OBJECT RECOGNITION AND MEASUREMENT FROM MOBILE MAPPING IMAGE SEQUENCES USING HOPFIELD NEURAL NETWORKS: PART II

Dr. Rongxing Li, Weian Wang, Zhuowen Tu and Hong-Zeng Tseng
Department of Civil and Environment Engineering and Geodetic Science
The Ohio State University, Columbus, OH 43210, USA
Tel: (614) 292-6946, Fax: (614) 292-2957
E-mail: li.282@osu.edu, www:http://gscrl.eng.ohio-state.edu

KEY WORDS: Hopfield neural networks, object recognition, light poles, back projection, image sequence, mobile mapping

ABSTRACT

Spatially referenced mobile mapping image sequences contain rich information for applications such as transportation and utility management. Automatic object recognition and measurement of objects from the images to reduce human operations and to enhance efficiency is a challenge in mobile mapping data processing. Part I of the paper, published in proceedings of ASPRS 1998 annual convention, introduces the principle of Hopfield neural networks and recognition of objects in stereo pairs. This paper deals with object recognition and measurement from stereo image sequences.

1. INTRODUCTION

This paper presents research results of recognition and measurement of objects, particularly street light poles, from mobile mapping image sequences using Hopfield neural networks. Part I of the paper deals with literature survey and object recognition from stereo pairs and was published in Li et al. (1998). Part II introduces a modified Hopfield neural network and presents application results of the model using image sequences. The mobile mapping systems and data used in this research are depicted in Li (1997) and Li et al. (1998). The neural network makes a good use of priori knowledge of the object (light poles) and cameras (exterior orientation parameters) when recognizing objects. The developed algorithm a) recognizes all light pole features in an image sequence, b) finds corresponding light pole features among the recognized poles in the sequence, and c) calculates locations of the poles from multiple recognized corresponding pole image features by a bundle adjustment. Results of an experiment using an actual image sequence are presented. Discussion and conclusion are given based on the promising results.

2. A BRIEF DISCUSSION ON THE NEURAL NETWORK

Suppose that a 3-D street light pole model is back projected onto an image as two edges numbered 0 and 1 (rows) in Figure 1. The objective is to find image features of light poles from the extracted edges in the image, which are numbered from 0 to 9 (columns) in Figure 1. The one-layer Hopfield neural network solves this recognition problem by minimizing the following energy function (Li et al. 1998):

$$E = -\sum_i \sum_k \sum_j \sum_l C_{ikjl} V_{ik} V_{jl} + \sum_i (1 - \sum_k V_{ik})^2 + \sum_k (1 - \sum_i V_{ik})^2 \qquad (1)$$

C_{ikjl} represents the interconnection between neurons (i,k) and (j,l) and is evaluated to describe the similarity (or

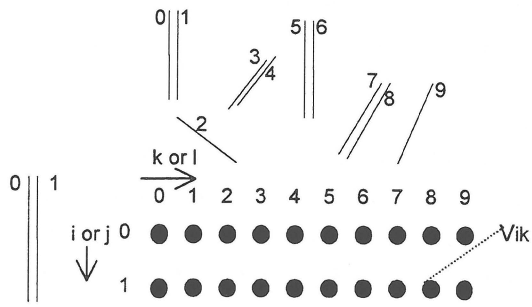


Figure 1. Hopfield neural network for finding street light pole image features using back projected edges of a 3-D light pole model

dissimilarity) between the model edges and image edges (Lin et al. 1991, Li et al. 1998). The neuron state V_{ik} converges to 1.0 if model edge i matches image edge k perfectly, otherwise, it is greater than or equal to 0. The second term $\sum_i (1 - \sum_k V_{ik})^2$ and third term $\sum_k (1 - \sum_i V_{ik})^2$ in Equation (1) apply uniqueness constraints so that each row (model edge) matches only one column (image edge) and vice versa.

