

Interpolation in multispectral data using neural networks

V. Tsagaris, A. Panagiotopoulou and V. Anastassopoulos
Electronics and Computers Division, Physics Department,
University of Patras, Greece, 26500

ABSTRACT

A novel procedure which aims in increasing the spatial resolution of multispectral data and simultaneously creates a high quality RGB fused representation is proposed in this paper. For this purpose, neural networks are employed and a successive training procedure is applied in order to incorporate in the network structure knowledge about recovering lost frequencies and thus giving fine resolution output color images. MERIS multispectral data are employed to demonstrate the performance of the proposed method.

Keywords: Interpolation, super-resolution, multispectral images, neural networks.

1. INTRODUCTION

Remote sensed imagery derived from aircraft or satellite mounted sensors is characterized by their spectral and spatial resolution. High spatial resolution is necessary for an accurate description of shapes, features and structures. On the other hand, high spectral resolution is needed for accurately describe the area covered by each land cover type, as well as to locate different terrain types and signatures.

Image super-resolution (SR) reconstruction algorithms aim at improving the spatial resolution of an image [1]. In this approach a sequence of oversampled low-resolution (LR) images is used to produce a high-resolution (HR) image [2-3]. Most of the super-resolution or interpolation methods have been developed in order to improve the resolution of images acquired by digital cameras [4]. Classical prediction techniques such as interpolation methods [5] or adaptive approaches have been used to restore the lost frequency components in various cases as in RGB cameras and instrumentation [6], or for recovering lost pixels in color or grayscale images [7]. Recently, super-resolution techniques for improving the spatial resolution of multispectral images have been reported [8-9]. The combination of super-resolution with image fusion methods [10-11] leads to HR images that fully exploit the high spectral resolution of the source multispectral data set.

The main purpose of this work is to provide a combined super-resolution and image fusion scheme. The final goal is to create an RGB image of improved spatial resolution (150m) from low-resolution (300m) MERIS multispectral data. A feed-forward neural network (NN) is employed for this purpose. The NN is built using as input training data 15 MERIS bands. The training procedure comprises three consecutive training steps. Decimated multispectral data, of finer resolution from step to step, are used as input to the network for the training process. The corresponding desired outputs in RGB representation are formed by means of a specific fusion technique using each time the data of higher resolution than those employed for the input of the network in the specific step. This training procedure, provides the neural network with information on the way frequencies have to be unfolded in order to obtain the final high-resolution RGB representation. The proposed method presents for the first time, simultaneously super-resolution and color image fusion techniques. Furthermore, the successive training of the network architecture with data of finer resolution provides the network with improved training and generalization performance.

This paper is organized as follows. Section 2 discusses the most commonly used observation model in the SR problem and also reviews in brief existing neural architectures that have been used for image interpolation purposes. The proposed combined scheme is thoroughly presented in Section 3. A description of the data along with the experimental procedure and results can be found in Section 4. Finally, Section 5 provides the conclusions.

2. OBSERVATION MODEL AND EXISTING ARCHITECTURES

In the next subsections the observation model that relates the HR image to the observed LR images along with the existing architectures that employ neural networks for the SR problem are presented. This brief study provides the necessary background as well as the limitation of the existing approaches.

2.1 Observation model in super-resolution image reconstruction

It is necessary to express clearly and exactly an observation model, which connects the original HR image to the observed LR images in order to comprehend the SR reconstruction problem. The observation model for still images is presented in the following [2]. A similar model can be applied in video sequences. We assume that the requested HR image of size $L_1 N_1 \times L_2 N_2$ pixels is denoted by the vector $x = [x_1, x_2, \dots, x_N]^T$, where $N = L_1 N_1 \times L_2 N_2$. This image, x , is the ideal HR image which is sampled at or above the Nyquist rate from a continuous scene supposed to be bandlimited. For the LR images, the horizontal and vertical directions are down-sampled by the factors L_1 and L_2 , respectively. The k -th LR image is represented by the vector $y_k = [y_{k,1}, y_{k,2}, \dots, y_{k,M}]^T$, with $k = 1, 2, \dots, p$ and $M = N_1 \times N_2$. The observed LR images results from warping, blurring and subsampling operators which degrade the HR image x . If we suppose that each LR image becomes corrupted by additive noise, the observation model [2, 13-14] is represented as

$$y_k = DB_k M_k x + n_k \quad \text{for } 1 \leq k \leq p \quad (1)$$

where M_k and B_k are warp and blur matrices respectively, D is a subsampling matrix and n_k denotes the noise vector. This observation model is illustrated in Figure 1.

