A Hierarchical Neural Network System for Signalized Point Recognition in Aerial Photographs

Veton Z. Këpuska and Scott O. Mason

Abstract
A system for automatically recognizing signalized points in digitized aerial photographs is described. Signalized points take on a highly variable appearance in terms of radiometry, shape, scale, and orientation. Additionally, they appear as very small features, making reliable feature extraction infeasible. Using a hierarchical neural network architecture accompanied by suitable image pre-processing—recognition invariant to the orientation of the pattern and robust with respect to shape, scale, and radiometric variations—was achieved. The hierarchical neural network architecture consists of (1) a Self-Organizing Feature Map network employed for the learning of reference signalized point patterns and the handling of their variable orientation, scale, and shape; and (2) a Feed-Forward Network employed for the classification of similarity tests between the reference patterns and a given test patch. Experimental results demonstrate that the system achieved a near 100 percent recognition rate for classes of patterns for which it was trained, with a very low false-alarm rate. A gradual fall-off in performance was recorded for "unlearned" signalized point patterns, resulting in an overall recognition rate of 74 percent.

Introduction
The application of artificial intelligence (AI) technology in the analysis of photogrammetric image sequences was set to be one of the primary goals of a recent project at the Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology, Zurich. The measurements of signalized points, a fundamental task in aerotriangulation, was focused upon because it contains the elements of general image analysis and understanding problems. In this paper, a system consisting of image pre-processing, and a hierarchical neural network architecture for recognizing the patterns cast by signalized points (SP) in digitized aerial images is presented. Recognition here entails both detecting the presence of an SP pattern and estimating its position and orientation within an image.

Signalized points (SPs) are coordinated landmarks "signalized" in such a manner as to appear with a shape and size that permits their manual measurement in aerial photographs. The measurement of these points in aerotriangulation consists of three basic operations: (1) recognition, entailing the detection of the presence of SPs and estimates for their coarse locations in the photographs; (2) identification, in which each of the recognized SPs is labeled; and (3) precise localization, in which precise image coordinates for the landmark in the photographs are determined. The photogrammetrist commonly draws on a number of knowledge sources in order to simultaneously identify and recognize the occurrence of SPs in the photographs, including topographical maps, sketches from the survey of the SPs, and the stereo cue in transferring the location of the points from overlapping photographs. Given the desirability of removing the reliance on human operators in all stages of the mapping process, the automation of each of these operations demands attention. To this end, the precise localization of SPs has been practically solved with various least-squares matching (LSM) algorithms, providing image coordinate accuracies on the order of 0.02 pixels for well-defined points (Grün and Baltsavias, 1988). It is well-recognized, however, that LSM requires good approximations of the location and orientation of the targets. If the SP measurement process is to be automated, these approximate values should be provided by the recognition operation.

The recognition of specific patterns in images, such as SPs, falls under the general field of study of automatic target recognition (ATR). While template and feature-based methods have been successfully used in many pattern recognition applications (Duda and Hart, 1973), they generally fail to provide adequate solutions in cases where the pattern to recognize exhibits large variability, has poor resolution, and varies in scale—characteristics commonly the case in ATR and, in particular, in SP recognition (SPR). In contrast, even primitive biological organisms are perfectly suited to perform such cognitive tasks (e.g., bees). Inspired by such biological systems, Artificial Neural Networks (ANN) models have been studied for many years in the hope of developing artificial systems that provide comparable performance. The development of new ANN topologies and algorithms, as an attempt to model the processes of perception and learning in biological systems and the belief that massive parallelism is essential for achieving high performance of algorithms employed for cognitive tasks, may form the basis on which such tasks can be tackled successfully. The importance of artificial neural networks (ANNs) for ATR applications, most of which have been of a military nature to date, was confirmed in the study

Institut für Geodäsie und Photogrammetrie, ETH-Hönggerberg, 8093 Zürich, Switzerland.

V.Z. Këpuska is presently with Voice Processing Co., One Main St., Cambridge, MA 02166.

Photogrammetric Engineering & Remote Sensing,
Vol. 61, No. 7, July 1995, pp. 917–925.

0099-1112/95/6107–917$3.00/0

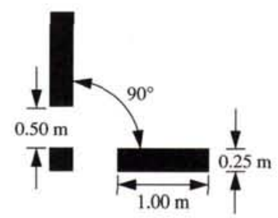

Figure 1. Dimensions and layout of an ideal target as used in the test data.

conducted by the United States Defence Advanced Research Projects Agency (see DARPA, 1988).

Artificial neural networks form the basis of the SPR strategy outlined in this paper. The potential of ANNs presented only with the patterns the signalized points cast in digitized aerial images is explored. In the following sections, the characteristics of the SPR task are described, approaches to target recognition are reviewed, the components of the developed system and the training of the neural networks are described, experimental results are reported, and, finally, conclusions and recommendations for future work are discussed.

The aerial photographs used in these investigations were taken from a 4 by 4 sub-section of an aerial triangulation test-block flown in 1986 over the Heizenberg area of eastern Switzerland. The average image scale for these 16 photographs is 1: 15,000, although, because the terrain height ranges from 650 to 2150 m, the effective image scale ranges between 1: 19,300 and 1: 10,700. In this block, a single right-angled signalized point type was used, as described below. A number of patches, each containing one signalized point, were digitized at a size of 592 by 574 pixels from the photographs in this sub-block (for more details, see Wilkins (1990)). A total of 81 images provided the data for training and testing the ANN-based recognition system.

