Neural Network Method for Sharpening Landsat Thermal Data from Higher Resolution Multispectral Data

By George Lemeshewsky

Open File Report 97-301

U.S. Department of the Interior
U.S. Geological Survey
National Mapping Division

Neural Network Method For Sharpening Landsat Thermal Data From Higher Resolution Multispectral Data

George Lemeshewsky
U.S. Geological Survey
521 National Center
Reston, VA 22092
Open-file Report 97-301

ABSTRACT

In this study, a sharpening method based on a neural network (NN) approximation technique is described to increase the spatial resolution of thematic mapper (TM) thermal-infrared (T-IR) data. Sharpening is derived from a learned input-output mapping of edge contrast patterns between T-IR and higher resolution TM bands. This method is similar to a reported adaptive least squares (LS) method to estimate TM T-IR data at a higher resolution. However, there are two major differences: use of NN approximation instead of LS estimation, and application or reported multiresolution technique to adaptively combine spatial information from the original image and its high-resolution estimate. With training examples from reduced resolution data, a multilayer feedforward NN is trained to approximate T-IR data samples on the basis of a possible nonlinear combination of data samples from three other TM bands. The trained NN output for full-resolution input data is an estimate of T-IR image at 30-m resolution. A potential benefit of this sharpening method is that the NN approximation technique can be developed from only a subset of image scene samples, and yet be applied to the entire scene. Preliminary examples show sharpening at four times higher resolution, but further evaluation with a high-resolution reference image is recommended. Although results are encouraging, a different training strategy to improve network generalization is suggested as a way to improve the sharpening process.

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

I. Introduction

Image sharpening has been defined as a resolution enhancement process where information from a higher spatial resolution image is used to increase the resolution of a second lower resolution image (Iverson and Lersch, 1994). This process is also generally referred to as image merging or image fusion (Schowengerdt and Filiberti, 1994). For many applications, such as image display and interpretation, it is important that the enhanced, merged image incorporate as much spatial information as possible from the individual images, perhaps at the expense of spectral fidelity. As much as possible, sharpening techniques should preserve the spectral characteristics of the original image while increasing the spatial resolution. However, with applications that rely on machine classification of multispectral data, spectral fidelity can be the most important criterion of a sharpening process. Some reported techniques directed toward increasing the spatial information include intensity-hue-saturation, principal component, and high-pass filtering techniques (Iverson and Lersch, 1994).

The Landsat thematic mapper (TM) is a dual spatial resolution sensor that records reflected and emitted energy in seven spectral bands, which range from visible (blue) to thermal-infrared (T-IR). Six of the spectral bands have a 30- by 30-m instantaneous field of view (IFOV). The seventh band (labeled band 6) has a 120- by 120-m IFOV and records, in the 10.4- to 12.5-μm spectral band, the infrared radiant heat flux from surfaces. These data are important for vegetation stress analysis, vegetation classification, and soil moisture studies (Jensen, 1986).

Unfortunately, the difference in resolution of band 6 causes many difficulties in analyzing TM data. With discriminate analysis of TM data, often band 6 data are omitted, resulting in loss of information (Nishii, Kusanobu, and Tanaka, 1996). Spectrally accurate, enhanced spatial resolution T-IR data at 30-m resolution might help to identify vegetation communities and distinguish vegetation density in areas consisting of mixed land cover (for example, open water and vegetation) like that found in the Everglades of southern Florida.

This report describes initial results of an image-sharpening method that makes use of neural network (NN) approximation techniques. This method is similar to a reported adaptive least squares (LS) method for sharpening TM T-IR data at a higher resolution (Tom, Carlotto, and Scholten, 1985) because the latter approach to sharpening suggests an implementation with NN approximation and other techniques that, in turn, simplify the overall process and possibly improve performance. The intended application of this technique is to improve the automated classification of vegetation cover from TM data.

Two previously reported adaptive methods for sharpening TM T-IR data from higher resolution TM data, which are similar to the method reported here, are described in the next section. (A local statistical predictor approach to enhancing TM thermal band image data that is based on fixed, not moving or sliding, windows was reported by Nishii, Kusanobu, and Tanaka (1996). In passing, it is noted that adaptive methods are needed to improve spectral fidelity of sharpened multispectral data. As pointed out by Tom (1986), Schowengert (1980) first described a highpass filter sharpening technique that used adaptive techniques in order to deal with the edge contrast reversals frequently seen in visible and infrared images (Landsat multispectral scanner data). These opposite contrast patterns correspond, for example, to vegetation-soil boundaries on the image scene.

It is believed that because of these localized opposite contrast effects, adaptive sharpening algorithms are required to improve spectral fidelity. On the other hand, numerous reported sharpening techniques based on global characterization of the data also can be quite effective when, because of scene content, there are few opposite contrast patterns.

The process reported here is adaptive to the local edge patterns and also incorporates a backpropagation-trained (to minimize a sum of squares approximation error) NN approximation method to estimate, with minimal error, sharpened data.