The warp matrix M_k represents the motion that takes place when acquiring the image, such as global or local translation and rotation. The matrix B_k denotes the blurring that arises because of the optical system, the relative motion between the original scene and the imaging system and the point spread function (PSF) of the LR sensor. In the case of SR image reconstruction, the fact that a physical dimension in LR sensors is limited represents a significant cause of blur. In Figure 2, this LR sensor PSF is generally modeled as a spatial averaging operator. Furthermore, aliased LR images are generated by the subsampling matrix D from the warped and blurred HR image. Figure 3 shows the relationship between LR image and HR image in the frequency domain. The frequency domain approach is based on three principles: a) the shifting property of the Fourier transform b) the aliasing relationship between the original HR image and the observed LR images, as shown in Figure 3 and c) the assumption that the original HR image is bandlimited. The above observation model can be expressed into a simpler form as follows

$$y_k = W_k x + n_k \quad \text{for } k = 1, 2, \dots, p \quad (2)$$

where matrix W_k represents blurring, motion and subsampling effects. According to this model, the goal of SR image reconstruction is the estimation of the HR image x from the LR images y_k . Non-uniform interpolation approaches, frequency domain approaches, deterministic and stochastic regularization approaches and projection onto convex sets approach are among the existing SR algorithms which are based on this observation model [2].

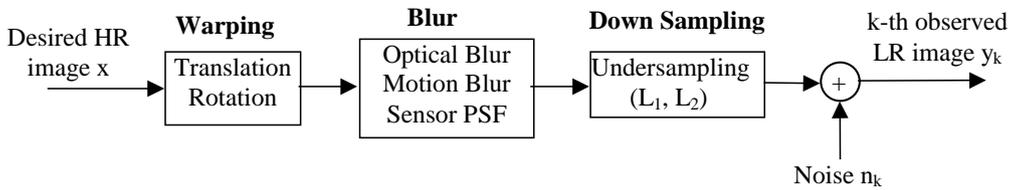


Figure 1. Observation model relating LR images to HR Images

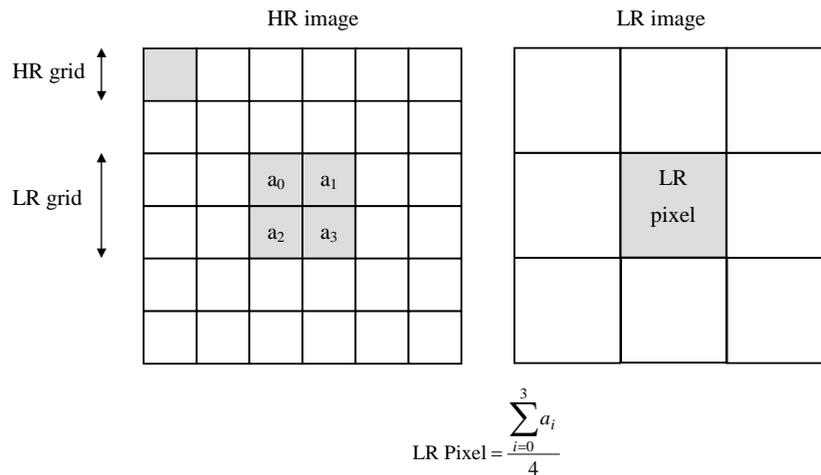


Figure 2. Low-resolution sensor PSF is usually modelled as a spatial averaging operator.

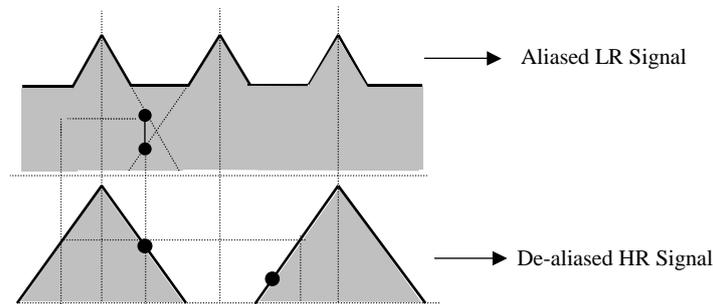


Figure 3. Aliasing relationship between LR image and HR image

2.2 Existing Architectures

In the following, the existing neural network architectures that have been employed in several types of images in order to perform interpolation and increase the resolution are briefly described.

