral or geometric attributes of regions). These hypotheses then have to be checked until they are rejected or until enough supporting evidence has been collected, such as by using the invocation network and the relations between the objects, so that they may be treated as stable hypotheses and stored in the long term memory.

Operator

The complexity of natural scenes as well as the great variety of tasks makes a human operator indispensable when building an interpretation system. His responsibility lies in guiding the perceptual interest, in helping to substantiate good hypotheses or in supporting the location inference in the case of difficult matching situations. The necessary man–machine interface has to link the operator internal with the computer internal modelling of the system. This reveals the *control* of the system (not shown in Fig. 1) to be the central problem.

This unsolved problem, which was the motivation for the interactive set-up of the image understanding environment (IUE) (cf. Mundy *et al.*, 1992), is under active current research in the artificial intelligence community.

There are several systems which show this kind of set-up.

(a) The Visions system (Bruce *et al.*, 1989) is probably the most advanced and contains parts of hierarchies and appearance models. It uses bottom up and top down strategies and works on a blackboard architecture. It is built for analysing outdoor scenes for visual navigation.
(b) The Condor system (Strat and Fischler, 1991) aims at interpreting natural scenes. The object models contain rich structural context relations between likely neighbouring objects. The strategy efficiently selects consistent cliques and uses the "core knowledge system" (Strat and Smith, 1987) and is also the basis for a digital photogrammetric workstation (Hanson and Quam, 1988).
(c) The Ernest system (Niemann *et al.*, 1990) is a shell for pattern analysis using

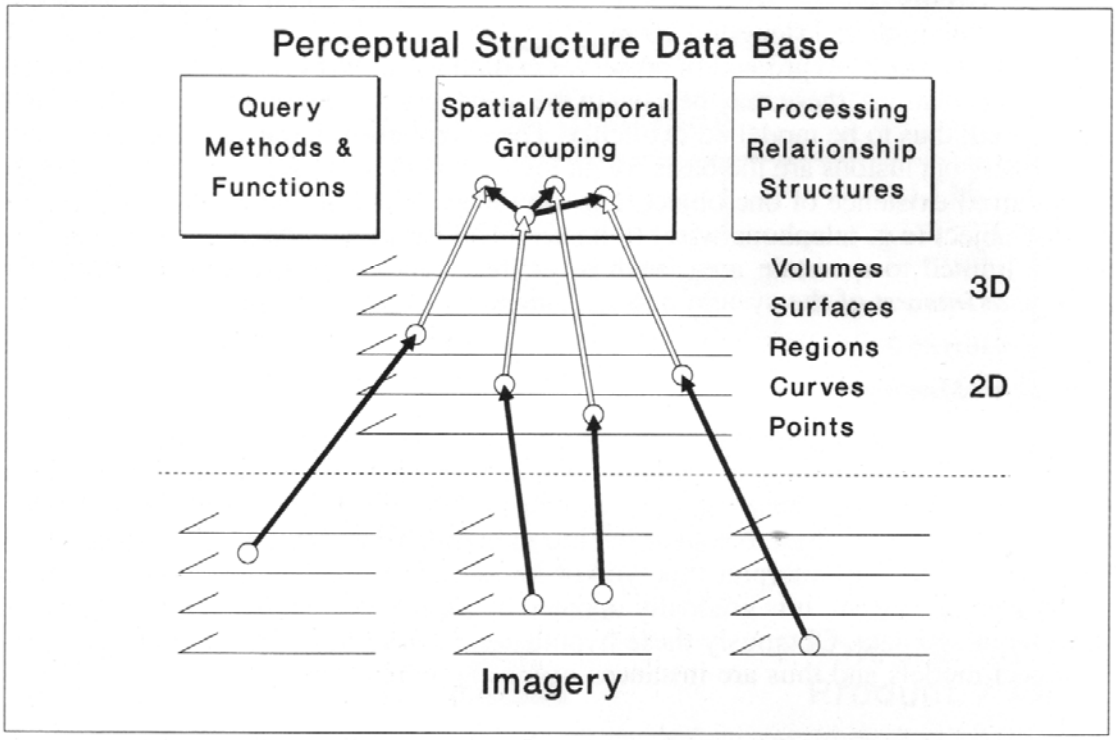


FIG. 2. The structure of the perceptual structure data base (from Lawton *et al.*, 1987).

semantic networks. It allows handling of specialization and containment hierarchies using the A^* algorithms as search strategy. Applications occur in both image and speech analysis (Sagerer, 1985).

(d) The "hierarchical structure code" developed by Hartmann (1983) is a highly organized perceptual data base, containing an iconic and a symbolic image pyramid. Using the Ernest system, it is used for identifying machine parts (Mertsching *et al.*, 1990).

Without going into more detail, we want to discuss the perceptual image structures as the basis for a clear understanding of the feature extraction problem.

PERCEPTUAL STRUCTURES AND FEATURE EXTRACTION

The Perceptual Structure Data Base

The perceptual structures obviously form the link between the sensor data and the appearance models. The perceptual structures have to be derived from the images by a bottom up procedure by using no knowledge or only low level knowledge about the scene. They consist of distinct points, edges and regions in the two dimensional images; these are the basic image features which, after being matched with structures of other images, lead to surfaces or volumes. They may be grouped based on their proximity, their symmetry and their spatial or temporal coherence (Fig. 2).

The perceptual structure can only partly be built in a bottom up manner. During the interpretation, further structures may be derived based on hypothesised expectations in the image (for example, in case a house has to be found with a certain orientation, special filters may be used to find low contrast edges with a prespecified orientation). Due to the amount of structures which have to be processed and because the perceptual structures are the basic primitives for the whole interpretation process, they may be organised in a data base with special tools for queries or methods for changes and for processing the perceptual structures. Fig. 2 suggests the data base to be extensible in a way that hierarchical image structures are made explicit. On the other hand only the lower levels of such structural pyramids are well understood unless the structure of the pyramid is homogeneous through several levels as in the hierarchical

structure code of Hartmann (1983). In general the aggregation criteria should be found by the system itself, based on the available object knowledge leading to a variable structure with increasing aggregation level (Lindeberg, 1990).

The Role of Feature Extraction

Features, playing the linking role between image and object, are characteristic or striking parts of objects. Describing objects is identical to describing their features. It also is the basic paradigm of artificial intelligence that the real world may be modelled and described this way.