To enhance the uniqueness constraints, two additional terms $\sum_i \sum_k \sum_{l \neq k} V_{ik} \times V_{il}$ and $\sum_k \sum_i \sum_{j \neq i} V_{ik} \times V_{jk}$ are appended to Equation(1). The energy function can then be expressed as

$$E = -A \sum_i \sum_k \sum_j \sum_l C_{ikjl} V_{ik} V_{jl} + B \sum_i (1 - \sum_k V_{ik})^2 + C \sum_i \sum_k \sum_{l \neq k} V_{ik} \times V_{il} + D \sum_k (1 - \sum_i V_{ik})^2 + E \sum_k \sum_i \sum_{j \neq i} V_{ik} \times V_{jk}, \qquad (2)$$

where A, B, C, D and E are weights. It is expected that Equation (2) will accelerate the convergence of the iterative computational procedure. Furthermore, it will increase the contrast of neuron states V_{ik}. Thus, neuron states representing matches will have high values (close to 1) and those for nonmatches will have low values (close to 0). This helps interpret the recognition results. Details of solving the energy function is presented in Li et al. (1998).

Equation (2) can be rearranged as a Lyapunov function (Hopfield and Tank 1985, Nasrabadi and Choo 1992)

$$E = -\frac{1}{2}\sum_i\sum_k\sum_j\sum_l T_{ikjl}V_{ik}V_{jl} - \sum_i\sum_k I_{ik}V_{ik}, \tag{2a}$$

where $I_{ik} = 2$ in this case and $T_{ikjl} = A \times C_{ikjl} - B\delta_{ij} - C\delta_{ij}(1-\delta_{kl}) - D \times \delta_{kl} - E \times \delta_{kl}(1-\delta_{ij})$. $\delta_{ij} = 1$ if $i = j$, otherwise $\delta_{ij} = 0$.

3. RECOGNITION OF LIGHT POLES IN ALL IMAGES OF THE SEQUENCE

This is to find all light pole features in all image of the sequence. It is carried out by building a neural network, such as Figure 1, for each image. As a result, image edges matching the pole model are indicated by high neuron states. An algorithm is developed to find image edge pairs among the recognized pole edges, which form poles in the image. Once this procedure runs on the entire image sequence, light poles in all images are recognized.

This object recognition task requires that one model (row 0 and row 1) allows to match multiple image edges (columns), while one image edge (column) should only match one model edge (row). Therefore, the weights in Equation (2) are set as A=1, B=0.1, C=0, D=0.65, and E=1. The solution of Equation(2) provides two sets of edges that match model edge 0 and 1, respectively (Figure 1): the first set is $S_0 = \{edge_{0k} \mid V_{0k} \geq threshold\}$, and the second set contains $S_1 = \{edge_{1k} \mid V_{1k} \geq threshold\}$. Only two edges from different sets of S_0 and S_1 can be combined to form the edge pair of a pole. The procedure of determination of the edge pairs is described by the following steps:

1) Choose recognized image edges by examining V_{0k} and V_{1k} and form S_0 and S_1.
2) For any $edge_{0k} \in S_0$, compute its strength of connection M_p with the p-th $edge_{1p} \in S_1$
   $$M_p = C_{0k1p} + C_{1p0k}. \tag{3}$$
   M_p will be calculated for all $edge_{1p} \in S_1$. The edge in S_1 that has the maxim value M_p and $M_p \geq threshold$ is combined with the edge in S_0 to build a pole edge pair $(edge_{0k}, edge_{1p})$. Otherwise, $edge_{0k}$ does not have a partner edge and is not a pole edge.
3) Repeat step 2) for all $edge_{0k} \in S_0$, so that all possible edge combinations are examined. The output of the procedure is the recognized edge pairs of light poles in the image.
4) Repeat steps 1) and 3) to find all edge pairs of poles for all images in the sequence.