Characteristics of the Signalized Point Recognition Task

Signalized points are man-made, planar structures consisting of a square center-mark placed directly over a coordinated feature. In the case of the Heizenberg block, they are accompanied by two rectangular strips placed equidistant from the center mark and perpendicular to one another (see Figure 1). The latter provide the SPs with some degree of uniqueness and thereby assist in their recognition in the aerial photographs. For a recognition approach based solely on this pattern, it is important to describe the characteristics of the appearance of SP patterns in digitized images, as these characteristics define the properties required of an automatic recognition strategy. As exemplified in Figure 2 SPs appear with

- *Random orientation within the aerial photographs.*
- *Variable size.* The size of the SP pattern is influenced by the physical dimensions of the SP itself, the image scale, and the photograph digitization resolution. Table 1 summarizes estimates for the variability of SP pattern size in the test data. With an average maximum dimension of 13 pixels, SP patterns appear with very low resolution, practically excluding the application of feature-based recognition techniques, as outlined below.
- *Variable structure.* As a consequence of irregular SP construction caused by, e.g., the topology of the area surrounding the SP, the structure of the SP pattern can vary, i.e., the strips of the SPs are ideally placed at 90°, although this is rarely the case as can be seen in Figure 2. A further contributing factor is the non-uniformity of digitization: the pixel size of the Sony XC-77CE inter-line transfer device used to digitize the test photographs is 13.6 by 11.1 μm, leading to a slight orientation-dependent distortion of SP shape.
- *A variable number of components.* As a consequence of the other characteristics mentioned here, the number of distinct features comprising the SP pattern can vary as the patterns of the SP components (point and strips) become merged.
- *Variable radiometric properties.* The intensity and contrast of SP images is influenced by numerous factors, including the illumination conditions at the time the photographs were taken, the context (or background) in which the SP is found, the gain of the photograph scanner, as well as disturbances of the SP both on the ground and in the photographs, e.g., by dust.
- *Variable context.* SPs are ideally located in a position free from obstruction, particularly from the air. This is not always achieved, as the examples in Figure 2 testify.

As a consequence of the above characteristics, SPs appear in digitized photographs as highly variable patterns, making formulation of a general strategy for SPR extremely difficult. In order to achieve recognition robustness, the strategy needs to be invariant, or at least insensitive, to these pattern variations. To this end, three levels of invariance are desired:

- *Orientation invariance:* the orientation of the SP in an image should not influence it's recognition. This is particularly important for asymmetric SPs.
- *Radiometric invariance:* the recognition strategy should be insensitive to the SP's context and its absolute intensity.
- *Shape-deformation invariance:* the recognition strategy should be insensitive to variations in SP shape (scale, structure, number of components) and context.

As mentioned in the introduction, context plays an important role in the process by which humans recognize SPs. In the strategy reported in this paper, recognition is solely based on the pattern of the SP projected into the photographs (low-level, as compared to the high-level knowledge of the

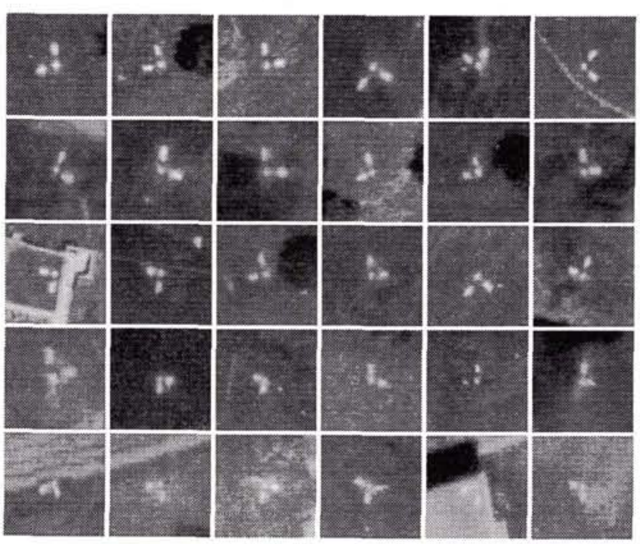

Figure 2. Typical patterns for a right-angled signalized point extracted from digitized aerial photographs. The variability in the appearance of signalized points poses a significant problem for automating their recognition.

TABLE 1. VARIABILITY OF SIGNALIZED POINT PATTERN SIZE.

Largest SP pattern dimension (pixels)			SP pattern area (pixels)		
minimum	mean	maximum	minimum	mean	maximum
9	13	20	~35	~60	~85

context). Consequently, the context of an SP is considered here to be a random disturbance. In other words, the SP must be discriminated with respect to an infinite number of backgrounds.

Approaches to ATR

ATR is a general problem, relevant to many applications in image analysis and understanding. Numerous approaches and algorithms can be found in the literature. These can be classified into three broad classes: area-based, feature-based, and model-based approaches. In model-based vision, the target to be recognized is stored as a general model. The components, features, etc., describing this model must be found on/in the test pattern for a match to occur. While a model-based target representation may be efficient, the complexity of this approach arises in the means by which a correct model for a given local situation can be derived from the general model. If the situation is difficult to interpret, the computational search space may be large (Thoet *et al.*, 1992). In feature-based matching, attempts are made to extract stable characteristic features from segmented patterns in images and match them against the equivalent features derived from representative target patterns. The features commonly reported in the literature (based on, e.g., Method of Moments, Fourier Descriptors, stochastic models of contours, etc.) vary in the degree of their invariance and/or insensitivity to orientation, scale, and shape distortions in the target patterns (Këpuska, 1992). Their suitability in this particular application is called into question when considering the relatively small size of the SP patterns (see Table 1) and the variability of their bounding contours and number of components (see Figure 2). Robust recognition using such features generally demands that the patterns be easily segmentable and/or consist of long, distinct edges. That these conditions cannot be reached easily in SPR is clearly illustrated in the example in Figure 3. Here, it was not possible to determine a set of global parameters for the Canny edge detector providing satisfactory segmentation for a range of SPs: usable results could only be obtained by optimizing the algorithm parameters for each individual SP image.