Feature extraction in this context may now be seen to be identical to filling the perceptual structure data base. Depending on the complexity of the features, we may distinguish three types:

(a) low level features which are attributes of the pixel arrays of the images. Examples are spectral features used in multispectral classification, textural features used for segmentation or temporal features used for change detection. No labelling of the features is assumed to be available;
(b) mid level features which are either geometric primitives such as points, edges or regions or they are aggregates of these primitives including their relations. As mentioned above, they may be further grouped into more complex structures. Again, it is essential that no meaning needs to be attached to these primitives or relations; and
(c) high level features which are already interpreted, usually quite complex structures derived from the images. Examples are "a roof" or "a tree trunk" and are thus parts of the images where a meaning has already been attached. Cartographic features derived from an aerial image belong to this group.

Strictly speaking, spectral features which are also labelled, for example "grass land", should be termed high level as a meaning is attached to the feature. This reveals that the above distinction is no rigorous classification. It is meant, however, to help classifying the different notions to which the term "feature" refers.

We now want to concentrate on the extraction of mid level features which are understood best and which allow the derivation of rich enough image descriptions which are useful for various photogrammetric tasks. We therefore refer to points, lines, regions and their relations as features without further specification.

A Concept for Feature Extraction

As features are the observable elements on which any image analysis procedure is based, modelling feature extraction is equivalent to modelling the observation process. This requires a specification of the object model used from which the appearance or image model may be derived using knowledge about the imaging process employed. The image model itself should then contain the necessary components which define the goal of the feature extraction procedures (cf. Förstner, 1993).

The Object and the Image Model

We use a very much simplified object model in the sense that no high level or semantic knowledge is involved. However, it is complex enough to cover the geometry and physics of a larger class of objects. We assume the object to have the following properties:

(a) the surface of the object consists of piecewise smooth patches;
(b) the boundaries of the surface patches consist of piecewise smooth boundaries;
(c) the albedo of the object consists of piecewise smooth patches;
(d) the boundaries of the albedo patches consist of piecewise smooth boundaries; and
(e) the surface and albedo patch boundaries may or may not coincide.

Instead of the surface or the albedo, any locally computable function of the surface geometry or of the albedo may be assumed to be piecewise smooth. This is essential when modelling textured objects, where the texture may result from local geometric or physical variations. For special applications we restrict the surface to consist of planar patches and thus the objects to be polyhedra.

This object model is motivated by the coherence of matter. No objects with fractal type or ill defined boundaries are included. Here the strong dependency of the modelling on the scale of aggregation becomes obvious, which depends heavily on the application.

Based on this object model and assuming a homogeneous illumination, we obtain the image model. We assume the image function to have the following properties:

(a) the image consists of segments;
(b) the image intensities or locally computable functions are assumed to be piecewise smooth within the segments; and
(c) the segments show piecewise smooth boundaries.

The image model explicitly refers to segments with piecewise smooth boundaries. Therefore one may describe the image with a set of basic features, namely points, edges and regions and their mutual relations. The points may either be boundary points of high curvature or nodes where three or more regions meet. Boundaries may result from depth discontinuities (contours), orientation discontinuities (break lines), albedo discontinuities (material changes) or, in the case of shadows, illumination discontinuities (Binford, 1981).

The image model is observed to contain a region overlay of projections of the surface and the albedo patches. Therefore the interpretation of the result of a segmentation needs to recover this overlay. Also no guarantee is given for the observability of the patch boundaries. They may very well disappear in the image, making the segmentation an ill posed problem which requires regularization based on higher level knowledge about the scene to be recovered.

Formalization of the Image Model

In the most simple case of a grey level image this model may be formalized in the following manner. We start with the continuous image. The image area $\mathcal{I}$ is segmented into three sets:

(a) homogeneous *segments* $\mathcal{S}_i$
(b) smooth *boundaries* $\mathcal{B}_j$ and
(c) *points* $\mathcal{P}_k$.

Points are positions where the boundary is not smooth, or where three or more segments meet. The segments $\mathcal{S}_i$ are assumed to be open two dimensional sets, the boundaries $\mathcal{B}_j$ are one dimensional sets, embedded in two dimensions, whereas the corners are single points $(x, y) \in \mathcal{I}$. Of course the boundaries may also be assumed to be part of the segments, but then neighbouring segments, theoretically, share a boundary and no classification of *all* image points into three disjoint sets is possible anymore.

Thus we obtain the decomposition of the image area:

$$\mathcal{I} = \mathcal{S} + \mathcal{B} + \mathcal{P} = \bigcup_{i=1}^{n_s} \mathcal{S}_i + \bigcup_{j=1}^{n_b} \mathcal{B}_j + \bigcup_{k=1}^{n_p} \mathcal{P}_k. \tag{1}$$

We now have to choose a homogeneity or smoothness measure, say $h(x, y)$. It should be suited so as to distinguish the three different classes of image points.

One possible choice is based on the average squared gradient of the image function f:

$$\overline{\Gamma_\sigma f}(x, y) = G_\sigma(x, y) * \Gamma f(x, y) \tag{2}$$

with

$$\Gamma f(x, y) = \nabla f(x, y) \nabla f(x, y)^T = \begin{pmatrix} f_x^2(x, y) & f_x(x, y) f_y(x, y) \\ f_x(x, y) f_y(x, y) & f_y^2(x, y) \end{pmatrix} \tag{3}$$

and $G_\sigma(x, y)$ being the centred normal density function with standard deviation σ.

We first distinguish between segments on one hand and boundaries and points on the other hand by investigating the *trace* of the average squared gradient being a meaningful homogeneity measure:

$$(x, y) \in \begin{cases} \mathcal{S} & \text{if } tr\ \overline{\Gamma_\sigma f}(x, y) < T_h(\sigma) \\ \mathcal{B} \cup \mathcal{P} & \text{else} \end{cases}. \tag{4}$$

The threshold may be chosen to lead to $3 - \sigma$ wide boundary regions. In the limit $\sigma \to 0$, these regions then become thin lines.

In the second step, we separate boundaries and points by investigating the curvature

$$\kappa_\sigma^2(f) = \frac{1}{\sigma^2}\frac{\lambda_2}{\lambda_1} \tag{5}$$

of the isolines (isophotes) in the non-segment area. It depends on the ratio λ_2/λ_1 of the two eigenvalues of the homogeneity measure $\overline{\Gamma_\sigma f}(x, y)$. We therefore obtain

$$(x, y) \in \begin{cases} \mathcal{B} & \text{if } \kappa_\sigma(f) < T_\kappa(\sigma) \\ \mathcal{P} & \text{else} \end{cases}. \tag{6}$$

The *true digital image* $f(r, c)$ (r = rows, c = columns) is obtained by sampling, that is by discretization and quantization of the continuous image leading to an array of integers. The observed image g is a contaminated version of the true signal f with possibly signal dependent noise n:

$$g(r, c) = f(r, c) + n(r, c). \tag{7}$$

The proof of (5) uses the function $g(u, v) = au + bv^2$. It is an approximation of the image function $g(x, y)$ at places with $|\nabla g| \neq 0$, with u being the gradient direction and v perpendicular to it. It has constant slope $g_u = a$ in u direction and constant curvature $\kappa = -g_{vv}/g_u = -2b/a$ of the contour lines (isophotes) at all points of the u axis and especially at (0, 0). With $\nabla g = (a, 2bv)^T$, the average squared gradient

$$\Gamma g(0, 0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \begin{pmatrix} a^2 & 2abv \\ 2abv & 4b^2v^2 \end{pmatrix} G_\sigma(u, v)\,du\,dv = \begin{pmatrix} a^2 & 0 \\ 0 & 4b^2\sigma^2 \end{pmatrix} \tag{8}$$

has eigenvalues $\lambda_1 = a^2$ and $\lambda_2 = 4b^2\sigma^2$ from which $\lambda_2/\lambda_1 = \sigma^2\kappa^2$ follows.