II. Background

The earliest report on adaptive sharpening of TM T-IR data that explicitly included optimal estimation methods was given by Tom, Carlotto, and Scholten (1985). They exploit the observation that edge contrast boundary information appears locally correlated across bands by developing an adaptive LS method to optimally estimate T-IR data at higher resolution using a moving window of samples from the higher resolution TM bands. A frequency domain replacement process was used to combine the estimated and original data. Using simulated low-resolution data, they reported a relatively low mean-square gray-level error (2.21) between true and sharpened data (band 7 as a function of bands 5, 4, 3, and 1). Basic steps to this process included the following:

1. reducing resolution by lowpass filtering and subsampling,
2. calculating LS estimator coefficients for T-IR data (at 120 m) using data from the other reduced-resolution bands,
3. estimating T-IR data at 30 m from interpolated estimator coefficients and full-resolution data, and
4. combining the high-resolution estimate with the original data by a frequency domain, highpass and lowpass filtering process.

Recently, Iverson and Lersch (1994) described a TM T-IR sharpening technique that was based on NN estimation methods and the multiresolution image pyramids of Burt and Adelson (1983). The NN was trained to estimate the required higher resolution, bandpass pyramid data needed for sharpening. The TM images were first decomposed into a multiresolution image pyramid of Gaussian (lowpass) and Laplacian (bandpass, or detail) images at sample distances of 30, 60, 120 m, and greater. An NN was trained to estimate reduced-resolution band-pass data from the other detail images. Output of the trained NN, an estimate for the T-IR bandpass images at the next two higher resolutions (60 and 30 m) of the pyramid, was used to reconstruct a sharpened image. Iverson and Lersch observed that because of the local correlation between the band-pass images, the NN was able to learn the input-to-output mapping. Although no quantitative measure of spectral fidelity was reported, their results indicate that a multilayer NN can estimate multiresolution (highpass) detail information over a 4:1 scale.

Aside from the important multiresolution approach, one difference between the two previously described sharpening methods is that for the LS method, the estimated image was a function of image data, without bandpass and lowpass decomposition. Conversely, for the NN-based process, highpass data, and at multiple resolutions, was estimated from highpass data. Comparative evalu-

ation of the performance of these two techniques would be beneficial. In the following experiment, a multilayer NN is used to estimate T-IR data as a function of image data, not multiresolution derived bandpass data, from three higher resolution bands.

III. NN Sharpening Method

The multiple processing steps for the NN-based sharpening technique described in this report are similar to those for the sharpening technique reported by Tom, and others (1985), with these major differences: (1) T-IR data at higher resolutions are estimated by a trained multilayer NN instead of an adaptive LS estimator, and (2) estimated and original data are adaptively combined by a reported multiresolution, image pyramid merging technique instead of a frequency domain replacement technique.

The processing steps below (and shown in figure 1) are for the case where the original 120-m IFOV T-IR data have been interpolated to a 30-m sample distance and coregistered with the other image bands (that is, TM level 1 scene). In the following, an estimator for band 6 is developed from the arbitrary combination of bands 7, 5, and 2; other band combinations are the subject of further study.

1. Bands 7, 6, 5, and 2 image data are reduced from 30- to 120-m sample distance by a sequence of lowpass filtering and down sampling; a Gaussian image pyramid process (fig.1a.)
2. Reduced-resolution data and supervised learning are used to train an NN to approximate band 6 data (10.4-12.5 μm) from the arbitrary combination of bands: 7 (mid-infrared: 2.08-2.35 μm), 5 (mid-infrared: 1.55-1.75 μm) and 2 (green 0.52-0.60 μm) (Jensen, 1986) (fig. 1b).
3. Output of the trained NN provides an estimate of band 6 data at 30 m when given full-resolution data from bands 7, 5, and 2.
4. The higher spatial frequency information from the NN-estimated high-resolution image is transferred to the band 6 image by means of a reported multiresolution image pyramid-based technique (fig. 1c).

A potential advantage of the NN approximation of steps 2 and 3 is that the NN can be trained on a small, representative subset of the image data and then applied to the entire image. In contrast, the adaptive linear LS sharpening method requires computing and then interpolating linear (LS) estimation coefficients for each local window of the entire image.

As shown in figure 1, the reduced-resolution data of step 1 were derived from the third level of the Gaussian image pyramid sequence after two iterations of lowpass filtering and down sampling by a factor of 2 (that is, discarding alternate rows and columns) (Burt and Adelson, 1983). Separable lowpass filter coefficients were 1/16 [1 4 6 4 1] (Burt, 1985). This multiresolution process was used because of its computational efficiency and implementation with small-kernel, one-dimensional filters.

Step 4 is the application of a reported Laplacian pyramid transform technique (Ogden, Adelson,

Bergen, and Burt, 1985) for adaptively combining spatial detail from two coregistered images that differ only in sharpness. (The article by Li, Manjunath, and Mitra (1995) has additional examples of this fusion technique.) This step is included to preserve the lowpass, gray level value of the raw T-IR data. As applied here, step 4 (fig. 1c) includes the following:

1. decomposition of both the TM-IR image and its high-resolution estimate into two Laplacian (bandpass) image pyramids (two bandpass resolution levels),
2. selectively replacing individual Laplacian image samples from band 6 with those from the estimated band 6 on the basis of a maximum selection rule, and
3. reconstructing a sharpened T-IR image from its modified bandpass-data.

In effect, the maximum selection rule causes excess spatial detail to be transferred to the lower resolution image because the sharper contrast boundaries of the estimated image have larger amplitude.