4. OBJECT RECOGNITION AND LOCATION IN THE OBJECT SPACE

So far, the light poles are recognized in the image space. Corresponding edge pairs of poles, which are image features of light poles in the object space, have to be recognized and used to locate the poles in the object space photogrammetrically. For example, a detected a pair of edges in one image is selected. The image coordinates of the pole bottom and pole top are known as (x_b, y_b) and (x_t, y_t). Their corresponding coordinates in the object space are (X_B, Y_B, Z_B) and (X_T, Y_T, Z_T). Assume that the pole is vertical and the pole length is l, we have $X_T = X_B$, $Y_T = Y_B$, and $Z_T = Z_B + l$. Since the camera orientation parameters are known, there are four collinearity equations from two spatial lines, namely, the line from the exposure center (X_o, Y_o, Z_o) to the pole bottom in the image (x_b, y_b,-f) and to the pole bottom (X_B, Y_B, Z_B), and another line from the exposure center (X_o, Y_o, Z_o) to the pole top in the image (x_t, y_t,-f) and to the pole top $(X_B, Y_B, Z_B + l)$ (Figure 2):

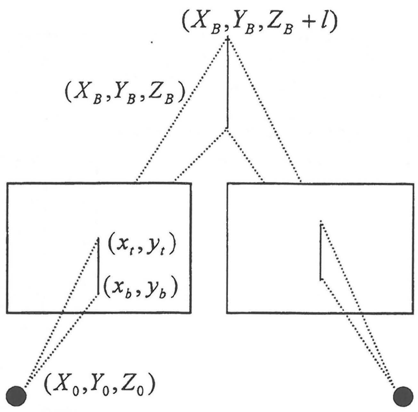


Figure 2. Finding corresponding pole image features in a stereo image pair

$$x_b = -f\frac{a_{11}(X_B - X_0) + a_{12}(Y_B - Y_0) + a_{13}(Z_B - Z_0)}{a_{31}(X_B - X_0) + a_{32}(Y_B - Y_0) + a_{33}(Z_B - Z_0)},$$

$$y_b = -f\frac{a_{21}(X_B - X_0) + a_{22}(Y_B - Y_0) + a_{23}(Z_B - Z_0)}{a_{31}(X_B - X_0) + a_{32}(Y_B - Y_0) + a_{33}(Z_B - Z_0)},$$

$$x_t = -f\frac{a_{11}(X_B - X_0) + a_{12}(Y_B - Y_0) + a_{13}(Z_B + l - Z_0)}{a_{31}(X_B - X_0) + a_{32}(Y_B - Y_0) + a_{33}(Z_B + l - Z_0)}, \text{ and}$$

$$y_t = -f\frac{a_{21}(X_B - X_0) + a_{22}(Y_B - Y_0) + a_{23}(Z_B + l - Z_0)}{a_{31}(X_B - X_0) + a_{32}(Y_B - Y_0) + a_{33}(Z_B + l - Z_0)}. \tag{4}$$

In the above four equations, there are three unknowns (X_B, Y_B, Z_B) which can be solved by a least squares adjustment. The coordinates are then used to back project the 3-D pole model onto the second image. The two back projected model edges are then used to match the image pole edges in the second image by a neural network (position specific, Li et al. 1998). Once the corresponding image pole feature in the second image is found, the precise location of the pole in the object space can be calculated by a photogrammetric triangulation. To obtain a more precise location of the pole, the 3-D pole model can be projected onto multiple images and all corresponding image pole features in the images can be recognized. A bundle adjustment using all corresponding image pole features gives an precise 3-D location of the pole. Once the above procedure is performed for all images in the sequence, a spatial database of street light poles can be generated.

5. RESULTS AND ANALYSIS

5.1 Recognition of Light Poles in the Image Sequence

To detect all light poles in the image (Figure 3), regardless of their positions, measuring features that are not significantly affected by object positions in the object space should be applied. It is clear that lengths of pole edges in the image space vary because of pole positions and the perspective projection. The selected measuring features include azimuth of edges (close to vertical), width-length ratio (photo-scale invariant), and relative gradient (compatibility of a line pair to form a pole, photo-scale invariant). Other measuring features are excluded by setting the corresponding weights to 0. For the image data, the applied parameters are:

Threshold of λ=0.00	Weight of λ=0.00
Threshold of α=3.00	Weight of α=0.25
Threshold of δ=0.00	Weight of δ=0.00
Threshold of δ/λ =0.05	Weight of δ/λ =0.50
Threshold of gradient=0	Weight of gradient=0.25

The modified energy function (2) was used. Table 1 gives the final states (V_{ik}) of the neurons for detecting all light poles in the image. The neural network recognized all light poles correctly by indicating matches of model edges 0 and 1 with image pole edges of 4 and 5, 8 and 9, 12 and13, and 44 and 45 with the maximum state value 1.0.