Feature Extraction

We are now able to describe the extraction of the basic features of a digital image. Homogeneous and non-homogeneous regions are distinguished by testing the homogeneity measure, here $trace\,\overline{\Gamma_\sigma g}$, thus the gradient magnitude, to be significantly larger than 0, must be larger than a threshold $T_h(\sigma)$. Boundary and corner regions are distinguished by testing the curvature $\kappa_\sigma(f)$ of the boundary, to be significantly larger than 0, that is larger than a threshold $T_\kappa(\sigma)$. Both thresholds $T_h(\sigma)$ and $T_\kappa(\sigma)$ can be made dependent on the noise characteristics, the width of the used kernel G_σ and a prespecified significance level. Using $\sigma > 0$ leads to boundary regions and point regions. The thin boundaries (one dimension) and points (zero dimension) then lie within these regions.

Fig. 3 shows the principle of this classification of the image content. The low level feature $tr\,\overline{\Gamma_\sigma g}$ (Fig. 3(b)) and $\kappa_\sigma g$ (Fig. 3(c)) of the house image (Fig. 3(a)) are grouped into regions (Fig. 3(d)) which obviously reflect the image structure correctly. A precise location of the boundary (that is edge and corner elements) may be achieved by a two step procedure which first searches for optimal windows and then estimates an optimal position of the boundary or the corner within these windows. For locating corners we proceed as follows. The centres of the most likely windows for determining the point location are found by searching for the relative maxima of $tr\,\overline{\Gamma_\sigma g}$. Then the optimal position $\mathrm{p}_o = (r_o, c_o)^T$ within these windows may be determined from the equation system

$$[G_\sigma * \Gamma g] \cdot \mathrm{p}_o = G_\sigma * [\Gamma g \cdot \mathrm{p}] \tag{9}$$

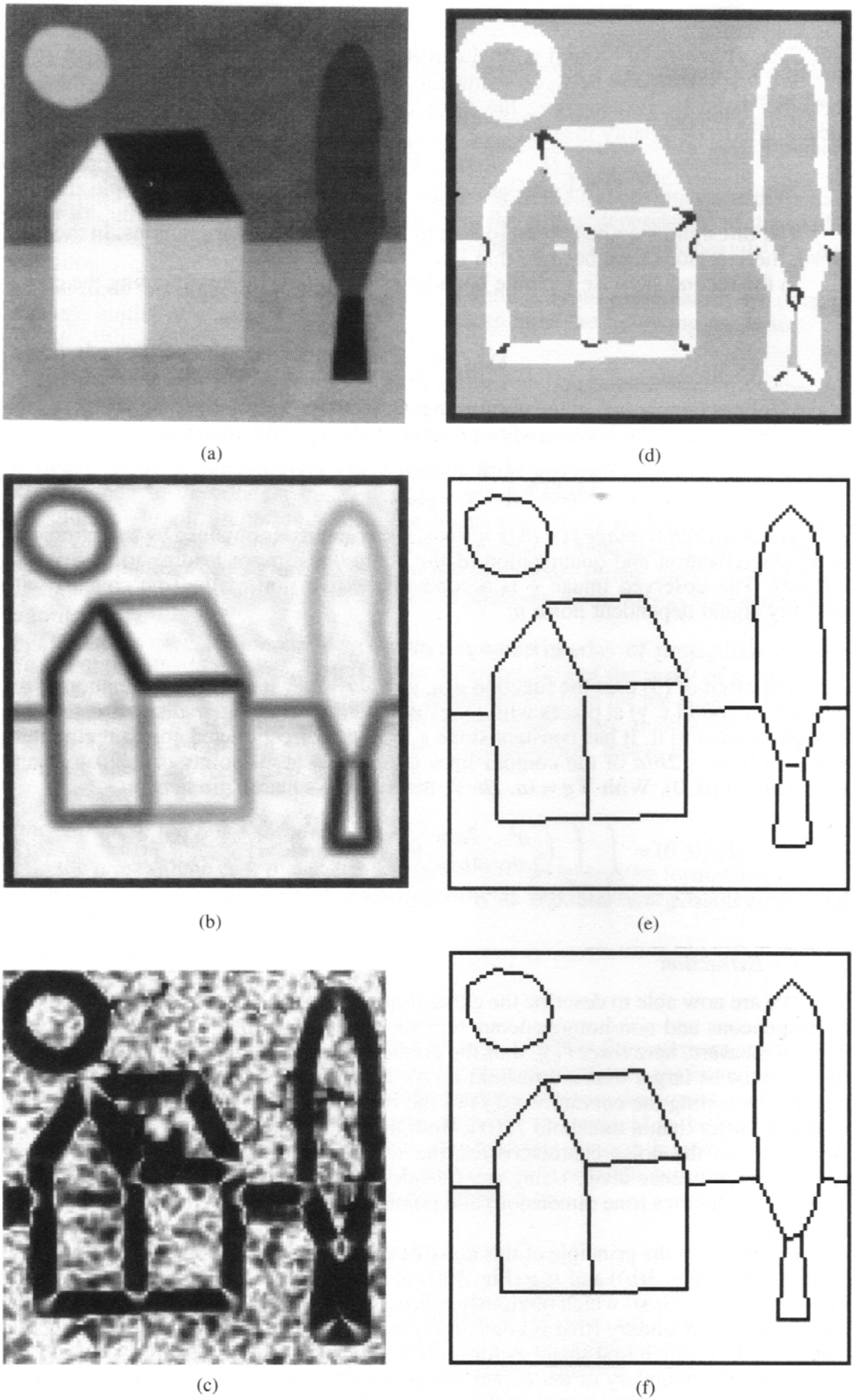


FIG. 3. A house (a), the homogeneity and the curvature measure (b) and (c), the three classes of image points (d), the extracted edges/boundaries (e) and the final segmentation (f).