IV. Approximation by Neural Network

The function of the multilayer NN is to estimate band 6 T-IR data at a higher resolution on the basis of a learned, potentially highly nonlinear, relation between edge contrast patterns of band 6 and bands 7, 5, and 2. The estimator was developed in a two-step process. Supervised backpropagation learning techniques (Rumelhart, Hinton, and Williams, 1986) with reduced-resolution (120 m) data were used to train a single output NN to approximate individual band 6 data samples, as a function of a 3- by 3-sample neighborhood from an arbitrary combination of bands 7, 5, and 2 (27 input samples). Backpropagation training attempts to minimize an LS objective function (the sum of squared residuals) using a gradient descent technique (Chen and Jain, 1994). Once trained, NN output for higher (that is, original) resolution input data is taken as an estimate of T-IR data at a higher resolution.

This output is, to some degree, an accurate estimate of true 30-m data because of the network's property of generalization; that is, similar input (to the training data) produces similar output; thus the network approximates the correct output for an input not in the training data (Wasserman, 1993). A network that does not generalize, but produces the correct output for training data input, would be of little value; it could be replaced by a simple lookup table. In this application, generalization allows network learning based on a limited number of input-output example pairs from a subarea of the image, if the examples are representative of patterns throughout the entire image and over several scales of resolution. The subject of further study, this process assumes a self-similarity exists between edge contrast patterns of the reduced and full images.

V. Network Configuration and Training

A multilayer artificial feedforward NN consists of layers of interconnected neurons or processing elements (PE). The interconnections, or weights, are repeatedly modified during a training phase so that the NN performs a desired mapping between input-output example pairs. Figure 2a shows the type of single hidden layer network used in this application. The output (Out) of each PE

(fig. 2b) for its inputs x_i and input interconnection weights w_i is

$$Out = \mathrm{f}(net) = \mathrm{f}\left(\sum_{i=0}^{n} x_i w_i \right) \tag{1}$$

where x_0, w_0 are input bias (fixed value) and bias weight, respectively; f(·) is the nonlinear sigmoidal function, and *net* is the net sum of $x_i w_i$ for $i = 0,...,n$ (Wasserman, 1993). For back-propagation (BP) learning, f(*net*) must be differentiable. Two widely used activation functions are f(*net*) = 1/(1 + exp(-*net*)) and f(*net*) = tanh(*net*). Tanh(.) was used because its derivative is largest when | *net* | is small. Thus during BP training, stronger weight modification is made to PE's having small | *net* |; that is ". . . those which are "in doubt" about their output" (Hertz, Krogh, and Palmer, 1991).

The process of training an NN with BP learning requires the determination of several interrelated network parameters: network size (number of hidden layers and number of PE's per layer,) training set size, learning rate parameters, numerical range for initial values of the weights, and amount of training needed.

The NN used here has 27 inputs and 1 output (estimate of band 6 data sample). One hidden layer was used here because this is sufficient to produce an arbitrary continuous function mapping of input-to-output to any desired degree of accuracy given a sufficient, but unspecified, number of hidden layer PE's (Hornik, Stinchcombe and White, 1989). However, this reference does not address network training to produce this mapping nor performance relating to unseen data that are not part of the training data set; that is, the generalization problem.

One way to obtain good network generalization is to use the smallest NN that will fit the problem. This may be done by successively training smaller networks until the smallest one that meets the required performance is found. Another approach is to train a larger than necessary network and then remove PE's that are not needed (Reed, 1993). Often, a large network can be a poor choice because too many hidden layer PE's (and interconnection weights) lead to poor generalization. There are methods for improving generalization by 'pruning' or removing selected PE's or weights (Hertz, Krogh, and Palmer, 1991; Reed, 1993).

For these initial results, training and validation data sets (7,938 samples each) were simply from the lower half and upper half, respectively, of a 128- by 128-sample (120-m sample distance) image, less border samples. The generation of a more representative set of training patterns might be based on a process for selecting edge contrast patterns throughout one or more scenes.

A simple process was used to find a small number of hidden layer PE's. Two NN's (15 and 10 PE's in the hidden layer) were trained for an arbitrary 10 presentations of the training data, and their RMS error to the training data set compared. For this and the next training, momentum was always 0.0 and learning rate was 0.1 for the first 7,938 samples, then 0.01 for the remaining training. The range for initial random values of the weights was -0.1 to 0.1. Because each had comparable error, the smaller network was used. Presumably this less complex NN would be less prone to overlearning or overfitting to the learning data.

Next, the 27 input, 10-PE hidden layer, 1output NN was trained for an arbitrary 120 presentations of the training data (973,800 weight updates). The amounts of RMS error for network rescaled training and validation data sets were 0.0375 and 0.0315, respectively, after 20 presentations of the training data, and at the end of training 0.0321 and 0.0354, respectively.

A poor assumption with these first tests was that this training would not lead to overlearning and poor generalization. Thus, a modified training strategy is recommended to minimize overtraining, or overfitting to the training data, whereby training proceeds until network generalization, as measured by network performance to the validation data, starts to increase. This approach is motivated by the fact that during the early stages of training, both training and test-validation error decrease as the network generalizes to the underlying data. However, with further training, training error will continue to decrease, but error to the validation data will eventually reach a minimum and then increase (Reed, 1993). Training should stop at this minimum.

VI. Image Test Results

Figure 3a shows the TM T-IR band 6 data reduced to 120-m resolution. (With the level 1 scene, TM-IR data are interpolated to a 30-m sample distance.) Training and test data were from the upper half and lower half of the image, respectively. The NN-estimated band 6 image as a function 3- by 3-sample neighborhood (27 total) from bands 7, 5, and 2 is shown in fig. 3c. The band 5 image (fig. 3b) is given for comparison. Note the improved spatial detail of the estimated, rather than the true, band 6 image. This unexplained detail, which may be due to overfitting to the training data, is subject to further study.