A greyscale image interpolation method based on a multilayer perceptron is introduced in [15]. This network has one hidden layer each cell of which uses as nonlinearity the tanh function and the output layer uses only a linear sum. The method is tested in noise-free as well as noisy line doubling and image expansion problems. The neural network that performs line doubling is composed of a 30 cell input layer, a 20 cell hidden layer and a 3 cell output layer, while the network performing image expansion consists of a 24 cell input layer, a 16 cell hidden layer and a 5 cell output layer. It should be mentioned that two different back-propagation algorithms were used. According to the test results, the proposed neural network architecture gives improved performance under noisy conditions. However, this method addresses only the super-resolution or interpolation problem and does not provide dimensionality reduction that is needed in the case of multispectral data and fusion procedures.

A high-resolution multi-neural network (MNN), based on the local variance, is proposed in [16]. Here, an image is separated into two regions according to the local variance. It is expected that the signal region with high local variance will need resolution improvement, while the region with low local variance will require smooth interpolation. This specific network is composed of two NNs, the NN for low local variance and the NN for high local variance. The weighted sum of the two NN outputs represents the result of the enlargement. The above method has the advantage of not causing noticeable artefacts even in the case of a large magnification factor and reducing the performance variation when using different training images.

Another approach of resolution improvement of remotely sensed images is that of a fully interconnected neural network model [8]. The network used during testing is composed of three layers and a global interconnection mapping exists. This mapping ensures a fast convergence to the most correct answer. The input layer neurons contain the

information of the pixel's values in various bands, while the unique neuron at the output layer shows the desirable pixel value. As far as the number of hidden units is concerned, it is about twice the number of input neurons.

The network architecture employed in this work is a simple feed-forward neural network with a 25 cell input layer, a 25 cell hidden layer and 9 cell output layer. The novelty of the method is based on successive training the network and simultaneously using a fused color image as target image.

3. PROPOSED METHOD

In this section the proposed method for obtaining a high resolution RGB image (150m resolution per pixel) from low resolution (300m) multispectral imagery is analytically described. The method is based on the following two procedures which contribute to develop the final neural network architecture.

- A successive training of a specific neural network structure so that the network acquires all necessary information for creating the final high resolution RGB image.
- A segmented PCT fusion procedure which is used to create an RGB image from 15 multispectral bands of the same resolution.

These two procedures are comprised in the block diagram shown in Figure 4. Actually, the figure demonstrates the whole content of the proposed method. In the following sub-sections the two procedures are described in detail.

3.1 Neural network architecture and training

The network training procedure presupposes that the training data are all available and well organized. For this purpose, the original 15 multispectral bands of 300m resolution are used to create, by means of a decimation process, three new groups of bands with resolution 600m, 1200m and 2400m respectively as it is shown in Figure 4. Furthermore, each of the groups is used to create an RGB image of the same resolution by means of a segmented PCT fusion technique described in the next subsection.

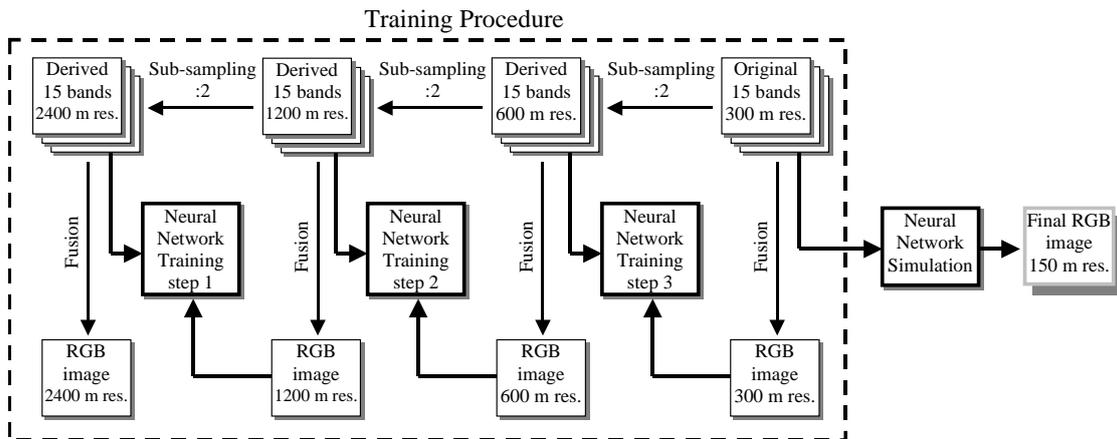


Figure 4: The proposed method for creating a high resolution RGB image is based on successive training of the same neural structure with increasing in resolution multispectral images and the corresponding fused RGB images with higher resolution.