where $p = (r, c)^T$ denotes all pixel positions in the used window $\mathcal{W}$, which when replacing G_σ by a simple box filter yields $(\sum \Gamma g) \cdot p_o = \sum(\Gamma g \cdot p)$ or explicitly (Förstner and Gülch, 1987)

$$\begin{pmatrix} \sum_{\mathcal{W}} g_r^2 & \sum_{\mathcal{W}} g_r g_c \\ \sum_{\mathcal{W}} g_r g_c & \sum_{\mathcal{W}} g_c^2 \end{pmatrix} \begin{pmatrix} r_o \\ c_o \end{pmatrix} = \begin{pmatrix} \sum_{\mathcal{W}} (g_r^2 r_o + g_r g_c c_o) \\ \sum_{\mathcal{W}} (g_r g_c r_o + g_c^2 c_o) \end{pmatrix} . \tag{10}$$

Observe that p_o in (9) is a weighted centre of gravity with the weights being proportional to Γg and decreasing with increasing distance from the centre of the window. (9) and (10) are especially useful for locating corners of polyhedra without explicitly extracting edges.

Similarly we obtain the procedure for detecting and locating edges. The centres of the most likely windows for the determination of the edge location are the relative maxima of $tr\,\Gamma_\sigma g$ across the edges, that is in the gradient direction. An optimal position p_o may be determined from

$$(G_\sigma * \Gamma g + \sigma_o^{-2} I) \cdot p_o = G_\sigma * (\Gamma g \cdot p) \tag{11}$$

where the term $\sigma_o^{-2} I$ is necessary to stabilize the edge position along the edge. Other approaches for determining the precise edge position are available (Canny, 1986).

The amount of averaging is determined by the choice of σ which may be different for edge and corner detection and should be determined automatically or given by the object model.

The result of this step consists of lists of edge and corner elements, possibly with further attributes such as contrasts or degree. The edge elements then have to be grouped according to their neighbour relations involving chains of edge elements. These may be approximated by piecewise smooth curves or polygons. The relations of the edge and corner regions with respect to the homogeneous regions may be used to finally arrive at a rich symbolic description.

The line segments (Fig. 3(e)) of the house image (Fig. 3(a)) are grouped (Treutler, 1992) and lead to the final segmentation (Fig. 3(f)). All relations between points, lines and regions may be collected in the perceptual structure of that image.

The next section discusses the use of image features and their relations for photogrammetric tasks.

Photogrammetric Applications of Feature Extraction

This section intends to give an impression of the potential of the use of automatic feature extraction for classical photogrammetric tasks. Aerial triangulation (AT) and digital elevation model (DEM) generation are typical for using point type features. The identification of buildings as natural control points for efficient exterior orientation can be based on sets of straight line segments. The mutual links between features may be used for both image to map matching based on structural descriptions and for interpretation.

Point Features

Points are the basic elements for classical photogrammetric tasks. The manual selection of points and the numbering are difficult, time consuming and error prone procedures. They can be fully automated. In most cases the numbering is only of internal value as long as the selected points do not carry a meaning. This holds for point transfer in AT and, to a certain extent, for point mensuration in DEM generation.

Automation of point selection and numbering can be done very efficiently. Therefore the strategy for selecting points can be changed completely. Instead of selecting a small number of well defined points which carry rich information with respect to precision and possibly for interpretation but show unfavourable local redundancies, one can now select a huge number of points which, when seen individually, may not carry much information with respect to precision and reliability, but allow us to exploit the resultant extremely high redundancy to achieve very robust and stable results. Formally

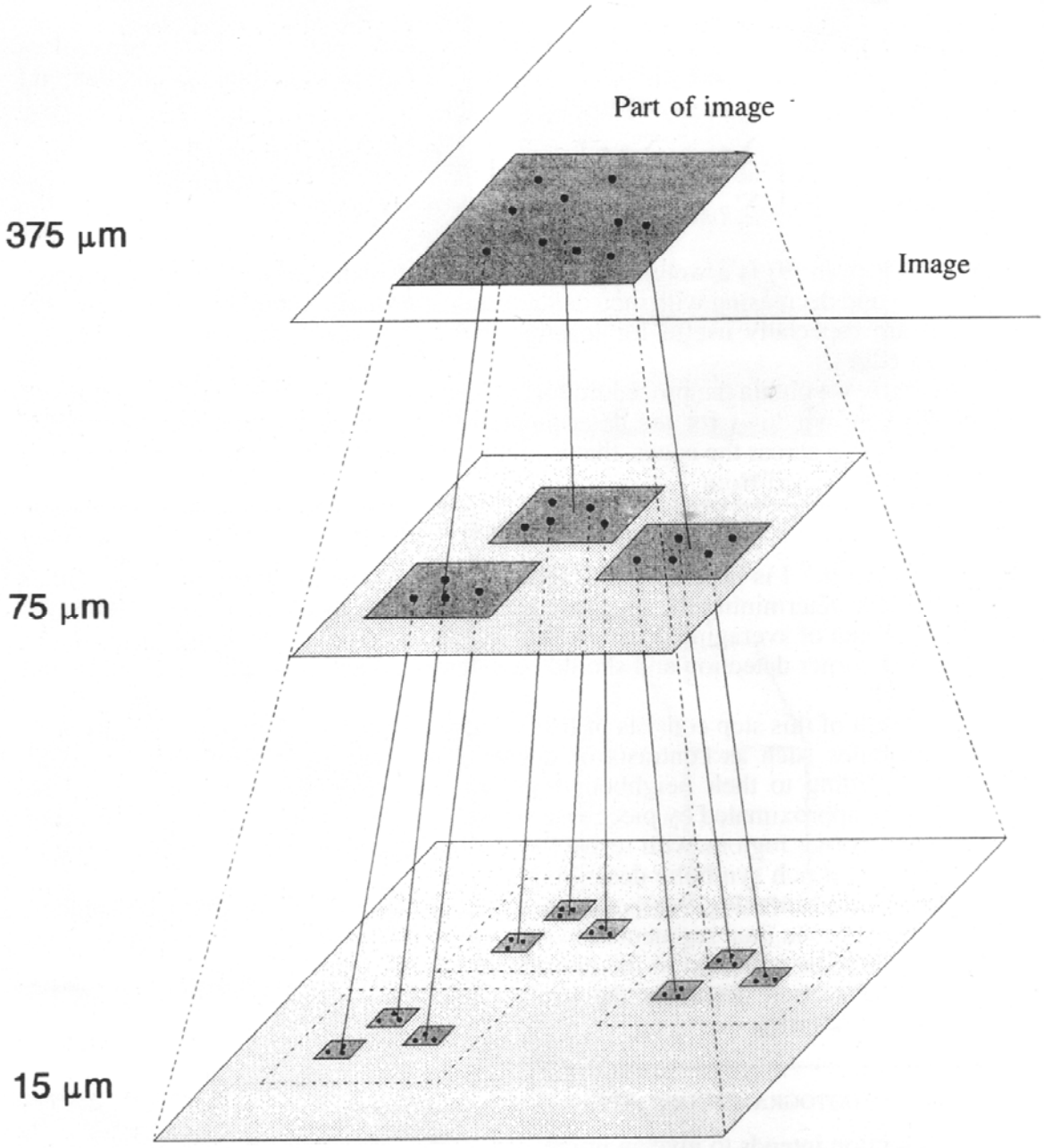


FIG. 4. The coarse to fine strategy of point transfer using a point feature pyramid. The matching results of the coarser levels serve as approximate values for the matching in the finer levels (from Tsingas, 1992).