Figure 4 depicts the 120-m IFOV T-IR band 6 image, interpolated to a 30-m sample distance (that is, TM level 1 scene) and figure 5 is the high-resolution NN estimate for band 6 as a function of 30-m bands 7, 5, and 2. This estimated image was used (step 4) to adaptively sharpen the raw band 6 image (fig. 1c) by means of a reported multiresolution technique. Figure 6 is the final NN-based, sharpened thermal infrared band 6 image. Shown for comparison is the 30-m band 5 image (fig. 7) and part of a U.S. Geological Survey 7.5-minute topographic quadrangle (Opa-Locka, Fla.) map corresponding to the upper third of this image. Figure 6 has many sharpened, well-defined boundaries; for example, the grass-ponds boundaries in the upper center area. With discriminate analysis of TM multispectral data, this sharpened T-IR data may provide additional information to improve classification.

Because of the limited number of images available for these first tests, the performance of this sharpening technique with opposite contrast edge pattern data has not been evaluated. These are the first examples produced by this NN method for sharpening TM T-IR data. It should be noted that these promising results were derived from an NN that was trained with only a relatively small number of edge contrast pattern examples.

VII. Conclusion

An NN-based technique for sharpening TM thermal band data was used to develop an NN method to approximate thermal data as a learned function of data from three other TM multispectral bands. Although these first experimental examples are promising, they were only based on a simple NN training strategy with limited training examples. Further study and tests are needed to ensure that the NN is not overtrained and in this process has good generalization properties. An evaluation of radiometric fidelity is also needed and can be accomplished by sharpening a simulated low-resolution image computed from another TM band (Tom, Carlotto, and Scholten, 1985). This would allow quantitative comparison with a known true (but not thermal band) image.

VIII. References

Burt, P.J., 1985, Smart sensing within a pyramid vision machine: Proceedings IEEE, v. 76, no. 8, p. 1,006-1,015.

Burt, P.J., and Adelson, E.H., 1983, The Laplacian pyramid as a compact image code: IEEE Trans. on Communications, v. 31, no. 4, p. 532-540.

Chen, D.S., and Jain, R.C., 1994, A robust backpropagation learning algorithm for function approximation: IEEE Trans. on Neural Networks, v. 5, no. 3, p. 467-479.

Hertz, J., Krogh, A. and Palmer, R., 1991, Multilayer Networks, chapt. 6 *of* Introduction to the Theory of Neural Computation: Addison-Wesley, p. 115-162.

Hornik, K., Stinchcombe, M., and White, H., 1989, Multilayer feedforward networks are universal approximators: Neural Networks, v. 2, p. 359-366.

Iverson, A.E., and Lersch, J.R., 1994, Adaptive image sharpening using multiresolution representations: Proceedings SPIE, v. 2,231, p. 72-83.

Jensen, J.R., 1986, Remote sensing data acquisition alternatives, chapt. 2 *of* Introductory digital image processing: Prentice-Hall, N.J., p. 11-49.

Li, H., Manjunath, B.S., and Mitra, S.K., 1995, Multisensor image fusion using the wavelet transform, Graphical Models and Image Processing, v. 57, n. 3, p. 235-245.

Nishii, R., Kusanobu, S., and Tanaka, S., 1996, Enhancement of low spatial resolution image based on high resolution bands, IEEE Trans. Geoscience and Remote Sensing, v. 34, no. 5, p. 1,151-1,158.

Ogden, J.M., Adelson, E.H., Bergen, J.R., and Burt, P.J., 1985, Pyramid based computer graphics: RCA Engineer, v. 30, no. 5, p 4-15.

Reed, R., 1993, Pruning algorithms-a survey: IEEE Trans. on Neural Networks, v. 4, no. 5, p. 740-747.

Rumelhart. D.E., Hinton, G.E., and Williams, R.J., 1986, Learning Internal Representations by Error Propagation, chap. 8 of Parallel Distributer Processing 1: Foundations, D.E. Rumelhart and J.L. McCelland, MIT Press, Cambridge, Mass., p. 318-362.

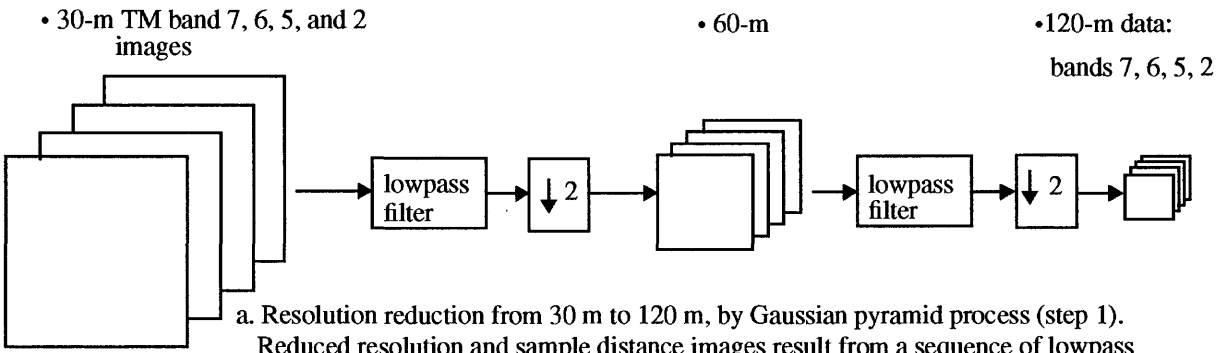
Schowengerdt, R.A., 1980, Reconstruction of multispectral image data using spatial frequency content: Photogrammetric Engineering and Remote Sensing, v. 46, no. 10, p. 1,325-1,334.