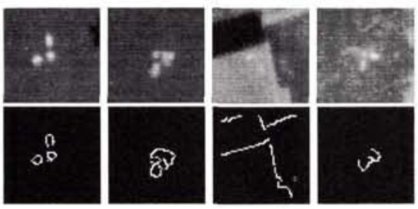

Figure 3. Example of the difficulty of using conventional pattern analysis techniques for SPR. The edge images for the four signalized points shown here were derived from the Canny edge detector using a common set of processing parameters.

In traditional area-based (e.g., template) matching, an entire test patch (a small window of an image) is compared to a patch representative of the target pattern. If the match between the test and representative patches is "good enough," the system claims recognition of the target at the test patches location. Under "open-world" conditions, template matching possesses a number of limitations, i.e., changes in pattern scale, orientation, and appearance (due to shadowing, occlusions, and the context) can cause mismatches. This limited robustness requires that a large set of representative patches be employed, thereby increasing the computational load (Thoet *et al.*, 1992). Nevertheless, given the limitations of feature- and model-based recognition, an area-based approach appears to be the only most feasible approach for addressing the SPR task and was adopted for this work.

Suitability of Artificial Neural Networks for Target Recognition

Due to their intrinsic properties, ANNs appear to provide a framework for complex recognition tasks. An ANN is a system composed of many computationally simple processing elements (units) operating in parallel, whose function is determined by the network architecture, the strength of the connections between the units, and the actual processing performed at these units (DARPA, 1988). These units are generally non-linear; the output is a function of the weighted sum of its inputs. The architecture of an ANN is specified by the topology of the interconnection links between the units. In order for an ANN to be able to perform a task, these interconnection weights must be suitably adjusted, typically using a learning algorithm. Its learning capacity provides an ANN with the capability of adapting to new situations, e.g., in ATR, to additional targets and environments, and depending on the type of network, may follow a supervised, unsupervised, or self-supervised procedure. For a general introduction to neural network techniques and tools, see Rumelhart *et al.* (1987), Lippmann (1987), Kohonen (1988b), and DARPA (1988). ANN techniques have recently found numerous practical application in ATR. An excellent overview of this subject can be found in Roth (1990). A number of attempts have been also made to use ANNs for ATR in photogrammetric applications. Toth and Schenk (1990) used a parallel associative memory architecture for the recognition of synthetic cross-shaped targets overlaid in aerial images. Chiu and Hines (1990) applied a feed-forward network with back propagation (FFN-BP) to recognize synthetic reference points overlaid on processed photogrammetric plates. To the authors' knowledge, however, there are no reports on the recognition of signalized points for photogrammetric applications in an "open-world" environment.

Earlier Work

As with other recognition approaches, it is important to investigate the means by which an ANN can be made invariant (or at least insensitive) to the variability of the target patterns. There are a number of possible approaches: (1) by training the ANN with a set of patterns that cover the range of pattern appearance variations, or (2) by embedding invariance into the ANN architecture itself through use of an invariant feature space. In earlier work by Këpuska and Mason (1991a; 1991b), the potential of the first approach was investigated. A single feed-forward network was trained by back propagation with a set of patterns which encompassed

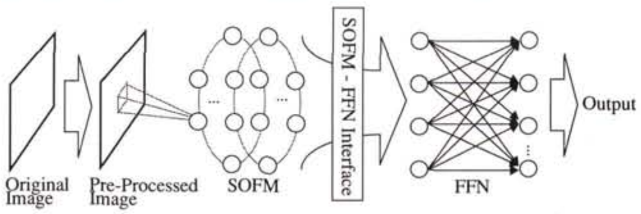

Figure 4. The signalized point recognition system components and data flow: an image pre-processing step followed by a hierarchically structured ANN consisting of an SOFM and a two-layered FFN network. The output for each recognized signalized point includes estimates for its location and orientation.

a wide range of appearance variations in orientation, scale, and context. This set was comprised of both manually selected SP and non-SP patterns. In conjunction with an iterative training and testing paradigm, this initially small training set was gradually expanded to include those patterns the network falsely classified, thus overcoming the difficulty of selecting an optimal training set. While a recognition rate of 85 percent[1] over 32 images containing well-defined SP patterns was achieved with this paradigm, this simple architecture appeared to be limited in its capacity to be simultaneously invariant to large pattern variations. The experiences made in these investigations lead to the conclusion that at least part of the invariance must be embedded into the ANN architecture itself. In addition, a pre-processing scheme was deemed necessary to suppress the SP background and contrast variations.

ANN-Approach to Signalized Point Recognition

The experience gained by early investigations led to the formulation of the SPR system in which the three desired levels of recognition invariance are tackled individually. First, robustness with respect to radiometric variations is achieved by pre-processing the images to suppress each SP's context and to enhance its absolute grey level; second, orientation invariance is achieved through selection of a suitable ANN architecture; and third, shape-deformation robustness is achieved through suitable training of the networks. The ANN architecture consists of two hierarchically organized networks—a self-organizing feature map (SOFM) network and a feed-forward network (FFN). These networks were trained by first presenting a set of example SP patterns to the SOFM for the learning of SP reference patterns, and then training the FFN based on the output of the trained SOFM for the example patterns. As output from this system, coarse estimates for the location (around 1 to 2 pixels) and orientation of the SP pattern in each tested image are determined. The components of the SPR system, as shown in Figure 4, are detailed below.

Image Pre-Processing for SP Enhancement and Context Suppression

The fact that SPs are white, regular structures permits exploitation of three sources of knowledge in image pre-processing. First, the SP can be expected to contrast with its immediate context; second, the grey-level (intensity) value of SP image patterns should lie within the upper-bounds (brighter) of the image's intensity histogram; and third, SPs are small, compact patterns. Based on these expectations, a preprocessing scheme was developed through empirical testing, that consists of

Step 1. Wallis Filtering. The purpose of this filter is to force the mean grey value, and especially the contrast of an image, to fit some given, in this case globally set, values (Baltsavias *et al.*, 1990). It is applied here to suppress SP context.