The neural network architecture employed in this work is the one shown in Figure 5. It consists of a 25 cell input layer, a 25 cell hidden layer and 9 cell output layer. Its implementation and training was carried out using MATLAB. The training of the network starts with the first step shown in block diagram of Figure 4. The network takes as source vectors the 36 light shaded pixels shown in Figure 6, that are coming from the same position in the 15 multispectral bands of 2400m resolution (total amount of 540 pixels). Additionally it accepts as target vector the corresponding 3 dark-shaded pixels coming from each of the RGB components of the image with resolution 1200m (total of 9 pixels). The network is trained using a large number of these vectors coming from a specific region of the multispectral scene. After many training attempts the network that corresponds to the lower minimum is obtained and each weights are

saved for the next training step. The size of the network was determined experimentally, i.e. larger architectures did not yield better performance.

The training of the network continues in step 2, using the trained architecture obtained from the first step and training vectors from multispectral bands of 1200m resolution and RGB image of 600m resolution. Finishing the second step the network has now achieved a better performance (lower minimum) while the third step is coming for further training. At the end of step 3 the network training is complete and the network is ready to create the 150m improved resolution image employing both the 15 multispectral bands of 300m and the fused RGB image of 300m resolution.

Employing the proposed training procedure, demonstrated in Figure 4, we achieve the following two goals:

- The network is trained and learns to unfold the frequency components from a specific resolution to a higher one.
- The network learns how the frequency unfolding must be carried out as the resolution increases and generalizes its behaviour.

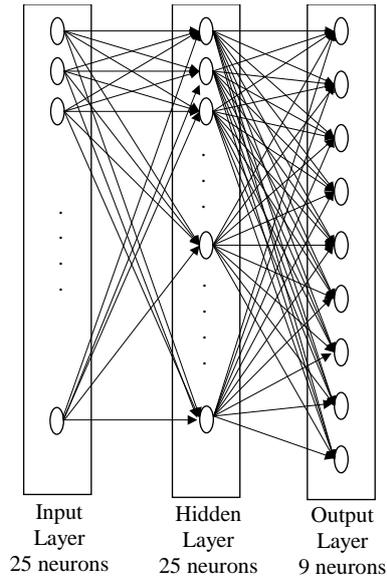


Figure 5: Neural architecture used in the proposed super-resolution method.

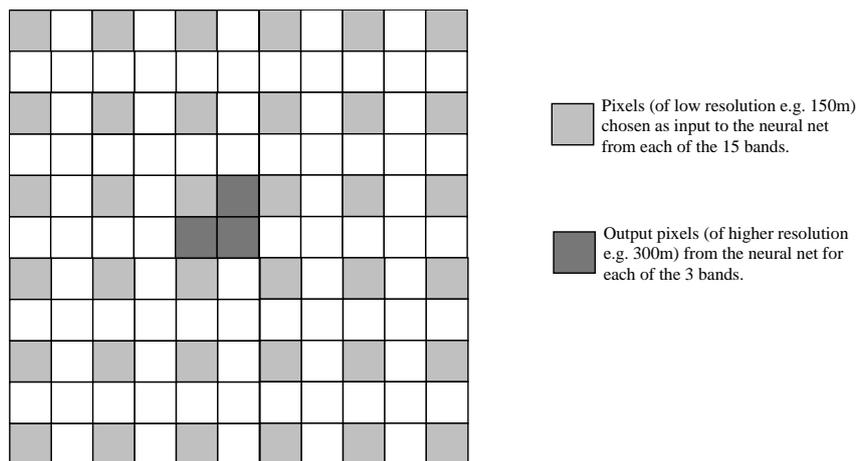


Figure 6: The arrangement of the pixels that are used for training the neural network. The light-shaded pixels are coming from the low resolution multispectral bands while the black-shaded pixels are taken from the one step higher resolution RGB images. These three pixels are created for each of the RGB components in the simulation procedure.