Schowengerdt, R.A., and Filiberti, D., 1994, Spatial frequency models for multispectral sharpening: Proceedings SPIE v. 2,231, p. 84-90.

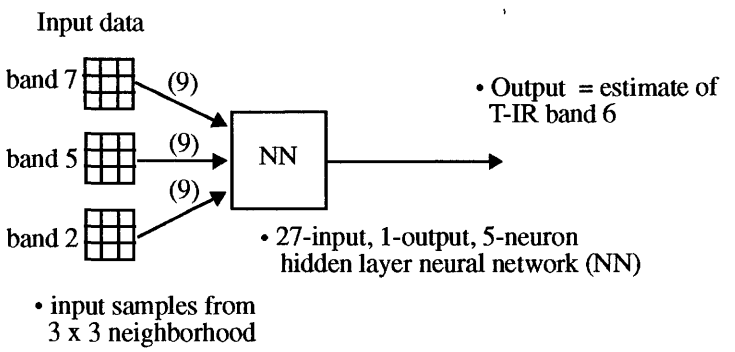
Tom, V.T., 1986, A synergistic approach for multispectral image restoration using reference imagery: Proceedings of IGARSS Symposium, Zurich, ESA Publications SP-254, p. 559-565.

Tom, V.T., Carlotto, M.J., and Scholten, D.K., 1985, Spatial sharpening of thematic mapper data using a multiband approach: Optical Engineering, v. 24, no. 6, p. 1,026-1,029.

Wasserman, P.D., 1993, Advanced Methods in Neural Computing, New York, VanNostrand-Reihold.

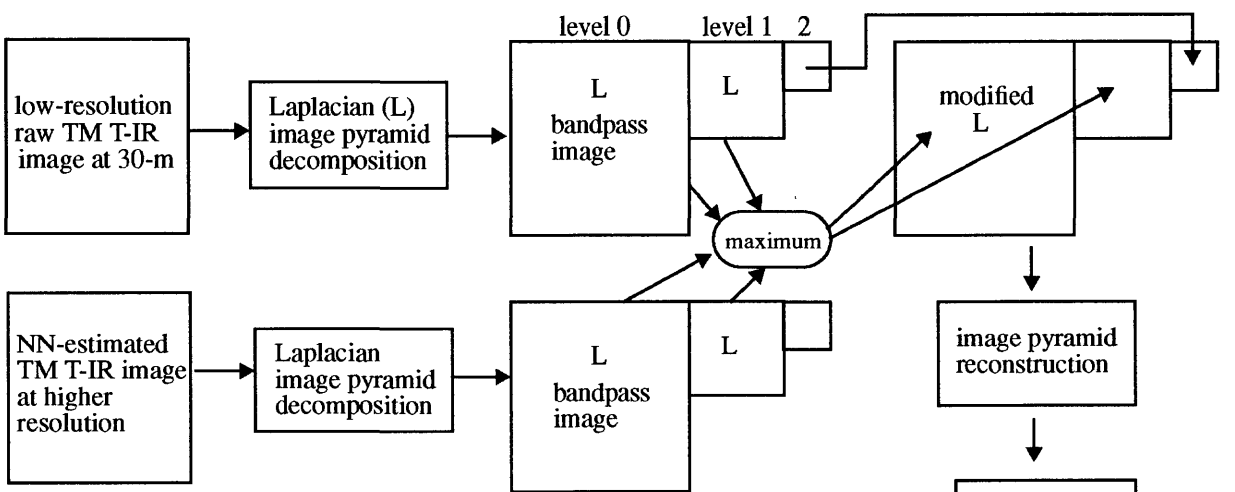


a. Resolution reduction from 30 m to 120 m, by Gaussian pyramid process (step 1). Reduced resolution and sample distance images result from a sequence of lowpass filtering and down sampling along rows and columns by a factor of 2. This process begins with 120-m IFOV T-IR band 6 data, interpolated to 30-m sample distance



b. Multilayer neural network (NN) approximation

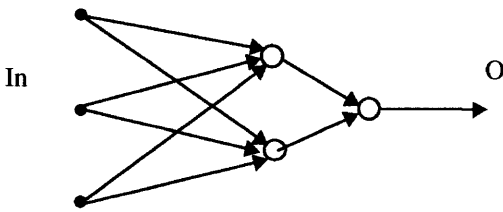
The NN is trained to approximate T-IR data at 120 m, given input at 120 m (step 2). Output of the trained NN is an estimate of T-IR data at 30 m when given 30-m input data (step 3).



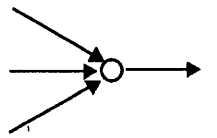
c. Adaptive merging of high-resolution estimate into raw T-IR image (step 4).

Excess spatial detail of estimated T-IR is adaptively transferred to raw image by a reported multiresolution technique.