Step 2. Median Filtering (3 by 3 window) to eliminate high frequency noise possibly enhanced by Wallis Filtering.

Step 3. Histogram Transformation (illustrated in Figure 5) to further enhance the contrast between the SP patterns and their contexts.

Step 4. Low-Pass Filtering (σ = 1.0) for smoothing of the SP shape to overcome possible artifacts created by the histogram transformation.

This pre-processing scheme was found to be generally applicable to the test images—in only two cases out of 81 tested images, did the scheme fail to enhance the SPs due to very weak feature-background contrast. The effect of each of these steps is illustrated in the example in Figure 6.

Hierarchical ANN Architecture

Invariance with respect to orientation, and robustness with respect to shape and scale variations, is handled in this system by means of a hierarchical ANN architecture. This architecture consists of a self-organizing feature map (SOFM) network and a feed-forward network (FFN) (see Figure 4). The variations in SP scale, orientation, and shape are handled by the SOFM, while FFN performs classification and orientation estimation based on the responses output from the SOFM.

Self-Organizing Feature Map

Self-organizing networks, such as SOFMs, partition a set of training examples into groups or clusters using unlabeled training data, i.e., unsupervised training. This type of clustering, or vector quantization, is an efficient means for compressing information that is to be processed at higher levels with little loss of information (DARPA, 1988). The number of clusters formed can be either pre-specified or automatically determined from the training data. Moreover, the units forming the SOFM network can be structured in one, two, or higher dimensions and in a number of different topologies

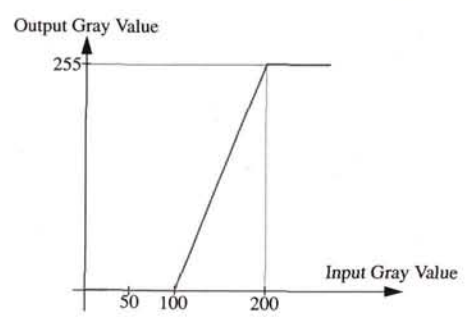

Figure 5. Histogram transformation employed for enhancing signalized point contrast.

[1]This recognition was based on the condition that there was only one SP per image.

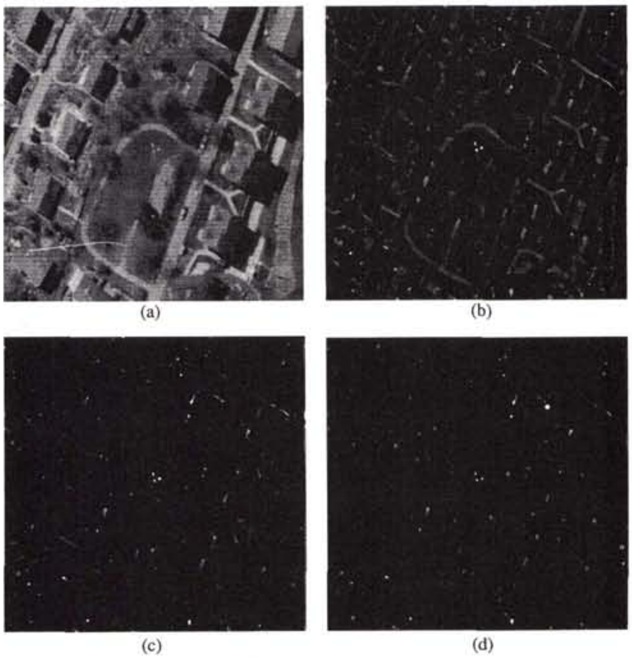

Figure 6. Pre-processing of images for signalized point recognition: (a) illustrates the original image; (b) after Wallis filtering; (c) after median filtering and a histogram transformation; and (d) after low-pass filtering.

(Kohonen, 1988a; Këpuska, 1990). The motivations for employing an SOFM in an SPR recognition system are twofold: (1) a network topology can be selected that permits a natural handling of the SP orientation variations; and (2), through appropriate clustering, the variability in SP shape and scale can be covered. The clusters in this application consist of a set of reference SP patterns.

In the SPR system, the SOFM units are structured in a "cylindrical" architecture (see Figure 7). Each of the rings is associated with a different SP pattern cluster and is comprised of 32 units. Each unit in a ring corresponds to a different primary orientation of the ring's reference pattern, i.e., from 0° through 348.75° in 11.25° intervals. By testing an equally dimensioned input pattern for similarity against the reference pattern at each of the units in a given ring, orientation invariance is achieved. The choice of 32 orientations was reached through experimentation. (Note that, for symmetric SP patterns, e.g., crosses, eight orientations would be sufficient).