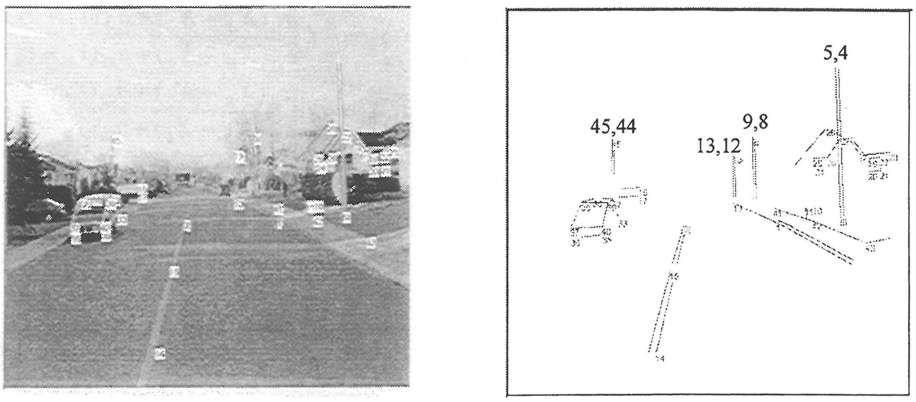
This recognition process can proceed to detect all light poles in the image sequence.

ID of extracted line	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Neuron state (Matching with model line 0)	0	0	0	0	1.0	0	0	0	1.0	0	0.27	0	1.0	0	0	0
Neuron state (Matching with model line 1)	0	0	0	0	0	1.0	0	0	0	1.0	0	0.27	0	1.0	0	0

ID of extracted line	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Neuron state (Matching with model line 0)	0	0	0	0.27	0	0	0	0	0	0	0	0	0	0	0	0
Neuron state (Matching with model line 1)	0	0	0	0	0	0.28	0	0	0	0	0	0	0	0	0	0

ID of extracted line	32	33	34	35	36	37	38	39	40	41	42	43	44	45		
Neuron state (Matching with model line 0)	0	0	0.28	0	0.28	0	0	0	0	0	0	0	1.0	0		
Neuron state (Matching with model line 1)	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0		

Table 1. Final states (V_{ik}) of the neurons for detecting all light poles in Figure 3



(a) (b)

Figure 3. Image (a) and extracted and labeled edges (b)

5.2 Location of Light Poles in the Object Space

Figure 4 depicts a sequence of 25 stereo image pairs from a mobile mapping survey line. Exposure stations of stereo images taken simultaneously by the mobile mapping system are indicated by + + symbols. The exposure station is numbered from 102 to 127. The two stereo images at each station are distinguished by its station number with an L (left) or R (right) extension, for example, 112L and 112R. Six light poles, from LP1 to LP6, are marked by symbols. The exposure stations and/or images are listed for each light pole that appears in the images. The algorithms discussed above are able to detect and locate a position specific light pole or to detect all light poles in the image sequence in the image space. In order to locate the detected poles in the object space, corresponding pole image features in stereo images have to be identified. Subsequently, their locations in the object space can be triangulated from the detected corresponding image features. Thus, the major task at this stage is to find the corresponding image pole features in the image space. Such corresponding features are mostly found in stereo image pairs with "hard" baselines at the same exposure stations, but they may also be found in stereo image pairs with "soft" baselines formed by images of preceding and/or following exposure stations. Considering the effective baseline (baseline component vertical to the line that links the pole and the middle point of the baseline) of possible combinations of stereo pairs with hard/soft baselines, for each pole, only three stations that are close to the pole are used for location of the pole. For example, light pole LP2 in Figure 4 is covered by images of exposure stations 102, 103, 104, 105, 106, 107, and 108. However, only three stations (106, 107, and 108) are used. For example, the location of pole LP2 is first calculated from Equation (4). It is then back projected onto multiple images (106L, 106R, 107L, 107R, 108L, and 108R). The neural network finds all corresponding image pole features in the selected images. A bundle adjustment is applied to estimate the optimal location of the pole in the object space using all corresponding image pole features. This process is performed on the entire image sequence, so that all light poles in the sequence are located.