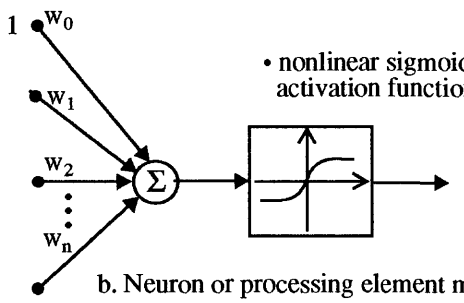
Figure 1. Neural network based method for sharpening low-resolution TM thermal infrared (T-IR) data from higher resolution TM data.



a. An example of a multilayer feedforward neural network with three inputs, one output, and two processing elements (PE) in hidden layer,



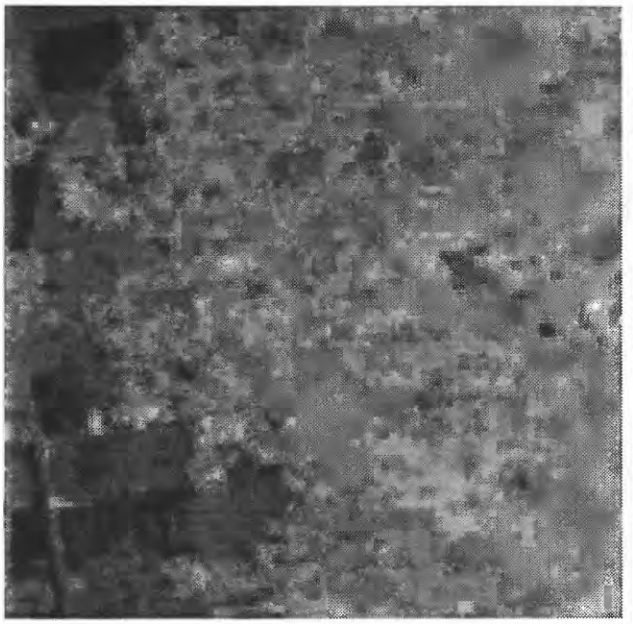
• Symbol for processing element, or neuron.



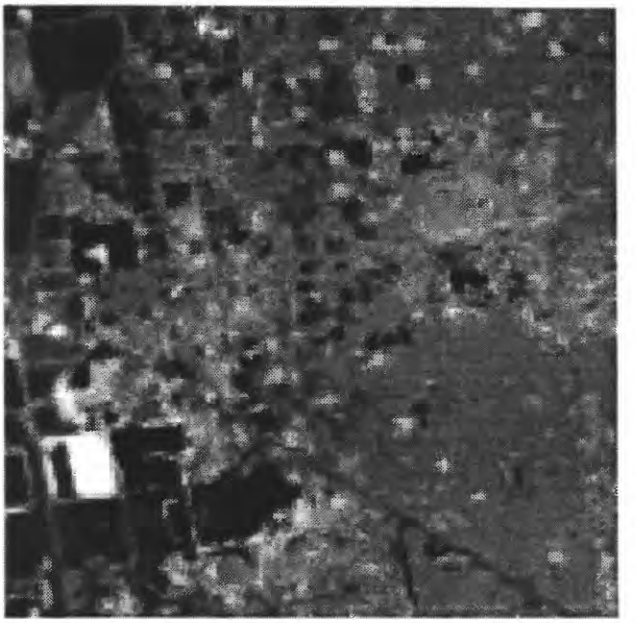
• nonlinear sigmoidal activation function

b. Neuron or processing element model.
w_0 is bias weight, with a fixed input of 1.

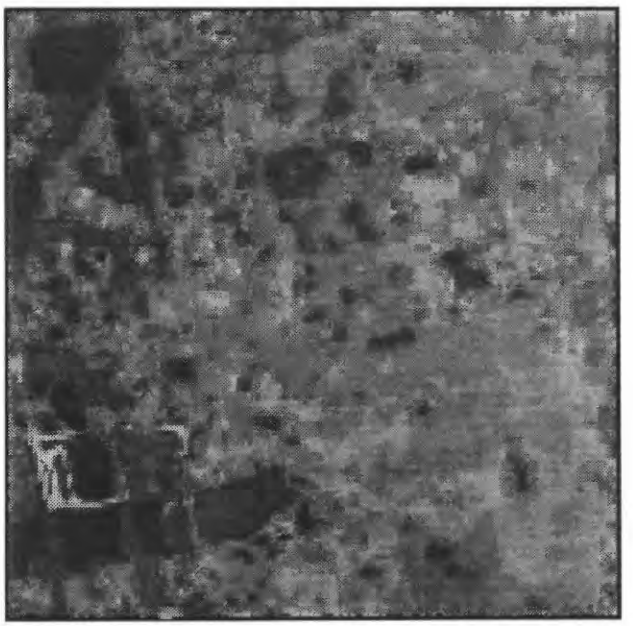
Figure 2. Neural network and neuron models.



a.



b.



c.

Figure 3. Reduced-resolution, 120-m sample distance training data and low-resolution NN approximation for T-IR data.

3a shows T-IR data. 3b is band 5, an input for the NN approximation of T-IR. 3c is the NN approximation of 3a. NN training data (from lower half of these images) were 3a (desired output) with input from 3b and, not shown, bands 7 and 2. A linear contrast stretch has been applied to the images of figs. 3 to 7 to improve their display.

Figure 4. Original-resolution, 120-m IFOV TM T-IR image at 30-m sample distance.