Computing the Similarity between Reference and Test Patterns

The final classification of an input pattern as either an SP or non-target is made by the FFN on the basis of the similarity of this pattern to the rotated reference patterns located at each unit of the SOFM. For a given orientation o, the Euclidean distance d^j_o between the intensity of pixels in the test patch $I(x,y)$ and the l reference patterns $R^j_o(x,y)$ located at units of the SOFM for this orientation, is computed as follows:

$$d^j_o = \left(\sum_{x=1}^{X}\sum_{y=1}^{Y}(I(x,y) - R^j_o(x,y))\right)^{0.5} \qquad (1)$$

where X and Y define the size of the input test patch (here, $X = Y = 35$), and $o = \{0, ..., 31\}$. Accepting the reference pattern $R^i_o(x,y)$ with the distance d^i_o, where

$$d^i_o = \frac{1}{n} min\ (d^j_o),\ j = 1, \ldots, l \qquad (2)$$

with n being a scale factor ($n = 255XY$), as that patch most similar to $I(x,y)$, the similarity response for $I(x,y)$ for orientation o is computed as follows:

$$S_o = \begin{cases} e^{-(d^i_o - \tau)} & (d^i_o - \tau) > 0 \\ 1 & (d^i_o - \tau) \le 0 \end{cases} \qquad (3)$$

The parameter τ (here, $\tau = 2.0$) ensures that, for small average differences between $I(x,y)$ and $R^i_o(x,y)$, the maximum similarity score is obtained. For differences greater than τ, this score decays exponentially. The similarity responses are computed for each of the 32 reference patch orientation, i.e., $S = \{S_0, ..., S_{31}\}$, and have values on the range [0, 1], where $S_o \approx 0$ indicates a low similarity response and $S_o \approx 1$ indicates a high similarity. This set S of similarity responses is fed into the FFN for final classification. In other words, the input to the FFN is the highest similarity response recorded for each orientation of the SOFM.

Feed-Forward Network

Feed-forward networks (also known as multi-layer perceptrons) are typically used for classification tasks and are trained under supervision using labeled training data to partition input patterns into a pre-specified number of classes. FFNs consist of a number of layers of units, typically the input, hidden, and output layers. A three-layer FFN can form any decision region and can approximate any desired nonlinear input/output mapping (DARPA, 1988). Two-layer networks (i.e., without hidden layer(s)) are capable of solving

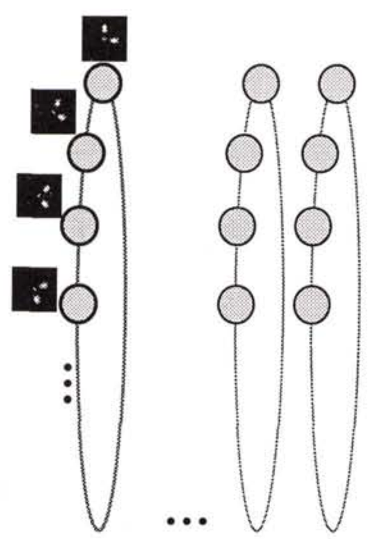

Figure 7. The SOFM consists of six rings (one for each reference pattern), each comprised of 32 units (one for each orientation of the reference pattern).

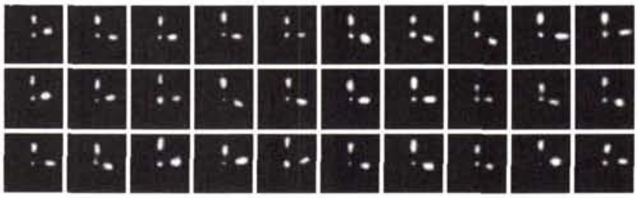

Figure 8. Set of pre-processed signalized points patterns employed for training the neural networks.

linearly separable problems. The units of each layer are typically connected only to the units of the succeeding layer. The weights, i.e., strengths of the unit connections, are most often adjusted using a form of the back-propagation (BP) training algorithm (Rumelhart *et al.*, 1987). BP is a supervised training algorithm requiring specification of the network behavior (desired output) for each given input. Training proceeds iteratively until the actual output is equivalent to the desired to within a given tolerance.

In the SPR system, a simple two-layer FFN was considered sufficient to perform the necessary classification of the similarity responses outputs from the SOFM. This network consists of an input and an output layer, each containing 32 units. Based on these measures, the FFN classifies the input pattern as either an SP or a non-target. In doing so, if the input pattern is an SP, the similarity measures should peak around the unit corresponding to the SP pattern's orientation and should be minimal for all other units. On the other hand, if the input pattern is not an SP, the responses at all output units should be low.

Training the Neural Networks

Neural networks must learn or be taught the tasks that they should perform. This is achieved through a training phase in which the unit connection strengths (or weights) within the network are adjusted. In the SPR system, training is divided into two steps. The SOFM is trained first, and then the responses of the trained SOFM are used for training the FFN network. The algorithm employed for training the SOFM is essentially a vector quantization procedure, i.e., Learning Vector Quantization (LVQ) (Kohonen, 1988a; Kohonen *et al.*, 1988). The aim of LVQ is to obtain a small number of reference patterns that adequately describe the continuous distributions of the actual SP patterns. In this case, the SOFM units "learn" weights equivalent to oriented SP reference patterns. To perform this training, a set of high-quality SP patterns were manually selected. By "high-quality" is meant SP patterns that can be easily recognized by humans without the assistance of additional knowledge, e.g., context information. The motivation in selecting such patterns was that the SPR system should recognize good SP patterns well, with a gradual fall-off in performance as the quality of the pattern deteriorated.

There exists a trade-off between choosing a smaller number of reference patterns for the sake of computational efficiency, and choosing a larger number for the sake of more a accurate representation. From empirical testing, six reference patterns were deemed sufficient for this application, although it is possible to automate the process of setting an optimal number of reference patterns, as discussed in the conclusions section of this paper. In Figure 8, the set of SP patterns used for training is illustrated. Each of these SP patterns has been extracted from the test images and pre-rotated to an orientation of 0° (based on the direction of the left-most point strip). The training of the SOFM, carried out for this single orientation, produced the six reference patterns shown for the 0° orientation in Figure 9. These reference patterns were then incrementally rotated by 11.25° into the additional 31 orientations of the units in the network. The compression effect is evident: the reference patterns are generalizations of classes of training patterns.