Figure 5 presents locations of light pole LP2 estimated by 11 combinations of stereo image pairs and a bundle adjustment. Ideally, the locations should be at the same point or within a very small area. However, because of the relatively short effective baselines (small intersection angles) the locations are spread along the track direction. In the middle of the distribution are points calculated from the stereo pairs with larger effective baselines (e.g., 106R&108L and 106L&108L) and those close to the pole. Two image pairs, namely 107L&108R and 107R&108R, have very small effective baselines and the locations calculated thereby are far away from the average location. A bundle adjustment using all detected corresponding image features of LP2 from three close exposure stations is performed and leads to the location that matches the results of triangulations with larger effective baselines, instead of the average location from all individual locations. This confirms the conclusion from a previous study on optimization of photogrammetric triangulation using mobile mapping data (Li et al. 1996).

6. CONCLUSIONS

Algorithms based on Hopfield neural networks for object detection and location, especially for street light poles, from mobile mapping image sequences have been researched and developed. A software system N^2M^2 (Neural Networks for Mobile Mapping) is developed based on the C++ programming language in the Microsoft Windows 32bits environment. Major contributions of this research are:

- Establishment of a Hopfield neural network for object recognition from mobile mapping image sequences using 3-D object models,
- Application of the developed algorithms for detection and location of light poles from a single image pair and/or from an image sequence,
- Understanding of the behavior of the neural network when applied in various mobile mapping situations, and
- Development of the N^2M^2 system.

A further challenge is to develop a systematic learning process of the neural network for handling mobile mapping data. We believe that such research will result in a generic method for the optimal determination of thresholds and weight values in the network for different objects.

7. ACKNOWLEDGEMENT

This research was supported by the OSU Center for Mapping (CFM) and NASA Center for the Commercial Development of Space. Mobile mapping data are provided by TRANSMAP Corp. and VISAT Technologies Inc. Discussions and collaboration with Dr. J. Bossler and colleagues at CFM are appreciated.

REFERENCES

Li, R., 1997. Mobile Mapping - An Emerging Technology for Spatial Data Acquisition. *Photogrammetric Engineering and Remote Sensing,* Vol.63, No.9, pp.1085-1092.

Li, R., M.A. Chapman, and W. Zou, 1996. Optimal Acquisition of 3D Object Coordinates from Stereoscopic Image Sequences. *Intl. Archives of PRS,* Vol. XXXI, Part B3, pp.449-452.

Li, R., W. Wang, and H.-Z. Tseng, 1998. Object Recognition and Measurement from Mobile Mapping Image Sequences Using Hopfield Networks: Part I. *Proceedings of ASPRS/RT Annual Conference* 1998, CDROM.

Li, W. and N.M. Nasrabadi, 1989. Object Recognition Based on Graph Matching Implemented by a Hopfield-Style Neural Network. *Int. J. Conf. Neural Networks II,* Washington, DC, June 18-22, pp. 287-290.

Lin, W.-C., F.-Y. Liao, C.-K. Tsao, and T. Lingutla, 1991. A Hierarchical Multiple-View Approach to Three-Dimensional Object Recognition. *IEEE Transactions on Neural Networks.* Vol.2, No.1, pp.84-92.

Hopfield, J.J. and D.W. Tank, 1985. "Neural" Computation of Decisions in Optimization Problems. *Biol. Cybern.* Vol. 52, pp.141-152 .