using groups of 50 points of five times higher standard deviation yields the same result as highly accurate point pairs, but allows simple, automatic and therefore internal error checking. As, on average, the automatically selected points are comparable or even better than manually selected points, we can expect an overall increase of the performance when using automatic feature extraction procedures.

Automatic Point Transfer for Aerial Triangulation. Tsingas (1992) developed a concept for full automation of AT based on digitized aerial images. In case approximate values for the orientation parameters (for example, by GPS) and some manually measured control points are available, the selection and the numbering of the tie points can be done without human interaction. He extended the point based matching procedure (Paderes *et al.*, 1984; Förstner, 1986) to an n stage matching procedure, allowing each automatically selected point to appear in any number of overlapping images (Fig. 4). The results are automatically selected and points numbered with their image coordinates. For efficiency reasons, the matching or the point transfer is performed in a coarse-to-fine manner using images of different resolution. He used 375 μm, 75 μm and

15 μm (Fig. 4). The result of the matching in the lower resolution serves as an approximation for the match in the next finer resolution.

Tsingas tested the approach with a close range 60 per cent sidelap bundle block with five strips consisting of five CCD images each. The system automatically selected 1060 tie points leading to 5934 observations. The block with 3330 unknowns and a redundancy of 2604 was highly overconstrained. The estimated accuracy of the tie points was $\sigma_0 = 0{\cdot}29$ pixel = $4{\cdot}5$ μm which is extremely good for natural tie points. The absolute accuracy was $\mu_{xy} = 1{\cdot}5$ μm in planimetry and $\mu_z = 0{\cdot}5$ pixel = $7{\cdot}5$ μm measured at image scale which corresponds to the theoretical expectation. The main fact, however, seems to be that there were no outliers in the block adjustment. This proves the robustness and reliability of the automatic approach.

DEM Generation. Manual mensuration of a DEM contains several steps. These are the mental match of the images on the retina of the operator, which results in a dense (internal) surface description, the interpretation of this three dimensional model, an application based selection of representative object features, a precise location of the selected feature in three dimensions and finally a computer match exploiting the perspective geometry. Automation has to start from the digitized images and in principle must follow these steps. As the derivation of a dense three dimensional model, the interpretation of the model with respect to the definition of a "topographic surface" and its definition under general conditions are unsolved problems up to now, only simplified solutions for DEM generation have been realized. However, they cover quite a large percentage of applications.

In cases when the visible surface is piecewise smooth and can be assumed to be identical to the topographic surface almost everywhere, powerful matching algorithms are available. Algorithms based on features have proved to be more robust and efficient in contrast to intensity based matching procedures. The Match-T system (Krzystek and Wild, 1992; Krzystek, 1991 and 1992) contains the feature based matching procedure specifically designed to exploit the knowledge of the relative orientation. Also a coarse-to-fine strategy is used to gradually refine a finite element description of the topographic surface (Ackermann and Hahn, 1991). Depending on the image content some 1 million to 3 million points per image are selected, yielding some 0·3 million to 0·8 million matched three dimensional points which finally result in a very dense DEM with some 50 000 to 80 000 grid points. The high redundancy of about 10 points per grid mesh, in conjunction with a grid which is a factor 3 to 5 times finer than available with manual methods, is the reason for the high precision and reliability of the system.

The results in extensive controlled tests showed mean deviations of less than 0·1‰ of the flying height. Due to the robust procedure incorporated in the algorithm, single trees or houses do not disturb the shape of the topographic surface (Krzystek, 1992). The computing times of 30 minutes to 1 hour per model on a Silicon Graphics workstation (4D35) demonstrate the feasibility of the use of automatic feature extraction procedures for an important photogrammetric task.

Absolute Orientation using Line Segments

Point transfer, DEM mensuration as well as relative orientation are based on the matching of features derived from data sources which have the same representation, namely a raster. This simplifies the procedures as systematic effects of the feature extraction disappear. For example, features detected in one image are likely to be detectable in the other image. Absolute orientation, locating control points and more general object location require the matching of features derived from data sources of different representation in general. In cases when one is able to find a representation structure which is common to image and object features, a matching procedure may enable correspondence to be established. However, it has to cope with the generally large amount of background features which do not belong to the object to be detected and located.

Schickler (1992) developed a system for automating the exterior orientation of a single image. It is based on control "points" which consist of a list of straight three dimensional line segments, whose co-ordinates are known in an object centred