In addition to the patterns employed for training the SOFM shown in Figure 8, three uniformly toned patterns—white, black, and average grey, respectively—were included into the set for training the FFN. These were necessary in order to establish a proper classification for non-SP patterns. The training set was then further expanded by reproducing each of the SP patterns—initially with an orientation of 0°—at the remaining 31 discrete orientation steps supported by the system. Thus, the full training set contained typical SP patterns at all orientations, as well as a few non-target patterns. The FFN was then trained using, as mentioned above, the back-propagation algorithm with the desired outputs for each of the system inputs (i.e., the training patterns) defined as follows:

- for SP patterns, the output should be ≈ 1 at the unit with orientation, shape, and scale matching that of the input pattern, and ≈ 0 or less at all other output units; and
- for non-target patterns, the output should be ≈ 0 or less at all output units.

The manner in which the SPR system performs recognition can be seen as a form of surface evaluation. The set of similarity measures computed at the units in all layers of the SOFM can be interpreted as forming a surface with axes: SOFM ring, orientation, and similarity response. An example of such a surface for an SP pattern is shown in Figure 10. Because the input to the FFN consists (only) of the set of similarity measures belonging to the layer in which the highest similarity measure was obtained, classification is based on the profile containing the similarity surface peaks. In end effect, it is the shape of this profile on which the pattern recognition decision is based. The profile corresponding to the surface shown in Figure 10 is depicted in Figure 11.

Experimental Results

As mentioned in the Introduction, the data used for testing the SPR system consisted of 81 images (dimension 592 by 574 pixels), each containing a single SP pattern. Each image was processed by the system in the following manner:

(1) Pre-processing was carried out.
(2) All possible 35- by 35-pixel patches were then sequentially extracted from the image. (For the test data used, this amounts to some 3×10^5 patches per image!).
(3) For computational reasons, a simple heuristic test was carried out to eliminate patches not containing sufficient

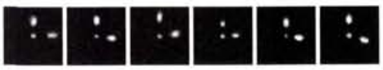

Figure 9. Signalized point reference patterns generated by the SOFM network.

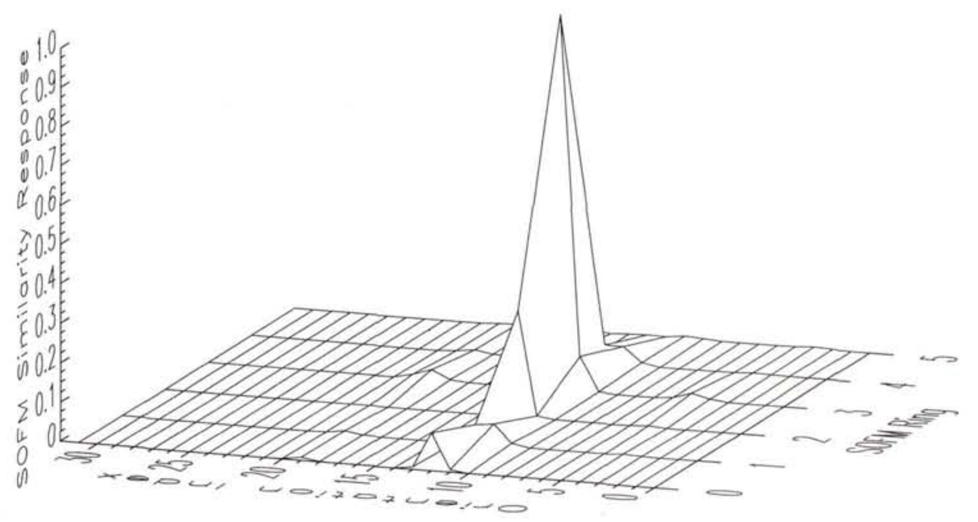

Figure 10. The complete set of similarity responses recorded by each unit in the SOFM for a signalized point pattern.

"bright" pixels from further consideration. It is improbable that such patches contain an SP.

(4) Patches passing the test in (3) were presented to the SOFM. Similarity measures between the test patch and the reference pattern at each unit of the SOFM were computed.
(5) Results from the SOFM were then presented to the FFN-BP for classification. Output from the FFN-BP consists of a set of scores for each orientation. These require interpretation. If all scores are "low," the input pattern is considered to be a non-target; alternatively, if a single "high" score is recorded, an SP is considered to have been recognized. The coarse location of the SP in the input image is then given by the center pixel coordinates of the successful patch, and an estimation for it's orientation is given by the orientation of the associated unit.

The results of applying this procedure to the test images are summarized by the two recognition response histograms in Figure 12 and Figure 13. In Figure 12, the distribution of the recognition scores for 81 targets is given, while in Figure 13, the distribution of all response scores triggered by non-target patterns is shown. The images accompanying these histograms exemplify the type of patterns triggering different response levels. Ideally, the SPR system should output scores of 0 for non-target patterns, and positive high scores (≈ 1) for SPs. In reality, however, SPs which differ significantly in appearance from the reference patterns will trigger lower recognition scores, while non-target patterns appearing similar to SPs can trigger high scores. If a decision between SPs and non-targets is made on the basis of a threshold set on the recognition response of the system, it must be anticipated that some SPs may be missed and false-alarms recorded by the system. Table 2 summarizes, based on the experimental results, the recognition and false alarm rates for three different thresholds.

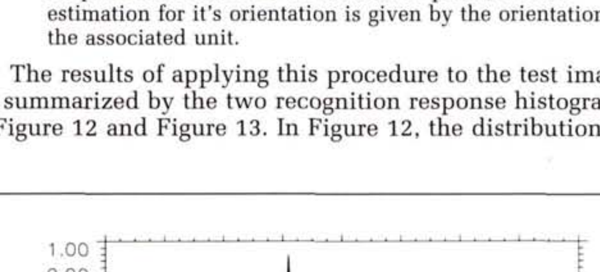

Figure 11. Input to the FFN network derived from the set of similarity responses illustrated in Figure 10.