3.2 Image fusion approach

The main aim of image fusion methods is to combine the source multispectral or multimodal images in a compact representation that provides improved interpretation capabilities. The fusion approach used in this work is based on partitioning of the multispectral data into subgroups of bands. Then, principal components transformation (PCT) is applied on each subgroup and the derived principal components are used for information representation. The selection of subgroups depends on the spectral range of the multispectral bands.

The use of PCT in remote sensing has been proven as effective in compressing information in multivariate data sets. It gives a new orthogonal vector space where the largest amount of energy is concentrated in a few components. PCT is optimal in the sense that the first principal component will have the highest contrast and thus it can be displayed as a grayscale image with the larger percentage of visual information. The main drawback of conventional PCT is its computational burden. Furthermore, the use of PCT in the entire data set may not reveal information that appears only in a few pixels of certain bands. For these reasons segmented PCT is employed in this work. In the general form, the segmented PCT scheme is applied on the complete data set after partitioning it into 3 subgroups with the bands in each group possessing high degree of correlation. The number of bands is denoted n_1, n_2 and n_3 in subgroups 1, 2 and 3 respectively, with $N = n_1 + n_2 + n_3$ being the total number of bands. The selection employed for partitioning the MERIS bands in subgroups and determine n_1, n_2 and n_3 is based on the spectral range of the bands shown in Table 1 and the primary colours of the human visual system.

Specifically, the first subgroup is comprised of the first three bands of Table 1. These bands cover the Blue region of the electromagnetic (EM) spectrum. The first principal component of this subgroup is used as the Blue component in order to form the final RGB image. The next two bands, namely bands 4 and 5, cover the green region of the EM spectrum and PCT is applied on them to evaluate the principal component that will be used as Green in the RGB fused image. Finally, the rest bands comprise the third subgroup and PCT in this subgroup provides the Red component of the fused colour image.

Table 1 : MERIS spectral channels

No.	Band Centre (nm)	Bandwidth (nm)	Applications
1	412.5	10	Yellow substance and detrital pigments
2	442.5	10	Chlorophyll absorption maximum
3	490	10	Chlorophyll and other pigments
4	510	10	Suspended sediment, red tides
5	560	10	Chlorophyll absorption minimum
6	620	10	Suspended sediment
7	665	10	Chlorophyll absorption & fluo. reference
8	681.25	7.5	Chlorophyll fluorescence peak
9	708.75	10	Fluo. Reference, atmosphere corrections
10	753.75	7.5	Vegetation, cloud
11	760.625	3.75	O ₂ R – branch absorption band
12	778.75	15	Atmosphere corrections
13	865	20	Vegetation, water vapour reference
14	885	10	Atmosphere corrections
15	900	10	Water vapour, land

4. EXPERIMENTAL PROCEDURE AND RESULTS

The MERIS multispectral data used in this work are acquired from the ENVISAT satellite. These data have been provided by European Space Agency (ESA) under a Category-1 project. The data set is comprised of 15 bands and their spectral range is shown in Table 1. The area covered by the multispectral images is the larger part of Greece, the whole Aegean sea and a big part of the Turkish shore. The smaller images, which are parts of the multispectral bands and used either for training or validation of the neural network, are chosen randomly. The original multispectral images have 300m spatial resolution and three subsampled versions of 600, 1200 and 2400m respectively are derived. The next step is to apply the segmented PCT fusion process described in subsection 3.2 in the original along with the subsampled versions of the image. The results can be found in the first column of Figure 7. These images are used in the successive training procedure described in subsection 3.1. The performance of the neural architecture employed, improves from step 1 to step 3 as the achievable total MSE decreases.