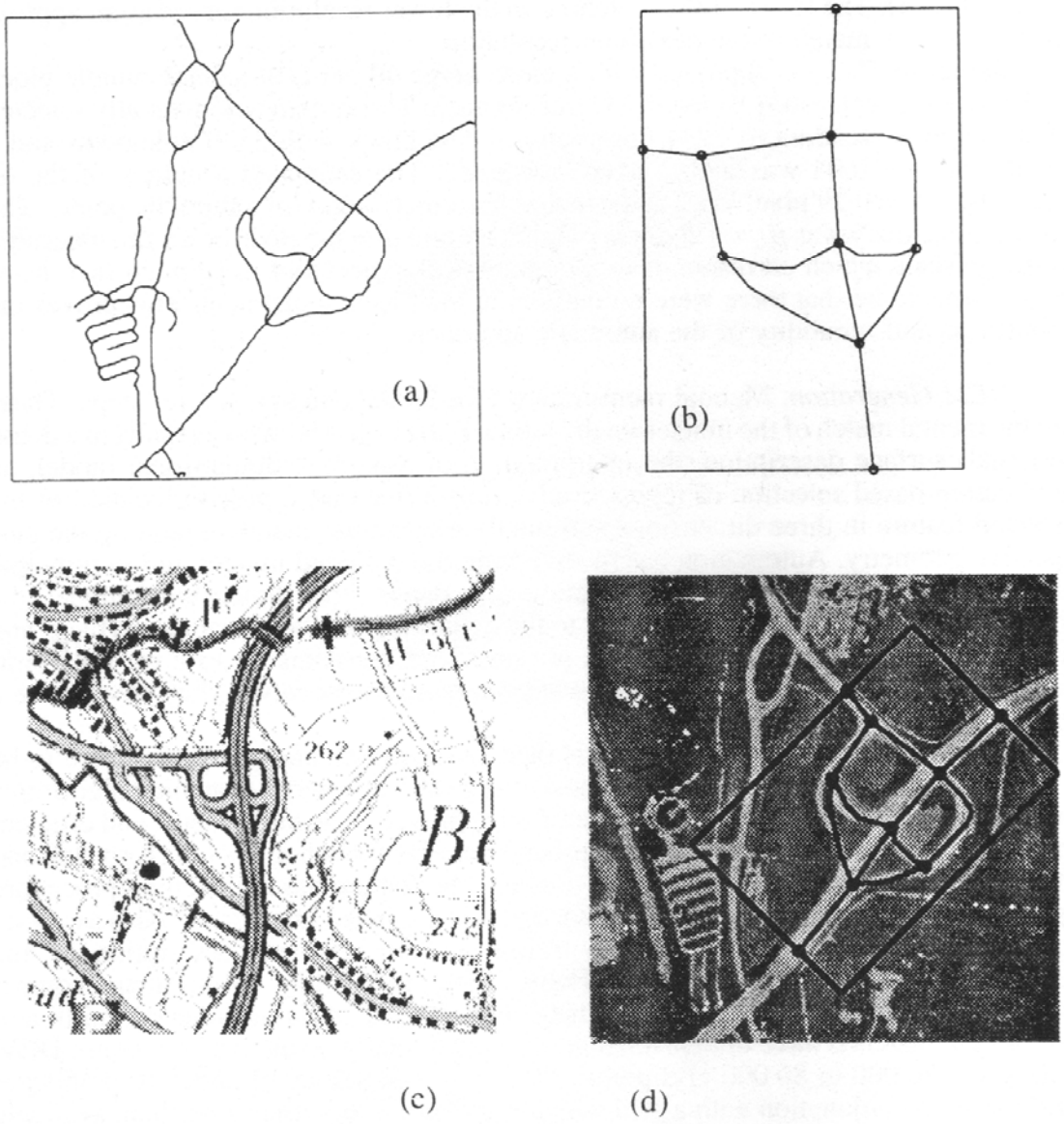


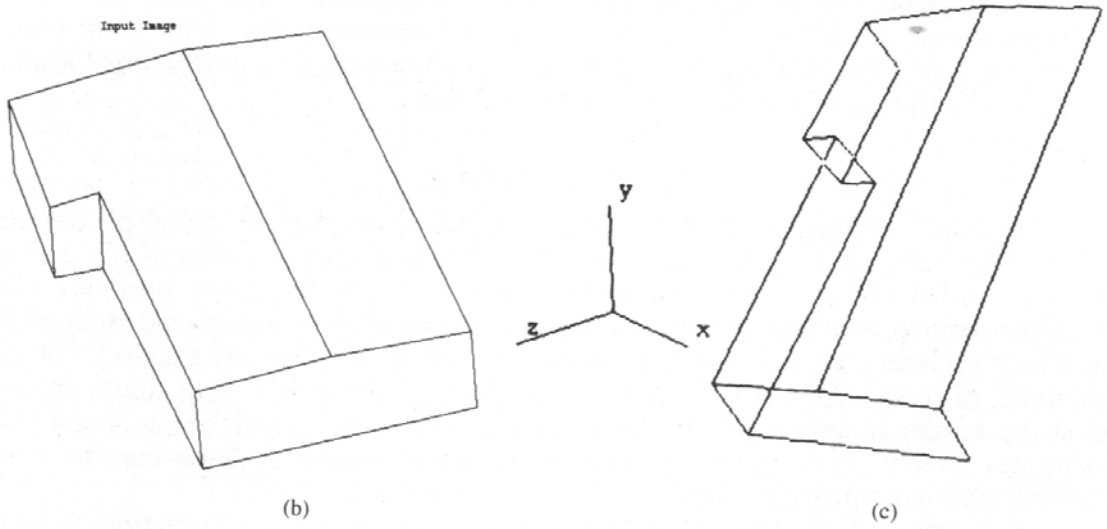
FIG. 5. The graph of the relational description derived from the hue component (a) of an aerial image, originally in colour, the road structure (b) manually derived from a map (c) and the result of the relational matching (d) (from Haala and Vosselman, 1992).

co-ordinate system and which mostly represent buildings. Using approximate values for the orientation parameters in order to increase the efficiency, the system matches the images of the object lines with straight line segments, as features, extracted from the image. This is done in several steps for all available control points. The final spatial resection is performed with the straight line segments and evaluated with respect to precision and reliability. An extensive test with 52 images shows the procedure, which takes 3 minutes to 5 minutes for the orientation of one aerial image on a Sun Sparc 1, to be successful in about 94 per cent of cases.

Relational Descriptions

Single features often do not carry enough information to be decisive for invoking specific hypotheses. Relations between the features result in richer attributes of the individual features (for example, a point being a node of a certain degree within a graph) and in information itself (such as a set of lines forming a closed polygon). Relational image descriptions are the necessary prerequisite for image interpretation where they are matched to the relational appearance models derived from the image models. The two following examples show that relational descriptions may also be

(a)



(b) (c)

FIG. 6. The section of an aerial image (a) and the semi-automatically extracted sketch of the building (b). The three dimensional reconstruction (c) is automatically derived using rules of inverse perspective (from Braun, 1992).

helpful for location and reconstruction tasks and thus for geometrically related tasks of photogrammetry.

Image to Map Registration. Relational matching (Shapiro and Haralick, 1981; Vosselman, 1992) establishes correspondence between individual features (points, lines and areas) by taking into account both the feature attributes as well as attributed relations between the features, such as angles between edges. Thus they use the full perceptual structure derived from an image.

This is especially useful for matching images and maps which reveal quite different appearance. An example is given in Fig. 5 taken from the work of Haala and Vosselman (1992). The structure of the road crossings can be derived from both a map and a colour image, by taking the skeleton of the hue component. The attributes of the point and line features are chosen to be rotation and, to a certain extent, are scale invariant. This allows the location of the road crossing without knowing the orientation or the exact scale of the image. The procedure is very robust with respect to spurious and missing image features, such as those resulting from occlusions or additional image detail, and to geometric distortions.

In spite of the high computation time of approximately 3·5 minutes on a VAX 3200 for this example (language POP-11), the procedure may be used very satisfactorily as a "boot strap" method for exterior orientation when complemented by a more appropriate method.