At first glance, the recognition rate even for the lower threshold (0.60) may appear disappointing at 74 percent. Recalling, however, that the SOFM was trained for high quality patterns, a closer inspection of the results reveals that a near 100 percent recognition rate was recorded by the system for such patterns, with performance deteriorating gradually as the magnitude of the shape distortions and scale differences of the pattern with respect to the reference patterns increased. Poor results were also achieved when a pattern's context failed to be adequately suppressed. Indeed, some 18 SP patterns (over 22 percent of the test set) suffered from these problems. The examples in Figure 14 illustrate patterns exhibiting these problems. It should be clear that these patterns are not well represented by the reference patterns in Figure 9, a direct result of the fact that the system was not trained with suitable examples (see Figure 8). As future work, it is planned to further test the system trained to explicitly handle a broader category of SP patterns. With respect to the false-alarm rate, it can be seen in Figure 13 that the type of non-target patterns the system responded to closely resemble the SP patterns and may easily be falsely recognized by a human operator in the absence of contextual information.

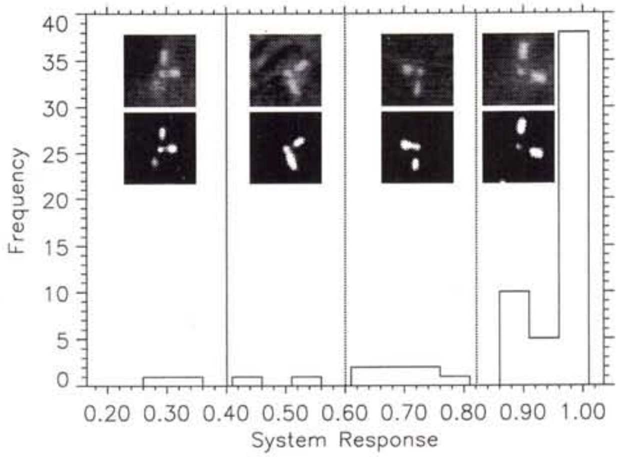

Figure 12. Histogram of FFN responses for the signalized point patterns in the test data, and corresponding typical original and pre-processed signalized point patterns that triggered it.

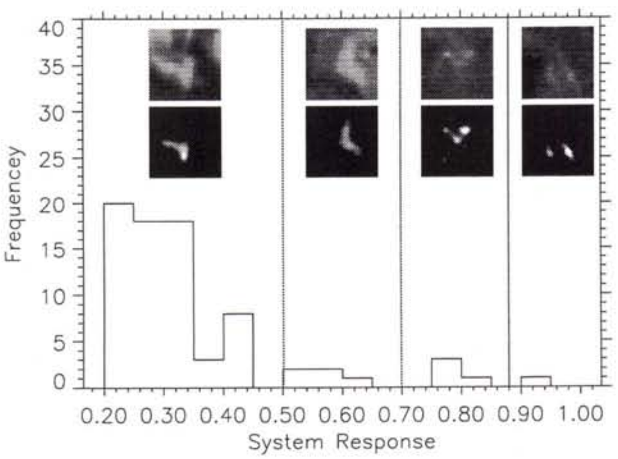

Figure 13. Histogram of FFN responses for distinct non-signalized point patterns in the test data, and corresponding typical original and pre-processed patterns that triggered it.

Discussion and Conclusions

The primary goals of this work were, generally, to investigate the application of AI technology to solving photogrammetric tasks, and more specifically, to develop a system for signalized point recognition based solely on the patterns these objects cast in digitized aerial imagery. With respect to the second goal, a neural network-based SPR system was developed and trained to recognize well-defined SP patterns. Testing demonstrated that this system performs its task robustly for such patterns, i.e., with 0 well-defined SP patterns being wrongly classified as non-targets and a very low false-alarm rate. For poorly defined targets—for which the system was not specifically trained—reasonable recognition rates were still achieved, although accompanied by a gradual inflation of the false-alarm rate. It should be noted that, once a number of well-defined SPs have been recognized and identified, knowledge in the form of the relative orientations of the models in the block and of the ground locations of SPs can be exploited to (1) narrow the search space for "missed" targets by estimating their image coordinates, and (2) detect recognition blunders and false alarms by checking that the rays defined by candidate SP patterns intersect in object space.

Table 2. Recognition and False-Alarm Rates Achieved in Testing.

Threshold	Recognition Rate	False Alarm Rate
≥0.95	40/81 (49%)	0/81 (0%)
≥0.85	53/81 (65%)	1/81 (1%)
≥0.60	60/81 (74%)	6/81 (7%)

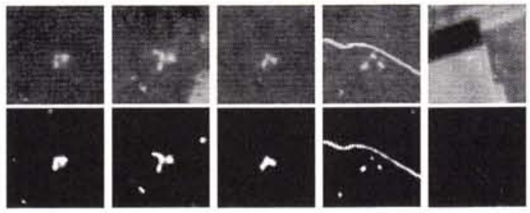

Figure 14. Examples of poorly defined signalized point patterns the system failed to recognize. A comparison with the training patterns in Figure 8 indicates that the system was not explicitly trained to recognize such patterns.

A number of potential improvements to the current system are also foreseen which can lead to improved performance:

- The range of SP patterns the system is able to recognize can be increased by the inclusion of example patterns into the training of the SOFM network.
- For the current system, the number of SP reference patterns was chosen, by experience, to be six. An improved SP pattern representation may be obtained if, during SOFM training, the number of reference patterns is automatically determined by the training algorithm. This can be achieved by incrementing the number of SOFM reference patterns until a certain minimal similarity score is achieved for each training pattern.
- The current similarity measures employed for both training the SOFM network and later for testing images is based on the simple Euclidean distance between the intensities of reference and testing patterns. Alternative measures, e.g., normalized cross correlation coefficient, may be more suitable.
- Currently, only a sample of the similarity scores computed by the SOFM are presented to the classification network (FFN). While this data compression leads to a more efficient implementation, the associated information loss may contribute to a degradation in performance. It is to be expected that, if the complete set of similarity responses obtained from the SOFM (i.e., the responses at each unit of each ring of the network) were to be used, system performance could be improved along with a reduction in the false-alarm rate.