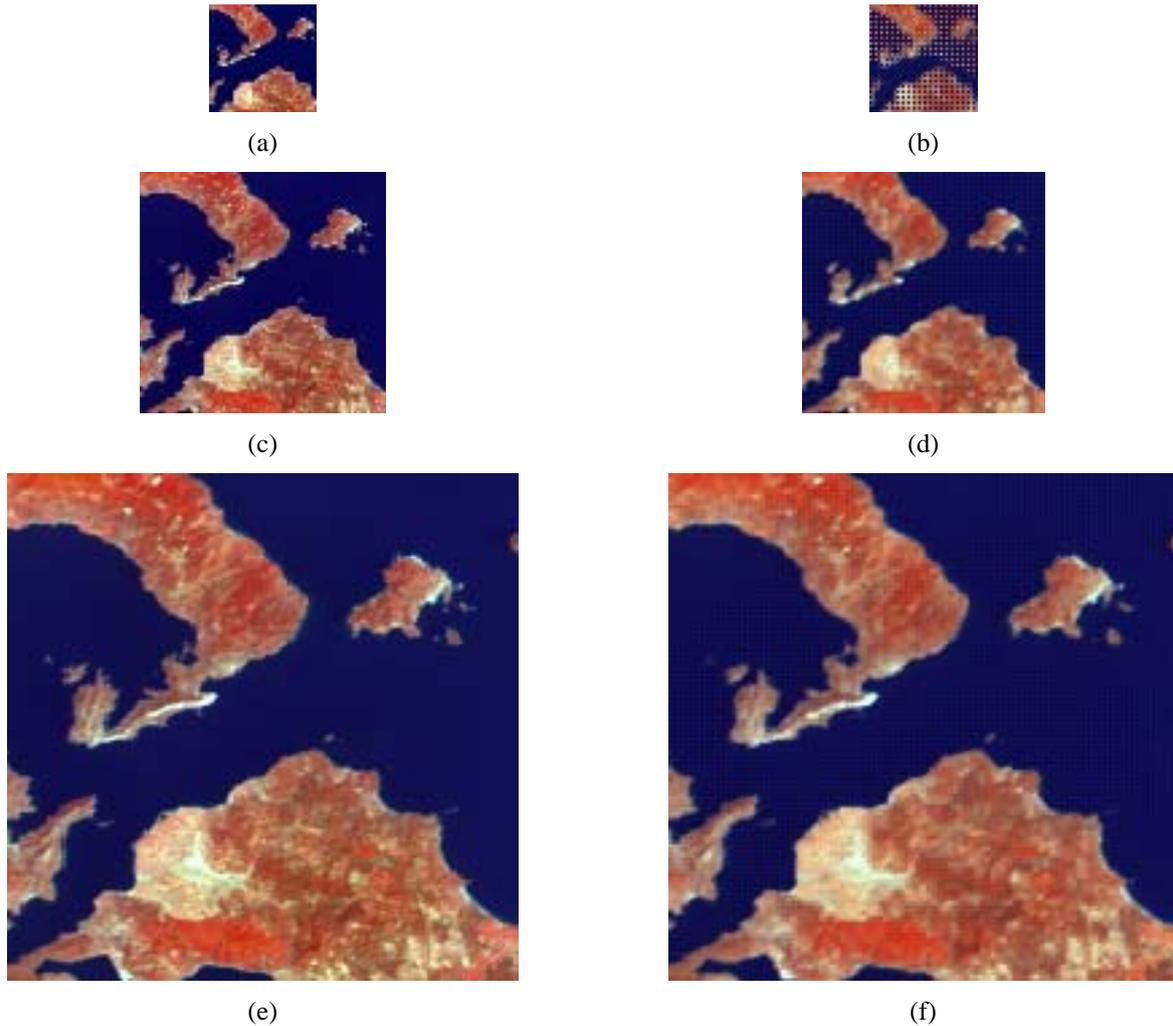


Figure 7: Area used in the training of the neural network (a) fused image at 1200m (b) network result at 1200m (c) fused image at 600m (d) network result at 600m (e) fused image at 300m (f) network result at 300m.

The successive training of the network results in improved performance as far as its generalization is concerned [17]. The part of the data used for training was selected in order to contain all types of land cover as well as part of Aegean Sea. Moreover, the neural network resulted from the successive training is tested in different regions of the broader area. The corresponding subsampled and fused images are evaluated and the MSE between the results of the

fusion process and the image derived from the trained neural network is calculated. The results, shown in Table 2, indicate that the neural network performs in the same way or sometimes even better in other randomly selected areas (Figure 8). The final improved resolution image of 150m resolution, derived from the neural network for the area used in the training process, is shown in Figure 9. The similar result, corresponding to the area used for cross validation, is depicted in Figure 10. All images shown in Figure 7 to 10 are in real size and no scaling was performed.