Nasrabadi, N. M. and C. Y Choo, 1992. Hopfield Network for Stereo Vision Correspondence. *IEEE Transaction on Neural Networks,* Vol. 3, No. 1, pp. 5-12.

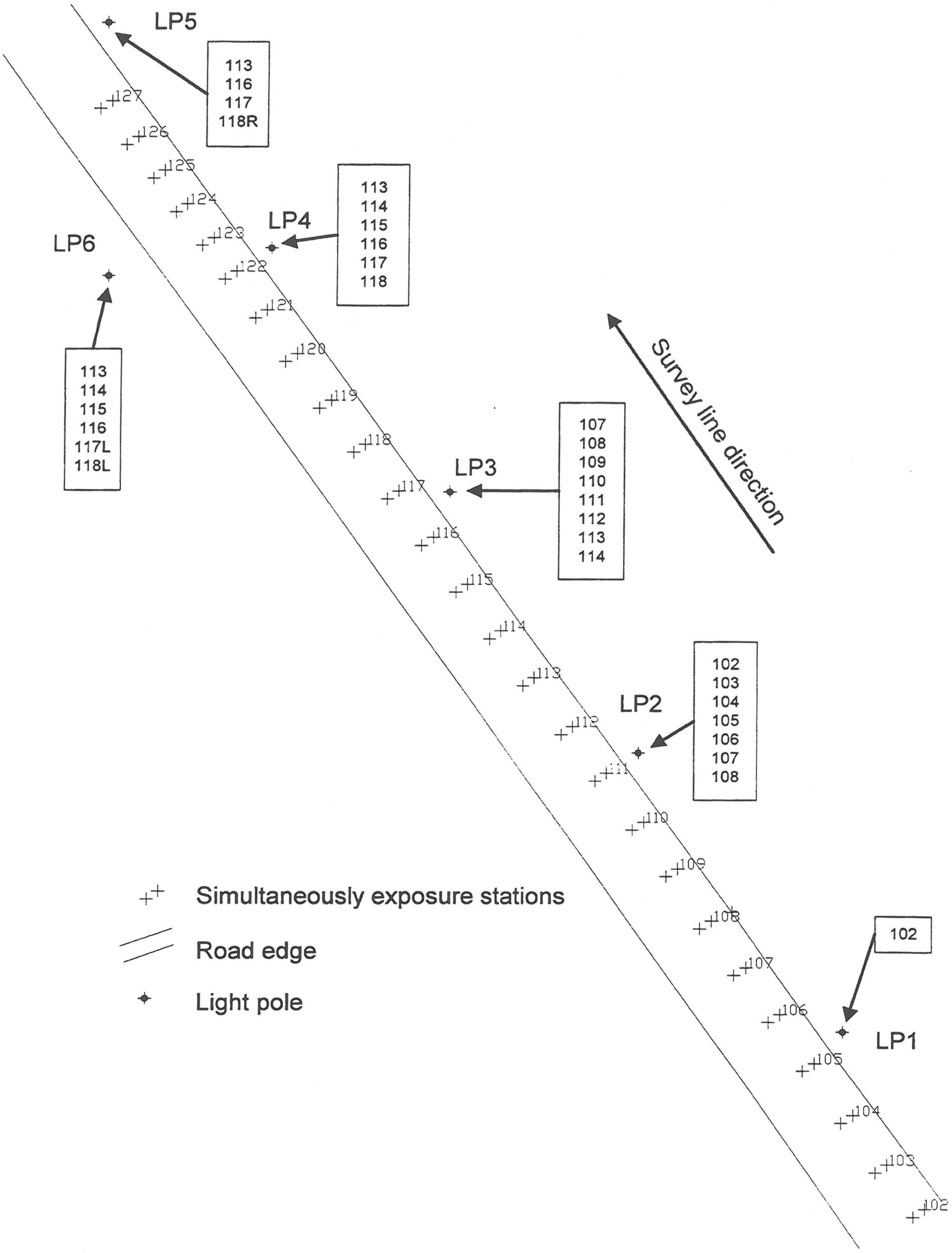