In developing this system, greater emphasis was placed on optimizing the recognition rate than on efficiency, in particular, processing times for images. Currently, the complete testing of an image of 592 by 574 pixels requires about 10 to 15 minutes on a Sparc-2 workstation, depending on image content. It is anticipated that, with optimized software, dedicated hardware, and an improved strategy for pre-sorting image patches such that only those patches with promising feature content are tested, these times can be drastically reduced. Note further that only a single, asymmetric SP layout was considered in this work. Although alternate layouts are used in practice, e.g., crosses, the generality of the developed recognition strategy is not compromised. The ANNs simply need to be re-trained for the new pattern. In the case of crosses, the difficulty of the task can be simplified by taking into consideration the symmetry of the feature insofar as a reduced number of SP patterns orientations need to be catered for.

Finally, despite the promising results achieved by this work, and the potential offered by future system improvements, the authors remain skeptical that 100 percent robust SP recognition without false-alarms can be achieved purely on the basis of patterns in digitized aerial images. To achieve this, either an improved signalization procedure that increases the scale and uniqueness of the SP patterns in the aerial images is required, or a strategy capable of fusing all the sources of knowledge employed by the human operator needs to be developed.

Acknowledgments

The support of this work by the Swiss National Research Program 23 "Künstliche Intelligenz and Robotik" is gratefully acknowledged.

References

Baltsavias, E., H. Beyer, D. Fritsch, and R. Lenz, 1990. *Fundamentals of Real-Time Photogrammetry*, Tutorial Notes, Institute of Geodesy and Photogrammetry, ETH, Zurich.

Chiu, W.C., and E.L. Hines, 1990. Artificial Neural Networks for Photogrammetric Target Processing, *SPIE*, 1395:794–801.

DARPA, 1988. *Neural Network Study*, AFCEA International Press, Fairfax.

Duda, R.O., and P.E. Hart, 1973. *Pattern Recognition and Scene Analysis*, Wiley, New York.

Grün, A., and E.P. Baltsavias, 1988. Geometrically Constrained Multiphoto Matching, *Photogrammetric Engineering & Remote Sensing*, 54(5):633–641.

Këpuska, V.Z., 1992. *Literature Survey on the Feature Extraction for the Signalized Point Recognition*, Internal Report, September.

———, 1990. *Neural Networks for Speech Recognition Applications*, Dissertation, Department of Electrical and Computer Engineering, Clemson University.

Këpuska, V.Z., and S. Mason, 1991a. Automatic Signalized Point Recognition with Feed-Forward Neural Network, IEE Second International Conference on Artificial Neural Networks, Bournemouth, United Kingdom.

———, 1991b. An Artificial Neural Network Approach to Signalized Point Recognition in Aerial Photographs, First Australian Photogrammetric Conference, Sydney, 7–10 Nov.

Kohonen, T., 1988a. *Self-Organization and Associative Memory, Second Edition*, Springer Verlag, New York.

———, 1988b. An Introduction to Neural Computing, *Neural Networks*, 1:3–16.

Kohonen, T., G. Barna, and R. Chrisley, 1988. Statistical Pattern Recognition with Neural Networks: Benchmark Studies, *IEEE International Conference on Neural Networks*, San Diego, 1:61–68.

Lippmann, R., 1987. An Introduction to Computing with Neural Nets, *IEEE Acoustics, Speech, and Signal Processing Society Magazine*, 4(2):4–22.

Roth, M., 1990. Survey of Neural Network Technology for Automatic Target Recognition, *IEEE Transaction on Neural Networks*, 1(1): 28–43.

Rumelhart, D., J. McClelland, and the PDP research group, 1986. *Parallel Distributed Processing—Exploration in the Microstructure of Cognition*, Vol. 1, The MIT Press, Cambridge.

Thoet, W.A., T.G. Rainey, D.W. Brettle, L.A. Slutz, and F.S. Weingard, 1992. ANVIL Neural Network Program for Three-Dimensional Automatic Target Recognition, *Optical Engineering*, 31(12):2532–2539.

Toth, C.K., and T. Schenk, 1990. Pattern Recognition with Parallel Associative Memory, *SPIE Conference*, 1395:558–563.

Wilkins, D., 1990. Digital Photogrammetric Applications with the Prime-Wild S9 Analytical Plotter, ISPRS Symposium of Commission II, Dresden.

(Received 6 April 1993; accepted 3 August 1993; revised 8 October 1993)

Veton Z. Këpuska

Veton Z. Këpuska received his Bachelor degree in Electrical Engineering from the University of Kosova in Prishtina, in 1981. He served as an Assistant Lecturer at the Electrical Engineering Faculty of the University of Kosova from 1981 until 1984. In 1984 he continued his Graduate Studies at Clemson University, South Carolina, as a Fulbright Fellow. In 1986 he obtained his MSc degree, and in 1990 his Ph.D. degree, both in Computer Engineering. From 1990 until 1992 he joined Swiss Federal Institute of Technology, Institute of Geodesy and Photogrammetry, Zürich, as a Post-Doctoral Research Associate. Currently he is with Voice Processing Co. Cambridge, MA.

Scott Mason

Scott Mason received his Bachelor of Surveying degree from the University of Newcastle, Australia, in 1988 and the MSc degree from the University of Illinois at Urbana-Champaign in 1990. He is currently completing his doctoral thesis at the Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology, Zurich.