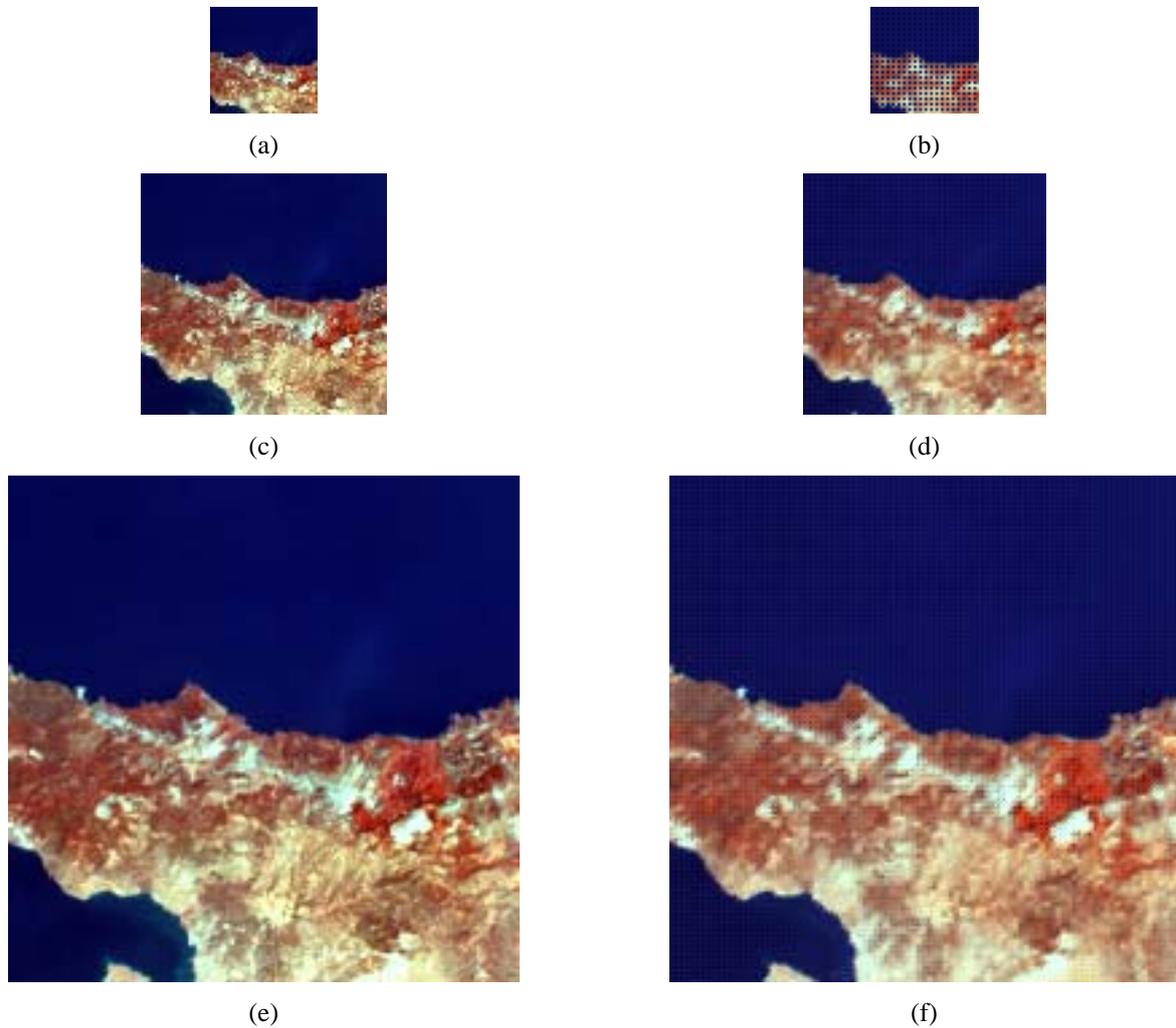


Figure 8: Randomly selected area (a) fused image at 1200m (b) network result at 1200m (c) fused image at 600m (d) network result at 600m (e) fused image at 300m (f) network result at 300m.

Table 2 : Cross-validation results in terms of MSE

	Step 1: 2400m to 1200m	Step 2: 1200m to 600m	Step 3: 600m to 300m
Training Area	0.187	0.045	0.022
Area 1	0.238	0.104	0.056
Area 2	0.084	0.057	0.033
Area 3	0.162	0.060	0.035
Area 4	0.092	0.056	0.040
Area 5	0.045	0.026	0.020



Figure 9: Improved resolution fused colour image with 150m pixel resolution for the original area used in the training process.