Reconstruction of Polyhedra from Single Images. The last example is taken from current work on building extraction (Braun, 1992). It seeks to demonstrate that exploiting geometrical *and* relational information extracted from a single image may be sufficient for reconstructing the form of objects, provided that a rich enough object model is available.

Fig. 6(a) shows a subsection of an aerial image with a building which is not too complex. Semi-automatic extraction of the lines and their incidence relations yields the sketch (Fig. 6(b)). Using knowledge of the object to be approximately a polyhedron and the sketch derived from an aerial image with known interior and exterior orientation, the three dimensional form of the object can be derived *automatically* (Fig. 6(c)). The reconstruction process uses well known facts from perspective geometry. A set of parallel three dimensional lines leads to a set of lines in the image intersection in a vanishing point which is given by the intersection of the image plane with a line through the projection centre being parallel to the set of three dimensional lines. Knowing the orientation of two lines of a plane allows derivation of its orientation and, assuming an arbitrary scale, enables derivation of all other three dimensional points or lines in that plane. The incidence relations between points and lines, which are derived from the image, are decisive for this kind of spatial reasoning. The result of the partial reconstruction of the building may be fused with other partial reconstructions leading to a full description of the object in three dimensions.

Conclusions

The paper attempts to show the role of feature extraction in digital photogrammetry. The feature extraction process itself may be based on an image model derived from a simple object model, but being general enough for many purposes. The examples from production systems and recent research all refer to classical photogrammetric or geometric tasks: point determination, orientation and reconstruction. The last example, of recovering the three dimensional form of polyhedra from single images, is at the border of image interpretation as a generic object model is assumed. The examples already show the great possibilities which digital systems may have for solving photogrammetric tasks.

The framework discussed for image interpretation and feature extraction has to be developed further and is not yet exploited. It will be the task of photogrammetric research to develop the theoretical and conceptual basis for extracting semantic features which may then be used in photogrammetric inspection systems or for topographic mapping.

References

Ackermann, F. and Hahn, M., 1991. Image pyramids for digital photogrammetry. *Digital photogrammetric systems*. Edited by H. Ebner, D. Fritsch and C. Heipke. Wichmann Verlag, Karlsruhe. 344 pages: 43–58.

Binford, T. O., 1981. Inferring surfaces from images. *Artificial Intelligence*, 17(1–3): 205–244.

Braun, C., 1992. Interpreting single images of polyhedra. *International Archives of Photogrammetry and Remote Sensing*, 29(B3): 514–521.

Bruce, A., Draper, R. T., Collins, J. B., Hanson, A. R. and Riseman, E. M., 1989. The Schema system. *International Journal of Computer Vision*, 2(3): 209–250.

Canny, J., 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6): 679–698.

Förstner, W., 1986. A feature based correspondence algorithm for image matching. *International Archives of Photogrammetry and Remote Sensing*, 26(3/3): 150–166.

Förstner, W. and Gülch, E., 1987. A fast operator for detection and precise location of distinct points, corners and circular features. *Proceedings of ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, Interlaken. 437 pages: 281–305.

Förstner, W., 1993. A future of photogrammetric research. *NGT Geodesia*, 93–8: 372–383.

Haala, N. and Vosselman, G., 1992. Recognition of road and river patterns by relational matching. *International Archives of Photogrammetry and Remote Sensing*, 29(B3): 969–975.

Hanson, A. J. and Quam, L. H., 1988. Overview of the SRI cartographic modeling environment. *Proceedings of the Image Understanding Workshop*, Cambridge, Mass. 1065 pages: 576–582.

Hartmann, G., 1983. Erzeugung und Verarbeitung hierarchisch codierter Konturinformation. *5. Symposium der Deutschen Arbeitsgemeinschaft für Mustererkenung*, VDE-Fachberichte 35: 378–383.

Krzystek, P., 1991. Fully automatic measurement of digital elevation models with Match-T. *Schriftenreihe des Institut für Photogrammetrie der Universität Stuttgart*, 15: 203–215.

Krzystek, P., 1992. Automatic DEM generation. *First course in digital photogrammetry*. Landesvermessungsamt Nordrhein-Westfalen, Bonn. 13 pages.

KRZYSTEK, P. and WILD, D., 1992. Experimental accuracy analysis of automatically measured digital terrain models. *Robust computer vision*. Edited by W. Förstner and S. Ruwiedel. Wichmann Verlag, Karlsruhe. 395 pages: 372–392.

LAWTON, D. T., LEVITT, T. S., MCCONELL, C. and GLICKSMANN, J., 1987. Terrain models for an autonomous land vehicle. *Readings in computer vision*. Edited by M. A. Fischler and O. Firschein. Kaufmann. 802 pages: 483–491.

LINDEBERG, T., 1990. Scale-space for discrete signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3): 234–254.

MERTSCHING, B., BÖHMER, I. and HARTMANN, G., 1990. Kontrollalgorithmen für auf hierarchischen Datenstrukturen arbeitende wissensbasierte Bildanalysesysteme. *Informatik Fachberichte* 254. Springer, Heidelberg. 682 pages: 98–105.

MUNDY, J., BINFORD, T., BOULT, T., HANSON, A., BEVERIDGE, R., HARALICK, R., RAMESH, V., KOHL, C., LAWTON, D., MORGAN, D., PRICE, K. and STRAT, T., 1992. The image understanding environments program. *Proceedings of the Image Understanding Workshop*, San Diego, California. 1062 pages: 185–214.

NIEMANN, H., SAGERER, G. F., SCHRÖDER, S. and KUMMERT, F., 1990. Ernest: a semantic network system for pattern understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(9): 883–905.

PADERES, F., MIKHAIL, E. M. and FÖRSTNER, W., 1984. Rectification of single and multiple frames of satellite scanner imagery using points and edges as control. *Proceedings of the Second Annual NASA Symposium on Mathematical Pattern Recognition and Image Analysis*, Houston. Edited by L. E. Guseman. 309 pages: 65 *et seq*.

SAGERER, G., 1985. Darstellung und Nutzung von Expertenwissen für ein Bildanalysesystem. *Informatik Fachberichte* 104. Springer, Heidelberg. 270 pages.

SHAPIRO, L. G. and HARALICK, R. M., 1981. Structural descriptions and inexact matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3(5), 504–519.

SCHICKLER, W., 1992. Feature matching for outer orientation of single images using 3-D wire frame control points. *International Archives of Photogrammetry and Remote Sensing*, 29(B3): 591–598.

STRAT, T. M. and FISCHLER, M. A., 1991. Context-based vision: recognizing objects using information from both 2-D and 3-D imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10): 1050–1065.