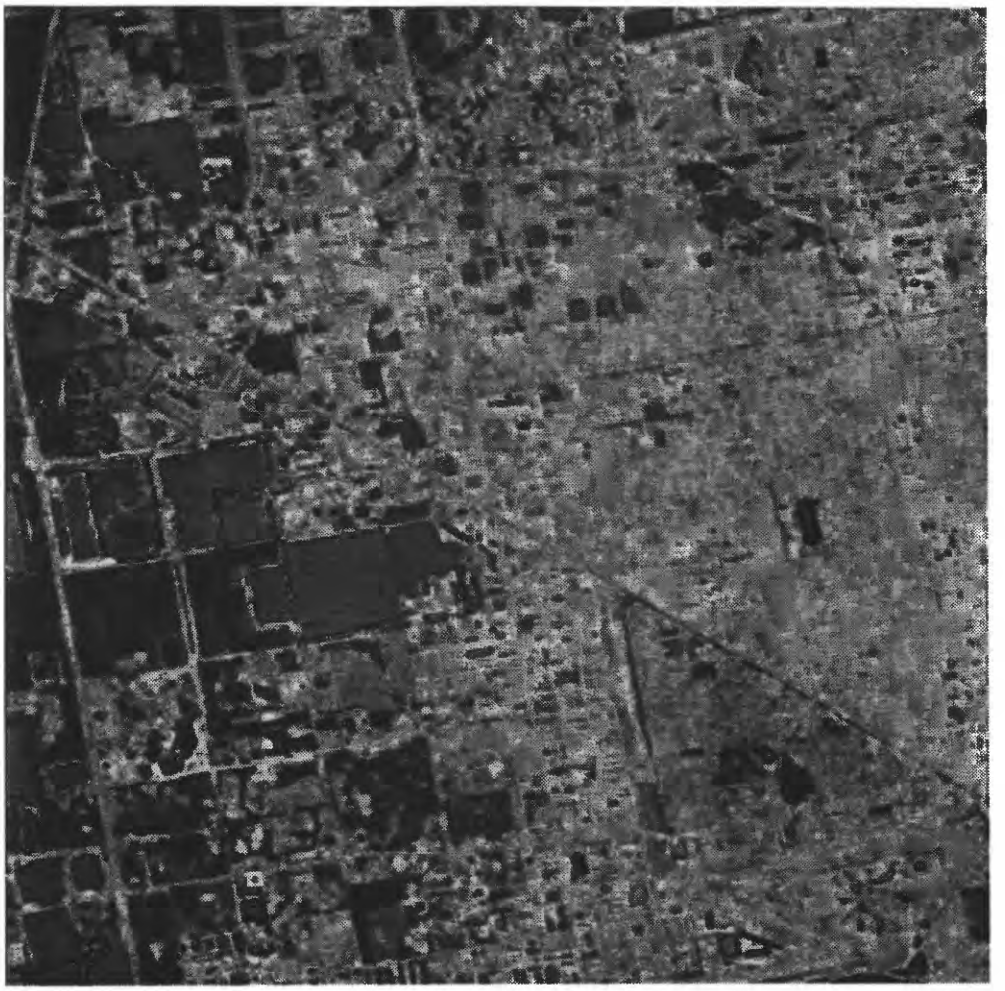


Figure 5. Neural network estimate of T-IR at 30-m (high) resolution.