Figure 10: Improved resolution fused colour image with 150m pixel resolution for a randomly selected area.

5. CONCLUSIONS

A simple feed-forward neural network was proposed for creating a high-resolution RGB image from low-resolution multispectral data. Actually, super-resolution as well as image fusion are carried out simultaneously using the proposed network. The basic idea behind the proposed approach is to train the network not only to learn and unfold frequencies from a specific resolution to a finer one, but to generalize this knowledge to even finer resolution image. To achieve this, successive training of the network with increasing resolution is carried out. Decimation and fusion procedures are employed to create and organize the data.

ACKNOWLEDGMENTS

This work was supported by the European Social Fund (ESF), Operational Program for Educational and Vocational Training II (EPEAEK II), and the Program IRAKLEITOS of the Ministry of Education and Religious Affairs, Greece.

REFERENCES

- [1] S. Chaudhuri, "Super-resolution imaging", Kluwer Academic Publishers.
- [2] S. Ch. Park, M. K. Park and M. G. Kang, "Super-resolution image reconstruction: A technical overview", IEEE Signal Processing Magazine, vol. 20, no. 3, pp. 21-36, 2003.
- [3] S. Borman and R.L. Stevenson, "Super-resolution from image sequences – A review", in Proc. 1998 Midwest Symposium Circuits and Systems, pp. 374-378, 1999.
- [4] J. Go, K. Sohn and Ch. Lee, "Interpolation using neural networks for digital still cameras", IEEE Transactions on Consumer Electronics, vol. 46, no. 3, pp. 610-616, 2000.
- [5] T.M. Lehmann, C. Gonner and K. Spitzer, "Survey: Interpolation Methods in Medical Image Processing", IEEE Trans. on Medical Imaging, vol. 18, no. 11, pp. 1049-1075, 1999.
- [6] S-C. Pei and I-K Tam, "Effective Color Interpolation in CCD Color Filter Array Using Signal Correlation", IEEE trans. on Circuits and Systems for Video Technology, vol. 13, no. 6, pp. 503-513, 2003.
- [7] B.K. Gunturk, Y. Altunbasak and R. M. Mersereau, "Color Plane Interpolation Using Alternating Projections", IEEE Trans. on Image Processing, vol. 11, no. 9, pp. 997-1013, 2002.
- [8] M. Del Carmen Valdes and M. Inamura, "Spatial resolution improvement of remotely sensed images by a fully interconnected neural network approach", IEEE Transactions on Geoscience and Remote Sensing, vol. 38, no. 5, pp. 2426 – 2430, 2000.
- [9] A.J. Tatem, H.G. Lewis, P.M. Atkinson and M.S. Nixon, "Super-resolution land cover pattern prediction using a Hopfield neural network", Remote Sensing of Environment, vol. 79, pp. 1-14, 2002.
- [10] T. Ranchin, B. Aiazzi, L. Alparone, S. Baronti and L. Wald, "Image fusion – the ARSIS concept and some successful implementation schemes", Photogrammetry and Remote Sensing, vol. 58, pp. 4-18, 2003.
- [11] M. Gonzalez-Audicana, J. L. Saleta, R.G. Catalan and R. Garcia, "Fusion of multispectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition", IEEE Transactions on Geoscience and Remote Sensing, vol. 46, no. 6, pp. 1291-1299, 2004.
- [12] V. Tsagaris and V. Anastassopoulos, "Multispectral image fusion based on perceptual attributes", Proc. of SPIE, vol. 5238, pp. 357-367, 2003.
- [13] M. Elad and A. Feuer, "Restoration of a single super-resolution image from several blurred, noisy, and undersampled measured images", IEEE Trans. Image Processing, vol. 6, no. 12, pp. 1646-1658, 1997.
- [14] N. Nguyen, P. Milanfar, and G. Golub, "Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement", IEEE Trans. Image Processing, vol. 10, pp. 1299-1308, 2001.
- [15] Nathalie Plaziac, "Image Interpolation Using Neural Networks", IEEE Transactions on Image Processing, vol. 8, no. 11, pp. 1647-1651, 1999.
- [16] Daisuke Sekiwa and Akira Taguchi, "Enlargement of Digital Images by Using Multi-Neural Networks", Electronics and Communications in Japan, part 3, vol. 84, no. 11, 2001.
- [17] Dan W. Patterson, "Artificial Neural Networks: Theory and Applications", 1996, Prentice-Hall.