STRAT, T. M. and SMITH, G. B., 1987. Information management in a sensor-based autonomous system. *Proceedings of the Image Understanding Workshop*, Los Angeles, California. 1000 pages: 170–177.

TSINGAS, V., 1992. *Automatisierung der Punktübertragung in der Aerotriangulation durch mehrfache digitale Bildzuordnung*. Deutsche Geodätische Kommission, München. Reihe C, 392. 109 pages.

TREUTLER, B., 1992. *Gruppierung von Knoten in vektorisierten Skizzen*. Diplomarbeit am Institut für Photogrammetrie der Universität Bonn (unpublished). 132 pages.

VOSSELMAN, G., 1992. Relational matching. *Lecture Notes on Computer Science* 628. Springer, Heidelberg. 190 pages.

Résumé

L'extraction automatique de silhouettes est la première étape fondamentale dans tout procédé d'analyse d'images.

Ces silhouettes constituent, en même temps que leurs relations mutuelles, des structures identifiables à partir desquelles on peut établir des hypothèses et procéder à la reconstitution. On donne dans cet article un cadre général pour l'extraction de silhouettes dans un système d'interprétation et l'on y discute d'une modélisation possible de l'image et de l'objet pour obtenir des structures identifiables. On fournit enfin des exemples sur l'emploi de diverses silhouettes d'objets dans les travaux photogrammétriques courants.

Zusammenfassung

Die Kantenextraktion stellt den 1. wesentlichen Schritt aller Bildanalyseverfahren dar. Gemeinsam mit ihren wechselseitigen Beziehungen bilden die Kanten wahrnehmbare Strukturen, auf denen Hypothesen und Rekonstruktionsprozesse aufgebaut werden können. Im Artikel werden ein Grundgerüst zur Kantenextraktion angegeben, mögliche Objekt- und Bildmodelle zur Entwicklung von Wahrnehmungsstrukturen diskutiert und Beispiele zur Nutzung verschiedener Kanten für photogrammetrische Standardaufgaben angegeben.

DISCUSSION

Chairman (Mr. J. E. Farrow): We have a few minutes left after that masterly exposition for anyone who has some searching questions.

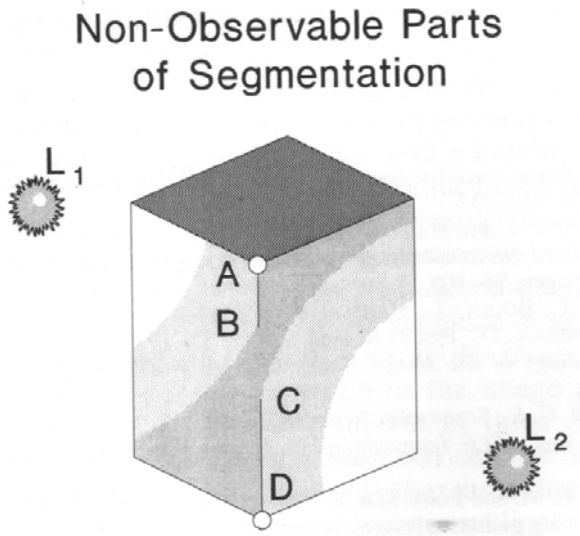


FIG. 7. As a result of illumination by the two light sources, L_1 and L_2, parts of the edge AD between B and C are not visible.

Professor Muller: I'd like to ask you about pyramidal feature based stereomatching. There has been a great deal of work done in manual photogrammetry with progressive sampling techniques. Is this meant to try to replace that work with an automatic system?

Professor Förstner: Yes, fully.

Professor Muller: If that is the case, how reliable are the interest operator based features that you are going to acquire as being representative of the underlying surface? In other words, you're sometimes going to capture the tops of roofs, sometimes going to capture the tops of trees; how representative of the surface is it going to be?

Professor Förstner: There are two points to make here which are similar to some extent. The progressive sampling in its original set-up is not able to catch features which fall between the grid. So if the initial grid looks very smooth and there is something in between this first grid, it will not be found. Of course the operator might see it, but that is a different matter. On the other hand, automatic systems assume a certain smoothness, at least piecewise, and that means it will work well in open terrain. It will even work well if there are single trees because it includes a robust estimation which cancels out wrong correspondences which come from the points on top of trees. If there are single houses, they are also cancelled because the higher level matching results are smoothed to a certain degree. If suddenly, there is a house at the next level, it's rejected. But in the case of a wood, of course, you obtain the top of the wood and if there is a group of houses, you obtain the top of this group. The problem is that you don't know whether it is one or the other. So the system can not decide automatically whether this is the real surface which is high or whether the true surface is below. So the system is silly in the sense that it does not interpret the features and has no idea what are houses or what are trees. It just says that these are feature points and there are no more characteristics. Of course, this is a weakness, but on the other hand if you want to do that at the same time, ten times larger computing times are required and an interactive process is still required. If you really want to have a full DEM for a whole area, you have to exclude the residential areas or places where you might expect that the system will produce failures, not in the sense that you don't obtain correlation because a surface will always be produced. It doesn't fail in that sense. But it would fail to give the right surface for the topography.

Dr. Gugan: Could you briefly define your use of the phrase segmentation? Exactly what do you mean by segmenting the image?

Professor Förstner: I don't like the word segmentation. I use it because it is a word which has a certain tradition. It reflects the model for idealised images. The image is partitioned into regions which are not overlapping. But reality is more difficult. You may miss features if you force the system to carry out segmentation. I can give you a simple example where the segmentation notion is not good as a basis for feature

extraction. If you take a block (Fig. 7) which is illuminated from the upper left side and the lower right side, you will find an edge at the top and an edge at the bottom. You will also have some point where the illumination is the same from right and left of the vertical edge. Therefore segmentation, in the sense that you have boundaries and no lines which intrude into areas, which is the classical definition of a segmentation, is not present here. Therefore I would rather derive a relational description of the image and then these effects are allowed.

Mr. Nwosu: If this DEM system is automatic, how are you going to exclude these problematical areas? Does the matching algorithm offer interactive possibilities or is it always automatic?

Professor Förstner: There are two possibilities. You assume that the system is good enough whatever it does. This means that, even if you have residential areas, you accept the result. On the other hand, if you really want to have contour lines which refer to the topographic surface in residential areas, you have to do it yourself, possibly supported by some local matching procedure measuring the heights. To make a meaningful DEM of the London area you really have to exclude buildings or use a different technique. But in open areas outside of cities, I think it will work reasonably well. Moreover, if you want to make orthophotomaps, it's good to have the DEM passing over the trees because then you don't have the wrong geometry at these places. So it's even better than if you were to take the DEM from maps.

Chairman: With no more burning questions, I think that you'd like to join in thanking Professor Förstner for a stimulating and interesting talk.