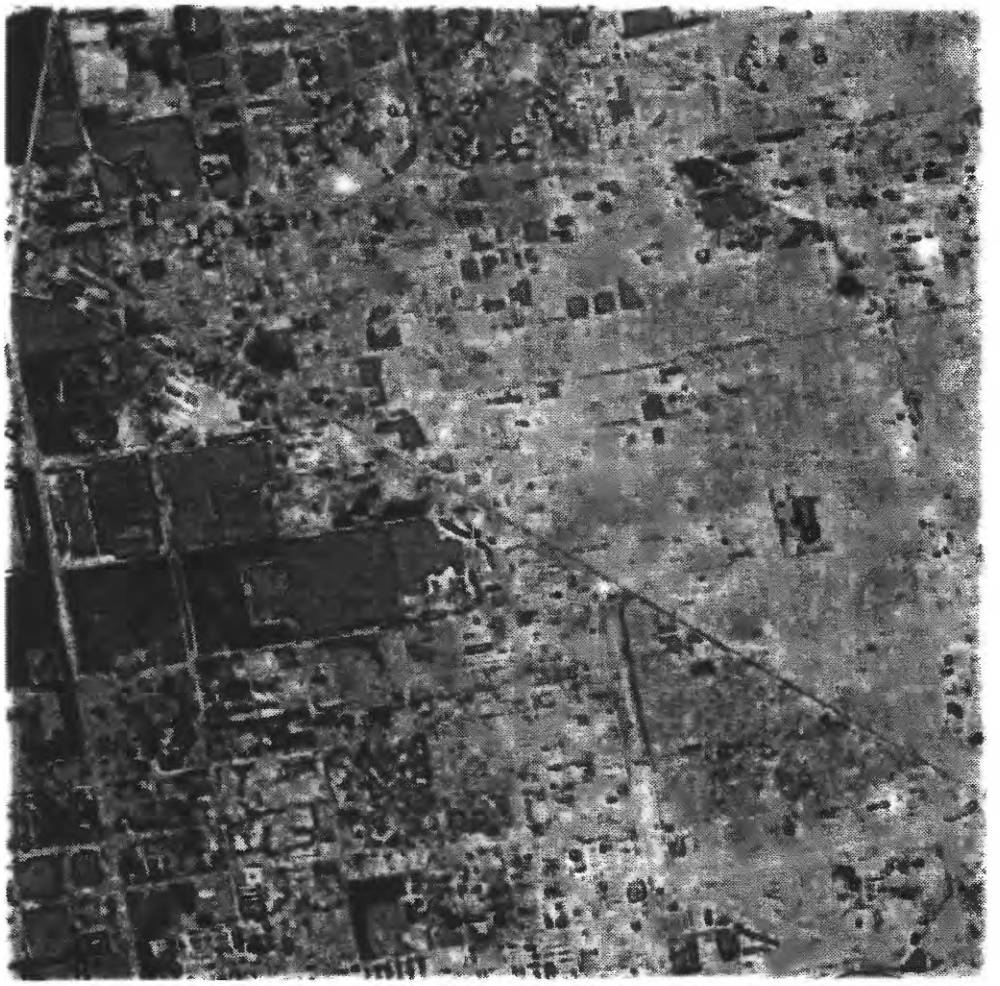


Figure 6. Final result: sharpened TM thermal infrared (band 6) image.

This is the result of adaptively combining the thermal image data of figure 4 with its high-resolution, NN-based estimate (fig. 5).

Figure 7. TM mid-infrared band 5 image; 30-m resolution.

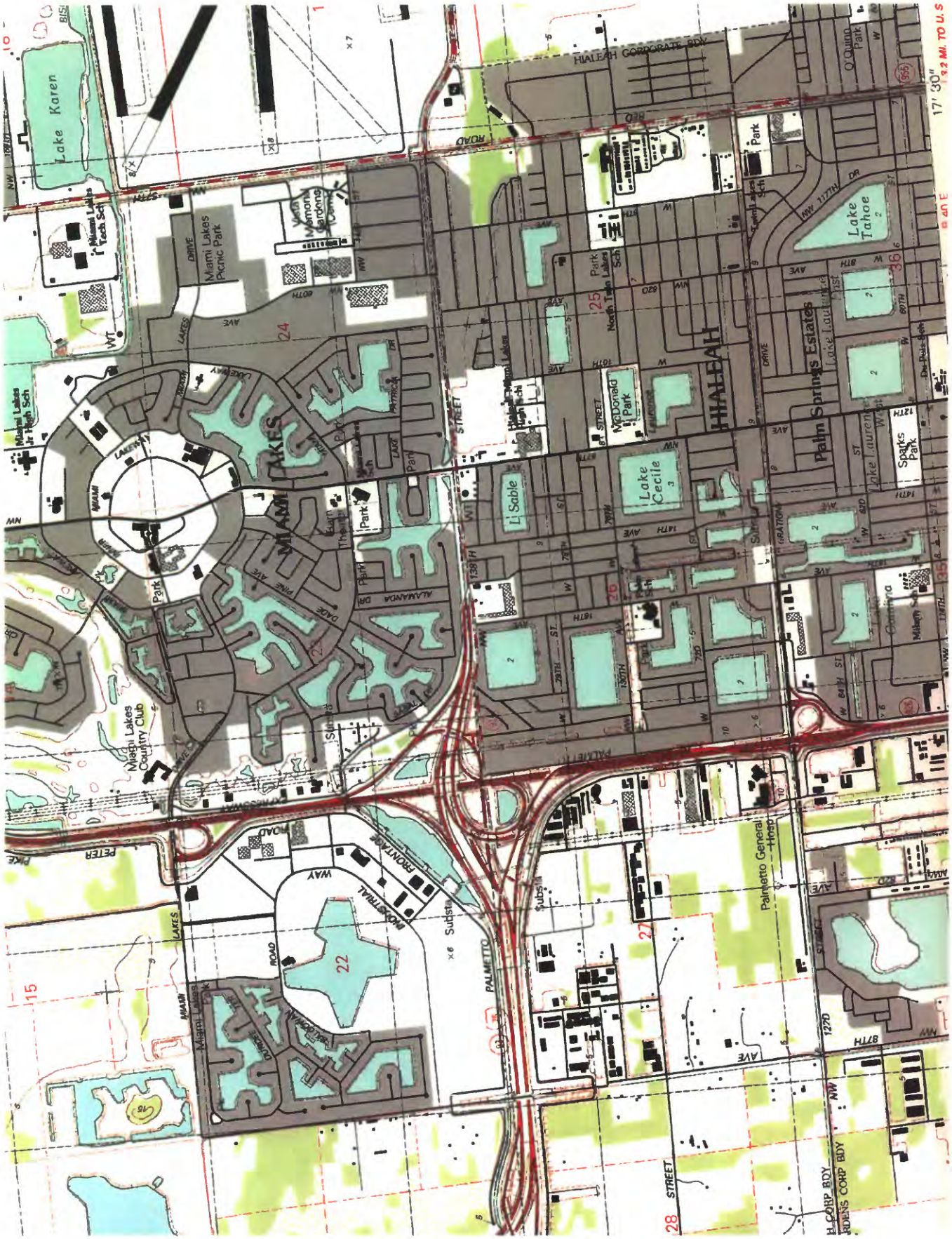


Figure 8. Image of USGS 7.5-minute series topographic map: Opa-Locka, Fla., quadrangle.