Figure 4. Light pole recognition and location from an image sequence

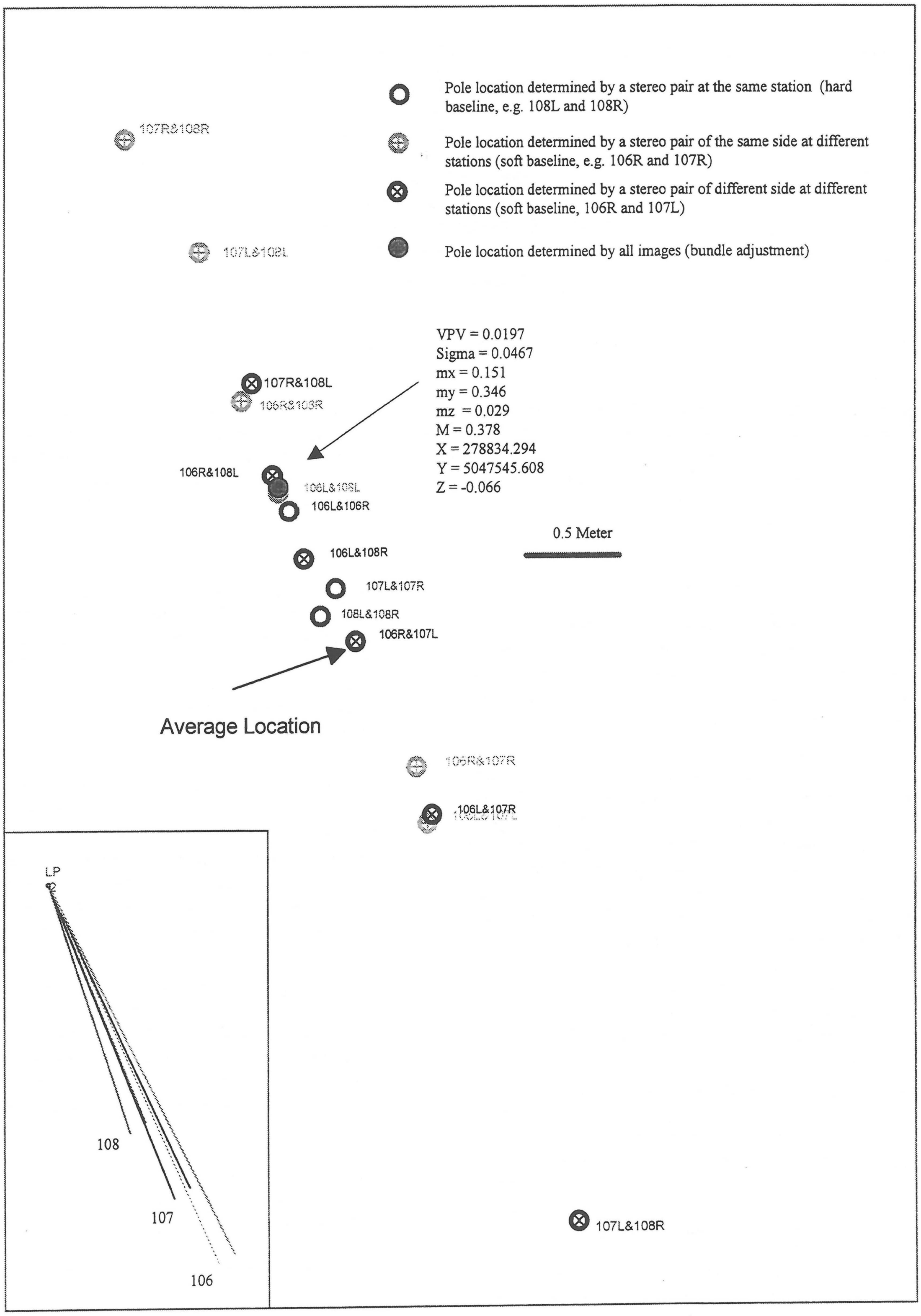


Figure 5. Distribution of locations of light pole LP2 estimated by various Combinations of stereo pairs and a bundle adjustment